LEVEL

Effect of the Ionosphere on Spaceand Terrestrial Systems

Editor

John M. Goodman

AD A100313

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

Jointly Sponsored by
Naval Research Laboratory
and
The Office of Naval Research

JUN 5 1981
A

Effect of the Ionosphere on Space and Terrestrial Systems

Editor

John M. Goodman



Based on

Ionosphere Effects Symposium

Held in

Crystal City, Arlington, Va.
January 24-26, 1978

JUN 1 5 1981

A

Jointly Sponsored by
Naval Research Laboratory
and
The Office of Naval Research





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

ORGANIZING COMMITTEE

Chairman: Dr. J. M. Goodman

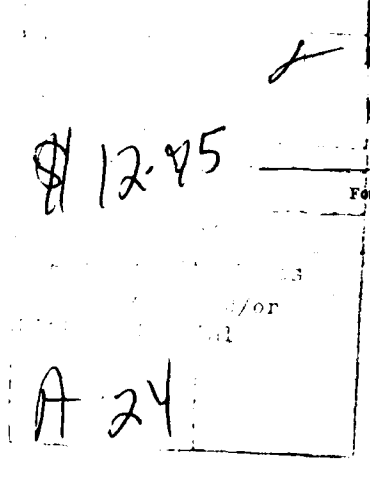
Coordinator: Mr. F. D. Clarke

Secretary: Mrs. R. A. Smithson

Session Chairman:

Dr. T. P. Quinn
Dr. R. Conley
Dr. A. Rosen
Mr. G. Hagn
Dr. W. Wasylkiwskyj
Mr. E. Peterkin

Library of Congress Catalog Card No. 78-600122



For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Stock Number 008-051-00069-1

☆ U.S. GOVERNMENT PRINTING OFFICE : 1978 O—277-162

PREFACE

Three years have passed since the last Ionospheric Effects Conference sponsored by the Naval Research Laboratory. We are happy that the Office of Naval Research has assisted us through co-sponsorship of IES '78, and we look forward to co-sponsorship by URSI or other organizations for conferences in the future.

As in the earlier IES conference, the central theme of IES '78 was embodied in solar-terrestrial physics but an emphasis upon environmental interaction with space and terrestrial systems was planned and, for the most part, was accomplished. Indeed, the prime consideration in conferences of this type, being constructed as combined sessions, is to develop and promote a dialogue between the several disciplines involved. These may be roughly divided (and I must take care not to offend anyone), into two camps; "environmental" scientists and customers of the products derived from science. I hesitate to employ the term "user community" since it may appear condescending and no one likes to be "used" anyway.

I should remark that hardly anyone at these conferences regards him(her)self as the ultimate customer of the products of environmental science. I suppose the ultimate "user" is the system operator who is pigeon-holed within the cockpit of an aircraft or similar environs. In general, this ultimate user is patently unaware of why or how his system reacts to variations in the solar-terrestrial environment. I'm not too familiar with attempts to circumvent this problem except in the U.S. Navy. An example is the mini-computer terminal developed by NOSC called PROPHET which by virtue of its application software and real-time displays of solar flux, ray tracings, and other options provides the HF operator with an inadvertent lesson in ionospheric physics. In general, however, the necessity to train operators to be familiar with the role which the environment plays in system performance is unclear.

It would be desirable to either design the systems properly to begin with so that they are "robust" or make them sufficiently adaptable that they can "ride out" the "storm" with minimal product delay or unreliability. In a sense, sub-systems such as PROPHET emphasize the adaptive role played by operators of systems which are highly vulnerable to the ionosphere. There are, of course, attempts to make such adaptive schemes fully automatic but the projected successes of these approaches is dependent upon the deleterious property to be thwarted.

One of the real needs which have surfaced, and generally expressed by all attendees, is adequate and timely modelling of the environment. Models may be used to determine the limits over which the more sophisticated systems must be designed to adapt, or the magnitude of effects which the less sophisticated systems must combat totally for a fixed grade of service or reliability. Models can also be employed operationally if driven by real-time observables. The development of models, either physical or quasi-empirical, involve a continuing R&D investment. Some operationally useful models have been discussed in this conference but the majority are models which are principally of use to system designers. Another approach to short-term prediction or real-time assessment, other than measurement of solar parameters and their application to somewhat imprecise physical models, is the observation and "tracking" of geophysical features such as the

midlatitude trough, travelling disturbances, scintillation zones, etc. This aspect has yet to be fully addressed but may have merit as an interim step.

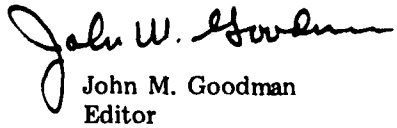
Among those topics covered in IES '78 included spacecraft charging which may be serious problem not only for synchronous satellites but also for systems such as NAVSTAR/GPS the components of which are to be located in some of the more virulent regions of the earth's radiation belts. These effects are apparently most pronounced during substorm periods, a fact which again implies a real need to address the modelling question with some seriousness, in this case emphasizing the magnetosphere. Various topical conferences will address this area in the near future and we look forward to those proceedings. Considerable information will shortly be forthcoming as participants of the International Magnetospheric Study (IMS) publish accounts of the various experiments which are being conducted to elucidate the magnetosphere and its interaction with the solar wind on the one hand, and the ionosphere on the other.

Due to space limitations, I will not attempt to summarize the wide latitude of topics covered in the IES '78 proceedings, since injustices would undoubtedly accrue from such an effort. I am confident, however, that the proceedings provide a comprehensive and up-to-date status report of the various topics included and deserves a place on the bookshelves of working scientists and engineers interested in the topics.

I would like to thank all of the Session Chairmen for their assistance to me and the Staff of IES '78. They were Dr. T. P. Quinn, Dr. R. Conley, Dr. A. Rosen, Mr. G. Hagn, Dr. W. Wasylkiwskyj, and Mr. E. Peterkin. Dr. A. Berman, NRL Director of Research, welcomed the attendees and Dr. H. Rabin, Associate Director of Research for Space and Communication Sciences and Technology at NRL, delivered the Keynote Address. Both of these addresses are contained within this volume.

I was quite pleased that Dr. R. Naka, Chief Scientist of the U.S. Air Force, agreed to deliver the Banquet speech as well as actively participate in the conference.

It is virtually impossible to acknowledge everyone who assisted us in the IES '78. However, I would especially like to thank Mr. F. D. Clarke in coordinating the sessions and for deriving solutions for a myriad of problems behind the scenes. We were fortunate to have an excellent and competent staff. My secretary, Mrs. Rhea Smithson is acknowledged for her painstaking efforts in behalf of IES '78. She was assisted by Mrs. Gailyn Nocente, Miss Gina Mattiello, Miss Becky DelCarmen, Ms. Mary Draugelis, and Mrs. Betty MacDonald. Document preparation and printing was provided by Mr. D. Darr and his staff of the Technical Information Division at NRL and Mr. Bill Nafey is acknowledged as the chief projectionist during the conference. Others involved in setup and pre-conference arrangements were Mr. J. Dowd, Mr. V. B. Richards, and Mr. P. L. Watkins; Mrs. Jane Clarke was also of great assistance. I would also like to thank my wife, Mrs. Jane B. Goodman, for her attempts in structuring a spouse's program and for her patience with me during the conference. Finally I would like to acknowledge the attendees of IES '78 for making the conference a worthwhile and memorable event.



John M. Goodman
Editor
Chairman, IES '78

TABLE OF CONTENTS

Preface iii

Welcoming Address xi
Alan Berman

Keynote Address

Perspectives in Ionospheric Research xii
Herbert Rabin

SESSION I
Radiowave Scintillation: Experiments and Mitigation Schemes

The Effects of Ionospheric Disturbances on C^3 Systems (full paper unavailable) 1
M.R. Epstein and J.H. Babcock

Amplitude and Phase Scintillation at High and Equatorial Latitudes as Measured by the DNA Wideband Satellite 2
C.L. Rino, R.C. Livingston, M.D. Cousins, B.C. Fair, and M.J. Baron

Upper Limit to the Bandspreading and Fade Rates Produced by Ionospheric/Magnetospheric Scintillations 10
C.W. Prettie

Equatorial Ionospheric Scintillations in the Indian Zone 14
R.G. Rastogi, M.R. Deshpande, and H.O. Vats

Amplitude Scintillation Studies in the Polar Region on 250 MHz 17
J. Buchau, J. Aarons, J.P. Mullen, E.J. Weber, J.A. Whalen, H.E. Whitney, and E.E. Crampton

Equatorial Scintillation Tests of Les 8/9 25
A.I. Johnson and P.K. Lee

Gigahertz Scintillation at the Magnetic Equator 31
J.P. Mullen, A. Bushby, J. Lanat, and J. Pantoja

Measurement Techniques for Long Persistence Fading Phenomena Applied to Satellite Communications Links 36
H. Hanft and J.P. Corrigan

Spaced-Receiver Investigations of Equatorial Scintillation Made at Guam During 1971 and 1976 41
M.R. Paulson and R.U. Hopkins

Spaced Receiver Measurements of Intense Equatorial Scintillations 48
H.E. Whitney

Mitigation for Airborne Satellite Receiver Platforms of UHF Ionospheric Fading by the Use of Spatial Diversity 59
C.W. Prettie and A.L. Johnson

A Time Diversity Experiment for UHF/VHF Satellite Channel with Scintillation 62
D.P. White

Effects of the Ionosphere on the Performance of the SEASAR Synthetic-Aperture Radar (full paper unavailable) 71
W.D. Brown

SESSION II
Radiowave Scintillation: In Situ Measurements, Modelling, and Channel Simulation

Error Probabilities for Digital Transmission in Scintillation Fading Channels Using Gauss-Quadrature Integration 72
R.E. Ziemer and Youngyearl Han

A Note of the Design of Filters for a Digital Fade Simulator for the Transionospheric Channel 77
R.W.D. Booth

An Analysis and Simulation of a Scintillating Time-Division-Multiplexed Teletype Channel 85
C.C. Kilgus, W.H. Goodnight, L.G. Smith, H.K. Stilwell, D.D. Stott, J.M. Whisnant, and T. Wyatt

A Review of Recent Studies of Equatorial F-Region Irregularities and Their Impact on Scintillation Modeling 90
S. Basu and M.C. Kelley

Large and Small Scale Properties of Nighttime Equatorial Irregularities from Scintillations and Radar Backscatter Measurements 107
S. Basu, H. Whitney, J. Aarons, and J.P. McClure

Radio and Optical Diagnostics Applied to an Isolated Equatorial Scintillation Event 115
J. Buchau, E.J. Weber, and J.P. McClure

Theoretical and Numerical Simulation Predictions of the Equatorial Spread F Environment 124
S.L. Ossakow, S.T. Zalesak, B.E.McDonald, and P.K. Chaturvedi

Chatanika Model of the High-Latitude Ionosphere (full paper unavailable) 130
R.Wondrak, R. Tsunoda, E. Hatfield, P. Perreault, and G. Smith

Modelling of Low Latitude Ionosphere 131
Y.V. Somayajulu and A.B. Ghosh

Topside Ionospheric Through Morphology at Mid- and High-Latitudes 140
M. Ahmed and R.C. Sagalyn

ATS-6 Observations of Ionospheric/Protonospheric Electron Content and Flux 150
L. Kersley, H. Hajeb-Hosseinieh, and K.J. Edwards

Satellite Measurements from Within Ionospheric Structures Responsible for Auroral Acceleration Process 159
R.D. Sharp, E.G. Shelley, and R.G. Johnson

Effects of Auroral Clutter on Space-Borne Microwave Transmissions 167
H.L. Groginsky, A.H. Katz, D.B. Odom, and G.D. Thome

SESSION III
Radio Noise and Radiowave Propagation Between ELF and HF

Worldwide Minimum Environmental Radio Noise Levels (0.1 Hz to 100 GHz) 177
A.D. Spaulding and G.H. Hagn

HF Noise in Space 183
C.M. Rush, D. Nelson, A.L. Snyder, V. Patterson, T. Tascione, and E. Ziemba

ELF Propagation Measurements Using Natural Sources of Radiation 188
M. Ishaq and D.L. Jones

Effects of Precipitating Energetic Particles on an ELF Communication Link 192
J.B. Reagen, W.L. Imhof, E.E. Gaines, T.R. Larsen, J.R. Davis, and W.R. Moler

A Low-Frequency Earth-Detached Communication System Using Balloon Hoisted Terminals 200
L.C. Humphrey and C.R. Roberts

Use of Omega Signals for Determining Group-Velocity Characteristics of the Earth Ionosphere Waveguide 205
R. Brown

Ionospheric Propagation Effects on HF Backscatter Radar Measurements 211
G. Millman

A Model for Pulse Transmission Over the Fading Ionospheric Reflection Channel ... 219
A. Malaga and R. McIntosh (full paper unavailable)

Rocket Observations of Ionospheric E-Region Irregularities: Implications for HF Modelling 220
E.P. Szuszczewicz, J.C. Holmes, and D.N. Walker

Angle of Arrival Using Continuous Wave Transmissions 228
R.N.E. Baulch and E.C. Butcher

HF Sky Wave Direction Finding Using a Fixed Crossed Space Loop Antenna Array... 234
T.C. Green, J.E. Hipp, and W.G. Guion

The Effect of Signal Modulation in the Application of a Wave-Front Linearity Test in HF Direction-Finding 243
D.W. Rice, G.O. Venier, and G. Atkinson

Measured Sea Path Sky Wave/Ground Wave Signal Power Ratios Over the 2-10 MHz Range..... 248
J.E. Hipp and T.C. Green

Digital Ionosonde Studies at Fort Monmouth, New Jersey 254
F.J. Gorman, Jr., and H. Soicher

SESSION IV
Spacecraft Charging and Space/Microwave Interactions

State-of-the-Art for the Assessment and Control of Environmental Charging of Spacecraft 262
A. Rosen

Differential Charging of Nonconducting Spacecraft 272
L.W. Parker

On the Perturbed Environment of Metallic Bodies in Space Plasmas 281
Jean-Pierre J. Lafon

Voyager Spacecraft Charging Model Calculations 285
N.L. Sanders and G.T. Inouye

Predictions of High-Voltage Differential Charging on Geostationary Spacecraft 293
J.G. Laframboise and S.M.L. Prokopenko

Charged-Area Effects on Spacecraft Dielectric Arc Discharges 302
K.G. Balmain, P.C. Kremer, and M. Cuchanski

A Proposed Mechanism for the Initiation and Propagation of Dielectric Surface Discharges 309
G.T. Inouye and J.M. Sellen

Reduction of Spacecraft Charging Using Highly Emissive Surface Materials 313
A.G. Rubin, P.L. Rothwell, and G.K. Yates

A Plasma Bridge Neutralizer for the Neutralization of Differentially Charged Spacecraft Surfaces 317
G.K. Komatsu and J.M. Sellen, Jr.

Spacecraft Charging on ATS-6 322
B. Johnson, J. Quinn, and S. DeForest

Active Modification and ATS-5 and ATS-6 Spacecraft Potentials 328
R.C. Olsen, E.C. Whipple, and C. Purvis

Electrical Interference to Satellite Subsystems Resulting from Spacecraft Charging ... 337
R. Shaw

Voyager Spacecraft Electrostatic Discharge Immunity Verification Tests 346
G.T. Inouye, A.C. Whittlesey, S.R. Ponamgi, B.D. Cooperstein, and A.K. Thomas

Ionospheric Effects of a High Power Space-Borne Microwave Beam 354
L.H. Holway, Jr., A.H. Katz, and G. Meltz

Solar Power Satellite and Its Interactions with Plasma and the Ionosphere 362
H. Oman

SESSION V
Forecasting Studies and Ionospheric Modification

Radio Propagation and Solar Cycle 21 371
P.E. Argo and J.R. Hill

Real Time Propagation Assessment: Initial Test Results 380
I.J. Rothmuller

Ionospheric Forecasting at Air Force Global Weather Central 385
C.S. VonFlotow

Prediction of Solar Proton Events at the Air Force Global Weather Central's Space Environmental Forecasting Facility 393
E.W. Cliver, J.A. Secan, E.D. Beard, and J.A. Manley

Remote Ionospheric Monitoring 401
J. Buchau, W.N. Hall, B.W. Reinisch, and S. Smith

Attempts to Predict Trough/Plasmapause Boundaries in Real Time 411
M. Mendillo, C. Chacko, F. Lynch, and P.J.L. Wildman

Digital Data Transmission Over HF Skywave Circuits: Prediction and Experiment 423
G.A. Clapp

Environmental Forecasting by Using Ionosphere Observations on Acoustic-Gravity Waves Associated with Severe Storms 431
R.J. Hung and R.E. Smith

Lagopedo — Two F-Region Ionospheric Depletion Experiments 438
M.B. Pongratz, G.M. Smith, C.D. Sutherland, and J. Zinn

Ion Composition Measurements of a Chemically Modified Ionosphere — An Initial Report 442
G.W. Sjolander, C.Y. Johnson, J.C. Holmes, and E.P. Szuszczewicz

Determination of Minimum Scale Sizes in Plasma Cloud Striations 448
B.E. McDonald, S.L. Ossakow, S.T. Zalesak, and N.J. Zabusky

SESSION VI
Ionospheric Effects: Ranging and Navigation Systems

Effects of the Ionosphere on Navigational and Positioning Systems Full Wave Solutions 458
E. Bahar and B.S. Agrawal

Mean Arrival Time and Mean Pulse Width of Signals Propagating Through an Inhomogeneous Ionosphere with Random Irregularities 470
Y.K. Wong, K.C. Yeh, and C.H. Liu

The Variability of Ionospheric Time Delay 479
J.M. Johanson, M.J. Buonsanto, and J.A. Klobuchar

The Contribution of the Plasmasphere to Total Time Delay 486
J.A. Klobuchar, M.J. Buonsanto, M.J. Mendillo, and J.M. Johanson

Temporal Variability of Ionospheric Refraction Correction 490
D.E. Donatelli and R.S. Allen

Ionospheric Range-Rate Effects in Satellite-to-Satellite Tracking 497
R.B. Bent, J.R. Lipofsky, S.K. Llewellyn, and P.E. Schmid

Preliminary Evaluation of a Newly Developed Satellite-to-Satellite Ionospheric Refraction Correction Model 506
J.A. Behuncik

Adaptive Correction of the Effect of the Ionosphere on Range Determination by Terrestrial Radars 512
A.H. Katz, M.D. Grossi, R.S. Allen, and D.E. Donatelli

Transionospheric Propagation Effects on Coherent Integration by Ground-Based Radars 522
R.W. Bush

Adaptive Mapping of Ionospheric Features 530
R. Leitinger, R.S. Allen, D.E. Donatelli, and G.K. Hartmann

Effects of Ionospheric Irregularities on Satellite Doppler Shift 538
P.L. Dyson and J.A. Bennett

A Proposed Experiment to Measure the Refractive and Coherence Properties of GHz Radio Wave Propagation in Turbulent Ionospheres Using Global Positioning Satellites 549
Y.T. Chiu and F.A. Morse

OTHER CONTRIBUTIONS

A Review of the Services Provided by the Space Environment Services Center 557
G.R. Heckman

An Objective Approach to Region Analysis for Flare Forecasting 565
J.W. Hirman and W.E. Flowers

An Update on the Solar Cycle 21 Forecast 572
J.R. Hill

UHF Scintillation at High Latitudes for a Brief Period During 1976 574
J.M. Goodman

Effects of Ionospheric Scintillations on Satellite Communication 579
Banshidhar, N.M. Vadher, Hari Om Vats, M.R. Deshpande, and R.G. Rastogi

Effects of the Jovian Ionosphere on Entry Probe Communications 586
J.H. Yuen, Richard Woo, and T. Grant

A Microcomputer-Based System for Acquisition and Analysis of Transionospheric Communication Signals 591
P.D. Stigall, R.E. Ziemer, and B. Williams

The Ionospheric Limitation to Coherent Integration in Transionospheric Radars 598
C.L. Rino, C.H. Dawson, R.C. Livingston, and J. Petriceks

Characteristics of Late-Time Striations Observed During Operation Stress 602
J.H. Wolcott, D.J. Simons, T.E. Eastman, and T.J. Fitzgerald

Effects of Local Site Scattering of HF Antenna Performance 614
F.N. Eddy

WELCOMING ADDRESS

DR. ALAN BERMAN
DIRECTOR OF RESEARCH
NAVAL RESEARCH LABORATORY

Good morning, it gives me great pleasure to welcome you to the 1978 symposium on the Effect of the Ionosphere on Space and Terrestrial Systems. About three years ago I welcomed a similar group of people to IES '75. Perhaps some of you were present for that symposium.

As you may recall the theme of that conference dealt to a great extent with the problems or characteristics imposed on satellite systems by the ionosphere. For example, amplitude and phase scintillations were of considerable interest. These problems are naturally still around and will be the subject of some papers in this symposium. However, new topics will be added. These include spacecraft charging phenomena, environmental forecasting, and the characteristics and effects of the ionospheric channel on systems operating between ELF and HF.

I understand that solar cycle 21 has now arrived on the scene. Although this might be good news to scientists who like to study the effect of enhanced solar emissions on the earth's environment, it may be bad news for system architects who have to contend with possible outages or performance degradation.

Thus it is conferences such as this one which will enable the system developers on the operation side of the house to become better apprised of the problems that they might encounter during the upcoming solar maximum and at the same time assist active scientists in their efforts to attack real problems.

My best wishes for a successful symposium. Thank you.

KEYNOTE ADDRESS

PERSPECTIVES IN IONOSPHERIC RESEARCH

HERBERT RABIN

ASSOCIATE DIRECTOR OF RESEARCH AND DIRECTOR
OF SPACE AND COMMUNICATION SCIENCE AND TECHNOLOGY

NAVAL RESEARCH LABOPATORY

In the distant future, as historians look back at the last half of the twentieth century, they are likely to record that this period marked the beginning of a major turning point in human thought and experience. They are likely to record that it was at about this time that an acute awareness for the finiteness of the natural environment began to develop, and that human beings began to organize their activities in ways which were commensurate with this awareness. The concomitant effects on society as a result were substantial.

The scientists and technologists who pursue environmental interests at the present time are not, and certainly should not be isolated from the prevailing attitudes and interests which society in the large projects on the subject. The mergence of societal needs with the growth of technological capabilities to influence the environment are, in my estimation, powerful forces which ultimately will produce major contributions. Ionospheric researchers are but a subset of the larger environmental research community. Our contributions, I am convinced, will have their fair share of significance across a spectrum of civilian and military interests.

The participants in this Symposium conduct their research under the auspices of university, industry, and government organizations. This diverse community is able to work in a professionally harmonious manner because we all share an overriding common interest; namely, we are attempting to understand the ionosphere, and to be able to deal with it intelligently. We know that this goal is only partially completed, although major progress has been made since the ionosphere was first conceived by Kennelly and Heaviside at the beginning of this century.

We are aware that many practical techniques and systems developed both within our military and civilian communities are subject to the vagaries of the natural ionosphere. This is the subject of our Symposium where we are focusing, in particular, on communications and space systems. The challenge to deal intelligently with miscellaneous types of ionospheric harrassment takes two principal forms.

First, we are primarily concerned about the natural ionosphere, its average behavior as well as its deviations and extremes. This range of interests is not too dissimilar from that of tropospheric weather interests. There are innumerable benefits to be derived from accurate weather prediction, and conversely, penalties for failing to achieve accurate prediction. Taking a crucial example from World War II, the records show clearly that the success of the D-Day landing was in no small measure the direct result of the Allies' accurate weather prediction capabilities. Total surprise was achieved by the most massive landing force in history because the adversary did not have a valid forecast of an impending lull in an adverse weather pattern on the Normandy beaches. I would suggest ionospheric weather may also have a significant role to play at some future time.

A second dimension of our interest is understanding the disturbed ionosphere produced by either directed or inadvertent acts of man. These disturbances again encompass both a range of civilian as well as military interests. In the same manner we are concerned with inadvertent contamination of the lakes and rivers, and the air we breathe, there is an analogue as pertains to the ionosphere. We clearly wish to be sufficiently knowledgeable as to be able to antici-

WELCOMING ADDRESS

DR. ALAN BERMAN
DIRECTOR OF RESEARCH
NAVAL RESEARCH LABORATORY

Good morning, it gives me great pleasure to welcome you to the 1978 *symposium on the Effect of the* Ionosphere on Space and Terrestrial Systems. About three years ago I welcomed a similar group of people to IES '75. Perhaps some of you were present for that symposium.

As you may recall the theme of that conference dealt to a great extent with the problems or characteristics imposed on satellite systems by the ionosphere. For example, amplitude and phase scintillations were of considerable interest. These problems are naturally still around and will be the subject of some papers in this symposium. However, new topics will be added. These include spacecraft charging phenomena, environmental forecasting, and the characteristics and effects of the ionospheric channel on systems operating between ELF and HF.

I understand that solar cycle 21 has now arrived on the scene. Although this might be good news to scientists who like to study the effect of enhanced solar emissions on the earth's environment, it may be bad news for system architects who have to contend with possible outages or performance degradation.

Thus it is conferences such as this one which will enable the system developers on the operation side of the house to become better apprised of the problems that they might encounter during the upcoming solar maximum and at the same time assist active scientists in their efforts to attack real problems.

My best wishes for a successful symposium. Thank you.

pate adverse ionospheric pollutants, to at least recognize them if they occur, and to be able to reverse them later. In the military context, the potential for ionospheric disturbance, and the potential for disruption of military systems, needs to be fully understood. These interests range from gross ionospheric disruption from nuclear detonations to more subtle effects possibly derived from ionospheric modification to achieve a particular objective. It is clearly necessary to be aware of the options open to an adversary who may choose to perturb the ionosphere to his advantage.

It is my judgement that environmental sciences, and in particular ionospheric research, will have major influences for many years to come. Some relatively elementary observations, I believe, are sufficient to sustain this point of view as pertains to military interests:

First, many system developments are being pursued along a technological spiral, dominated by ever greater sophistication, higher performance, and increased interdependence directed toward achieving multiple objectives to serve diverse user interests. As the state of the art is stretched by meet these objectives, system designs become more intricate and potentially more vulnerable. The resulting end product comes into confrontation with natural limitations more directly than it otherwise might. The ionosphere for many systems is indeed one limit which needs to be carefully considered in system planning and development.

Second, the use of an ionospherically dependent system can never ignore the fact that natural or man-made events may not only constrain, but may deny use of the ionospheric path, as was alluded to above. As new systems are developed, thoughtful analysis and review will constantly be required.

Third, the ionospheric medium which supports communications, navigation and surveillance systems of various types also will support electronic countermeasure systems which tend to degrade or preclude use of these capabilities. These antisystems may depend on a comparable, if not superior, knowledge of the ionosphere as the system to be countered. This is a cycle which will continue and will remain open-ended.

Fourth, the Ionospheric medium may provide great advantage to a system user, for example, a communicator. Intelligent use of the system, which includes sophisticated knowledge of the propagating path, may, in fact be the dominant contributor to the user's success in a particular instance. Conversely, it may contribute to his demise through compromise of his location. The trade-offs and alternatives to use or abstinence of use of communications or other electromagnetic channels require continuing contributions of the knowledgeable research community.

Fifth, and finally, the continuing military needs derived from ionospheric research will be sustained not only for the above cited reasons, but because there is a pattern of numerical growth and increased dependency on systems which in one way or the other are influenced by the ionospheric medium. The best example is, of course, satellite systems, all of which are dependent on a trans-ionospheric path.

Clearly there is wide latitude available to us in our approach to ionospheric research. Through sufficient depth and understanding of the underlying physics, we may ultimately be able to predict natural ionospheric events and deal with them in a manner which derives from a fundamental understanding of the root science. This is a goal which in my view is worth pursuing in the long term. Alternatively, there are approaches which are more empirically oriented, which provide attractive solutions, certainly in the near term. Time and experience will determine which of these will continue to be attractive for the long term as well. The use of ionospheric sounders, for example, provides an extremely valuable adaptive approach in satisfying certain propagation interests. This approach may circumvent need for understanding of the underlying phenomenology. At the same time we are aware that this approach has definite limitations. For example, it provides no predictive capability in itself, it may not be helpful in denied areas, and it is not conducive to covertness. A more fundamental approach may not only support solutions to these deficiencies in the long term, but might uncover yet new or unconventional techniques which circumvent the need for the sounder in the first place. In my view, both approaches are needed, and neither should be preemptive of the other until a much more comprehensive understanding of the ionosphere is obtained. Although there may be some attractiveness in employing empirical approaches, it is too early to compromise future opportunities in satisfying both civilian and military interests by truncating options at this stage. I believe we see evidences in the program of this

Symposium that there is healthy diversity in our approaches to maintain the needed balance.

As I surveyed the proceedings, which Dr. Goodman kindly made available to me in advance of this Symposium, I noted that over one-third of the submissions dealt in one fashion or another with ionospheric irregularities, their manifestations as they produce scintillation effects in various communications channels, and what might be done in a practical way to mitigate their adverse effects. Reflected in these reports is the fact that observational data of the phenomena is accumulating rapidly, giving us expanded perspective of geographic factors, time of observation factors - namely diurnal, seasonal or even longer periodic effects, and fundamental frequency dependency of the effects. It was interesting to note in one report that antenna pointing angle is significant as pertains to long, persistent fading at low elevation angles. As further data are accumulating, there is an increasing ability to relate the signal phenomenology to distinct ionospheric events and processes. The reports on the ability to model irregularities numerically will in the long run, I believe, prove most significant in being able to account for the observations, and ultimately obtain a scintillation predictive capability. This is perhaps one of the major challenges which this community will face in the next years. There were also a number of reports which began to evaluate the significance of various scintillation data as pertains to a specific systems interest or application. For example, an evaluation of the effect of L-band fading on NAVSTAR GPS performance was made, as were calculations of general bandspreading effects, signal statistics, and most important error analysis associated with scintillation phenomena. There are also reports on techniques for mitigation of scintillation problems based on so-called diversity approaches - either space, time, or frequency. A clever technique to use space diversity on an aircraft is described.

The second day of the symposium shifts focus to several important topics, some of which have been with us for a number of years and for which there is continuing interest, and at least one topic, spacecraft charging, is of more recent vintage. We will hear discussions of environmentally generated radio noise, the natural background which every practical use of the ionosphere must take into account. In the area of ELF and VLF there are discussions of propagation and, in particular, natural mechanisms for the inhibition of propagation. And of course, as might be expected, there is considerable interest in HF, with special emphasis in several papers on the general topics of ionosphericly-generated positional or bearing errors which significantly impact radar backscatter and direction-finding techniques. The important topic of distortion of digital HF signals is also addressed, as is the subject of E-region irregularities in general HF phenomenology.

Spacecraft charging is discussed in some dozen papers which encompass the entire spectrum from the mechanisms of charging in the space environment, the details of spacecraft materials both from the viewpoint of dielectric charging and discharging as well as concrete suggestions for the development of improved materials to diminish or avoid the charging effects, the important topic of criteria and techniques for qualifying spacecraft subject to charging and discharging effects, and very interestingly, approaches for negating deleterious effects by use of active, on-board, electric charge and discharge techniques. We also see in the papers presented that the charging problem is not only relevant to earth satellites at synchronous altitude, but similarly is of considerable concern to missions to other planets. One cannot help but be impressed that this subject has brought together a strong and synergistic combination of diverse talents from the worlds of satellite engineering, materials technology, plasma physics, and space science. This second day ends on a specialized topic of some interest, the effects and interactions of a solar power station in space. The topic of ionospheric heating is addressed as an intense microwave beam transmits energy from the satellite to ground. I think we should give the solar space station community high marks for doing their homework, by getting started early on a draft of their environmental impact statement.

The final day we will have the benefit of a number of reports in the area which are largely described by the terms prediction and forecasting. These important contributions are directed to improvement of models of ionospheric RF pulse delays, real-time updating to improve estimates of temporal variation of ionospheric refraction, and the improvement of radar range error models by real-time adaptive corrections. A final group of papers deals with the performance of specific systems, either deployed or projected for future deployment. For

example, there are several papers dealing with chemical releases presented, covering photographic analysis techniques of the disturbed ionospheric region, numerical modeling of scale sizes associated with releases, and results of F region depletion with releases of reactive gases. LORAN C signals are modeled to evaluate distortions and phase anomalies, satellite-to-satellite links are modeled to obtain refractive corrections over wide geometric changes between satellites, and modeling is reported of coherence intervals available for a ground-based radar for a variety of users.

As one surveys the presentations at this Symposium, one may justly conclude there is an active, imaginative and vibrant research community addressing relevant issues of ionospheric effects on space and communications systems. We have seen that the approaches are both diversified and reinforcing. In looking toward the Shuttle era, there will be enhanced opportunities for a variety of new observations and for major contributions to the understanding of ionospheric phenomenology. It is clear that much of what has been reported in this Symposium will form the basis of future Shuttle-based studies to further ionospheric understanding and predictive capability.

In a few concluding remarks I would like to comment briefly on the in interplay between the ionospheric research community and the user community. I believe these comments are generally appropriate to both civilian and military interests. It is clearly the intention that user groups should derive concrete benefits from the efforts of the ionospheric research community. These are measured in improved system performance, reliability, and other well known measures of effectiveness. There needs to be, in my view, an appreciation in each community of the range of interests, capabilities, and objectives of the other community for full benefit to be derived. The user needs to state his requirements as precisely as possible. He also needs to state them as early in the acquisition process as is possible to avoid complex and costly fixes later on, to avoid system degradations uncovered after deployment, and to provide, thereafter, compromises in operating modes to compensate for such difficulties. At the same time the user, in my view, needs to become less guarded and protective of potential weaknesses or deficiencies which exist in present or proposed systems, to extract the maximum benefit the research community can provide.

Even the most informed user and his agent, the manager in a program office, cannot do this entirely by themselves, they need the help of ionospheric researchers who are imaginative in translating the research product to the user format, and are able to articulate to the user community the significance of the research result. My experience has been that generally the researcher's help will be sought, but this is not always the case. It is incumbent on those of us in the ionospheric research community to step forward and be heard, even when not invited or asked. The researcher, thus, needs to keep a connection with the user community so he knows when, where, and how to gain access to transfer relevant information. It is not sufficient, in my view, to only be cognizant of the need to convert research results into engineering data for ready use, and to publish results promptly and in a clear manner. The researcher needs to be imaginative, to identify problems and to indicate opportunities that even the informed user community may have overlooked. The researcher should be generally cognizant of the impacting areas his research results may have on the performance of various systems. He should be prepared to enter into the discussion in system design and definition stages to ensure proper trade-offs are conducted. He should be generally aware of the kinds of operational environmental support which are provided, and how new ionospheric research can facilitate this support. Hopefully, this Symposium provides one approach in what needs to be a continuing effort to facilitate all levels of communication and interaction between the research and user communities.

THE EFFECTS OF IONOSPHERIC DISTURBANCES ON C^3 SYSTEMS

by

Mark R. Epstein
James H. Babcock
Office of the Secretary of Defense
Washington, D.C. 20301

ABSTRACT

Satellite communications systems are employed to assure effective C^3 to the forces because of their wide-area connectivity, high capacity and flexibility; however, these systems have limitations. In additions to their vulnerability to electronic interference and direct attack, satellite communications systems can be affected by ionospheric propagation distrubances. These disturbances can result in received signal amplitude environments. Following high altitude detonations of nuclear weapons, these deleterious effects may be enhanced and also occur in other geographic areas. The resulting effects on satellite communications can be particularly severe at UHF; less severe effects may occur at higher radio frequencies.

System architects and designers of C^3 systems that will employ satellite communications must take into consideration the effects of ionospheric radio propagation distrubances just as they consider other system threats, such as electronic interference. Higher radio frequencies can be employed to reduce the effects of the distrubances, but consideration must be given to the possibly increased terminal complexity and cost. Also, the effects of the propagation disturbances can be reduced through the use of specialized modulation and coding.

AMPLITUDE AND PHASE SCINTILLATION AT HIGH AND EQUATORIAL LATITUDES AS MEASURED BY THE DNA WIDEBAND SATELLITE

C. L. Rino, R. C. Livingston, M. D. Cousins, B. C. Fair, and M. J. Baron
SRI International
Menlo Park, California 94025

ABSTRACT

The Wideband satellite experiment has been in full operation for well over a year. A large data base on scintillation and total electron content has been acquired from a high-latitude station at Poker Flat, Alaska, and equatorial stations at Ancon, Peru, and Kwajalein Atoll in the Marshall Islands. The experiment is unique in that a phase-locked S-band signal is used to synchronously demodulate signals at L-band, UHF, and VHF.

In this paper we describe the general characteristics of ionospheric phase scintillation. The phase power spectral density has a power-law form with no identifiable outer-scale cutoff within data intervals that can reasonably be taken to be stationary. We show that the envelope of the phase spectrum in the low-frequency portion is essentially unaffected by diffraction when $S_4 \leq 0.6$. For larger scintillation levels, the phase spectrum asymptotes to f^{-2}. We characterize the phase spectrum by the strength parameter T and the spectral index p such that $\Phi(f) = Tf^p$. We also show that the intensity scintillation is completely characterized in terms of p and the corresponding strength parameters for the spatial wavenumber spectrum.

Finally, we briefly review the characteristics of the auroral zone and equatorial scintillation as deduced from the Wideband data. In the auroral zone, the phase scintillation relates directly to the precipitation phenomena, while strong UHF intensity scintillation occurs only under active auroral conditions. The equatorial data show a pronounced seasonal dependence with a high probability of occurrence for gigahertz scintillation during the local summer periods in the northern and southern hemispheres.

I INTRODUCTION

The Wideband satellite (P76-5) was launched on 22 May 1976 into a near polar, sun-synchronous, circular orbit (orbital period ~ 100 min). The satellite transmits phase-coherent CW signals at S-band, L-band, UHF, and VHF. At the receiver, the S-band signal is used to phase-lock a local reference oscillator, which is used in turn to synchronously demodulate all other signals.

Since the S-band (2891 MHz) signal is generally unaffected by scintillation, the data channel phase is well approximated by an expression of the form

$$\Phi = KN_T \frac{1}{f}[1 - (f/f_r)^2] + \text{terms that depend on } \sqrt{\lambda z} \quad . \qquad (1)$$

In (1), K is a constant, N_T is the integral of the electron density along the propagation path, and f and f_r are the respective data and reference channel frequencies. The terms dependent on $\sqrt{\lambda z}$ account for diffraction effects in the data channel signal. To the extent that the diffraction effects are negligible, the data channel phase scales with frequency as $1/f[1 - (f/f_r)^2]$.

We know that the diffraction effects are negligible for the largest scale sizes. Thus, if we filter out the rapid phase variations, the component that remains is proportional to the total electron content (TEC). The filtering procedure and the scheme for converting phase modulo 2π to absolute TEC are described in Fremouw et al. (1977), where a detailed description of the entire system is given.

The rapidly varying component of phase--the phase scintillation--is to some extent affected by diffraction. Unfortunately, no complete theory exists that accounts for diffraction effects in the signal phase. Nonetheless, one can obtain some intuitive guidelines by considering the signal phasor diagram shown in Figure 1. It is clear from the diagram that

deep fades are accompanied by rapid phase excursions that we expect to enhance the high-frequency components in the phase spectrum. Under conditions of severe fading, the rapid phase changes approximate discontinuities, and the phase spectrum asymptotes to f^{-2}. This effect has been demonstrated in both simulated data and actual data records.

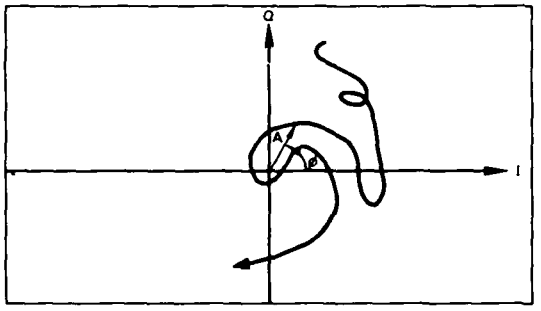


FIGURE 1 A FICTITIOUS TIME HISTORY OF SIGNAL AMPLITUDE AND PHASE SHOWING THE RAPID CHANGE OF PHASE THAT ACCOMPANIES FADES

To illustrate the diffraction effects in the signal phase, we consider the VHF detrended phase data shown in Figure 2. By taking the corresponding UHF phase data and applying the appropriate frequency scaling, we can reproduce the portion of the components of the VHF phase that is unaffected by diffraction. To isolate the diffraction effects, we have taken the difference between the directly measured and the frequency-scaled phase records.

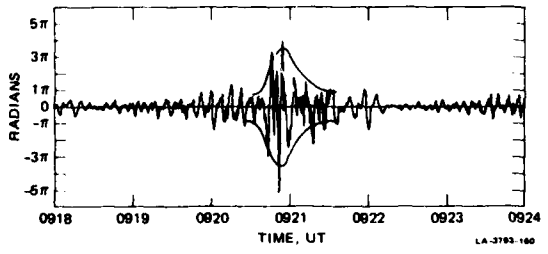


FIGURE 2 DETRENDED VHF PHASE FROM POKER FLAT PASS 6-36. The intensity scintillation index S_4 approaches unity during peak phase excursion before 0921 UT.

The power spectral density (PSD) of the phase difference for a 20-s segment taken near 0919 UT is shown in Figure 3 together with the corresponding VHF phase spectrum. At the low frequency end of the spectrum, there is ~ 20 dB of cancellation. The spectra cross near 1 Hz, with the phase difference spectrum ultimately settling at ~ 3 dB above the VHF phase spectrum, indicating that the most rapid phase fluctuations (which are partly noise) are uncorrelated. Since typically more than 80% of the area of the spectrum comes from frequencies below 1 Hz, it follows that the phase scintillation index $\sigma_\phi = \langle \phi^2 \rangle^{1/2}$ will scale with frequency very nearly as $1/f[1 - (f/f_r)^2]$. This frequency scaling is demonstrated in Fremouw et al. (1977).

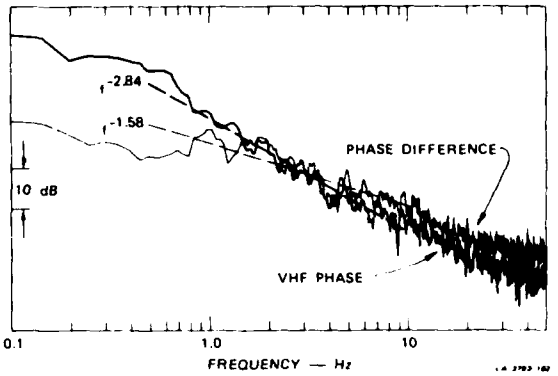


FIGURE 3 SPECTRAL DENSITY OF VHF PHASE AND PHASE DIFFERENCE FOR 20-s DATA RECORD FROM POKER FLAT PASS 6-36 AT ~0919 UT

We conclude that under conditions of weak to moderate fading the phase scintillation is well approximated by the expression

$$\delta\phi = r_e \lambda \int \delta\mu \, d\ell \quad , \tag{2}$$

where $\delta\mu$ is the perturbation to the index of refraction and the integral is taken along the propagation path. From (2), one can show that

$$\Phi_{\delta\phi}(\vec{\kappa}) = r_e^2 \lambda^2 L \sec^2\theta \times \Phi_{\Delta N_e}(\vec{\kappa}, -\tan\theta \, \hat{a}_{k_T} \cdot \vec{\kappa}) \quad , \tag{3}$$

where $\Phi_{\delta\phi}(\vec{\kappa})$ is the spatial wavenumber spectrum of phase, L is the layer thickness, θ is the zenith angle, and $\hat{a}_{k_T}$ is a unit vector along the horizontal projection of the propagation vector $\vec{k}$. Finally, $\Phi_{\Delta N_e}(\vec{\kappa}, \kappa_z)$ is the three-dimensional spatial wavenumber spectrum of the ionospheric irregularities.

To allow for the anisotropy of the irregularities, one assumes that the $(\vec{\kappa}, \kappa_z)$ dependence of $\Phi_{\Delta N_e}(\vec{\kappa}, \kappa_z)$ is given by a general quadratic form. One can show that

$$\Phi_{\delta\phi}(\vec{\kappa}) = r_e^2 \lambda^2 L \sec\theta \times Q[A\kappa_x^2 + B\kappa_x\kappa_y + C\kappa_z^2] \quad . \tag{4}$$

Expressions for A, B, and C are given in Rino and Fremouw (1977). It is generally accepted that Q(q) has the power-law form

$$Q(q) = \frac{C_s}{[q_o^2 + q^2]^{\nu + 1/2}} \qquad (5)$$

with the outer-scale wave number q_o corresponding to many tens of kilometers. Indeed, no outer-scale measure has yet been found in scintillation data (Crane, 1977).

Thus, the phase spectrum as it is measured with the Wideband satellite will have the form

$$\Phi_{\delta\phi}(f) = Tf^p \ , \qquad (6)$$

where

$$p = -2(\nu + 1) \qquad (7)$$

The temporal phase spectrum, (6), is obtained by performing a one-dimensional integration of (4) along the scan direction dictated by the combined satellite and irregularity drifts.

In our routine data analysis, the parameters T and p are obtained by performing a log-linear, least-squares fit to the phase data. The value of the phase scintillation index σ_ϕ, unlike T or p, clearly depends on the detrender cutoff frequency f_o. It is a good approximation to take

$$\sigma_\phi^2 \approx 2\int_{f_o}^{\infty} Tf^p \, df = \frac{2T}{(p+1)} f_o^{(p+1)} \ . \qquad (8)$$

That is, we ignore the small low-frequency distortion of the spectrum due to the detrending filter. For characterizing power-law processes, typically only the parameters p and T or C_s can be unambiguously deduced from actual data.

It is easily demonstrated that the intensity scintillation can be completely characterized in terms of C_s and p, although this fact is not widely used. If we apply the weak-scatter formula

$$\Phi_I(\vec{\kappa}) \approx 4\Phi_{\delta\phi}(\vec{\kappa}) \sin^2(\kappa^2 \lambda z/4\pi) \ , \qquad (9)$$

it follows from (4) and (5) with $q_o = 0$ that

$$S_4 \approx \left[4r_e^2\lambda^2 L \sec\theta\right]\left[C_s \left(\frac{\lambda z}{4\pi}\right)^{-\frac{p+1}{2}} \mathcal{I}\right] \ , \qquad (10)$$

where $\mathcal{I}$ is a geometry-dependent factor:

$$\mathcal{I} = \iint\left[A'q_x^2 + B'q_xq_y + C'q_y^2\right]^{-(\nu + 1/2)} \times \sin^2(q^2) \frac{d\vec{q}}{(2\pi)^2} \qquad (11)$$

The coefficients A', B', and C' are given in terms of A, B, and C in Rino and Fremouw (1977).

It follows from (11) that $S_4 \propto \lambda^{\eta}$, where

$$\eta = -(p-3)/4 \ . \qquad (12)$$

Thus, when $p = -3$, $S_4 \propto \lambda^{1.5}$, which is the generally accepted average value for ionospheric scintillation. From the multifrequency Wideband amplitude and phase scintillation data, only the parameters S_4, T, and p together with the TEC are potentially processor invariant and therefore fundamental. As we have already noted, the phase scintillation index, σ_ϕ, depends critically on the detrender cutoff.

In the next section we shall briefly review some of the more interesting results emerging from our studies of the Wideband satellite data. These are divided into three groups. The first involves general results that pertain to the analysis and relationships that have been presented in this section. The second and third groups pertain to the nature of the scintillation in the auroral and equatorial regions.

II PRELIMINARY RESULTS

A. General

As noted in Section I, we detrend all our data to extract the rapidly varying components of phase and intensity, which we identify as the scintillation. Phase and intensity scintillation indices are measured for every 20-s data segment. The spectral parameters T and p are calculated at VHF and UHF for all active passes.

In Figure 4, we show the behavior of the spectral index p measured at VHF and UHF as a function of S_4 for 50 representative passes taken at Ancon, Peru. The data points are obtained by averaging the p index values measured with $0.2n \le S_4 < 0.2(n + 1)$, where $n = 0, 1, 2...$ The absolute value of the p index is very near -3 when $0.3 \le S_4 \le 0.7$. For $S_4 \le 0.3$, the p estimate is affected by the noise floor, since a fixed frequency range is used for the curve fitting.

For $S_4 > 0.7$, there is a systematic reduction in the absolute value of the measured p index with increasing S_4. This is very probably the diffraction effect described in Section I.

Thus, the phase spectrum cannot be directly related to the in-situ spectrum via (3) unless $S_4 \leq 0.6$.

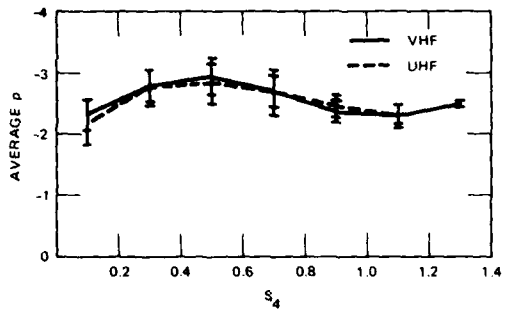


FIGURE 4 AVERAGE VALUE OF p FOR S_4 VALUES WITHIN THE INTERVALS $0.2\ n \leq S_4 < 0.2(n + 1)$. The UHF and VHF curves are computed separately.

In Figures 5 and 6, we have generated scatter diagrams for S_4 measured at UHF and VHF (Figure 5) and UHF and L-band (Figure 6). From Figure 5 we see a concentration of very weak S_4 values that follow the $\lambda^{1.5}$ curve very closely. The agreement is more striking in Figure 6. However, for large S_4 values, the data points systematically fall below the Rayleigh limiting value of $S_4 = 1$. This is clearly evident in both Figures 5 and 6.

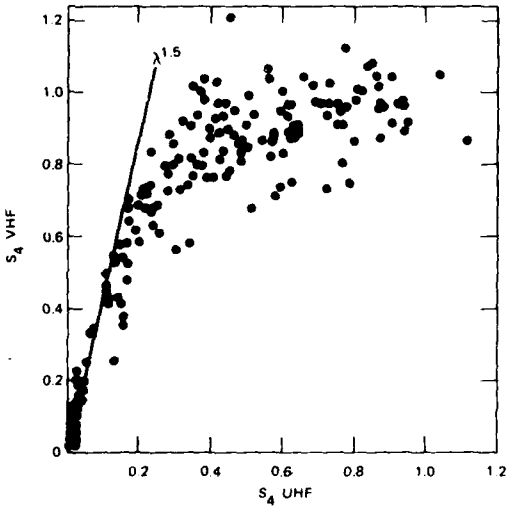


FIGURE 5 SCATTER DIAGRAM OF S_4 AT VHF VERSUS S_4 AT UHF FOR A SUBSET OF THE ANCON PASSES USED IN GENERATING THE CURVES IN FIGURE 4

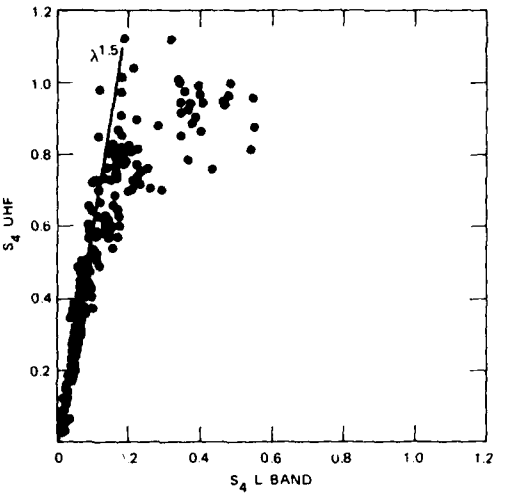


FIGURE 6 SCATTER DIAGRAM OF S_4 AT UHF VERSUS S_4 AT L-BAND FOR A SUBSET OF THE ANCON PASSES USED IN GENERATING THE CURVES IN FIGURE 4

The fact that S_4 does not saturate at the Rayleigh limit in the Ancon data is evidently due to our detrending operation, which is applied separately to intensity and phase. Because of the generally north-south orientation of the passes and the high degree of elongation of the equatorial irregularities, comparatively slow fading is observed at the equator.

The actual detrend filter cutoff was set at 10 s. This choice is due partly to practical considerations, since a longer detrend period could not be readily accommodated in the small, special-purpose computers that are used at the field stations. Selected equatorial passes showing strong scattering are currently being redetrended using a longer period to verify our hypothesis explaining the low saturation values of S_4.

The high-latitude data recorded at Poker Flat, Alaska, near Fairbanks, show p vs. S_4 behavior similar to the equatorial data, but with some conspicuous differences. The p vs. S_4 curves for VHF and UHF are shown in Figure 7. The VHF curve was derived from 50 representative passes; only 4 highly disturbed passes were used for the UHF curve because conditions of strong scattering at UHF are rare events.

We shall say more about the characteristics of high-latitude scintillation in Subsection II-B. Here we point out that the same tendency of the absolute value of p to decrease for large S_4 values is evident in the high-latitude data, although the extreme values fall below 2. Such deviations from theory are likely to be due to the high degree of inhomogeneity in the auroral ionosphere, particularly under disturbed conditions.

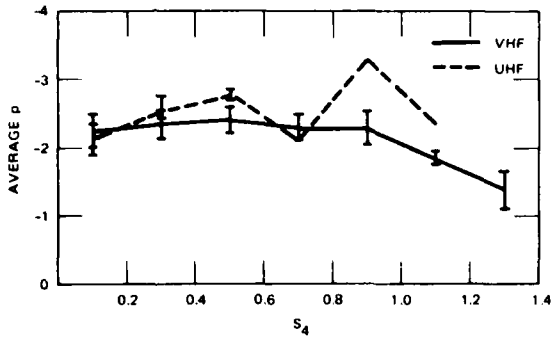


FIGURE 7 AVERAGE VALUE OF p VERSUS S_4 AS IN FIGURE 4. The VHF data are derived from 50 representative Poker Flat passes. The UHF data are derived from four disturbed Poker Flat passes.

The striking difference between the high-latitude and equatorial data is in the average value of p, which is closer to -2 than -3 for the high-latitude data. To verify the effect on the wavelength dependence of S_4 as given by (2), in Figure 8 we have plotted S_4 at VHF against S_4 at UHF. Although there is considerably more dispersion in the auroral scintillation data than in the equatorial data, there is clearly a better fit to the $\lambda^{1.25}$ curve, which corresponds to p = -2, than to the $\lambda^{1.5}$ curve.

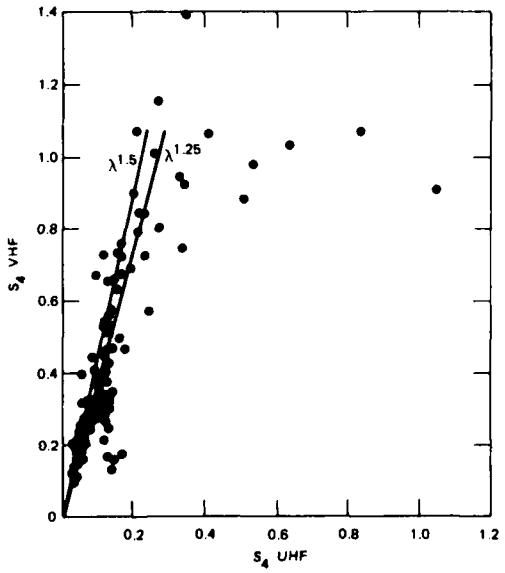


FIGURE 8 SCATTER DIAGRAM OF S_4 MEASURED AT VHF VERSUS S_4 MEASURED AT UHF FOR THE FOUR DISTURBED PASSES USED IN GENERATING THE UHF CURVE IN FIGURE 7

To summarize the general features of ionospheric scintillation, the phenomena are well characterized by a power-law spectral density function with an infinitely large outer-scale parameter. The power-law index can be deduced from the phase spectrum provided that $S_4 \lesssim 0.6$. We also note in passing that the spectral strength parameter T can be related to the spatial strength of turbulence C_s if the anisotropy and drift velocities are known. Finally, the theory can be generalized to allow for strong scattering (Rumsey, 1975).

B. Auroral Zone Scintillation

As one expects, the characteristic features of high-latitude scintillation are intimately associated with auroral phenomena. To illustrate the general features of auroral scintillation, the standard summary display for a typical disturbed pass is shown in Figure 9. The corresponding spectral index and strength parameters are shown in Figure 10.

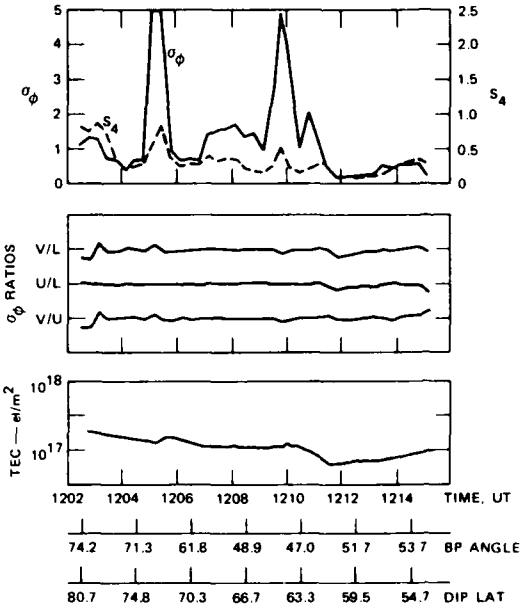


FIGURE 9 STANDARD SUMMARY DISPLAY FOR POKER FLAT PASS 6-35, 15 JULY 1976. The middle frame gives the indicated phase scintillation ratios corrected for the wavelength dependence ignoring diffraction.

The sharp drop in TEC at $\sim 63^\circ$ is the poleward boundary of the midlatitude electron density trough. Scintillation within the trough is generally weak. Poleward of the trough boundary we observe generally enhanced but highly variable phase scintillation. This is the region of the diffuse and discrete aurorae that

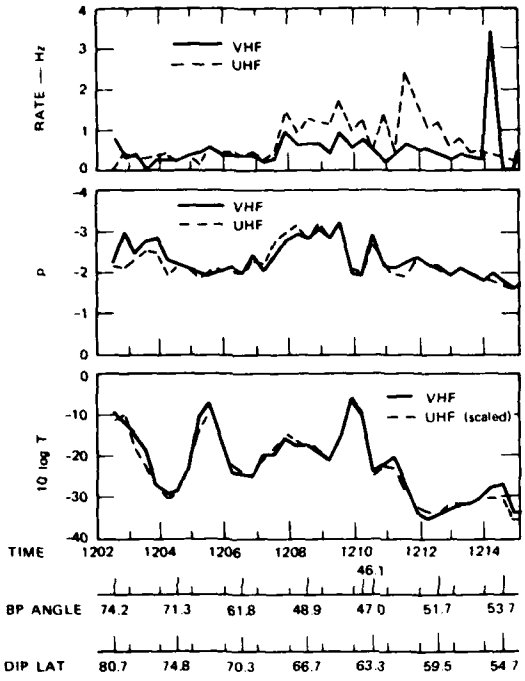


FIGURE 10 PHASE SPECTRAL SUMMARY FOR POKER FLAT PASS 6-35. The UHF T parameter is scaled to VHF by using wavelength dependence of phase squared.

characterize the instantaneous auroral oval (Akasofu, 1968).

The phase scintillation is clearly associated with the enhanced electron densities that are produced by the precipitating auroral particles. Thus, the phase scintillation follows the auroral morphology very closely, as is clearly reflected in Figure 11, where the TEC data are plotted against the L shell. The intensity scintillation by comparison is small, considering the large phase excursions that are present. Only during the pronounced phase scintillation enhancement at 72° does the VHF intensity scintillation approach saturated conditions.

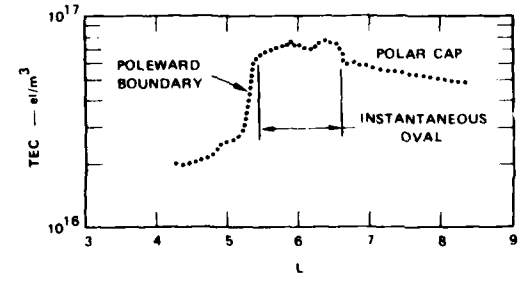


FIGURE 11 PLOT OF TEC FOR POKER FLAT PASS 6-35 CORRECTED FOR SLAB GEOMETRY AT 350 km

This event is evidently associated with an active auroral form. Kelley (1977) has suggested that velocity shears play a crucial role in generating ionospheric turbulence that can produce significant intensity scintillation. A less pronounced localized scintillation enhancement is evident in the data taken near 1210 UT, which is the point at which the propagation path to the satellite crosses the local L shell.

Such localized scintillation enhancements are prominent features in ~ 80% of the nighttime high-latitude data. We believe that the feature is a geometrical enhancement due to sheetlike structures within the auroral oval.

It is known that rodlike irregularities produce a geometrical scintillation enhancement for propagation paths that pass near the magnetic zenith. If, however, the irregularities have a high degree of coherence transverse to the magnetic field along the L shell, the enhancement occurs whenever the propagation path crosses the plane of the local L shell. In Figure 12 we show the location of this enhancement for a number of morning and evening passes. One can see the general alignment along the L shell at L = 5.5.

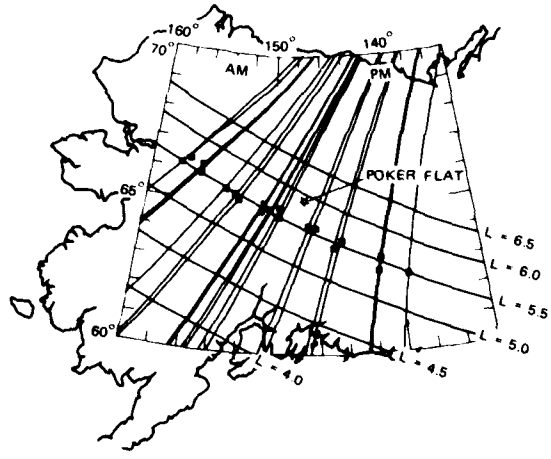


FIGURE 12 LOCATION OF GEOMETRICAL ENHANCEMENT FOR SELECTED EVENING AND MORNING PASSES

In addition to the geometrical enhancement, in ~ 20% of the passes there is a region of enhanced phase and intensity scintillation at the TEC boundary that we have associated with the poleward edge of the midlatitude plasma trough. This feature is described in Fremouw et al. (1977). In Figure 13 we show a dramatic example of this event, together with the meridional electron density distribution as measured by the Chatanika radar.

The scintillation enhancement is evidently associated with the F-region electron density

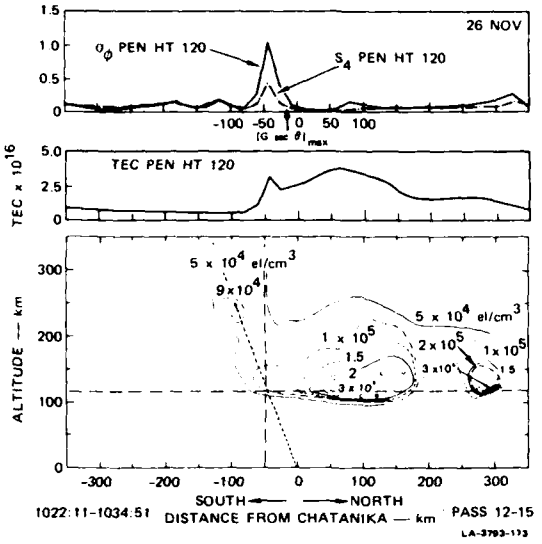


FIGURE 13 COMPARISON OF WIDEBAND SCINTILLATION AND TEC DATA FOR POKER FLAT PASS 12-15 AND CHATANIKA RADAR MERIDIONAL ELECTRON DENSITY MEASUREMENTS

enhancement above 200 km. Indeed, the TEC shows a local enhancement coincident with the passage of the raypath through the region of enhanced F-region electron densities. The extended contours peaking at 3 × 10^5 el/cm^3 are associated with the diffuse aurora. The small dark contour at 110 km altitude to the north is a discrete auroral arc.

To summarize, the auroral scintillation is highly variable and characterized by enhanced phase scintillation in the region of the instantaneous auroral oval. Discrete auroral arcs produce greatly enhanced phase scintillation, but they are only associated with enhanced VHF intensity scintillation under very active conditions. Localized scintillation enhancements that can be explained as a purely geometrical effect due to sheetlike irregularities are prominent in ~ 80% of the passes. A localized scintillation enhancement at the poleward midlatitude trough boundary is present in ~ 20% of the passes. These topics are covered in somewhat more detail in Rino et al. (1977).

C. Equatorial Scintillation

The most pronounced feature of the equatorial scintillation is its strong seasonal dependence. Scintillation activity diminishes rapidly at Ancon around April, which marks the end of the summar period in the southern hemisphere. Similarly, the scintillation activity at Kwajalein diminishes rapidly around September-October. The onset of the active period occurs around April at Kwajalein and October at Ancon.

The fading rates are typically slower for the equatorial data, as we noted in Section I. This effect can be explained in terms of the typically north-south trajectory of the passes and the high degree of elongation of the irregularities.

The occurrence of intense scintillation that persists into the gigahertz frequency range, however, is of primary interest both because of its practical ramifications for communication systems and because of the underlying physics, which are just beginning to be understood.

The occurrence of gigahertz scintillation events is most probable during the local summer active periods. This is shown in Figures 14 and 15, where we plot the times of occurrence for L-band S_4 levels greater than 0.2 at Ancon (Figure 14) and Kwajalein (Figure 15).

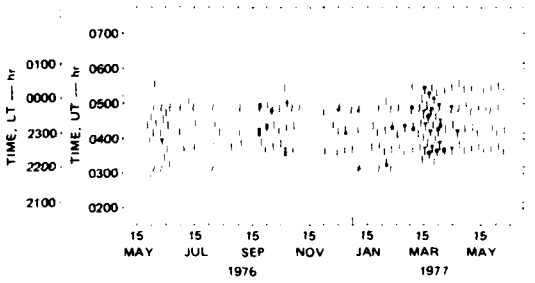


FIGURE 14 NIGHTTIME PASSES RECORDED AT ANCON. The shaded times indicate S_4 > 0.2 at 1239 MHz.

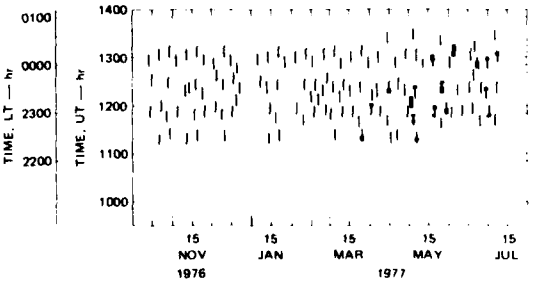


FIGURE 15 NIGHTTIME PASSES RECORDED AT KWAJALEIN. The shaded times indicate S_4 > 0.2 at 1239 MHz.

It is now generally accepted that the gigahertz scintillation events are associated with upward-propagating electron density depletions (McClure et al., 1977), which also produce enhanced VHF backscatter at their boundaries (Woodman and LaHoz, 1976). The mechanism of the "bubble" formation has been ascribed to a Rayleigh Taylor instability (Scannapieco and Ossakow, 1976; Kelley et al., 1976). The detailed mechanism that greatly enhances the small-scale irregularities, however, is not fully understood.

In previous reports (Rino et al., 1977) the flattening of the phase spectrums under conditions of strong scatter was noted as possibly due to a genuine enhancement of the small-scale irregularities. In light of the results presented in Subsection A, however, we believe that the flattening is purely a diffraction effect. Thus, the gigahertz scintillation must be accounted for, possibly in its entirety, by a pronounced enhancement in the level of turbulence, i.e., the C_s parameter in (5).

Presently there is no complete theory that links a set of initial conditions to C_s. However, theoreticians are attacking the problem (Weinstock and Williams, 1971; Sleeper and Weinstock, 1972; Rognlien and Weinstock, 1974), and simulation studies appear promising in this regard.

Acknowledgment

This research was supported by the Defense Nuclear Agency under Contract DNA001-77-C-0220.

References

Akasofu, S.-I., Polar and Magnetospheric Substorms, D. Reidel Publishing Company, Dordrecht, Holland, 1968.

Crane, R. K., "Ionospheric Scintillation," Proc. IEEE, Vol. 65, No. 2, pp. 180-199, 1977.

Fremouw, E. J., R. L. Leadabrand, R. C. Livingston, M. D. Cousins, C. L. Rino, B. C. Fair, and R. A. Long, "Early Results from the DNA Wideband Satellite Experiment--Complex-Signal Scintillation," accepted for publication, Radio Sci., 1977.

Kelley, M. C., private communication, 1977.

Kellev, M. C., G. Haerendel, H. Kappler, A. Valenzuela, B. B. Balsley, D. A. Carter, W. L. Ecklund, C. W. Carlson, B. Hausler, and R. Torbert, "Evidence for a Rayleigh-Taylor Type Instability and Upwelling of Depleted Density Regions During Equatorial Spread F," Geophys. Res. Lett., Vol. 3, No. 8, pp. 448-450, 1976.

McClure, J. P., W. B. Hanson, and J. H. Hoffman, "Plasma Bubbles and Irregularities in the Equatorial Ionosphere," J. Geophys. Res., Vol. 82, No. 19, p. 2650, 1977.

Rino, C. L. and E. J. Fremouw, "The Angle Dependence of Singly Scattered Wavefields," J. Atmos. Terr. Phys., Vol. 39, pp. 859-868, 1977.

Rino, C. L., E. J. Fremouw, R. C. Livingston, M. D. Cousins, and B. C. Fair, "Wideband Satellite Observations," Final Report, Contract DNA001-75-C-0111, DNA4399T, SRI International, Menlo Park, Calif., 1977.

Rognlien, T. D. and J. Weinstock, "Theory of the Nonlinear Spectrum of the Gradient Drift Instability in the Equatorial Electrojet," J. Geophys. Res., Vol. 79, No. 31, pp. 4733-4746, 1974.

Rumsey, V. H., "Scintillations Due to a Concentrated Layer with a Power-Law Turbulence Spectrum," Radio Sci., Vol. 10, No. 1, pp. 107-114, 1975.

Scannapieco, A. J. and S. L. Ossakow, "Nonlinear Equatorial Spread-F," Geophys. Res. Lett., Vol. 3, p. 451, 1976.

Sleeper, A. M. and J. Weinstock, "Nonlinear Theory of Density Fluctuations in Turbulent Plasmas," The Physics of Fluids, Vol. 15, No. 8, pp. 1507-1514, 1972.

Weinstock, J. and R. H. Williams, "Nonlinear Theory of Macroinstabilities and Enhanced Transport in Plasmas," The Physics of Fluids, Vol. 14, No. 7, pp. 1472-1480, 1971.

Woodman, R. F. and C. LaHoz, "Radar Observations of F Region Equatorial Irregularities," J. Geophys. Res., Vol. 81, p. 5447, 1976.

UPPER LIMIT TO THE BANDSPREADING AND FADE RATES PRODUCED BY IONOSPHERIC/ MAGNETOSPHERIC SCINTILLATIONS

Clifford W. Prettie
ESL Incorporated
Sunnyvale, California

ABSTRACT

In satellite communication system analyses of the effects of ionospheric irregularity induced fading, a parameter of importance is the maximum bandspreading of the signal to be expected. Amplitude and phase scintillations are responsible for the production of signal bandspreading; the more rapid the scintillations, the more severe the bandspreading. This paper brings to light the sometimes overlooked fact that there is a frequency dependent maximum to the signal bandspreading producible by ionospheric irregularities given a relative velocity between the medium and the propagation path. Based on angular spectrum techniques it is emphasized that the maximum spatial rate of change of a received signal is one cycle per wavelength. This limit implies a maximum bandspreading of ±100 Hz at 300 MHz for ground-based hemispherical antenna reception of synchronous satellite signals through equatorial ionospheric fading given a typical 100 meter per second medium velocity. Reduced upper limit values are found if the antenna directivity is increased or if the field of view of the irregularities from the ground-based receiver is limited. It is shown that an upper limit to the bandspreading is also possible if enough information is known about the plasma irregularities. An upper limit to amplitude fade rates is derived from the bandspread limit. A realistic upper bound to 300 MHz fading rates is 80 fades per second for omnidirectional signal reception and for a 100 meter per second diffraction pattern velocity.

INTRODUCTION

Ionospheric RF amplitude and phase scintillations caused by plasma density irregularities can adversely affect satellite communications, and for this reason, the statistics of scintillations are of interest. Two particular time statistics, signal bandspreading and average amplitude fade rate, are treated in this paper. This paper sets forth arguments to establish upper limits to their values.

Of the many assumptions underlying ionospheric plasma propagation theory, the "frozen-in" approximation is probably of greatest significance to this paper. In essence, this paper actually establishes upper limits to the spatial fade rate and spatial frequency bandspreading of ionospheric irregularity diffraction patterns. The "frozen-in" approximation allows one to ignore changes in the plasma medium with time when considering the time behavior of signals. Signal changes are assumed to be completely due to the relative movement of the propagation path plasma medium with some characteristic velocity. For ground station communications to synchronous satellites the conversion velocity which causes time variations to be observed due to diffraction pattern spatial velocities is that of the ionospheric plasma, typically, 100 m/sec. For airborne platforms or moving satellites the conversion velocity changes as appropriate for the geometry involved. Indeed, by flying an aircraft at the correct speed it is theoretically possible to completely null out fading variations, if the "frozen-in" approximation is valid. Such observations have been made in flight tests in the equatorial region by the Air Force Avionics Laboratory (24 March 1977) during AFGL's Equatorial Scintillation Campaign wherein fade rates as slow as one per minute were observed with proper adjustment of aircraft heading. The "frozen-in" approximation, thus seems sound for at least this set of equatorial data.

In the spirit of the equivalence of spatial rate and time rate for ionospheric scintillations, the next two sections will use the time related terms, such as fade rate and bandwidth, in place of spatial rate terms, such as spatial fade rate and spatial frequency bandwidth spreading. These sections discuss upper limits, but it should be remembered that upper limits require an upper limit to the spatial rate to time rate conversion factor, i.e., pattern velocity. While numbers such as 100 m/sec or 3 km/sec could be boldly suggested for this velocity, in view of the uncertainties and variances with geometry involved, the setting of an actual value of an upper limit for this number is left open.

The next section sets an upper limit to bandspreading (spatial frequency), the results of which are used to set an upper limit to fade rate in the following section. The concluding section gives numerical examples of the implications of these upper limits.

UPPER LIMITS TO THE BANDSPREADING INDUCED BY IONOSPHERIC PROPAGATION

A CW RF tone, in traversing ionospheric electron density irregularities, experiences phase perturbations attributable to varying phase velocities in the medium. Diffraction causes amplitude scintillation and additional phase sctintillations. The resulting amplitude and phase scintillations may be regarded as amplitude and phase modulation processes that spread the CW tone power over a non-zero bandwidth. This section presents the thesis that the bandwidth of a CW tone diffracted by the ionosphere is not only finite but also limited. As a consequence, the propagation caused spreading of the bandwidth of data signals transiting the ionosphere is similarly limited. Two methods of setting upper limits are suggested. The first method discussed uses the mathematical properties of the Helmholtz equation, appropriate to free space propagation outside the ionosphere, to set an upper limit that is independent of the strength of the environment. The second method using a heuristic argument based upon multiple phase screen techniques sets an upper limit depending upon how strong the ionospheric perturbations actually are.

The first method of setting an upper limit requires the noting of one fact, that, mathematically, harmonic solutions to the Helmholtz equation with wavefront variations faster than a wavelength propagate only evanescently. A brief justification can be given. Let a wave be propagating in the z-direction according to $\exp(-ik_z z)$ with wavefront variations along an x-y phase according to $\exp(-ik_x x)\exp(-ik_y y)$. The Helmholtz equation yields

$$-k_x^2 - k_y^2 - k_z^2 + k^2 = 0$$

which implies, for real k_z (and thus propagating waves),

$$k_x^2 + k_y^2 < k^2 .$$

This mathematical statement can be seen to be equivalent with the statement that variations in an arbitrary direction along a wavefront, in the above case with frequency

$$2\pi\sqrt{k_x^2 + k_y^2}$$

cycles per distance, cannot be faster than $2\pi k$, i.e., one cycle per wavelength. If these variations are faster than one cycle per wavelength, then k_z is imaginary and the wave propagation is evanescent. Thus, in a Fourier decomposition of a CW tone propagating through a disturbed ionosphere, components associated with variations smaller than a wavelength will be small because of the filtering through evanescence that occurs along the free space propagation distances that are typically involved.

The second method for establishing an upper limit is now discussed. Multiple thin phase screen methods treat the propagation effects of an irregular plasma in two steps. In the first step local phase velocity variations are lumped as a phase shift of the wavefront occurring at a representative central plane. The phase shift imposed on the wavefront is proportional to the local integrated electron content. In the second step the field is free space propagated the desired distance. The propagation performed in this second step is performed using angular spectrum techniques.

A diffracted wave at any plane can, in many instances, be decomposed into plane wave components each propagating at different angles (Stratton 1941). The coefficients of the different wave components comprise the angular spectrum of the wave. By virtue of the Helmholtz equation the angular spectrum is simply related to the Fourier transform of the field along the plane. For the two-dimensional propagation problem frequently applicable to ionospheric problems, the relation between the coefficient, $a(\theta)$, of the wave component propagating in the θ direction from a plane in an x,y coordinate system with constant y and the Fourier transform $S_y(k_x)$ of this field $E(x,y)$ is as follows:

$$a(\theta) = S_y(k_x = -k\cos\theta) = \int_{-\infty}^{\infty} E(x,y)\, e^{+i\,x\,k\cos\theta}\, dx$$

(The above expression is easily generalized to the three-dimensional form. The three dimensional relation gives the coefficient of the wave components propagating in the direction of the two polar angles, θ, ϕ, and the two dimensional Fourier transform of the field $E(x,y,z)$ at a plane of constant z.)

Note in the two-dimensional relationship that to each θ there corresponds a Fourier coefficient $k_x = k\cos\theta$. However, to each Fourier coefficient there corresponds either two angles, modulus 360 degrees, for $|k_x| < k$, or no real angles for $|k_x| > k$. The later fact reflects the mathematical property of Helmholtz wave solutions just discussed, that wavefront perturbations with sinusoidal variations with periods less than a wavelength are evanescent. (The former property, not important for this argument, reflects a degeneracy in the Helmholtz initial value problem in that waves can be propagating in either the plus or minus z direction.)

Free space propagation of a perturbed wavefront is performed by multiplying the coefficients of its Fourier transform by a (spatial) frequency dependent complex factor to find the

angular spectrum of the field at another plane. Important to this argument is the fact that the factor applied has unity magnitude for non-evanescent components. Thus, the power spectrum of the field variations is unaffected by free space propagation (a known result see Rino and Fremouw, 1973, for instance).

The propagation process can be regarded as a phase modulation process that increases signal bandwidth, followed by a propagation filtering process, that leaves the signal bandwidth unaffected (although its nonlinear phase does cause PM to AM conversion). The phase modulation of a CW tone by multiple local integrated contents is thus the only process responsible for the signal bandspreading. The bandspreading produced by multiple screens with phase shifts proportional to the local integrated electron contents is equivalent to the bandspreading produced by a phase shift proportional (and with the same proportionality constant) to the total integrated electron. Thus, knowing the integrated electron content, or less restrictively, knowing its power spectrum and that its statistics are Gaussian, one can find directly the bandspreading of a CW tone, and thus, an upper limit. Note that if this upper limit is greater than one cycle per wavelength the one cycle per wavelength upper limit will dominate.

The two upper limits just discussed do not account for antenna patterns; in essence, hemispherical coverage has been tacitly assumed. If it can be argued that wavefronts will not be received from angles beyond a certain value off beam axis, either because of antenna pattern or because of a propagation medium of limited angular extent, then a third upper limit can be set. This upper limit is independent of medium strength, similar to the first upper limit. Suppose the angular spectrum wave components with large angles of arrival, θ, such that $|\theta| > \theta_1$ are not received, then the CW tone components received will cause the tone to be spread less than $2\pi k \cos\theta$ cycles per distance unit away from its center frequency. This is the form applicable as a third upper limit. The $\cos\theta$ factor may be significant for high gain antennas.

UPPER LIMIT TO THE FADE RATE INDUCED BY IONOSPHERIC PROPAGATION

The previous section established an upper limit to the bandspreading producible by ionospheric irregularities, which, in itself, fulfills the requirements of many fast fading systems performance analyses. Another parameter that may be of interest is the fade rate below a given level. That there should be a relationship between an upper limit to the fading rate of a CW signal and its bandspread seems intuitively clear, for when the fade rate of a CW signal is fast, the bandwidth of its energy is expected to be correspondingly large. The results of S.O. Rice (S.O. Rice 1948), which gives a relationship between the fade rate of the envelope of a sine wave plus Gaussian noise in terms of the bandwidth of the noise process, are applied to the ionospheric fading problem. With some arm-waving, a fade rate upper limit is produced.

Rice's relationship gives the average rate, R that the envelope amplitude of a sinewave plus noise crosses a level, v, with negative slope, as a product of three terms: a constant, the rms bandwidth of the noise, f, and the amplitude probability density at v, p(v); i.e.,

$$R = (2\pi)^{1/2} f\, p(v)$$

where

$$f^2 = \int_0^\infty (x-f_c)^2\, w(x)\, dx \;/\; \int_0^\infty w(x)\, dx$$

and where f_c is the center frequency of the noise power spectrum w(x). For strongly diffracting ionospheric mediums this expression is useful for the specific case of Rayleigh fading in which the sine wave power is zero and the probability density is given as $p(v)=v\exp(-v^2/2)$. The above expression then gives $R=2\sqrt{\pi}\exp(-1)f = 1.304f$ as the Rayleigh fading rate below the amplitude characteristic of the average power level, $v=\sqrt{2}$. The expression also gives the rate of fades below other arbitrary levels with appropriate modification.

It is now suggested that the above expression for R gives an upper limit to the fade rate produced by ionospheric fading either with f chosen as the bandspread limit or with f calculated from the signal bandspreading power spectrum, if it is known. It is recognized that there are differences between the statistics of Gaussian bandpass noise and most diffracted waves in the nature of their amplitude and phase characteristics. A conceptual model of the development of fading sheds some light on these differences and why the fade rate relationship of Rice is applicable. Near strong ionospheric irregularities most of the plasma propagation effects are manifested as wavefront phase modulation, the amplitude fade rate being slow, with rates similar to the phase decorrelation rate. As a perturbed wavefront moves from the near field, strong amplitude fluctuations, with close to Rayleigh distribution develop. The fade rate increases, however, much of the signal bandspreading is still due to the wavefront phase modulation. As the wavefront approaches the far field the amplitude fade rate continues to increase but not indefinitely. The signal statistics approach Rayleigh (or Rician for weaker disturbances) more and more closely and the bandspreading of the signal approaches being "equally" spread by both the amplitude and phase modulation. It is suggested that in this far field Rayleigh limit, a diffracted wave is equivalent to bandlimited noise (bandlimited noise plus a sine-wave for Rician statistics) because of the "randomization" of amplitude and phase induced by the strongly dispersive far-field propagation filtering. The relationship

between fade rate and signal bandspreading finds application in this limit and, thus, an upper limit to the fading rate is obtainable. It is also suggested (again without proof) that because of the gradual PM to AM conversion of the propagation process that the far field limit Rayleigh fade rate provides an upper bound to the fade rate at all far and near field distances with amplitude distributions resembling Rayleigh.

RESULTS AND CONCLUSIONS

The previous sections have presented arguments for upper limits to signal bandspreading and fade rates for signals propagating through ionospheric plasma irregularities. Essential to the arguments establishing the limits is that the "frozen-in" assumption be fairly closely met. These limits imply the following numerical results if a 100 meter per second "frozen-in" pattern velocity is assumed:

1. The bandspreading limit of one cycle per wavelength implies a maximum spread of $\pm$100 Hz for 300 MHz signals and $\pm$1 kHz for 3 GHz signals. Note that while the upper limit of bandspreading at the 3 GHz frequency is higher, it requires a very much more intense environment to drive the fading to this upper limit value. These results are applicable to signals received by an omnidirectional antenna.

2. The bandspreading limit of 3 GHz signal received with a 3 degree (half power full width) beamwidth is significantly less than that received with an omnidirectional antenna. The bandspread limit is 1 kHz sin (1.5 degrees) = 26 Hz.

3. Signal bandspreading may be definable given an rms integrated electron content and the integrated electron content power spectrum. Suppose a Gaussian integrated electron content power spectrum is given with e-fold at 1 cycle per kilometer. Suppose also that the rms integrated content is bounded and such that it produces at most a 5000 degree rms phase shift of a 300 MHz signal; ignoring diffractive phase effects. Using Abramson's result (Abramson, 1963) that the rms bandwidth of a phase modulated wave is the product of the rms bandwidth of the phase modulation's power spectrum and the rms phase modulation in radians, an upper limit to the rms bandspread can be found. The rms bandwidth of the modulation is 0.7 cycles per kilometer or 0.07 Hz. The rms angle modulation is 5000 degrees or 87 radians giving a maximum rms 300 MHz signal bandspread of $\pm$3 Hz, a value significantly lower than the $\pm$100 Hz limit calculated for the general case in 1.

4. The Rayleigh fade rate below the average power level has a maximum value of 130 fades per second (fps) for a 300 MHz signal received with an omnidirectional and 1.3 fps for a 3 GHz signal. Again it requires a much stronger environment to produce fade rates close to the 3 GHz upper limits.

5. The average Rayleigh fade rate upper limit for fades 20 dB down from the average power level is a factor of 3.7 ($=p(\sqrt{2})/p(\sqrt{2}/10)$) down from the fade rate upper limit for fades below the average power level.

6. The maximum average Rayleigh fade rate for the 5000 degree 300 MHz environment is 4 fps.

REFERENCES

1. N. Abramson, "Bandwidth and Spectra of Phase and Frequency Modulated Waves," IEEE Transactions on Communications Systems, December 1963, p.409.

2. S.O. Rice, "Statistical Properties of a Sine Wave Plus Random Noise," Bell Systems Technical Journal, Vol. 27, January 1948, pp.125-126.

3. C.L. Rino and E.J. Fremouw, "Statistics for Ionospherically Diffracted VHF/UHF Signals," Radio Science Vol. 8, #3, March 1973, p.225.

4. J.A. Stratton, Electromagnetic Theory, New York, McGraw-Hill, 1941, Sec. 6.7 .

ACKNOWLEDGEMENT

The above work was sponsored in part by the Defense Nuclear Agency under Contract # DNA 001-76-C-0152.

EQUATORIAL IONOSPHERIC SCINTILLATIONS IN THE INDIAN ZONE

R.G. RASTOGI, M.R. DESHPANDE and HARI OM VATS
Physical Research Laboratory
Ahmedabad 380009
India.

TEXT

The phenomenon of scintillation is now well known as being the occasional changes in the amplitude, phase and angle of arrival of radio waves propagating through the ionosphere. These changes are known to be due to ionization irregularities present in the ionosphere. Earlier studies showed that these changes are more severe in equatorial and two auroral zones. The equatorial scintillation is basically a nighttime phenomenon with its maximum before midnight for any longitude and is largely due to spread F irregularities (Koster 1972, Mullen 1973, Chandra and Rastogi 1974). At high latitudes some scintillations are observed during daytime hours (Munro 1966, Frihagen 1971, Nielson and Aarons 1974). Recently Deshpande et al. (1977) reported that during a magnetic storm scintillations are found to be caused by the F region irregularities in the early hours of the day i.e., at post sunrise. McClure (1964) estimated the height of the irregularities responsible for the daytime scintillations to be embedded in the E region. Association of scintillations and Es patches has been noted at middle and low latitudes (Aarons and Whitney 1968, Rastogi and Iyer 1976).

Fig.1 shows some of the examples of the amplitude records of radio beacons (40, 140 and 360 MHz) from ATS-6 recorded at Ootacamund and simultaneous ionograms from a nearby equatorial station Kodaikanal. Note that whenever there is Es-q present in the ionograms the amplitudes do not show appreciable variation, whereas during Es-h type reflections severe scintillations have been observed. Chandra et al. (1977) have shown that daytime scintillations are caused by each of the irregularities responsible for Es-b, Es-h, Es-l, M-reflections and G-layer and never by Es-q irregularities. Similarly nighttime scintillations are caused either by spread F irregularities or by Es irregularities (Rastogi et al. 1977, Chandra et al. 1977).

Fig.2 shows the diurnal occurrence of ionospheric scintillations at 40, 140 and 360 MHz during two periods, i.e. Oct-Nov 1975 and May 1976. It can be noticed that during Oct-Nov 1975 there is a major peak occurrence of scintillations at night between 2000 hr and 2300 hr and two minor peaks during daytime around 0900 hr and 1400 hr. On the contrary May 1976 occurrence has a broader maximum during daytime than at night. Although at 40 MHz there seems to be appreciable scintillations at all times with a slight decrease in percentage around sunset and sunrise. The difference of daytime and nighttime occurrence of scintillations is quite clear at 140 and 360 MHz. This broader maximum during daytime in a summer month (May) is probably because of frequent occurrence of non q type of Es irregularities. A detailed study of this in comparison with ionosonde data seems to be necessary.

Acknowledgements

Authors are thankful to Dr.Kenneth Davies and his group for collaboration and to Professor G.Swarup for facilities for ATS-6 RBE project at Ootacamund. This project was partially supported by DOS and NASA.

References

Aarons J. and Whitney H.E., Planet. Space Sci., 16, 21 (1968).

Chandra H. and Rastogi R.G., Curr. Sci. 43, 567 (1974).
Chandra H. et al., Ann. Geophys.(1977) (in press).
Deshpande M.R. et al., Nature, 268, 6 614 (1977).
Frihagen J., J. Atmos. Terr. Phys., 33, 21 (1971).
Koster J.R., Planet. Space Sci., 20, 1999 (1972).
McClure J.P., J. Geophys. Res., 69, 2774 (1964).
Mullen J.P., J. Atmos. Terr. Phys., 35, 1187 (1973).
Munro H., Rad. Sci., 1, 1186 (1966).
Nielson E. and Aarons J., J. Atmos. Terr. Phys., 36, 159 (1974).
Rastogi R.G. and Iyer K.N., Curr. Sci., 45, 685 (1976).
Rastogi R.G. et al., Pramana, 8, 1 (1977).

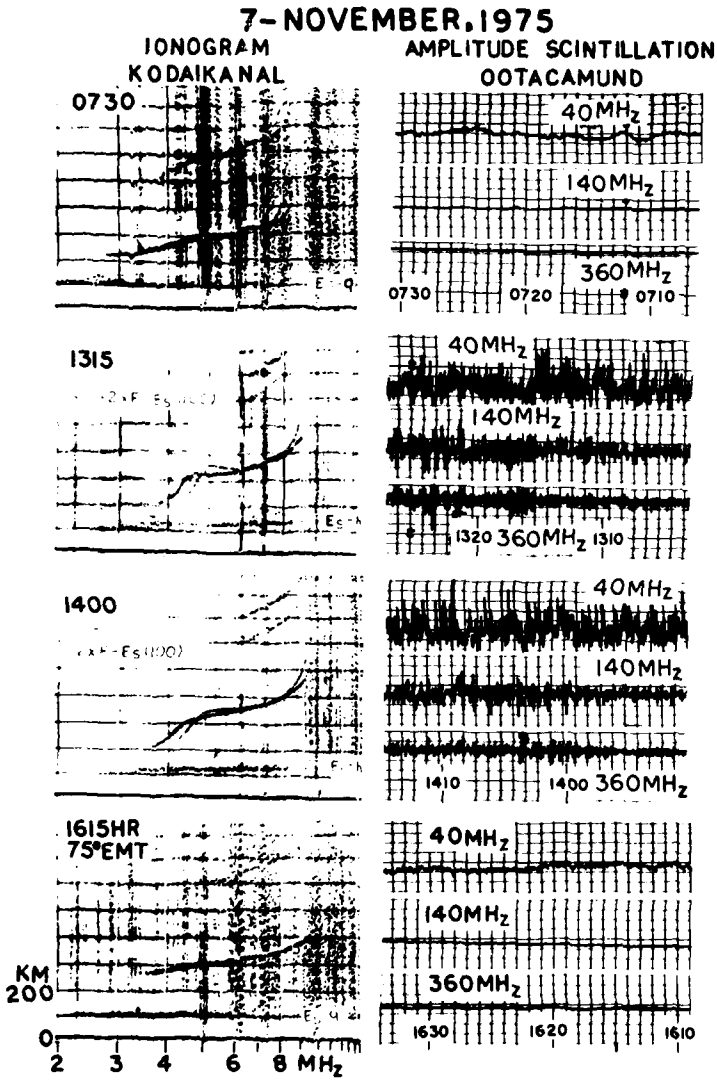


Figure 1

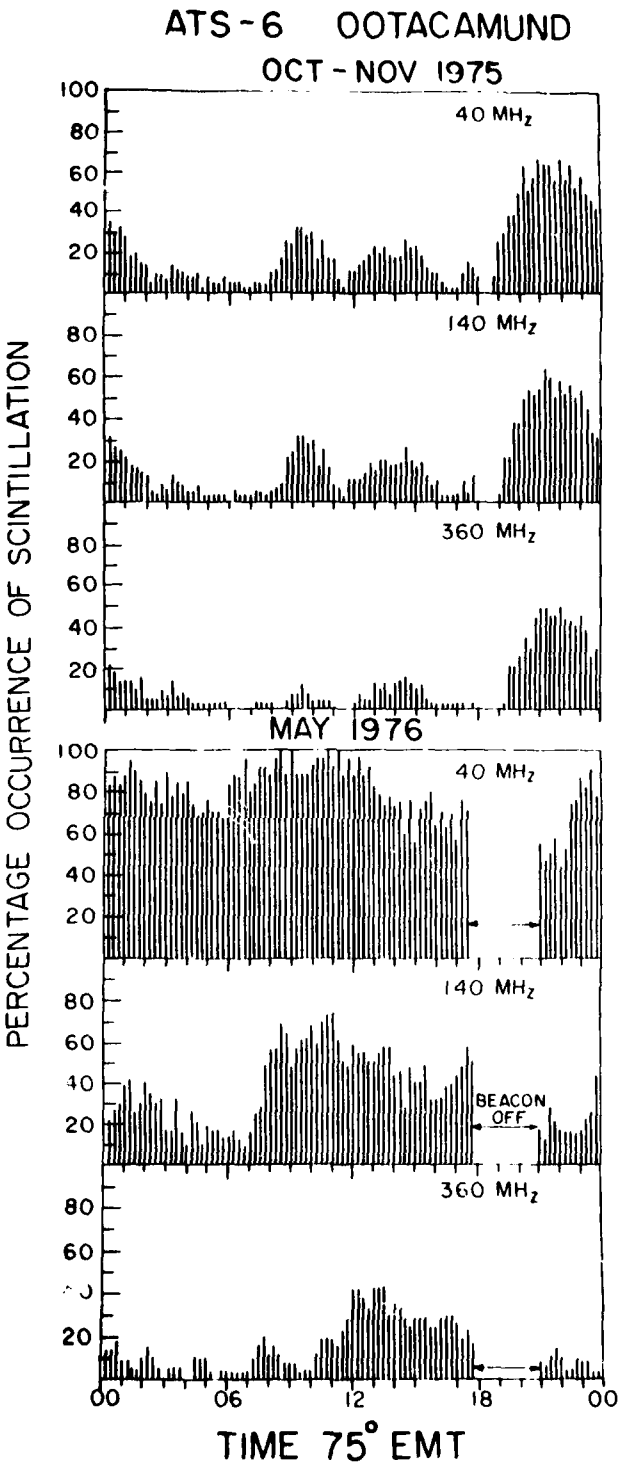

ATS-6 OOTACAMUND
OCT - NOV 1975
40 MHz
140 MHz
360 MHz
MAY 1976
40 MHz
140 MHz
BEACON OFF
360 MHz
PERCENTAGE OCCURRENCE OF SCINTILLATION
100
80
60
40
20
00
06
12
18
00
TIME 75° EMT

Figure 2

AMPLITUDE SCINTILLATION STUDIES IN THE POLAR REGION ON 250 MHz

By

J. Buchau, J. Aarons, J.P. Mullen, E.J. Weber
J.A. Whalen and H.E. Whitney
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

And

E.E. Crampton, Jr.
MITRE Corporation
Bedford, MA 01731

ABSTRACT

In April, July and September 1975 twelve C-141 flights (28 hours of data) were conducted to measure amplitude scintillations on VHF satellite transmissions in the polar region, under auroral conditions. A second aircraft, a KC-135, instrumented for ionospheric/auroral research, monitored the conditions of the auroral ionosphere in the ray path to relate scintillation events to auroral conditions in the April and September tests. Simultaneous ground measurements were made at Thule, Greenland; Goose Bay, Labrador, Canada; and Sagamore Hill, Massachusetts. For the first time in the polar region satellite 250 MHz transmissions could be monitored for an extended period (up to 4 hours) at high elevation angles.

The observed aircraft scintillations were separated into groups, describing the ionospheric situation along the ray path (trough, oval, or polar cap). The spatial distribution of the occurrence probability of scintillations indicates, that scintillations are more likely to occur in the northern half of the auroral oval (29% of the time) and in the polar cap (21%) than in the region of continuous (diffuse) aurora (17%). Scintillations were almost completely absent in the F-layer trough (4%). Even though aurora was present throughout most of the flights (with Kp ranging from 0+ to 4o) only weak to moderate scintillations were observed (mostly <9 dB).

Over 200 hours (34 during flights) of observations were made at each of three ground stations: Thule (polar cap), Goose Bay (auroral oval), and Sagamore Hill (mid-latitude ionosphere). For any distinct time period of simultaneous observations, there was considerable scatter of which station recorded maximum activity. During magnetically quiet periods (Kp = 0-3) the ground data show good latitudinal ordering. The highest probability of occurrence of scintillations and the strongest scintillations are found in the polar cap. During periods of moderate magnetic activity (Kp = 4-6) the probability of scintillations strongly increases at Goose, at times above the Thule level, which is not significantly changed by the increase in magnetic activity. In general, Sagamore Hill for moderate magnetic activity showed lower scintillation levels than either Thule or Goose Bay.

INTRODUCTION

In April, July and September 1975, a comprehensive program (Figure 1) was conducted to assess ionospheric and auroral conditions during AFSATCOM scintillation test flights and to determine the effects of the aurora on satellite communications. In this program coordinated flights were made by an Air Force 4950th Test Wing C-141, carrying an AFSATCOM communications terminal and by the AFGL Airborne Ionospheric Observatory, an NKC-135 instrumented for ionospheric and auroral research and VHF field-strength measurements. The Sagamore Hill Observatory at Hamilton, Massachusetts, the Goose Bay Ionospheric Observatory at Goose Bay, Labrador and the Geopole Observatory at Thule, Greenland, all operated by AFGL and equipped for routine satellite observations, enhanced their observation programs during the test period.

The AFSATCOM satellite, in a polar orbit and transmitting at 250 MHz, provided the signal source for the scintillation studies. These experiments resulted in a body of data that allowed the determination of the effects that auroral ionization and the high latitude ionosphere have on trans-ionospheric VHF signal propagation at relatively high elevation angles.

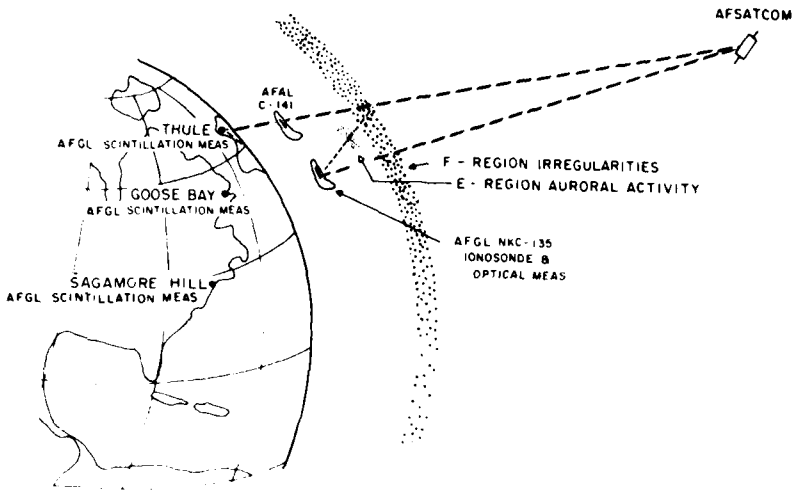


Figure 1. 250 MHz Scintillation Studies during April, July and September 1975 Involved Two USAF Research/Test Aircraft and Three AF Geophysics Laboratory Ground Stations.

The goal of these experiments was to determine the relation between ionospheric and auroral conditions and VHF scintillations, and to compare the results of the airborne experiments with results from the considerably more extensive ground observations.

AIRBORNE EXPERIMENTS

The purpose of the AFGL aircraft flights was to monitor the ionospheric environment in the region where the VHF signal path from satellite to receiver (C-141 aircraft) traversed the ionosphere. To enhance the scintillation data base, a satellite receive capability was installed in the AFGL aircraft for the September 1975 missions.

The high latitude ionospheric/auroral environment can be separated into 3 distinctly different regions, the F-layer trough region, the auroral oval and the polar cap (see Figure 2). The auroral oval is the location of a band of auroral forms around the magnetic pole, which separates the two other regions and provides the ordering frame for various ionospheric and auroral phenomena. VHF scintillations have been observed over the entire polar cap, in the auroral oval and at subauroral latitudes down to the scintillation boundary (Aarons and Allen, 1971; Aarons, 1973), also shown in Figure 2. It was the goal of the airborne studies to investigate the effects of irregularities in the different regions on the VHF signals. A flight track, shown in Figure 2, was chosen which permitted the consecutive scanning of F-layer trough, oval and polar cap regions in a single flight and also allowed for the known variability of the oval size. Since during the April missions only weak scintillations were observed in the trough the flights in September focused only on the study of oval and polar cap effects. Figure 3 shows schematically a cross-section of the ionospheric structures along the flight track and identifies the five possible ray path configurations. The path can encounter pure trough-, oval- or polar cap conditions and in addition two transition regions, where E-layer effects of one region and F-region effects of another jointly affect the ray path. Data were grouped

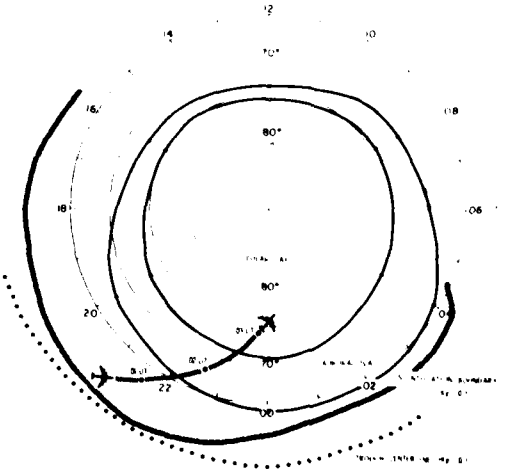


Figure 2. Corrected Geomagnetic/Local Time View of the Test Aircraft Flight Track in Relation to the Q=3 Auroral Oval, the Scintillation Boundary, the Mid-Latitude F-layer Trough, and the Polar Cap.

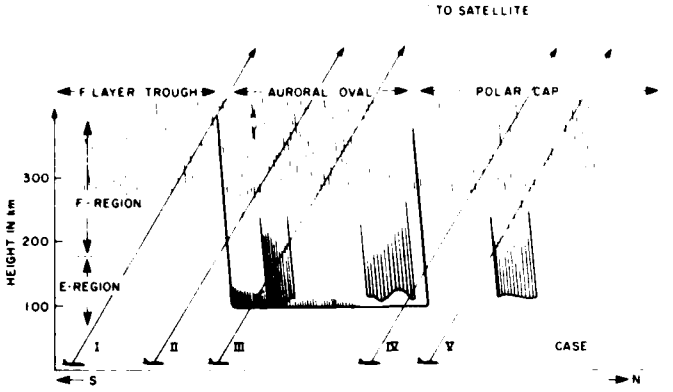


Figure 3. Schematic Cross-Section Through the High-Latitude Ionosphere/Auroral Structure along the Flight Track. The data are grouped according to the propagation conditions encountered: Trough (I), Oval (III), Polar Cap (V), and two Transition Regions (II and IV).

according to these five cases to assess the dependence of scintillations of the various morphological regions.

The Airborne Ionospheric Observatory's instrumentation consisting of ionospheric sounders, all sky cameras, photometers and spectrometers (Gowell and Whidden, 1968), was used to establish the location of the major geophysical boundaries previously described, with respect to the satellite ray path and also to assess the specific conditions, such as presence of auroral forms, in the ray path. These efforts were enhanced by vertical and backscatter soundings at the Goose Bay Ionospheric Observatory.

Figure 4 shows a typical flight track, along the 60°W meridian, flown by the C-141 carrying the satellite communications receiver. The majority of all flights discussed here followed this track. All September flights were made only to the North of Goose Bay concentrating on oval and polar cap studies.

In addition to the flight track the figure shows the observed scintillations given as scintillation index (SI) in dB (Whitney et. al., 1969). The scintillation index was determined for each one-minute interval. Following the international convention the ground data are analyzed using 15 minutes as time base for the data samples. The more dense determination of SI of the airborne data is required since geophysical conditions change rapidly due to the aircraft motion. The SI (in dB) is obtained by measuring the dB difference between the third peak down from the highest and the third fade up from the minimum within each time sample.

The satellite was in a northeasterly direction from the aircraft during the April flight tests and in a northeasterly direction in September. As Figure 4 (data taken on 23 April 1975) shows, scintillations tended to be enhanced when the aircraft was over water. Small land masses close by and in the direction to the satellite also reduced the scintillation level

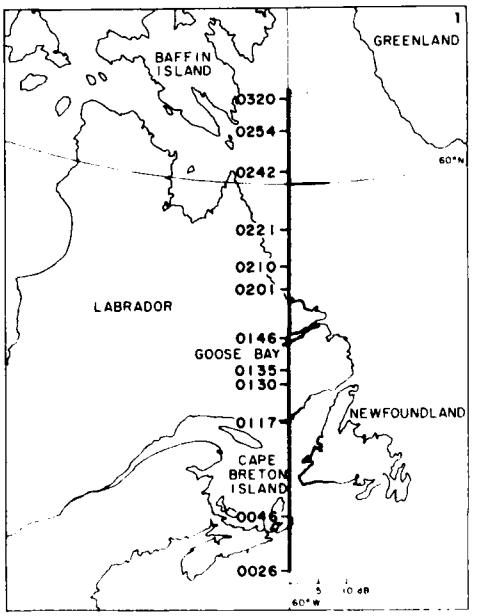


Figure 4. Flight Track of 23 April 1975. Times in UT. Scintillation Index (in dB) is plotted to the right of the track. Scintillations are enhanced over water due to ground reflected propagation mode.

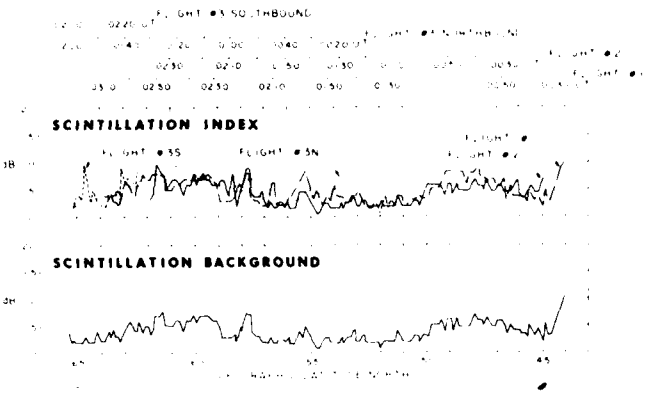


Figure 5. Superposition of Scintillation Data from Flights in April 1975 vs Geographic Latitude (upper panel). Minimum Scintillations (lower panel) derived from the superposed curves.

when the aircraft was over water. Land close to the flight track but not within the azimuth to the satellite did not affect the scintillation levels.

Figure 5 shows in its upper panel a superposition of the scintillation index determined during three April missions along this track. The consistency of regions of enhancement and decrease is evident and correlates well with presence of water and land below the aircraft. It was concluded, that a ground reflected signal, enhanced over water by the better reflectivity of water, added to the ionospherically produced scintillations. In order to isolate periods of ionospheric scintillations and to determine their magnitude, the ground effect had to be subtracted. The minimum observed signal fluctuations as a function of latitude were determined and are shown as the ground effect or scintillation background curve in the lower panel of Figure 5. Deviations on each individual flight from this background curve were determined minute by minute, using the time-reference scales shown in the upper part of Figure 5, and were plotted versus UT for correlation with the geophysical data. Figure 6 gives as an example of the differences of flight data and background for the 23 April 75 mission. An arbitrary 2 dB noise level was introduced and only signal fluctuations above this level were finally considerd to be clearly of ionospheric origin and were used for the analysis.

The scintillation data determined from the September 1975 C-141 flights showed similar behavior and thus were treated identically. The AFGL aircraft was equipped for the September missions with a Dorne Margolin VHF/UHF wide band antenna designed to reject ground reflected propagation modes and the data resulting from this instrumentation was free of a background-component.

The Table shown in Figure 7 gives the % of occurrence of scintillations by region for each flight which scanned through at least 3 adjacent regions, transition regions included. Even though the average conditions as indicated in the average column giving % of the total observing time during each mission that the signal was scintillated, changed over a wide range (6% to 71%), the individual flights show (with the exception of Flight 1) a clear similarity. The highest percentage of scintillations is found in region IV the transition between oval and polar cap. This result is also seen in the data averaged by expedition and shown in the top section of Figure 7. The graph shows the % of occurrence of scintillations by region, derived separately for the April 75 and the September 75 expeditions. The occurrence in the trough is a low 4% (small data base) and a moderate 14% in the auroral oval region, where both discrete and continuous auroras re generally present. (Discrete auroras are the well known curtains, folds and rays, visible to the eye. Continuous auroras is a more uniform glow, generally just above

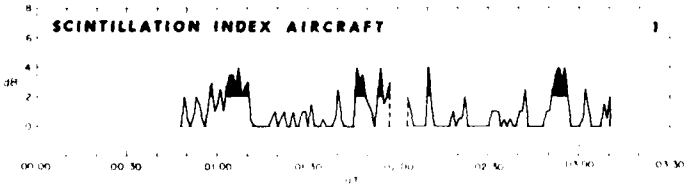
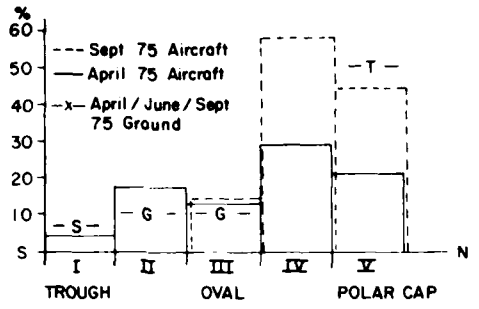


Figure 6. Example of Corrected Aircraft Scintillation Index vs UT for 23 April 1975 Flight. The black regions indicate scintillation peaks >2 dB.



INDIVIDUAL FLIGHTS

FLIGHT NO.	Regions I	II	III	IV	V	Average
1	0	26	12	23	—	15
2	7	0	7	10	—	6
3			23	26	21	23
4			4	38	—	22
5			33	100	87	71
6			19	63	16	30
7			2	54	27	18

Figure 7. Occurrence of Scintillations in % by Region Given Separately for the April and September Expedition (Top Panel) and by Flight (Bottom Table). Ground Results from the same period, measured at Sagamore Hill (S), Goose Bay (G) and Thule (T) are shown for comparison.

the level of visibility, *in which discrete forms* may be embedded.) The poleward edge of the oval belt, characterized by discrete auroras in the absence of continuous aurora, is the most probable region of scintillations (29% in April and 58% in September). Polar cap measurements show 21 and 44 percent scintillations respectively.

The conclusions on the relation between the occurrence of substantial scintillations and discrete auroral forms is summarized in Figure 8. This generalization is based on the somewhat

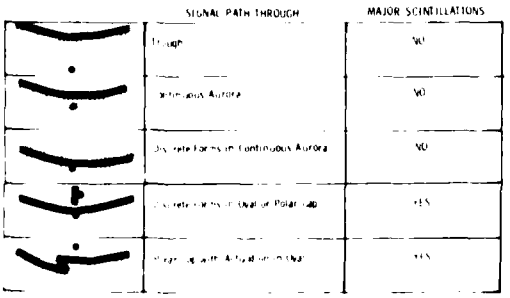


Figure 8. Schematic Representation of the Signal Ray Path (dot) and the Auroral/Ionospheric Features Leading to Scintillations. Shaded: Continuous Aurora. Heavy Lines: Discrete Auroral Forms.

subjective description of the nature of auroral forms by observers, onboard the AFGL aircraft, who made continuous visual observations in the satellite azimuth, and is supported by the statistical data shown in Figure 7. The only auroras definitely associated with high SI are the auroral arcs at the high-latitude edge of the auroral oval - hence at the oval-polar cap transition. Definitely not associated with high SI are the continuous aurora and discrete auroras embedded within the continuous aurora - even though this type of aurora was at times active and bright. The highest SI has been observed in the polar cap, even though the observed probability of scintillations is lower than in the oval/polar cap transition region.

GROUND BASED EXPERIMENTS

The limited amount of airborne measurements served to complement the ground based observations which were made throughout the test periods and resulted in a 34 hour data base from each station, compared to a total of 19.5 hours of data taken by both aircraft in the various regions. The recording equipment used was similar to the equipment used onboard the Airborne Ionospheric Observatory and consisted of Spectrum International Converters and R-390/URR receivers used as IF amplifiers. Signal strength was recorded both on strip charts and analog magnetic tapes. 10-element yagis with a gain of 12 dB and a 45° horizontal/58° vertical beamwidth were used. As an example of the results obtained during the flight tests the results of the ground observations for September 75 are shown in Figure 9. Intense scintillations were observed initially (5 dB on 1 September, 8 dB on 2 September), however the SI did not exceed 4 dB for the remainder of the measurements. Thule generally showed stronger scintillations than Goose Bay. Sagamore Hill except for 1 September is completely quiet. This behavior is typical for magnetically quiet (Kp = 0 to 3) conditions as will be shown later in the discussion of results from the complete ground data base. Inspection of individual days shows that the behavior of the scintillations is far from consistent. There are instances in the test series wherein the highest SI can be found at any of the three stations.

The levels of occurrence of scintillations >3 dB (in percent) at each of the three ground stations Sagamore Hill (S), Goose Bay (G) and Thule (T) determined as averages over the three test periods, April, July and September, are indicated for comparison with the airborne data in the bar graph in Figure 7. Results from each station are indicated by their respective symbols. They show general qualitative agreement with the airborne measurements considering the fact, the Sagamore Hill encounters trough-like conditions only occasionally, while Goose Bay's situation may vary from experiencing trough conditions to pure oval conditions depending on the time of day of the measurements.

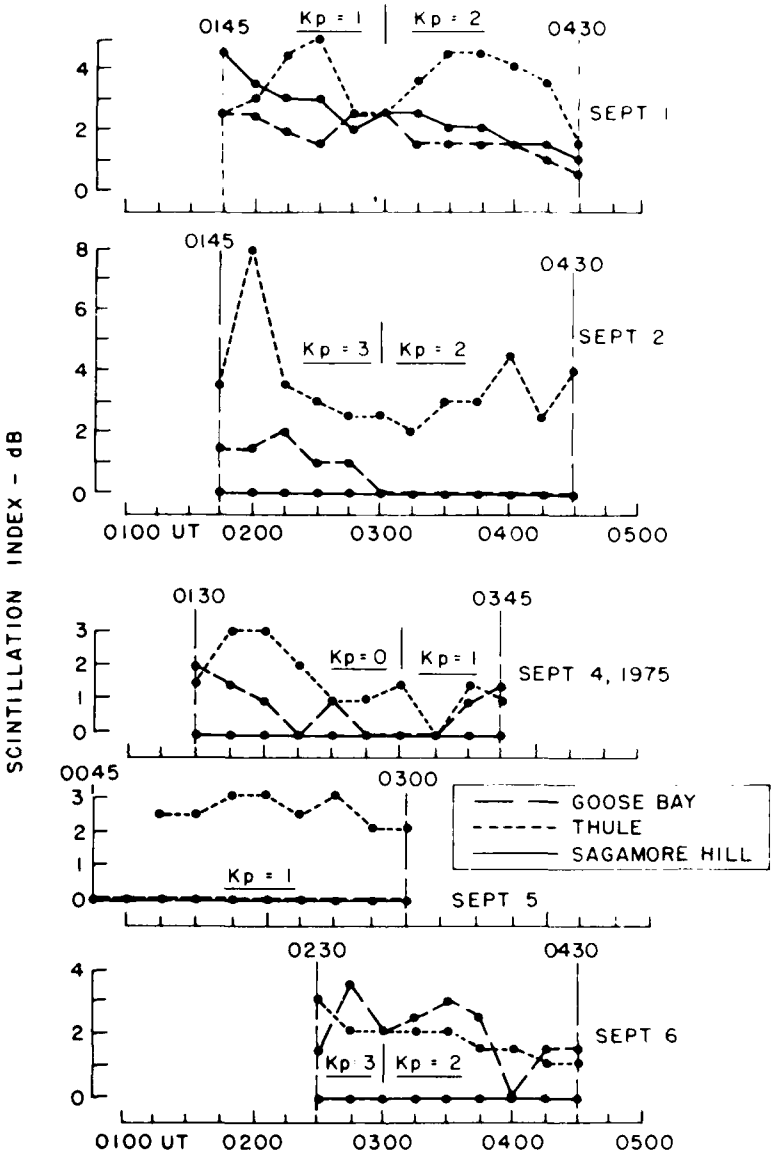


Figure 9. Ground Observed Scintillation as Seen at Goose Bay, Labrador, Thule, Greenland, and Sagamore Hill, Mass., During Tests of 1-6 September 1975.

A more consistent picture evolves if one considers the complete ground data base available. Besides transmissions during the airborne campaign there were many transmissions outside the flight schedule and every transmission was recorded at all three stations. This data base not only allows determination of the latitudinal dependence, but it is also sufficiently large to permit the assessment of the influence of magnetic activity on the occurrence of scintillations.

Figure 10 gives for the three stations the dependence of all scintillations, SI >2 dB, and of more substantial scintillations, SI >4 dB, on magnetic activity. The data are grouped for Kp ranges of 0 to 3 for magnetically quiet and 4 to 6 for moderately disturbed conditions and by 4 hour time intervals centered on the local times indicated on the ordinates. The data base for Kp = 4 to 6 in the time interval centered on 04 LMT was not large enough to be statistically significant and no value for the percentage

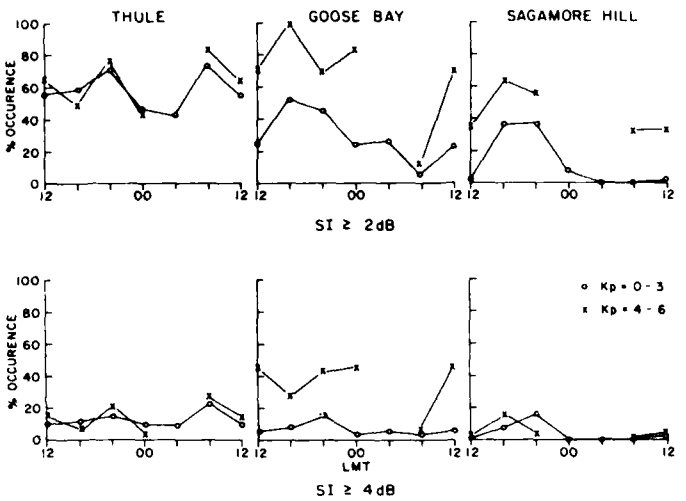


Figure 10. Scintillation Data Base April-September 1975. Grouped for magnetically quiet and moderately disturbed conditions, they show strong magnetic control at Goose Bay, moderate magnetic control at Sagamore Hill and independence at Thule.

occurrence was established. The graphs show a clear dependence of occurrence of scintillations with SI >2 dB on the magnetic activity for Sagamore Hill and Goose Bay, with the occurrence increasing by approximately 50% at Sagamore Hill and by 100% at Goose Bay. The local time dependence shows a well defined maximum in the afternoon to evening sector, which is evidence of the time asymmetry of the scintillation boundary shown in Figure 2. This result increases the confidence in the statistical significance of the accumulated data base. The Thule data show complete lack of magnetic control and a rather flat occurrence versus time distribution. Only a weak minimum in the post midnight sector is indicated. If one considers only stronger scintillations, SI >4 dB, the graphs show no evidence for magnetic control at Sagamore Hill and Thule, but a significant increase at Goose Bay. This may be the result of the oval being enlarged during Kp = 4 to 6 conditions bringing Goose Bay more often under the influence of discrete aurora. The continuous aurora with its lesser effects on VHF signals may under these conditions be found to the south of the station. Goose Bay and Sagamore Hill again show evidence of the afternoon to evening maximum with Thule again showing no temporal variability for the larger scintillations.

The latitudinal dependence of scintillations as a function of Kp is shown in Figure 11, a replot of the data shown in the previous figure. Scintillations with SI >2 dB show under quiet magnetic conditions a clear south to north increase in occurrence, a trend also evident though not quite as clear, if only scintillations with SI >4 dB are considered. Under more disturbed magnetic conditions the auroral oval station Goose Bay shows a strong enhancement, reaching 100% occurrence in the afternnon sector, while the scintillations at Sagamore Hill reach approximately the generally unchanged Thule levels except for a morning maximum. The same behavior is shown if only stronger scintillations with SI $\geq$4dB are considered.

CONCLUSIONS

Airborne and ground based scintillation measurements on 250 MHz signals conducted in 1975, a year of very low solar activity (smoothed sunspot numbers varied from a high of 18.6 in April to a low of 14.5 in September 1975), showed only small to moderate scintillations at subauroral, auroral and polar cap latitudes. Scintillations observed by the ground-stations rarely exceeded 10 dB and generally were in the order of 4 dB. The ground data base showed under quiet conditions good latitudinal ordering, the highest probability and the strongest SI was found in the polar cap. During magnetic disturbances, which were only moderate during the test period, Goose Bay scintillation occurrence increased strongly, and moderately at Sagamore Hill, with both stations showing large afternoon maxima in the occurrence distribution under all conditions. The Thule data show neither magnetic nor local time dependence. The aircraft data confirm that polar cap scintillations under magnetically quiet to moderately disturbed conditions are more prevailing than those in the oval. They do not corroborate the strong enhancement seen under magnetically disturbed conditions at Goose Bay, most likely since the aircraft data collected in the oval under disturbed conditions fell with Kp = 4o and 4- just above the lower limit of the disturbance criterion, encompassed

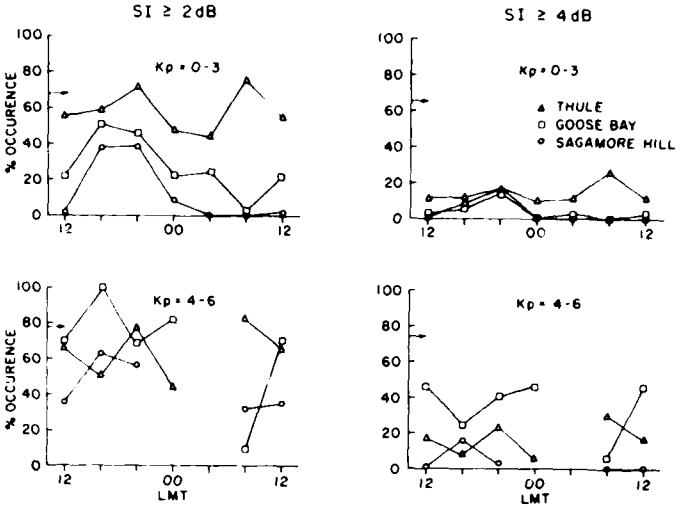


Figure 11. Scintillation Data Base April-September 1975. Presentation shows good latitudinal ordering for quiet magnetic conditions.

less than 3 hours of data and were not taken in the most disturbed 14-18 LT sector. The scintillation amplitude probability distributions determined from aircraft data for the various regions show good agreements with those determined from the ground data base, validating generalizations derived from the small airborne data base. The following are the major findings from the airborne data: The strongest scintillations (>9 dB) were found in the polar cap. During airborne measurements in the oval the highest occurrence of scintillations was consistently found in the poleward half of the auroral oval, at least twice as much as found in the southern half of the oval. Since no station has previously operated in a suitable location, this enhancement so far has been overlooked. The most surprising result is the lack of scintillations associated with that discrete aurora which is embedded in continuous aurora, a condition which is routinely found in the southern half of the oval. Even during auroral disturbances aurora in this location did not affect the signals. Even though the behavior described is not clearly understood, it is likely, that large-scale F-region irregularities are more prevalent at the poleward boundary of the oval, piled up there by convection of F-region ionization from the oval noon sector across the polar cap, and that the possibly more uniform F-region ionization in the southern half of the oval has little effect on the satellite signals.

ACKNOWLEDGEMENTS

All tests were conducted under the auspices of Lt Col R. Bothwell, ESD Test Director for the AFSATCOM System Evaluation, and without his support, this research could not have been performed. Major O.W. Hoering, SAMSO System Program Office, was responsible for the space segment during these tests. His help and interest are gratefully acknowledged. Engineering support by Messrs. T. Galanis, R.W. Gowell, L.C. James, J. Knapick, C. Perkins, P. Pickell, L. Ross, J.C. Spor, and J.B. Waaramaa was instrumental in providing this unique data base, collected under difficult airborne and ground arctic conditions. We thank Mr. Waaramaa especially for his strong efforts to establish a satellite receive capability on board the AFGL aircraft. The data analysis of Ms. A. Pauline Krukonis and Mr. W.N. Hall aided in the interpretation of the airborne and ground based ionospheric data. The support of Airborne Engineering groups and flight crews from the USAF 4900th Test Wing, Kirtland AFB, NM and the USAF 4950th Test Wing, Wright-Patterson AFB, OH is gratefully acknowledged. We also acknowledge the assistance of the staff of the AFGL Goose Bay Ionospheric Observatory and the AFGL Geopole Station.

REFERENCES

Aarons, J. and R.S. Allen (1971), Scintillation Boundary During Quiet and Disturbed Magnetic Conditions, J.G.R. 76, 1, 170-177.

Aarons, J. (1973), A Descriptive Model of F-Layer High Latitude Irregularities as Shown by Scintillation Observations, J.G.R. 78, 31, 7441-7450.

Gowell, R.W. and R.W. Whidden (1968), Ionospheric Sounders in Aircraft, AFCRL-68-0369, Instrumentation Papers, No. 144.

Whitney, H.E., C. Malik and J. Aarons (1969), A Proposed Index for Measuring Ionospheric Scintillations, Planet. Space Sci., 17, 1069-1073.

EQUATORIAL SCINTILLATION TESTS OF LES 8/9

ALLEN L. JOHNSON
AIR FORCE AVIONICS LABORATORY
WRIGHT-PATTERSON AFB, OHIO 45433

PAUL K. LEE
THE MITRE CORPORATION
BEDFORD, MA 01730

ABSTRACT

The effects of equatorial ionospheric scintillation on the communications signals from Lincoln Experimental Satellites (LES) numbers 8 and 9 were tested from 16 October through 1 November 1976 in Peru, and from Peru to Tahiti and Guam. Most tests were accomplished in flight with equipment on-board an aircraft of the Air Force Avionics Laboratory. A special flight pattern was carried out to examine the directional dependence of the fade rate. Descriptions of the pattern and the obtained results are presented. This paper also reports the overall amplitude statistics and time occurrences of the ultra high frequency (UHF) scintillation that was encountered during the tests.

INTRODUCTION

The effects of equatorial ionospheric scintillation on the communications signals transmitted from the synchronous Lincoln Experimental Satellites (LES) numbers 8 and 9 were tested from 16 October through 1 November 1976 in Peru, and from Peru to Tahiti and Guam. Most tests were accomplished in flight with equipment on-board an aircraft of the Air Force Avionics Laboratory (AFAL). In conjunction with the tests, scientists of the Air Force Geophysics Laboratory (AFGL) flew another aircraft with on-board optical and radio equipment to investigate the effects of equatorial F-region irregularities on trans-ionospheric UHF (near 250 MHz) propagation. This paper reports only the tests conducted on-board the AFAL aircraft.

The LES 8/9 communications system primarily employs Ka-Band (36-40 GHz) and UHF (225-400 MHz) frequencies for signal transmissions. No discernible scintillation was observed on the Ka-Band signal during the tests. The paper, therefore, discusses mainly the UHF scintillation effects which are derived from raw data collected near 250 MHz.

Because space diversity as a solution is restricted in an airborne platform, scintillation effects in the airborne communications terminal may be mitigated through employment of frequency and/or time diversity techniques. It has been reported that the time diversity technique with employment of error correction codes provides a method which could significantly improve UHF communications over a scintillation affected link [Bucher, 1975; Foshee, 1976; Johnson, 1974]. Depending upon the modulation rate over the radio channel, scintillation fading can cause burst channel errors if the margin of the signal power is not sufficient to overcome the fade. Most practical error correction codes, however, are designed to correct random errors. To fully exploit the effectiveness of error correction codes, an interleaver which would randomize the burst errors is essential. Consequently, knowledge of amplitude statistics and rates of scintillation fading is a key to a successful choice of error correction codes for combatting the scintillation induced signal fading. To explore the relationship of fade rates with the flight direction of an airborne terminal, a special flight pattern was carried out in two tests over the Lima, Peru, area. Prior to the discussion of the test results, such as the overall amplitude fade statistics, time occurrences of scintillation, and directional dependence of scintillation fade rates, a summary of test configurations as well as methods of data collection and reduction is appropriate.

TEST CONFIGURATIONS

LES 8/9 signals were employed for the tests. The LES 8 and LES 9 spacecraft are almost identical in communications functions. Each spacecraft is capable of receiving Ka-Band uplink signals which are digitally processed in the satellite and then retransmitted over UHF and Ka-Band downlinks. Depending upon the actual satellite configuration, the uplink signal can also be sent over UHF. If such occurs, the satellite downlink transmission will be over the Ka-Band. In most tests, the LES 9 UHF downlink signal, near 250 MHz, was available. The data reduced for this paper were mainly taken from the LES 9 signal.

The position of LES 9 during the tests was at 40 W longitude. At this longitude, the satellite had an apparent diurnal motion, swinging +25° north and south from the nominal 0° latitude, and +2.8° east and west from the nominal longitudes. The position of the satellite in its daily track shows a figure 8 pattern. During the time of local Peru tests, LES 9 was on the southeast-to-northwest track in the figure-8 pattern.

The test equipment was set up as shown in Figure 1 to process and record Ka-Band and UHF signals. The aircraft Ka-Band antenna which is radome enclosed, has a 3-foot diameter. Pointing is automatically accomplished

with predicted ephemeral data and inputs from the on-board navigation system. Hemispherical coverage UHF antennas were employed for receiving or radiating UHF signals. Both Ka-Band and UHF antennas were designed to work with circular polarized signals.

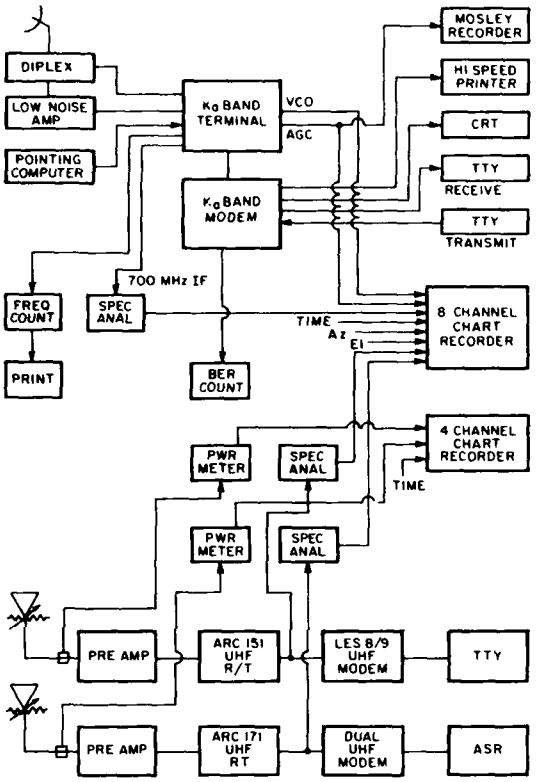


Figure 1. Test Equipment Setup

Two UHF receivers were employed to receive downlink signals from LES 8 and 9. The bandwidth of the 70 MHz intermediate frequency (IF) for both receivers was 3 kHz. Calibrated signal levels at the IF were displayed on spectrum analyzers as well as on an 8-channel strip-chart recorder. Other technical parameters recorded on the chart included elevation and azimuth angles of satellites, universal time, Ka-Band receiver IF amplitude, levels of Ka-Band voltage controlled oscillator receiver, and levels of Ka-Band automatic gain control.

DATA COLLECTION AND REDUCTION

The equatorial scintillation flight tests were made from 16 October through 1 November 1976. The tests were accomplished over point-to-point flights, except those which were conducted from 18 to 30 October over the Lima, Peru, area. These tests were accomplished either on the ground or on a flight over the local Lima area. See Figure 2. The dotted path in the figure shows the continued flight back to Dayton, Ohio.

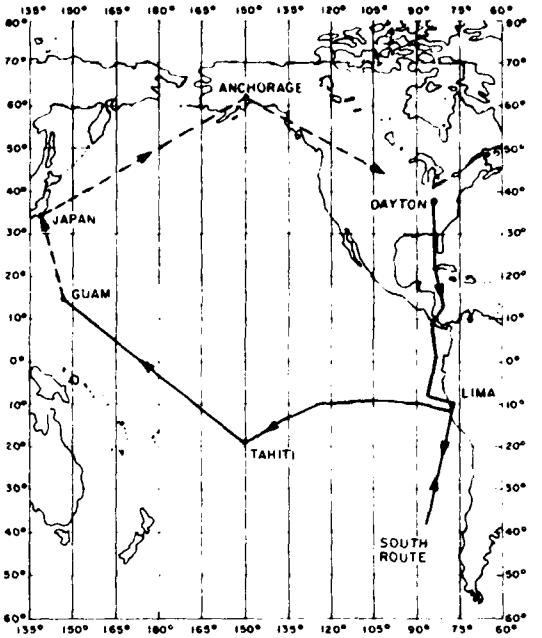


Figure 2. Equatorial Flight Test Route

The data on the signal amplitude were recorded on the 8-channel strip chart recorder and magnetic tapes as well. Calibrations of the complete receiving and recording equipment were accomplished prior to the start of each test. Periodic calibration checks were conducted appropriately during the test period to insure consistency of calibrations. Levels of a fading signal on a typical sample of the recording are shown in Figure 3. To maximize the certainty of scintillation data, raw data were considered insignificant in the

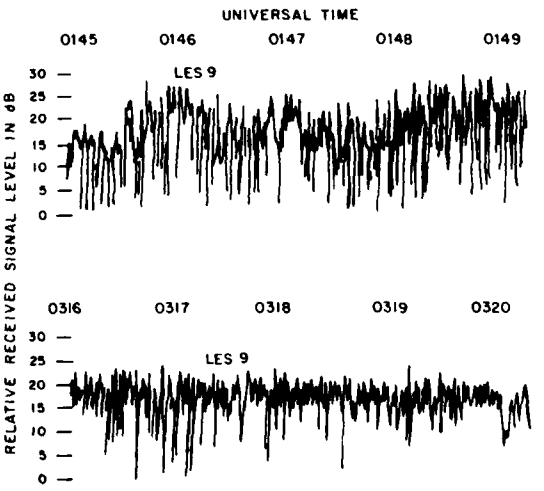


Figure 3. Samples of Scintillation Fading Data, 19 October 1976

process of data reduction if peak-to-peak signal levels were less than 5 dB. In the data reduction for the rate of scintillation fading, the level crossing technique was employed.

A cloverleaf flight pattern was implemented for two local flights over the Lima area to investigate the sensitivity of rates of fading as a function of flight heading, Figure 4. During the cloverleaf flight tests, LES-9 was arising from the southeast leg of the figure-8 pattern to the northwest leg. The straight legs of the cloverleaf pattern are diameters of a circle with 185 km in diameter centered at Ancon (11°-40' S, 77°-10' W), and Huancayo (12°-03' S, 75°-18' W). The flight time for each leg was about 25 minutes. Duration of the pattern was approximately six hours.

SCINTILLATION DATA

Time Occurrences of Scintillation

A summary of the scintillation tests from 16 October through 1 November is listed in Table 1. Date, test span, and availability of scintillation data for each test are indicated. The actual time occurrences of scintillation to the LES 9 signal are illustrated in Figure 5. In the time span of each test, 1800-0200 hours local time, the hour block 2100-2200 had the highest likelihood for the occurrence of scintillation. The result agrees with earlier observations [Aarons, Whitney and Allen, 1971; Paulson and Hopkins, 1973]. The overall amplitude statistics, expressed in terms of number of occurrences vs peak-to-peak fading in dB, is presented in Figure 6. It may become necessary in some applications to calculate the actual fade level from the peak-to-peak fading level. A quick conversion table is available [Whitney, 1974]. It is interesting to note in Figure 6 that the likelihood for strong scintillation to occur is almost the same as that for weak scintillation.

Table 1

Summary of Equatorial Scintillation Tests

Date	Universal Time		Remark	Scintillation
	Takeoff	Land		
16/17 Oct.	2021	0605	Dayton, Ohio to Lima	Strong
18/19 Oct.	2245	0706	Lima Local, Cloverleaf	Strong
19/20 Oct.	Ground (2245	Test 0610)	Aircraft Parked at Lima Airport	Strong
20/21 Oct.	2245	0614	Lima Local, Cloverleaf	Little
24/25 Oct.	1906	0257	Lima Local, South Route	No
28/29 Oct.	2240	0342	Lima Local, North Route	No
30 Oct.	0002	0946	Lima to Tahiti	Strong
1 Nov.	0500	1535	Tahiti to Guam	No

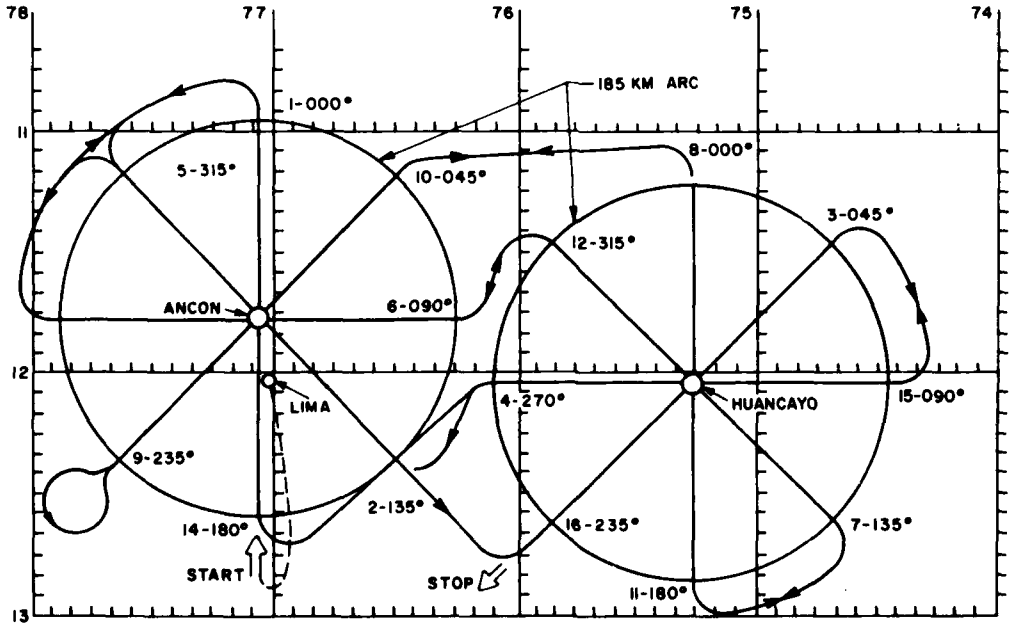


Figure 4. Flight Test Routes for Lima Locals

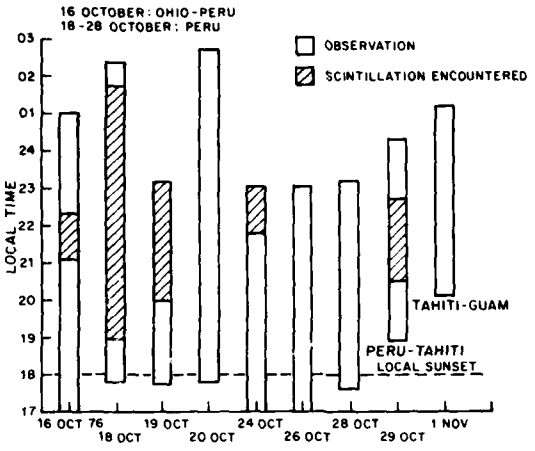


Figure 5. Summary of Ionospheric Scintillation Occurrence

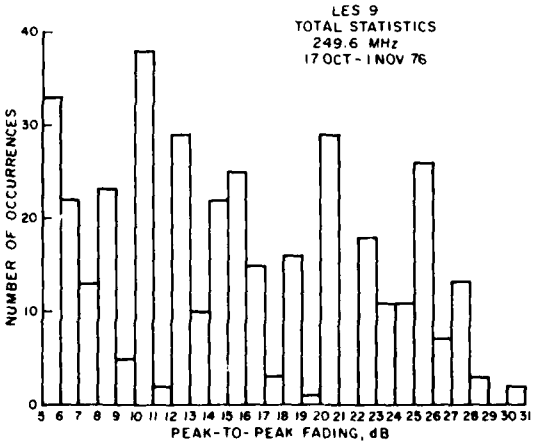


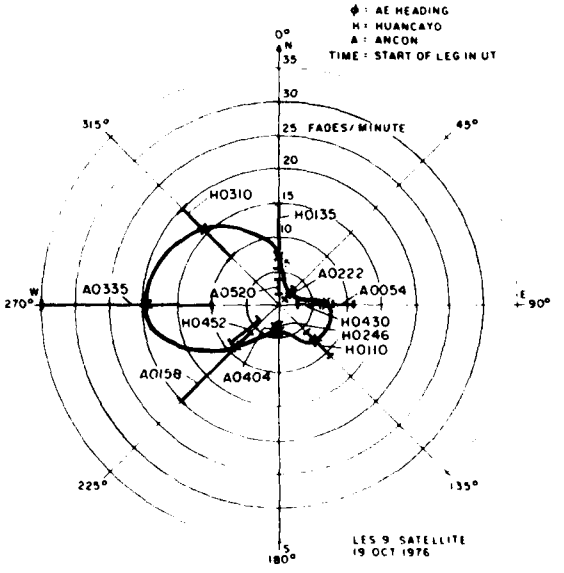
Figure 6. Number of Scintillation Occurrences vs dB P-P Fading

Directional Dependence of Scintillation Rates

An important area investigated during the tests was the dependence of scintillation rates on the aircraft flight heading. Both the 18/19 and 20/21 October flight tests were implemented for testing the directional dependence of the scintillation rate. A flight route following the clover-leaf pattern over Ancon and Huancayo, Peru, as shown in Figure 4, was carried out for the purpose. Since little scintillation was encountered during the 20/21 October test, the following discussion is based solely on the results obtained from the 18/19 October test.

Ionospheric scintillation occurring to the LES 9 signal was recorded at 0035 universal time and continued for the rest of the test. About six hours of data were collected. As normally observed [Paulson and Hopkins, 1973], the onset of scintillation was very rapid with high amplitudes and high rates. The amplitude remained relatively high for the entire test period, with some 10 to 30 minute periods of moderate amplitude.

The rate of scintillation fading was direction sensitive. Flying to the west resulted in the highest average fading rate observed, while flying to the east resulted in a very low average fading rate, Figure 7. At the onset of the scintillation, the aircraft was flying along the 315° leg of the clover leaf pattern. The fade rate was not included in Figure 7 because of incomplete data. The near east-west direction dependence of the fade rate as shown in Figure 7 agrees qualitatively with both a general observation of east-west drift of F-irregularities in the region [Aarons, Buchau et al., 1977; Koster, 1966] and the correlation of drift velocity with rates of fading [Crane, 1976]. A correlation of the fade rates and their corresponding fade amplitudes shows that there is no apparent proportionality between these two factors, Figure 8. A cumulative distribution of a 15-minute fading period (0309-0324 universal time) and a histogram of the percentage of time that the fading amplitude equaled the expected value are shown in Figures 9 and 10.



FADING RATE AS FUNCTION OF AIRCRAFT HEADING

Figure 7. Number of Fades vs Aircraft Direction

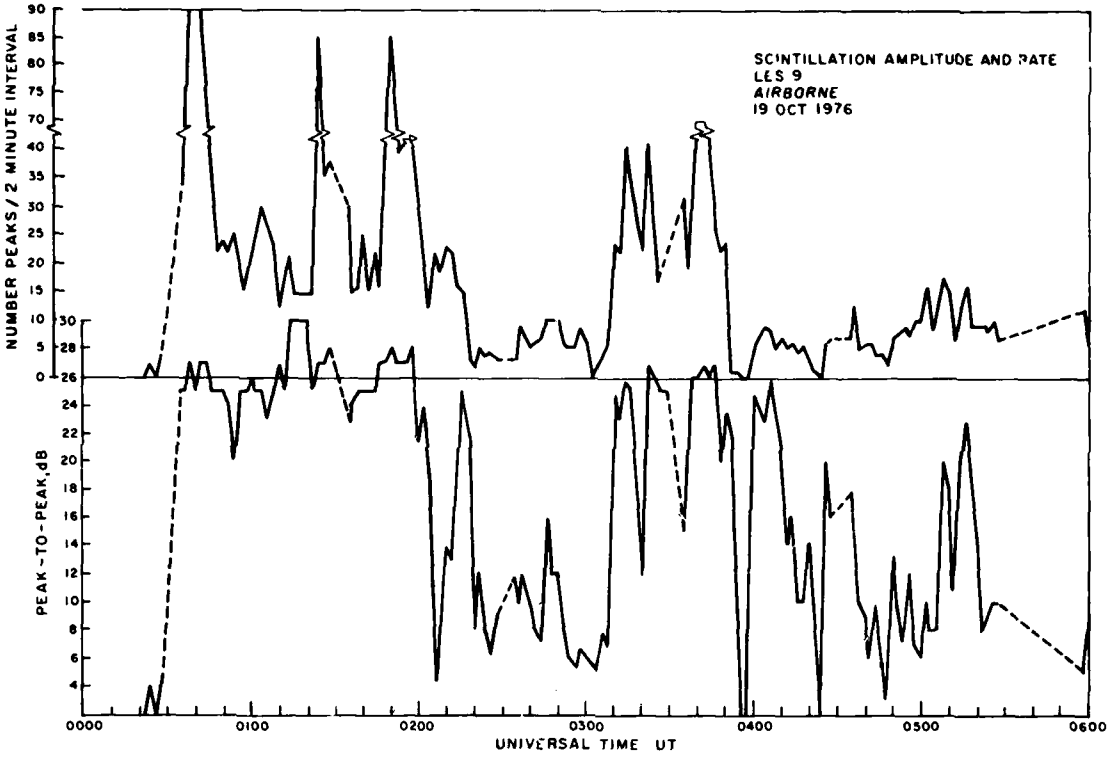


Figure 8. Correlation of Scintillation Amplitudes and Rates

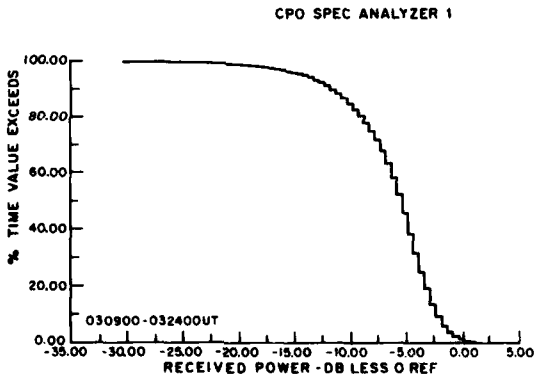


Figure 9. Cumulative Distribution Function of Scintillation Fading Data

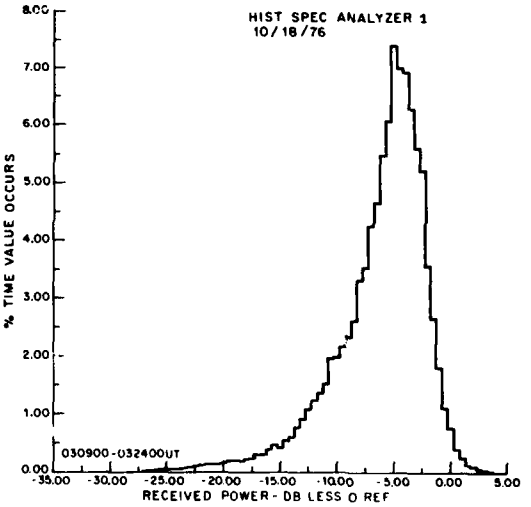


Figure 10. Histogram of Scintillation Fading Data

SUMMARY

During conduct of the equatorial ionospheric scintillation tests of LES 8/9, ionospheric scintillation on the Ka-Band signal was not discernible. UHF scintillation data with the signal frequency near 250 MHz have been discussed in this paper.

The onset of ionospheric scintillation generally took place after sunset. Within the test span, the local time 2000-2100 hour block represents the most probable time occurrence of the ionospheric scintillation. Peak-to-peak amplitudes of greater than 25 dB were encountered on numerous occasions. An interesting result shows that for the scintillation encountered, both strong levels and weak levels have almost the same likelihood of occurrence. The duration of scintillation varied from a few minutes to over six hours.

The rate of scintillation fading experienced by the aircraft was directionally sensitive. Flying to the west resulted in the highest average fade rate observed, while flying to the east resulted in a very low average rate.

ACKNOWLEDGMENTS

The scintillation test program was conducted jointly by AFAL and AFGL under the direction of the Electronic Systems Division of the Air Force Systems Command. Support to the AFAL tests includes personnel of 4950th Test Wing, MIT Lincoln Laboratory, AFAL, and The MITRE Corporation.

REFERENCES

Aarons, J. and J. Buchau, et al. (1977), "The Localized Origin of Equatorial Irregularity Patches," URSI, 22-24 June 1977, Stanford, California.

Aarons, J., H. Whitney, and R. Allen (1971), "Global Morphology of Ionospheric Scintillations," Proceedings of the IEEE, 59, 2, pp. 159-172.

Bucher, E.A. (1975), "UHF Satellite Communication During Scintillation," Technical Note, MIT Lincoln Laboratory, Lexington, MA (DDC AD A014801).

Crane, R.K. (1976), "Spectra of Ionospheric Scintillation," Journal of Geophysical Research, 81, 13, pp. 2041-2050.

Foshee, J.J. (1976), "A Bit Timing Extraction Technique," AFAL-TR-76-149, Air Force Avionics Laboratory, Wright-Patterson AFB, Ohio (DDC AD A031207).

Johnson, A.L. (1974), "Simulation and Implementation of a Modulation System for Overcoming Ionospheric Scintillation Fading," AGARD Conference Proceedings, No. 173 on Radio Systems and the Ionosphere (DDC AD A022439).

Koster, J.R. (1966), "Ionospheric Studies Using the Tracking Beacon on the 'Early Bird' Synchronous Satellite," Annales de Geophysique, 22, 3, pp. 435-439.

Paulson, M.R. and R.V.F. Hopkins (1973), "Effects of Equatorial Scintillation Fading on SATCOM Signals," Naval Electronics Laboratory Center, San Diego, California (DDC AD 910490).

Whitney, H.E. (1974), "Notes on the Relationship of Scintillation Index to Probability Distributions and Their Uses for System Design," Air Force Cambridge Research Laboratories, Bedford, MA.

GIGAHERTZ SCINTILLATION AT THE MAGNETIC EQUATOR

J. P. Mullen
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

A. Bushby, J. Lanat and J. Pantoja
Instituto Geofisico Del Peru
Apartado 3747, Lima, Peru

BACKGROUND

When a radio wave is propagated through the irregular ionosphere it is diffracted and caused to fluctuate in amplitude and phase. A wealth of experimental and theoretical analysis has been performed on this phenomenon; early experiments began in the high frequency range (≈ 20 MHz) and have progressed to the gigahertz region. It is the purpose of this note to consider results obtained at or near 1.5 GHz, the band to be used by the DoD NAVSTAR (Global Positioning System). Early observations of equatorial GHz scintillation were made by Golden (1970) who found 5 dB scintillation on 1.5 GHz at Quito, Ecuador. Skinner et al. (1971) reported 4 dB fading at SHF (6.3 GHz) at Kenya. Christiansen (1971) measured 20 and 25 dB at 2.2 GHz from Guam, Canary Islands and Ascension Island. Craft (1971) found up to 8 dB scintillation at 4 and 6 GHz at a number of equatorial locations. Other equatorial gigahertz scintillation measurements were reported by Crampton and Sessions (1971), Sessions (1972), Taur (1973 and 1976) and Mullen et al. (1977). Much of the early work was done using the ATS-5 satellite, which carried VHF and L Band (1.5 GHz) transmitters, the Intelsat space vehicles which transmitted at 4 and 6 GHz while more recent work (Mullen et al., 1977 and Paulson and Hopkins, 1977) used the UHF and 1541.5 MHz beacons on Marisat I and II. A very strong tool for investigation of scintillation at a number of coherent frequencies (137, 378, 390, 401, 413, 424, 435, 447, 1239, 2891 MHz) was provided by the DNA Wideband satellite (Fremouw, 1976).

THE DATA

Observations of scintillation at L Band (1.5 GHz) and at UHF (254.15 and 257.55 MHz) using the transmitters of MARISAT were begun at Huancayo in April 1976, and have continued without substantial interruption. Figure 1 shows the probability of occurrence of L Band scintillation greater than 2 dB. The month of October 1976 saw unusual activity, and it is shown separately, as well as lumped together with the months of April and December, 1976, January, February and March 1977. This shows that the scintillation occurs for the most part between 1900 and 2300 LMT, with a maximum at 2100 LMT. Figure 2, taken from recently published work (Mullen et al., 1977) shows the cumulative distribution function of a 15 minute period in which the L Band scintillation exceeded 7 dB. While this is far from a worst case, it does illustrate the problem of the systems designer. These results show that for operation at the 99th percentile, a system would require a fade margin of 2 dB at 1541 MHz. Figure 3 shows a common event; strong scintillation begins on the UHF signal, and L Band scintillation commences simultaneously. The UHF signal remains strongly scintillated after the L Band scintillation ceases. In this illustration is also seen one of the rare instances of substantial UHF daytime scintillation. Figure 4 shows an instance in which UHF scintillation begins, shortly thereafter follows L Band scintillation which quiets while the UHF strong scintillation continues for another two hours. Two hours later UHF scintillation re-

commences, followed by L Band, which ceases within an hour and finally the UHF scintillation ends. Figure 5 is an example of strong UHF scintillation associated with only moderate (≈ 2 dB) L Band scintillation. In these illustrations, L Band scintillation is found to vary from a nightly maximum of about 2 dB to a nightly maximum of about 8 dB. This last is the maximum L Band scintillation measured at Huancayo, where the elevation angle to the satellite is 21°. The occurrence pattern is of some interest; the UHF scintillation begins either before or at the same time as the L Band. It is also noteworthy that intense scintillation can develop at UHF with little or no activity at L Band (Figs. 3 and 5).

DISCUSSION

The Naval Ocean System Center (Paulson and Hopkins, 1977) has recently completed simultaneous observations of the L Band and UHF transmitters of Marisat II over a 6 months period. They found significant levels of amplitude scintillation (up to 5 dB) during the tests. From this they inferred that large phase scintillations also occurred.

One view of the causative mechanism for the gigahertz scintillation producing ionospheric irregularities is a thick layer (≈ 200 kilometers) extending upward from the base of the F region and containing irregularities varying from a few meters to several kilometers in size having density fluctuations on the order of 50-75% (Basu and Basu, 1976). This is corroborated by observations made at Huancayo, Jicamarca and Ancon, Peru in 1976 and 1977 by a joint task force which included personnel of AFGL, SRI, AFAL and University of Texas at Dallas. The program included participation by the Jicamarca radar, the AFGL flying laboratory and the AFAL airborne communications terminal as well as ground observations at Ancon and Huancayo.

The morphology of the irregular equatorial ionosphere has been explored first by bottomside soundings (Wright, 1959) by topside sounder (Calvert and Schmid, 1964) by transiting satellites (Sinclair and Kelleher, 1969), by a combination of in-situ and geostationary observations (Basu et al., 1976), by a world wide network of earth stations (Taur, 1973) and most recently by the combination of ground and airborne multi-media observations supplemented by an ionospheric backscatter radar (Aarons et al., 1977). The findings are in general agreement; the zone of maximum equatorial scintillation extends from about 20° geomagnetic north to 20° geomagnetic south.

The longitudinal morphology of equatorial scintillation has also been explored by Basu et al. (1976). Their findings suggest a pronounced longitude variation, with the scintillation zone width and percentage occurrence maximizing over the African sector. In the American sector, the width of the zone decreases without appreciable change in scintillation occurrence. In the Indian sector, both width and occurrence probability are decreased. The model based on in-situ measurements is consistent with ground based observations during the December solstice.

CONCLUSION

The nighttime equatorial ionosphere has been found to cause scintillation up to approximately 8 dB at L Band. Geographically, this occurs most frequently within ± 20° geomagnetic. Timewise, it can occur just about anytime between sunset and sunrise, with midnight ± 2 hours being most likely. Cumulative amplitude distributions have been reported earlier and found to correspond to the Nakagami distribution with m=22. It was found that the corresponding UHF distribution was a good fit to the Nakagami distribution where m=1.3, which is very close to a Rayleigh distribution, and which in our experience has been the limiting case.

REFERENCES

Aarons, J., J. Buchau, S. Basu and J. P. McClure, "The Localized Origin of Equatorial F-region Irregularity Patches", submitted for publication in J. Geophys. Res., 1977.

Basu, Sunanda, S. Basu and B. K. Khan, "Model of Equatorial Scintillations from In-Situ Measurements", Radio Sci., 11, 821, 1976.

Basu, Sunanda and S. Basu, "Correlated Measurements of Scintillations and In-Situ F-region Irregularities from Ogo-6", Geophys. Res. Lett., 3, 681, 1976.

Calvert, W. and C. W. Schmid, J. Geophys. Res., 69, 1839, 1965.

Christiansen, R. M., "Preliminary Report of S-Band Propagation Disturbance During ALSEP Mission Support", November 19, 1969 - June 30, 1976) Report X861-71-239, Goddard Space Flight Center, June, 1971.

Craft, H. D., Jr., "Investigation of Satellite Link Propagation Anomalies at 4 and 6 GHz, Preliminary Report", Comsat Laboratories, CL-9-71, 9 Feb 1971.

Crampton, E. E., Jr., and W. B. Sessions, "Experimental Results of Scintillation Measurement of Ionospheric Amplitude Variations of 136 and 1550 MHz signals at the Geomagnetic Equator", Report to the ASTRA Panel, International Civil Aviation Organization, 7 Jan 1971.

Fremouw, E. J., "Wideband Satellite Observations", Bimonthly Progress Report #2, Contract DNA 001-75C-0111, SRI Project 3793, Stanford Research Institute, Menlo Park, CA, 13 August 1976.

Golden, T. S., "A Note on Equatorial Ionospheric Scintillation at 136 MHz and 1550 MHz", Report X520-70-397, Goddard Space Flight Center, Greenbelt, MD, October, 1970.

Mullen, J. P., H. E. Whitney, Santimay Basu, A. Bushby, J. Lanat and J. Pantoja, "Statistics of VHF and L Band Scintillation at Huancayo, Peru", To be published in JATP, September, 1977.

Paulson, M. R., and R. U. F. Hopkins, "Spatial Diversity Characteristics of Equatorial Scintillation", Technical Report No. 3, Naval Ocean Systems Center, San Diego, CA, 2 May 1977.

Sessions. W. B., "Amplitude Fading of Simultaneous Trans-Ionospheric L Band and VHF Signals Received at the Geomagnetic Equator", NAS Report X810-72-282 GSFC, June, 1972.

Skinner, N. J., R. F. Kelleher, J. B. Hacking and C. W. Benson, "Scintillation Fading of Signals in the SHF Band", Nature (Phys. Sci.), 232, 19-21, 1971.

Sinclair, J. and R. F. Kelleher, "The F-region Irregularity Belt as Observed from Scintillation of Satellite Transmissions", JATP, 31, 201-206, 1969.

Taur, R., "Ionospheric Scintillation at Frequencies Above 1 GHz", Technical Memorandum CL-29-73, COMSAT Laboratories, Clarksburg, MD.

Taur, R. R., "Ionospheric Scintillation at 4 and 6 GHz", COMSAT Technical Review, 3, #1, Spring, 1973.

Taur, R., "Simultaneous 1.5 and 4 GHz Ionospheric Scintillation Measurement, Radio Sci., 11, 1029-1036, 1976.

Wright, R. W. H., J. Geophys. Res., 64, 2203, 1959.

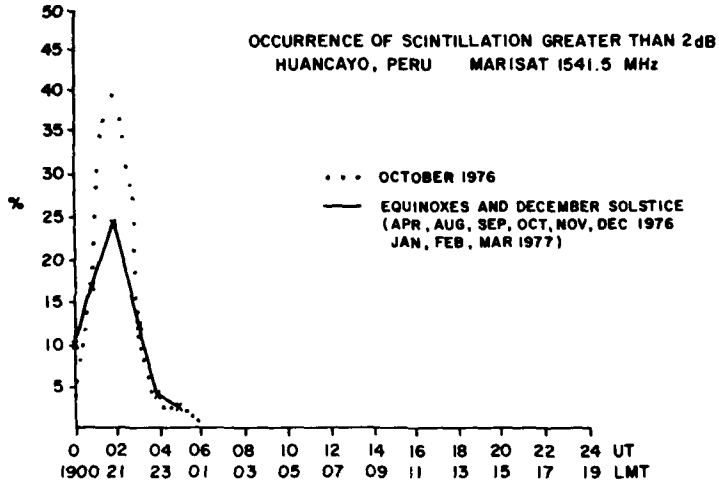

OCCURRENCE OF SCINTILLATION GREATER THAN 2dB
HUANCAYO, PERU MARISAT 1541.5 MHz
• • • OCTOBER 1976
EQUINOXES AND DECEMBER SOLSTICE
(APR, AUG, SEP, OCT, NOV, DEC 1976
JAN, FEB, MAR 1977)
%
50
45
40
35
30
25
20
15
10
5
0 02 04 06 08 10 12 14 16 18 20 22 24 UT
1900 21 23 01 03 05 07 09 11 13 15 17 19 LMT

fig 1

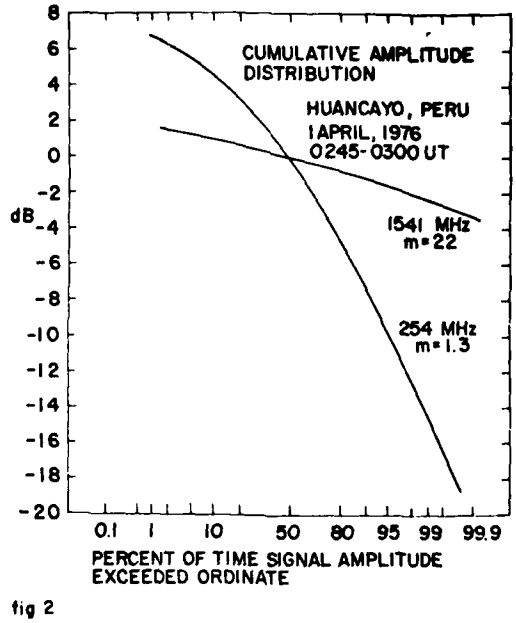

CUMULATIVE AMPLITUDE
DISTRIBUTION
HUANCAYO, PERU
1 APRIL, 1976
0245-0300 UT
1541 MHz
m=22
254 MHz
m=1.3
dB
8
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
0.1 1 10 50 80 95 99 99.9
PERCENT OF TIME SIGNAL AMPLITUDE
EXCEEDED ORDINATE

fig 2

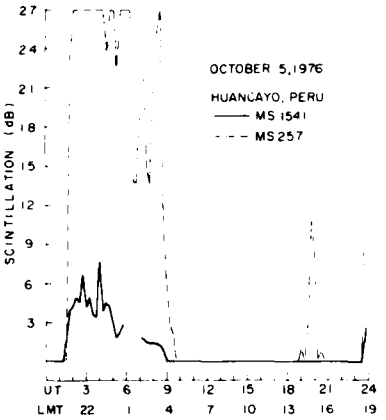


fig 3

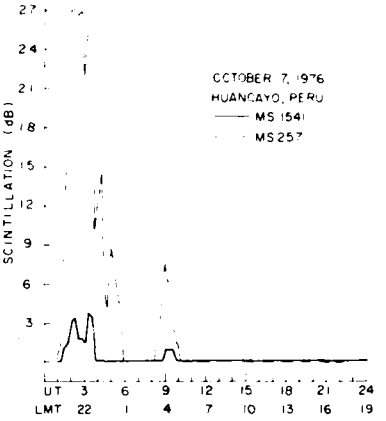


fig 4

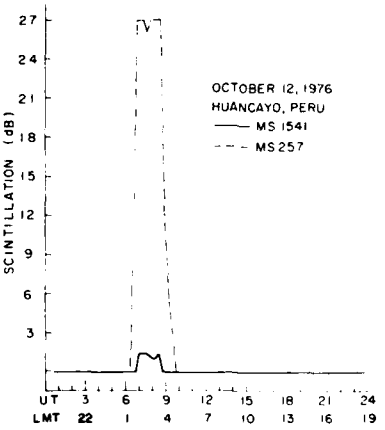


fig 8

MEASUREMENT TECHNIQUES FOR LONG PERSISTENCE FADING PHENOMENA APPLIED TO SATELLITE COMMUNICATIONS LINKS

H. Hanft
Westinghouse Electric Corporation
Baltimore, Maryland

J.P. Corrigan
Goddard Space Flight Center
Greenbelt, Maryland

ABSTRACT

The successful measurement of the effects of long persistence fading phenomena at C-band was performed during the drift period when ATS-6 moved from 35°E longitude to 140°W longitude. Time periods of signal fluctuation of long persistence fading for antenna elevation angles of 20° to lower than 1° were accurately measured and found to vary from 30 seconds to over 30 minutes. This new method of measurement permitted the acquisition of data in a constant manner. These measurements could provide a means of determining the actual extent of long persistence fading and point to methods of minimizing their effect.

INTRODUCTION

Long persistence fading phenomena was measured by using the ATS-6 satellite and two associated earth stations. This propagation phenomena is present in satellite communication links and becomes prominent when the ground station antenna is at lower elevation angles. Measurements of this type of radio frequency propagation effects were made during the drift period when ATS-6 moved from 35°E longitude to 140°W longitude. The satellite and the Ahmedebad earth station (AES) in India provided a unique means for the measrurement of C-band propagation over an essentially smooth earth surface. The unique location of AES on a hill overlooking a flat desert when pointing to ATS-6 was ideal for these measurements. A NASA operated earth station, called the Hybrid, was located near Madrid, Spain with a 4.6 meter diameter antenna, and was used to perform the tests and collect data. The NASA Rosman data acquisition facility (ROS), with its special equipment, provided a means of processing the data. AES in India recorded the downlink frequency from ATS-6 by means of a strip chart recorder. The time was manually entered on the resulting paper strip charts. Therefore, the technique used for measuring long persistence fading was proven to be possible by means of strip chart recorders.

DATA ACQUIRED

The reception of both downlink signals (AES and Hybrid) by the Hybrid were recorded on analog magnetic tapes as well as on paper strip charts as indicated in Figure 1, Ground Station Equipment for Recording Measurements.

At the AES, paper strip chart recorders were used to provide a record of the transmitted RF power level, received signal (their own transponder), and readout of the AES antenna position.

Test results collected at AES in India consisted essentially of three sets of data as follows:

- Teletype printout of antenna azimuth and elevation at 20 second intervals
- Transmitted power level on a paper strip chart with GMT manually entered
- Received signal level with GMT manually noted on the strip chart

All strip charts were carefully calibrated. The test schedule for making these measurements were fitted into an existing schedule that permitted this additional experiment with no interference to any other experiment. In order to get a reference base line of data, tests were performed for an AES antenna elevation of 20°. Measurements were made for elevation angles of 5.3° and 3.8° during test periods of approximately 2 hours each in duration. Extended periods of testing were made for elevation angles from 2.6° until AES was no longer in view of ATS-6 which was a total of more than 45 hours.

DATA PROCESSING

The data collected by the Hybrid, in the form of magnetic tape, was processed using special equipment located at the Rosman Earth Station (ROS). Figure 2 is a functional block diagram of this equipment. Use of the equipment was originally intended for Radio Frequency Interference (RFI) measurement. The RFI test data had several possible modes for data acquisition and processing. Use of one of these modes, designated "mode 8", was altered to suit the measurement of long persistence measurements, signal level vs. time was plotted. The program used for these measurements resulted in plots showing maximum, minimum, and average values of signal level in units of relative dB vs. time, as seen in Fig. 3, Signal Level vs. Time for an Antenna Elevation Angle of 20.8°.

MEASUREMENT TECHNIQUE REQUIREMENTS

Determination of long persistent fading requires the acquisition of radio frequency signal measurements and the processing of this data. These requirements include the following:

- Linear Spacecraft transponder
- Uplink of CW RF signal at constant power level
- Antenna pointing angles with respect to time
- Received signal recording vs. time
- Data processing to correlate and interpret results
- Ground station physical data

A spacecraft linear transponder may be maintained as a fixed linear transponder. Use of a constant level CW reference signal transmitted at a level 10 dB stronger than the signal being measured, as received at the spacecraft, will capture the spacecraft AGC. Both transponded signals may be received by the earth station and utilized in the data reduction process. Antenna pointing angles may be automatically recorded or alternatively recorded manually.

TEST DESCRIPTION

Tests were performed over substantial periods of time. The radio frequency propagation emanated from AES. AES uplinked a CW signal at a constant power level at 6 GHz to ATS-6. ATS-6 received the signal via its receive earth coverage horn (ECH) and translated the signal to 4 GHz through the transmit ECH. The Hybrid station transmitted the reference signal at a level 10 dB stronger and about 20 kHz lower than the AES signal, as received at the spacecraft, in order to capture the spacecraft AGC and maintain a fixed gain linear transponder for the AES signal. The Hybrid station received its own reference signal and the 10 dB weaker AES signal, down-converted both signals to 10 kHz and near 30 kHz respectively, and direct recorded both signals for off-line processing. Supporting data acquired include strip chart recordings of the AES received signal strength, transmitted power, and readout of the AES antenna position with respect to time recorded at 20 second intervals in India. The Hybrid station received signal strength was continuously recorded on a strip chart for both the reference and the AES signals. Fig. 1 is a functional block diagram of the earth station equipment used for recording measurements.

TEST RESULTS

Signal Level vs. Time was measured for an antenna elevation angle of 20.8°. Fig. 3 shows this response in the form of three curves that represent the maximum, average, and minimum values. These values are derived from the sampling of data measured which consisted of 39 samples for each 10 millisecond interval. The computer is used to derive the resulting maximum, average, and minimum values shown in Fig. 3.

Although the average signal level varied about 0.3 dB, a definite persistent fading is portrayed with a time period of fluctuation of 32.4 seconds. The time period of fluctuation is the time period between the signal peaks. The effect of diffraction, due to the curvature of the earth, was calculated from the parameters given in Fig. 4, Geometry of Test Parameters. The time period of fluctuation, defined above, is given by:

$$T = \frac{\text{Elapsed Time between signal peaks}}{\Delta n}$$

where $n = \frac{d}{\lambda}\left(\frac{1}{\cos\theta} - 1\right)$

for T in minutes, the elapsed time is in minutes, and the distance (d) is in the same units as the wavelength (λ).

Antenna elevation angles vs. Time for the AES were provided by the Goddard Space Flight Operation analysis section. A comparison of the test measurements and the calculated values is shown in Fig. 5, Diffraction effect for 5950.02 MHz. A representative example is shown in Fig. 6 in the results obtained at an antenna elevation angle of 1.4°. The overall time period of fluctuation is 10 minutes. The figure shows the distance in time to be nearly the same between the two maximum values as between the two minimum values.

At very low elevation angles the received signal level by ATS-6 from AES was enhanced. Even though the transmitted power levels from AES and the Hybrid remained constant, the signal received by the satellite from AES became much stronger. This phenomenon is shown in Fig. 7, where the comparison of signal levels simultaneously received by ATS-6 may be made. The signal from AES became strong enough to lower the signal from the Hybrid (due to power sharing). The signal from the Hybrid was reduced by as much as 3 dB while the signal from AES increased about 7 dB to reach its peak level. The responses, shown in Fig. 7, show the Hybrid signal minimum where the AES maximum occurs.

CONCLUSIONS

The results of the measurements described show the existence of long persistence fading. This phenomena is more prevalent at lower antenna elevation angles. The effect of propagation of C-band radio frequencies at low "grazing" angles is a potential source of interference when such signals are received by satellites because of the increase peak levels that occur due to refractive effects in addition to the diffraction effect. These effects are amplified with the propagation of signals over the horizon.

ACKNOWLEDGMENTS

A great deal of management support at NASA/GSFC and at Westinghouse made the performance of this experiment possible. The cooperation of the many

people at the ATS Control Center, ATS Hybrid, Rosman Ground Station, and especially at the Ahmedabad Earth Station in India is greatly appreciated. The authors wish to thank Joseph L. Day, whose invaluable assistance made these tests successful.

REFERENCES

1. The ATS-F Data Book, rev. May 1974.

2. Reference Data for Radio Engineers, Fifth edition ITT.

3. K. Bullington, "Radio Propagation Fundamentals" Bell System Technical Journal, Vol. 36, no. 3.

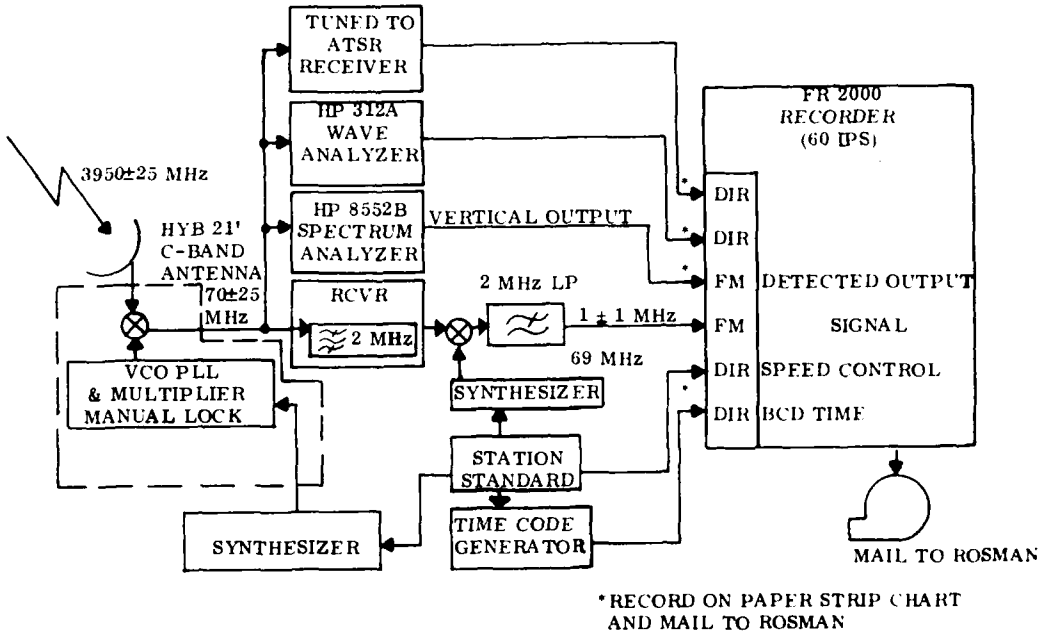


FIG. 1. GROUND STATION EQUIPMENT FOR RECORDING MEASUREMENTS

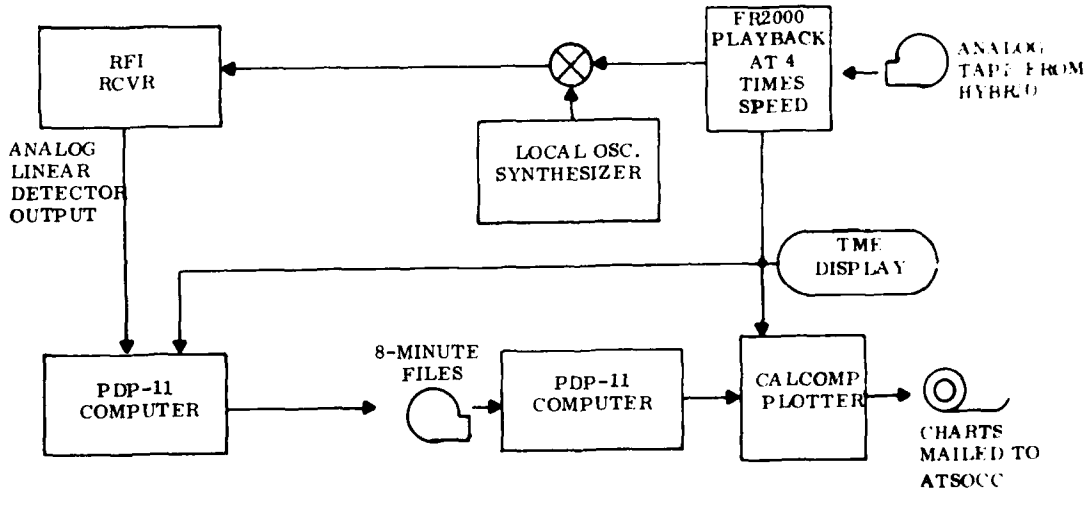


FIG. 2. GROUND STATION EQUIPMENT FOR DATA PROCESSING

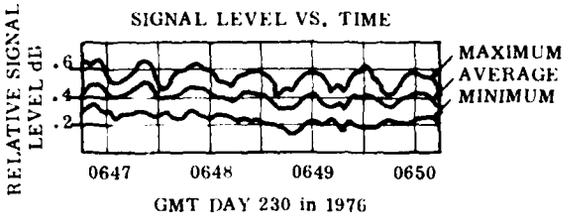
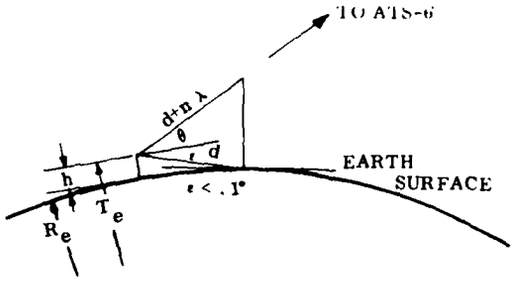


FIG. 3. SIGNAL LEVEL VS. TIME ANTENNA ELEVATION ANGLE of 20.8°



T_e Antenna height from Earth Center (6378.239 km)

θ Elevation angle

h Tower height (9.733 meters)

R_e Tower base from Earth Center (6378.229 km)

d Distance from antenna to horizon (11.114 km)

λ Wavelength of transmission (.050386 meters)

FIG. 4. GEOMETRY OF TEST PARAMETERS

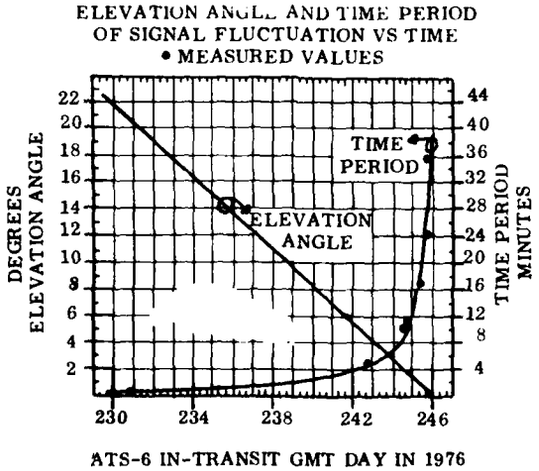


FIG. 5. DIFFRACTION EFFECT FOR 5950.02 MHz

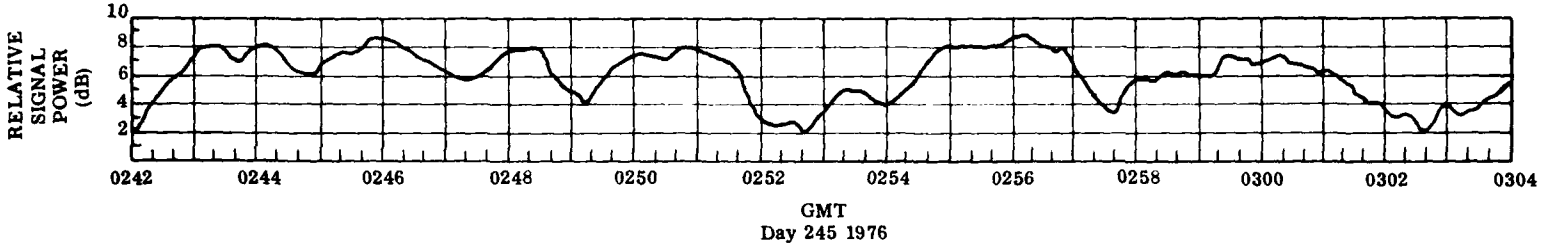


FIG. 6. SIGNAL LEVEL VS. TIME FOR AN ANTENNA ELEVATION ANGLE OF 1.4°

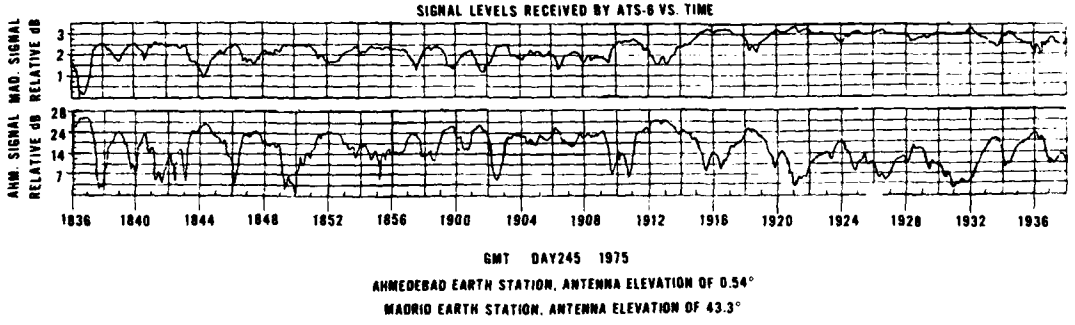


FIG. 7. COMPARISON OF SIGNAL LEVELS SIMULTANEOUSLY RECEIVED BY ATS-6

SPACED-RECEIVER INVESTIGATIONS OF EQUATORIAL SCINTILLATION MADE AT GUAM DURING 1971 AND 1976

By M. R. Paulson and R. U. Hopkins
Naval Ocean Systems Center
San Diego, CA, 92152

INTRODUCTION

The Naval Ocean Systems Center (NOSC), formerly the Naval Electronics Laboratory Center, has been actively investigating equatorial scintillation since 1970 to determine what adverse effects it might have on UHF satellite systems. The work done in the early 1970's is summarized in reference 1. One phase of this work was some limited spaced-receiver measurements made on Guam in 1971. The results of these measurements lead to a more extensive spaced-receiver investigation made on Guam in 1976 to evaluate space diversity as a means for overcoming the disruptive effects that equatorial scintillation has on UHF satellite signals. This investigation is detailed in reference 2. In this report we will consider some of the results from each of the two years.

1971 MEASUREMENTS

Figure 1 shows the receiver arrangement used for the 1971 measurements. Two receivers were separated 1100 metres in an approximate east-west direction. The third receiver was 750 metres approximately south of the east receiver. This system was operated at night from 27 through 31 October. The UHF beacon from TACSAT 1 was used as a signal source. During this time scintillation was observed only on the night of 28 October.

Cross correlations between pairs of receivers were calculated for five minute samples of the data. Figure 2 is an example of one of these. The delay times for maximum cross correlation were used to calculate the speed and direction of the apparent drift velocity of the diffraction pattern. These velocities, shown in figure 3, varied from about 125 to 180 metres per second in speed and 80 to 118 degrees in direction. Since the ionospheric irregularities are highly elongated in the direction of the magnetic field at the equator the variation in the direction of the velocity is probably reflecting a shifting about in the orientation of the magnetic field lines at F-region heights.

1976 MEASUREMENTS

The cross correlations calculated for the 1971 scintillation data showed a consistent negative correlation at zero time difference for the 1100 metre east-west separation. Because of this a program was planned to evaluate space diversity as a way to overcome the scintillation problems on Guam. TACSAT I failed, however, so it wasn't until the launch of the Pacific MARISAT/GAPFILLER satellite in 1976 that the program could be put into operation.

Figure 4 shows the receiver arrangement used for these tests. Two UHF receivers were placed on a magnetic east-west line with a 1000 metre separation. The third UHF receiver was placed on the same line 300 metres west of the east receiver so that three different separations could be evaluated at the same time. Since the MARISAT/GAPFILLER satellite had an L-Band beacon two L-Band receivers were co-located with the two UHF receivers at the 1000 metre separation to investigate L-Band scintillation at the same time. Although the primary objective was to evaluate space diversity several other aspects of scintillation were also observed.

UHF VS L-BAND

Figure 5 is an example of the amplitude scintillation recorded at UHF and L-Band. There were three nights when significant L-Band scintillation occurred although there were many more nights when only UHF scintillation was observed. Figure 6 shows the cumulative amplitude distributions for a fifteen minute sample at the two frequencies. While the UHF faded as much as 30 dB the L-Band fades were at the most about 3 dB. Another way of looking at this data is presented in figure 7. Here the standard deviation divided by the mean in dB is plotted for the UHF scintillation taken in five minute samples against the corresponding L-Band scintillation. It is seen that as the L-Band scintillation intensity increases the UHF scintillation tends to reach a limit. Since these data were recorded during a minimum in the solar cycle it appears that L-Band scintillation intensity may increase to a much greater extent than the UHF as solar activity increases.

Cross correlations were calculated for the various receiver separations using five minute samples. Figure 8 compares typical cross correlations at UHF and L-Band for the 1000 metre east-west separation. The maximum cross correlation values occur at the same delay time for the two frequencies. This would indicate that either the two frequencies are being affected by the same region in the ionosphere

or that two regions in the ionosphere affecting the two frequencies are drifting with the same velocity.

If we use the delay time to calculate a drift velocity we can use this velocity with the time difference between the maximum positive and maximum negative cross correlation to calculate an optimum separation for space diversity. This appears to be about 750 metres for UHF and 300 metres for L-Band in the case shown. Wittwer (3) has suggested that these distances may be a function of the first Fresnel zone. If this is so it may be a simple way of determining the ionospheric height where the irregularities causing the scintillation occur. It could also indicate that the irregularities causing scintillation were at a somewhat higher altitude in 1971 than they were in 1976 since this spacing was about 1000 metres in 1971 whereas it was consistently on the order of 700 metres in 1976.

SPACE DIVERSITY AND BIT ERROR RATES

Two parallel receiving systems were used to evaluate the effectiveness of space diversity in overcoming the effects of scintillation on received signals. In one system the signals from the three receiver sites were combined using a predetection combiner. The output of this combiner went to a demodulator and a bit-error-rate counter. In the second system signals from a single receiver site went to a demodulator and a bit-error-rate counter. During scintillation a known 2400 bit-per-second PN sequence was transmitted from San Diego over the satellite and bit error rates were determined for the two systems at Guam. Figure 9 shows about an hour of this data. The error rates for the single site approached 50 percent at times even though the undisturbed signal-to-noise ratio for the system was better than 30 dB. At the same time no errors were detected for the three site combination. Figure 10 shows the error rates for the single site in a different form. Here error rates are plotted against percentage of occurrence for the sample period considered. For example 27 percent of the time error rates were greater than one in 100.

We had planned to test two site diversity for the various receiver separations as well but not enough time was available when scintillation was occurring.

SCINTILLATION INTENSITY VS DRIFT VELOCITY

A noticeable decrease in the drift velocity as scintillation died out was observed in the 1971 data shown in figure 3. This suggested the possibility that the ionospheric irregularities which cause scintillation might be generated by turbulence resulting from zonal winds occurring at F-region heights. To further test this idea all of the 1976 UHF data were taken in five minute samples and the east-west component of the apparent drift velocity was calculated for each sample. The standard deviation divided by the mean was also calculated for each sample as an indication of the intensity of the scintillation. These two values, drift velocity and scintillation intensity, were plotted on the same graph as a function of time. Figure 11 is an example of this. A quite good correlation between drift velocity and scintillation intensity is evident. Figure 12 is another example. Here there is a gap in the drift velocity curve between about 1100 and 1145 when the scintillation died out and no cross-correlation maxima were observed.

A comparison of the 1971 and 1976 drift velocities shows that while the velocity in 1971 went as high as 180 metres per second it rarely exceeded 100 metres per second in 1976.

CONCLUSIONS

The primary conclusions of this investigation are as follows:

1. Three site space diversity using a predetection combining techniques can completely overcome the problem of high digital error rates caused by equatorial scintillation. Evaluation of the effectiveness of two site diversity using bit-error-rate measurements was not done in these tests.

2. Equatorial scintillation is considerably less intense at L-Band than at UHF, however, L-Band scintillation intensity may be expected to increase at a greater rate than UHF as solar activity increases.

3. The ionospheric region causing L-Band scintillation is either the same region causing the UHF scintillation or the two regions are drifting with the same velocity.

4. The good correlation between UHF scintillation intensity and drift velocity suggests that turbulence caused by zonal winds at F-region heights should be investigated as a possible generating mechanism for the ionospheric irregularities causing scintillation.

REFERENCES

1. MR Paulson, RUF Hopkins, Effects of Equatorial Scintillation Fading on SATCOM Signals, NELC/TR 1875, 8 May 1973.
2. MR Paulson, RUF Hopkins, Spatial Diversity Characteristics of Equatorial Scintillation, NOSC/TR 113, 2 May 1977.
3. LA Wittwer, Capt. USAF, Air Force Weapons Lab., Kirtland AFB, N.M., Private communication, 16 June 1977.

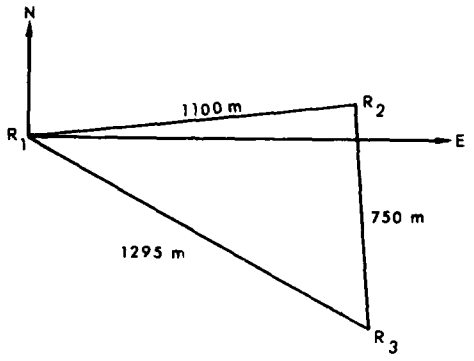


Figure 1. Dimensions and orientation of receiver triangle used at the Naval Communications Station, Guam in 1971.

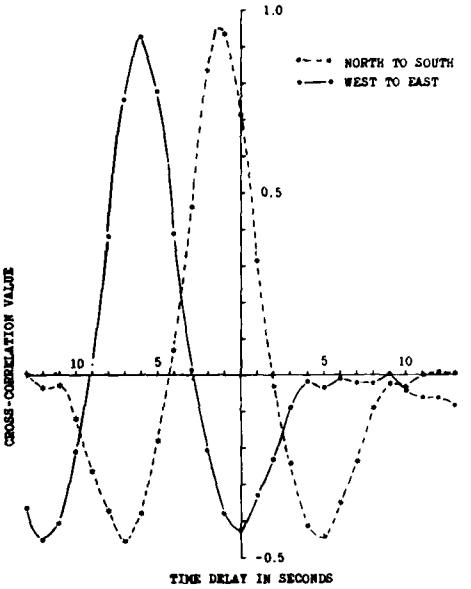


Figure 2. Cross-correlation values calculated for spaced receiver triangle on Guam for records from 1255 to 1300 GMT, 28 October 1971.

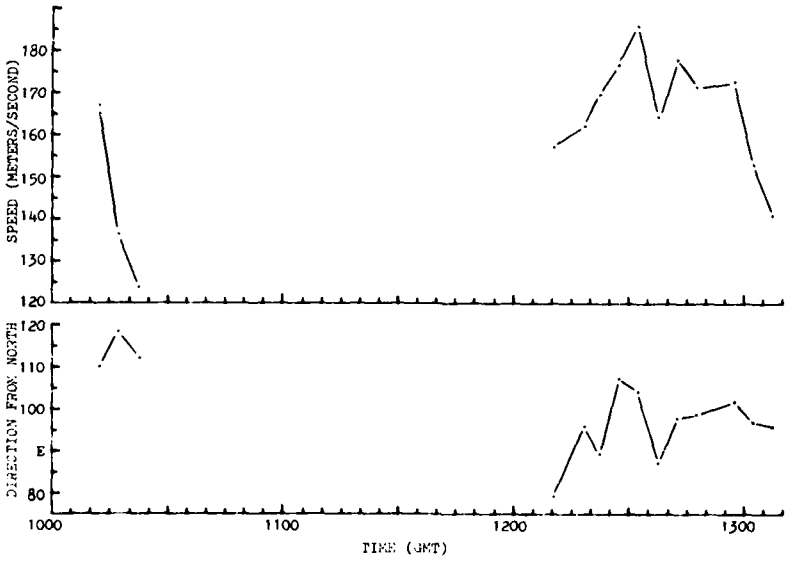


Figure 3. Speed and direction of movement of diffraction pattern for night of 28 October 1971 GMT at Guam.

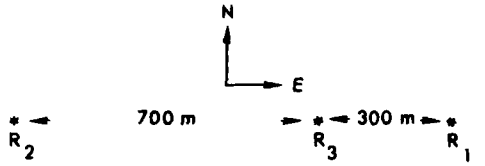


Figure 4. Dimensions and orientation of spaced receivers used at the Naval Communications Station, Guam in 1976.

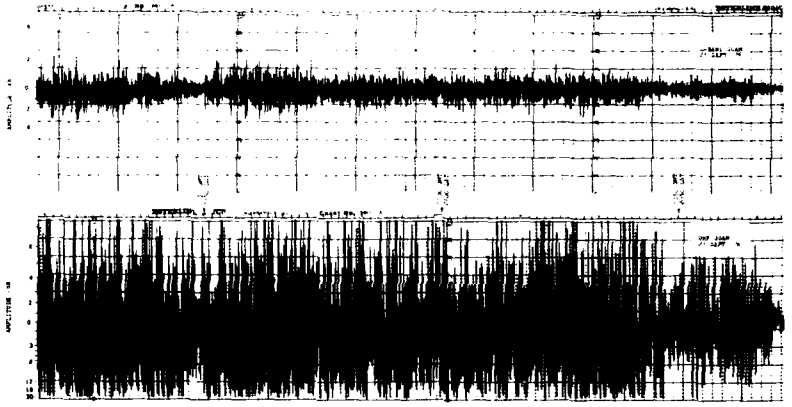


Figure 5. Chart record reproductions comparing L-Band and UHF scintillation fading.

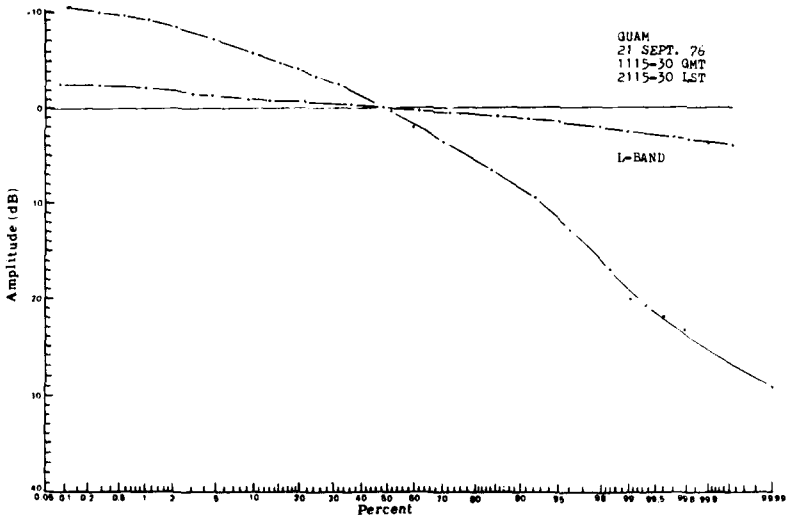


Figure 6. Comparison of UHF and L-Band scintillation intensity.

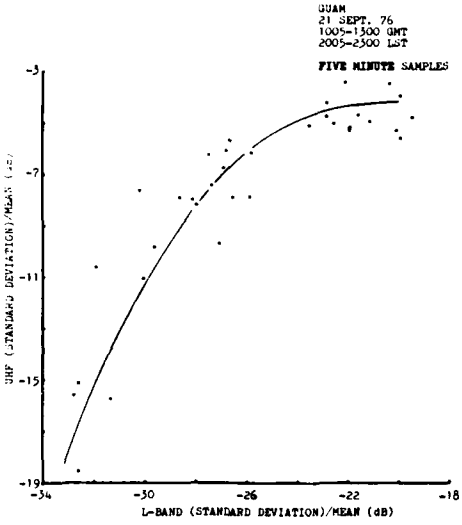


Figure 7. A comparison of UHF and L-Band scintillation intensity.

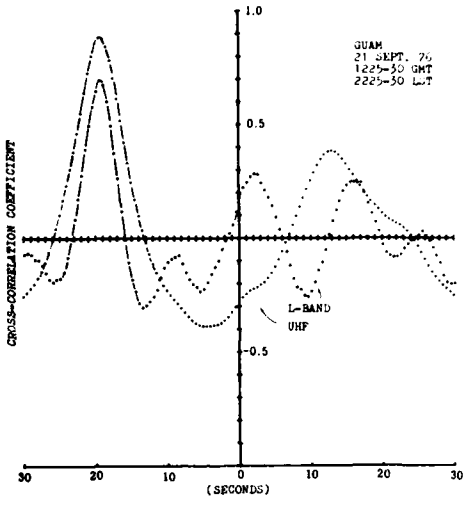


Figure 8. Cross-correlation functions for two wavelengths for a 1000 metre baseline.

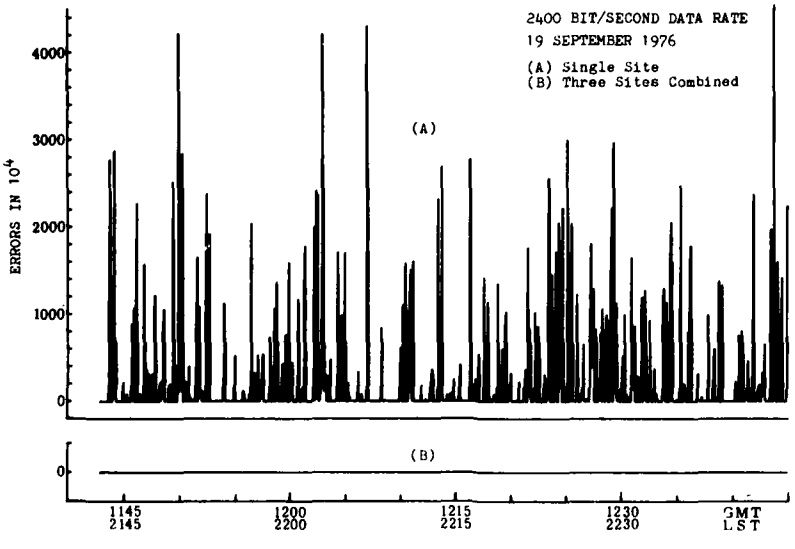


Figure 9. Down-link error rates recorded at Guam: A. No diversity. B. Using diversity.

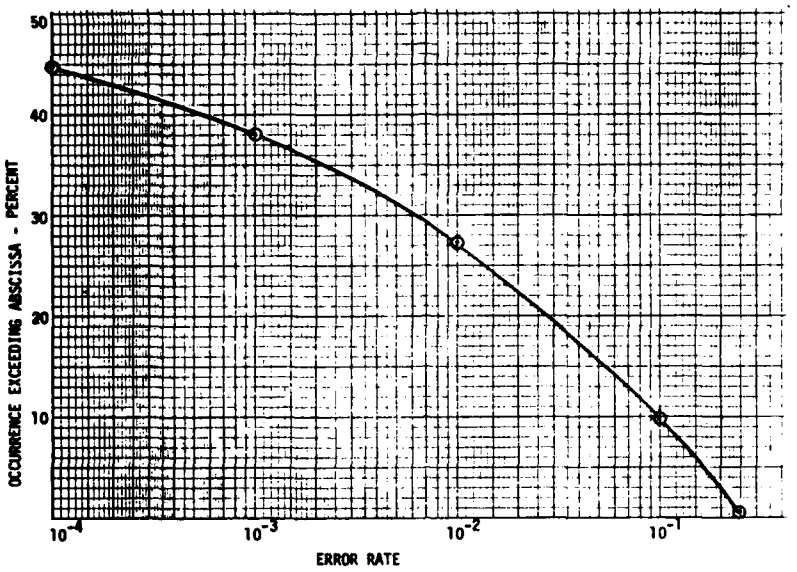


Figure 10. Data error rates during scintillation without the use of diversity.

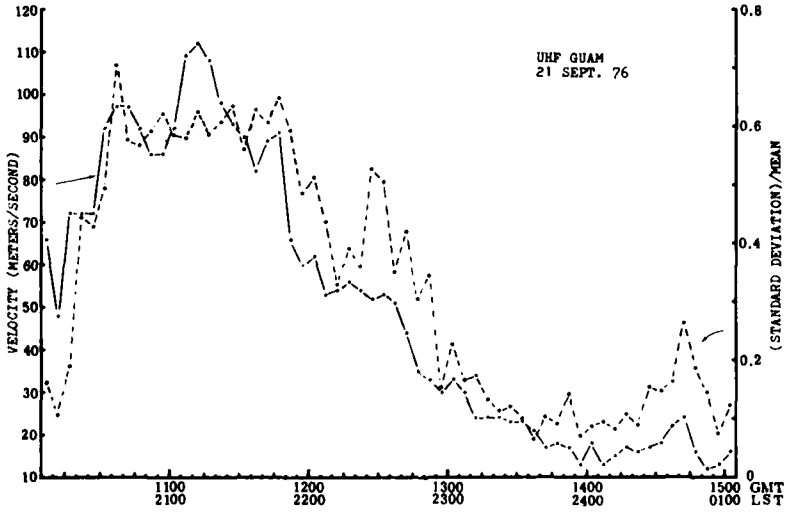


Figure 11. The east-west component of the apparent drift velocity, uncorrected for time decorrelation effects, is shown as a function of time and compared to the scintillation activity where the standard deviation divided by the mean is used as an indication of scintillation intensity.

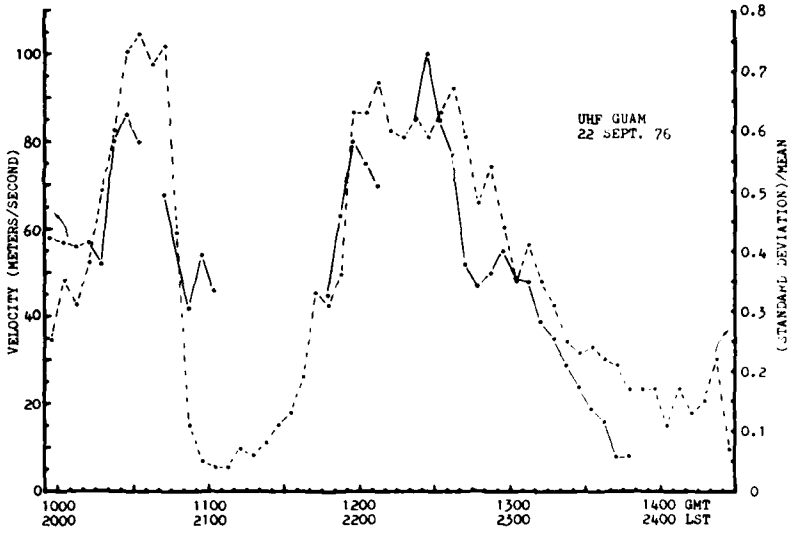


Figure 12. The east-west component of the apparent drift velocity, uncorrected for time decorrelation effects, is shown as a function of time and compared to the scintillation activity where the standard deviation divided by the mean is used as an indication of scintillation intensity.

SPACED RECEIVER MEASUREMENTS OF INTENSE EQUATORIAL SCINTILLATIONS

Herbert E. Whitney
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

INTRODUCTION

Satellite communication links can be affected by ionospheric scintillations particularly at frequencies less than 400 MHz and in the equatorial region. During the equatorial scintillation campaign of October 1976 and March 1977, spaced receiver measurements were made at Ancon, Peru to investigate several facets of scintillations which included a determination of the signal statistics of the scintillation events and a measurement of the temporal and spatial correlation functions. The signal statistics are necessary for evaluating the performance of communication links when affected by scintillations and the correlations functions are used to determine the effectiveness of diversity and coding techniques. The aim of this paper is to present some results which were obtained from the analysis of one of the tests, 19-20 Oct. 1976.

AMPLITUDE DATA

F-region irregularities develop shortly after local sunset in the equatorial ionosphere and show up as backscatter on the Jicamarca radar operating at 50 MHz and fluctuations of the LES-9 signal at 249 MHz as shown in fig. 1. The top panel is a reproduced range-time-intensity plot of backscatter of 3 m scale size irregularities detected by the Jicamarca radar. After sunset a localized depletion of electron density may form in the lower F layer and then rise and move into the middle and upper F layers. This upwelling results in "plumes" on the radar map and two "plumes" can be observed in fig. 1. The middle panel shows the time variation of scintillation index recorded on the propagation path from Ancon to the LES-9 satellite and the lower panel is the scintillation index recorded at Huancayo. The scintillation index was scaled in two minute intervals from the strip chart records. Once formed, the irregularity structure drifts eastward. Since the scintillations occurred first at Huancayo and then at Ancon, prior to the first "plume" on the radar map, irregularity structures must have formed first between Huancayo and Ancon, and then between Ancon and Jicamarca prior to the appearance of backscatter over Jicamarca.

SIGNAL STATISTICS

Magnetic tape recordings were made at Ancon and the results of the computer analysis for signal statistics during the time period 0030-0340 UT will be described. The strip chart record in fig. 2 shows the rapid development of scintillations at 0050 UT on two receivers spaced 366 meters on an east-west baseline. From an essentially undisturbed level, fluctuations reached more than 20 dB in about a one minute interval.

The data on magnetic tape was analyzed in 1.5 minute segments to obtain information on the S_4 index (square root of the variance of received power divided by the mean value of the received power), cumulative amplitude distribution (cdf), power spectrum, autocorrelation function and cross correlation function. To show examples of the analysis, the 1.5 minute section of input data shown in fig. 3 which occurred just after the developing phase of scintillations, will be used. Fading of the signal below the mean level was greater than 15 dB.

The cumulative amplitude distribution (cdf) for this period is shown in fig. 4. The cdf is a first order

statistic and is useful for defining the minimum margin requirements for the communications link of nondiversity systems. The S_4 index for this period was 1.17. Generally the S_4 index does not exceed unity except under foccusing conditions. The solid line is the experimental data and the dots are calculated for a Rayleigh distribution. Based on the analysis to date the Rayleigh distribution function provides a good fit to the observed data under the limiting conditions of intense scintillation. It can be noted that approximately 1% of the time the signal fades exceed 15 dB.

In addition to the information on the amplitude of the fades which is given by the cdf, a statistical description of the fading rate is needed to fully characterize the effects of scintillation on the communications channel. Information on the fading rate can either be produced by level-crossing techniques or by Fourier techniques which produce the power spectra and time correlation functions. The power spectrum for this data sample is shown in fig. 5. Generally the spectral shape consists of a relatively flat low frequency spectrum and a high frequency roll-off with a slope of approximately f^{-4} to f^{-5} for intense scintillations. The slope for the data in this sample period was -3.7

Fig. 6 shows the related autocorrelation function for this data sample. The autocorrelation function is another means of characterizing the rate of scintillation fading. It is the Fourier transform of the power spectrum and therefore has a width of the main lobe which is inversely proportional to the bandwidth of the power spectrum. The correlation interval or delay is listed in fig. 6 for various values of the correlation coefficient. The correlation interval was approximately 0.9 seconds for a correlation coefficient of 0.5. This parameter can be used to evaluate the effectiveness of time diversity techniques. Coding or interleaving can be an effective means of achieving time diversity improvement.

Since two antennas spaced on a 366 meter east-west baseline were used to record scintillation at Ancon, the spatial correlation function and drift velocity of the irregularities can be measured. A typical crosscorrelation plot is shown in fig. 7 for the same data sample on 20 Oct. 1977. The maximum coefficient was 0.88 at a delay of 2.39 seconds. Since the scintillation pattern at the west antenna leads the east antenna the velocity is eastward and has a value of 153 meters per second. While this is a typical value the drift velocity can vary from approximately 50 to 200 meters per second.

The variations with time of the S_4 index, autocorrelation interval (ρ=0.5), and the crosscorrelation coefficient for a 3¼ hour period (0030-0345) are shown in fig. 8. The S_4 index shows an abrupt rise at the onset of scintillations, and indicates the drift of several irregularity regions thru the antenna beam. The S_4 index reaches unity during the passage of the first two regions.

The autocorrelation interval was low ($\sim$ 0.5 seconds) following the onset of scintillations, but in general varied between 1 and 2 seconds. The bandwidth or rate of scintillation varied by more than a factor of 4. Generally the autocorrelation interval was lowest during the most intense scintillations. The data indicates that time diversity techniques would have to provide delays of a few seconds to significantly reduce the effects of scintillation.

The crosscorrelation coefficient also showed great variability, ranging from a low of approximately 0.2 following the onset of scintillations to almost unity under conditions of very intense scintillations. While the antenna spacing of 366 meters is sufficient to provide significant space diversity improvement for some periods of intense scintillation, a much larger spacing would be required to provide the necessary decorrelation for the entire period.

DIVERSITY TECHNIQUES

When correlation data are available it can be used to determine the advantages of using diversity techniques. Diversity techniques combine two or more signals that are fading independently to reduce the effects of scintillation fading. Autocorrelation data can be used to evaluate time diversity techniques; crosscorrelation data from multifrequency measurements can be used to evaluate the effectiveness of frequency diversity; and the crosscorrelation data from spaced receivers can be used to evaluate the effectiveness of space diversity.

When scintillations follow a Rayleigh distribution the signal will fade approximately 20 dB for one percent of the time. Depending on the degree of correlation between diversity signals some protection can be provided against the fading. Most of the diversity improvement is achieved by the time the correlation coefficient reaches 0.6. For example, the improvement at the 1% point is approximately 8 dB for $\rho=0.6$ and 10 dB for $\rho=0$ or complete decorrelation.

From a large collection of observations a limited amount of data has been presented on the amplitude and rate characteristics of intense scintillations in the equatorial region. Typical autocorrelation and crosscorrelation data was given which shows the feasibility of reducing the effects of intense scintillations.

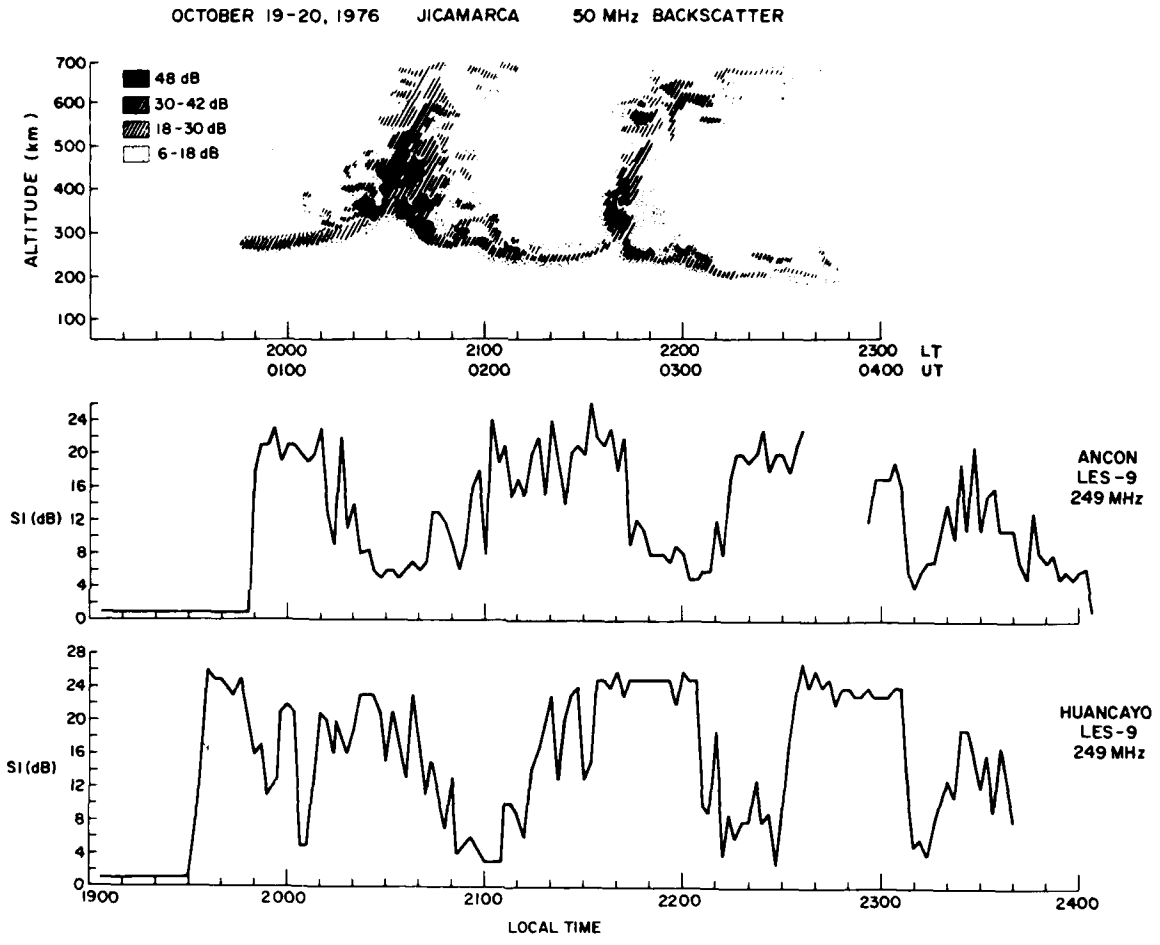


Fig. 1. Jicamarca radar backscatter map and corresponding amplitude scintillations observed at Ancon and Huancayo.

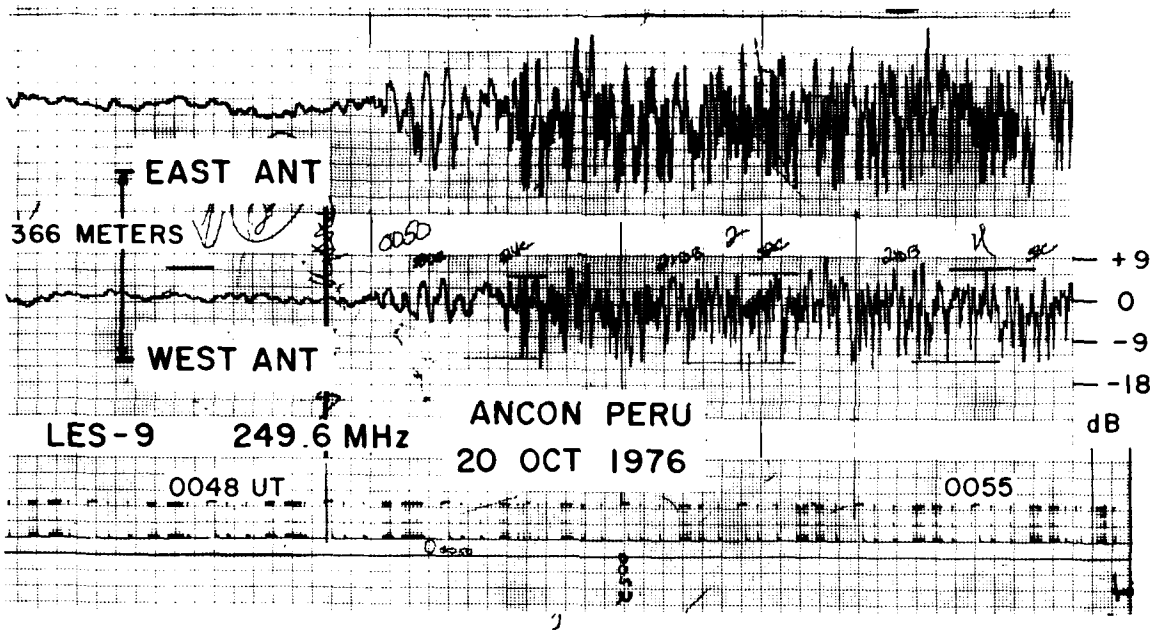


Fig. 2. Signal strength recording from spaced receivers at Ancon of LES-9 scintillations.

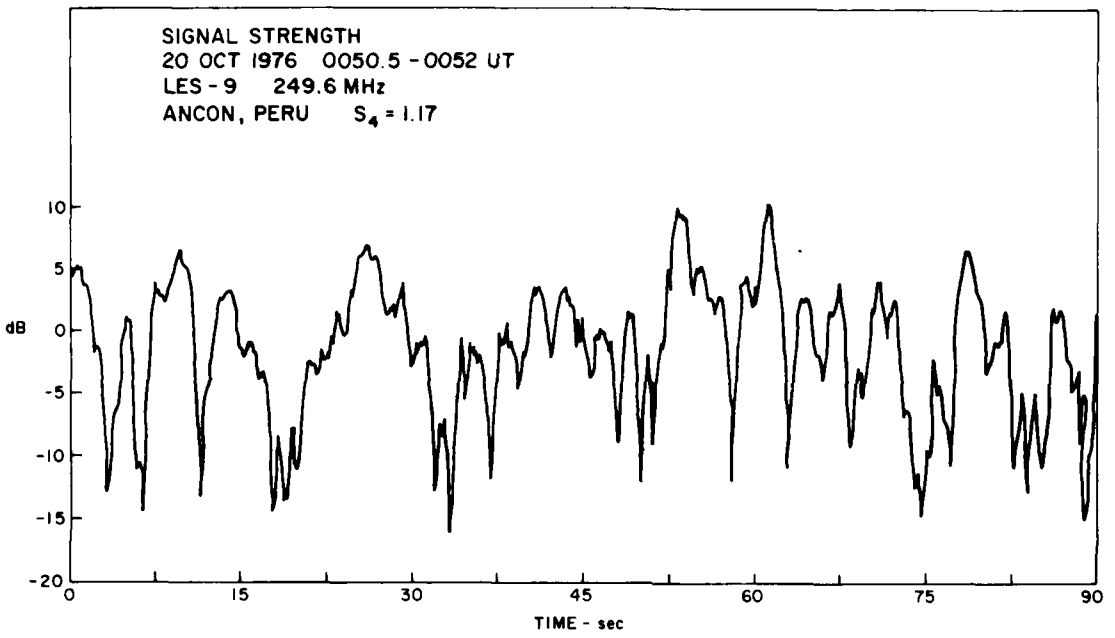


Fig. 3. Signal strength sample at start of scintillations.

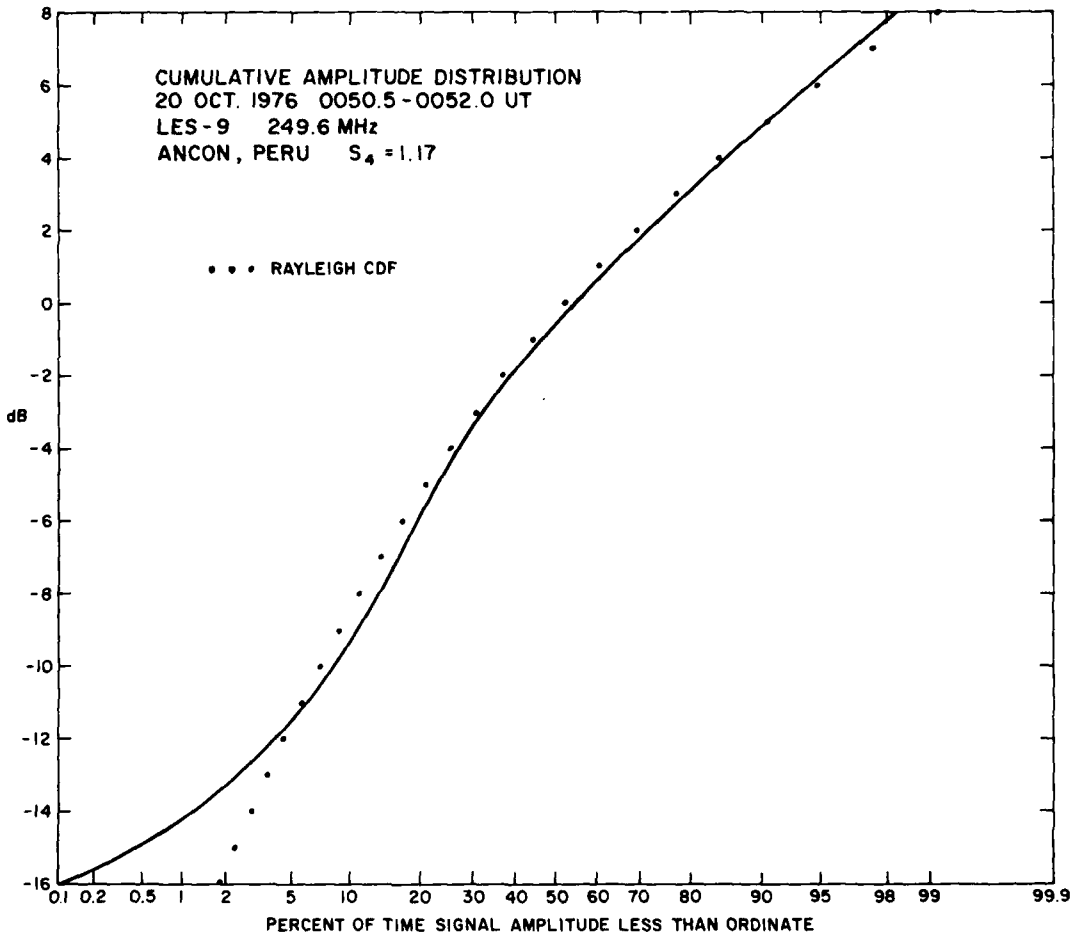


Fig. 4. Cumulative Amplitude distribution for input data shown in Fig. 3.

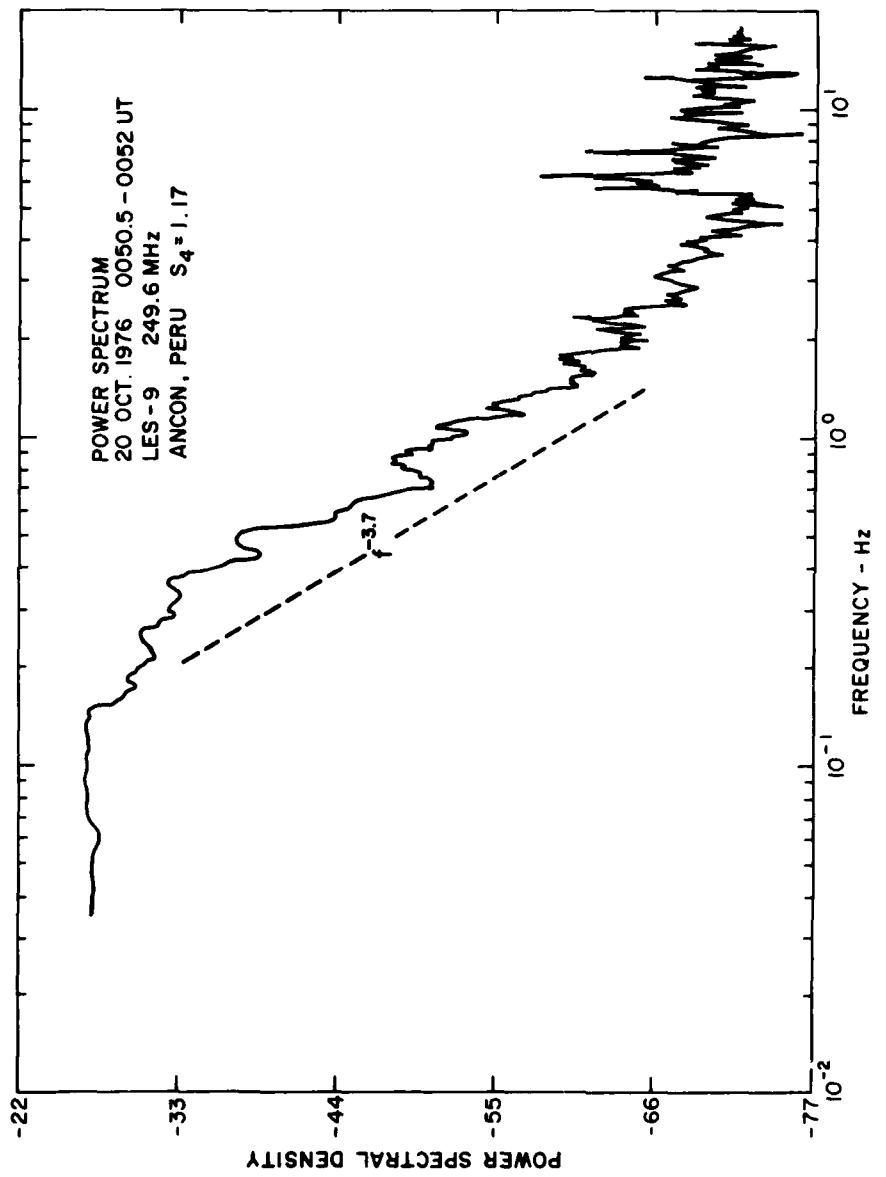


Fig. 5. Power spectrum for input data shown in Fig. 3.

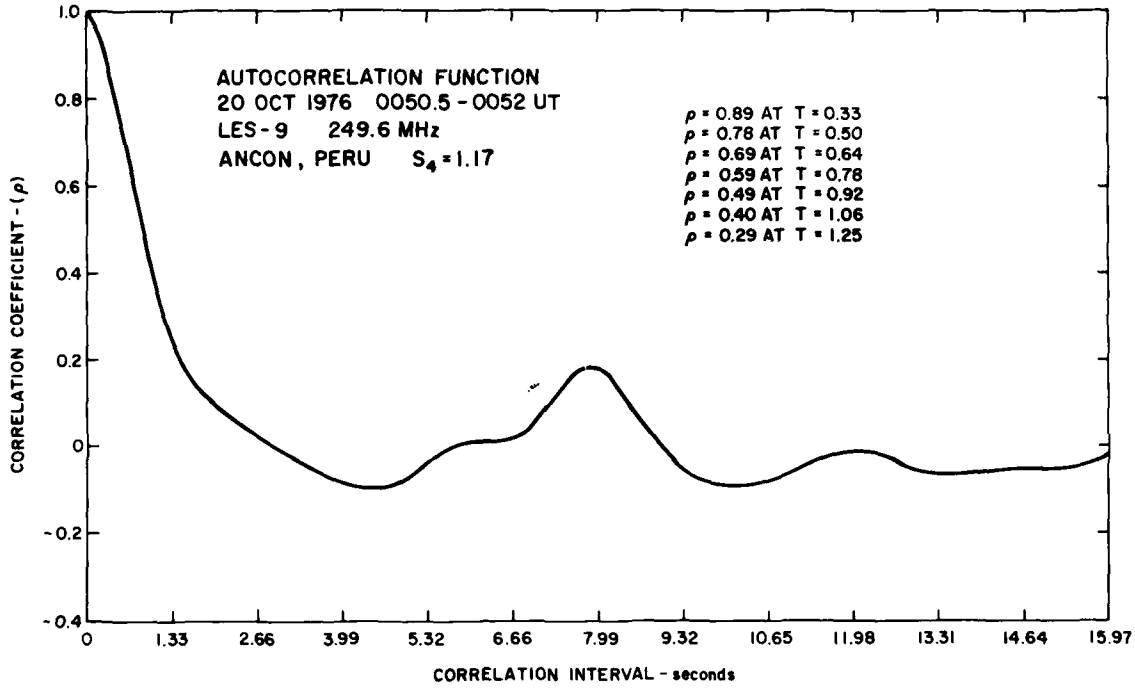


Fig. 6. Autocorrelation function for input data shown in Fig. 3.

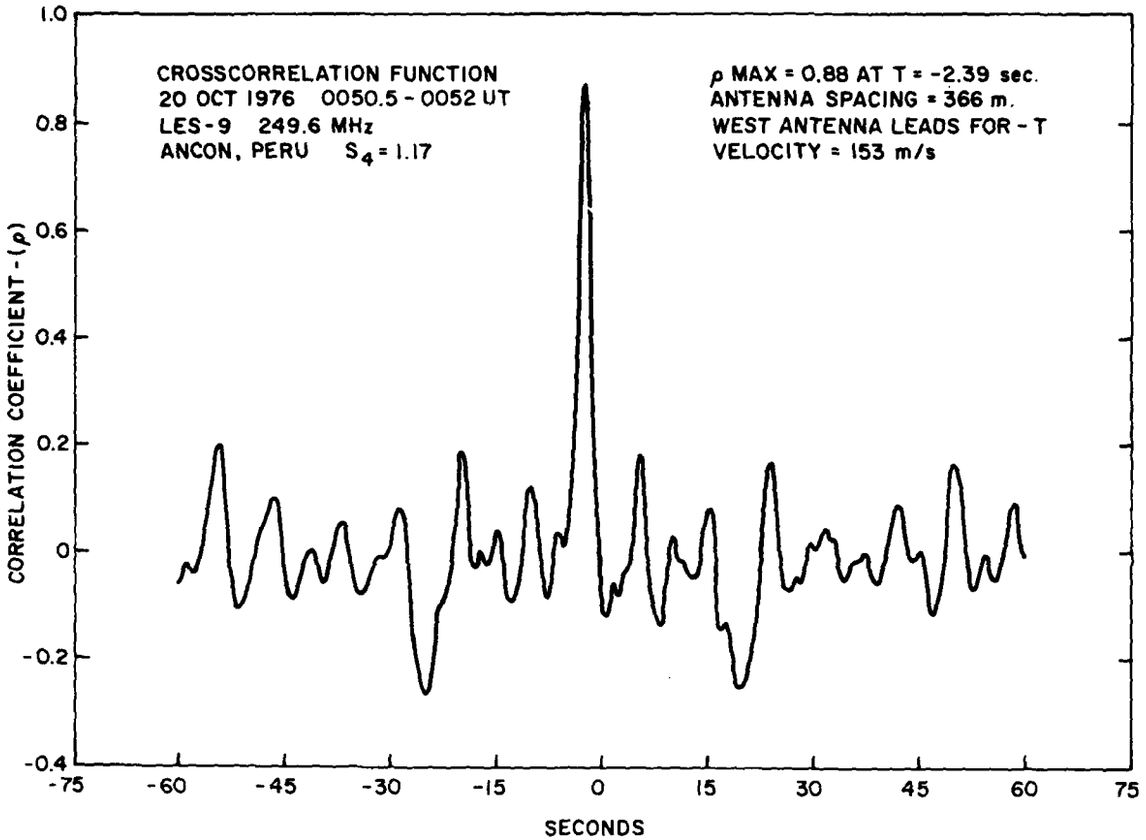


Fig. 7. Crosscorrelation function for data sample shown in Fig. 3 for two receivers spaced 366 meters on an east-west baseline.

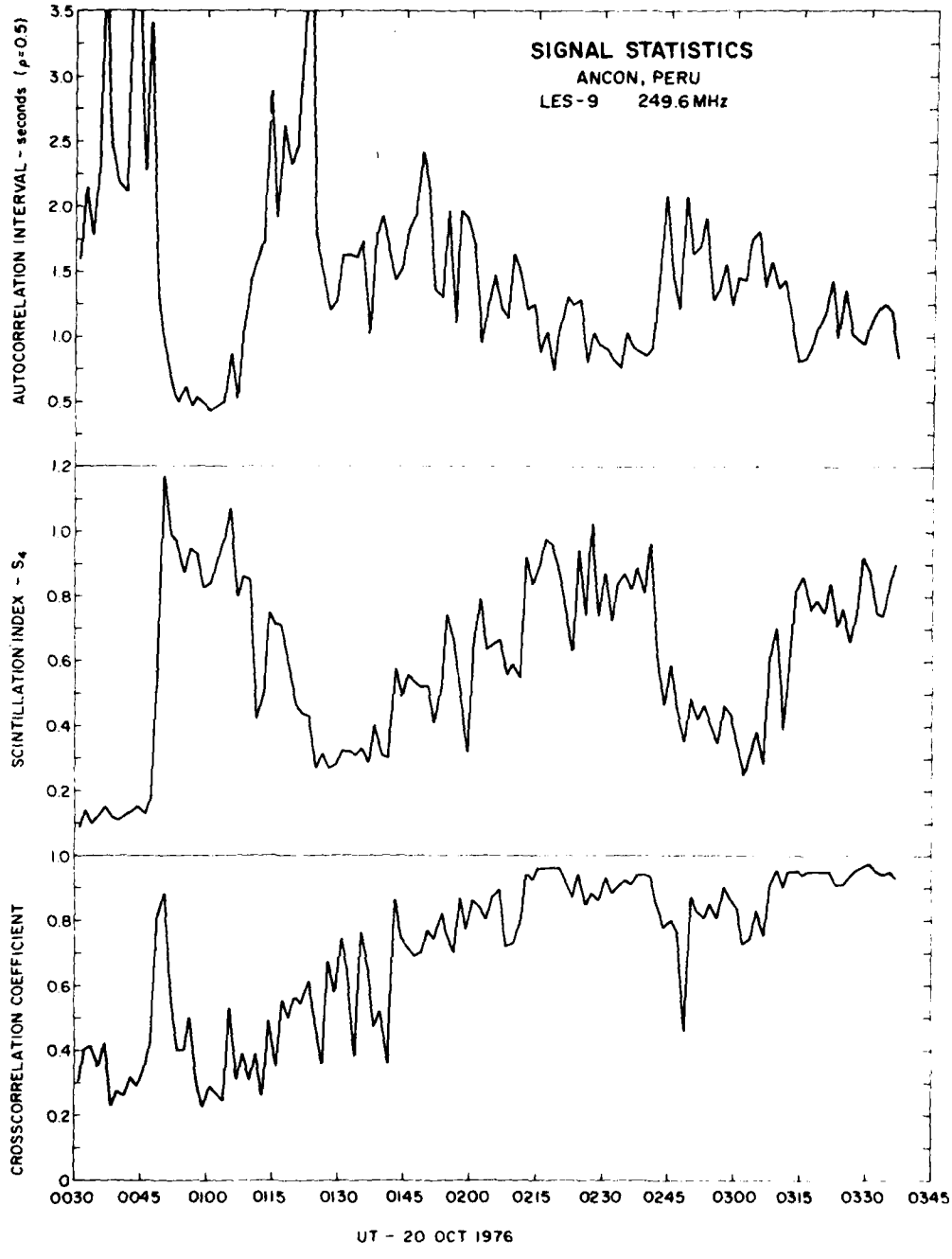


Fig. 8. Variation of S_4, autocorrelation interval, and crosscorrelation coefficient for an extended period of scintillations on 20 Oct. 1977.

MITIGATION FOR AIRBORNE SATELLITE RECEIVER PLATFORMS OF UHF IONOSPHERIC FADING BY THE USE OF SPATIAL DIVERSITY

Clifford W. Prettie
ESL Incorporated
Sunnyvale, California

Allen L. Johnson
Air Force Avionics Laboratory

ABSTRACT

Spatial diversity in the sense that two or more spaced antenna signals are combined, has been recognized as a viable mitigation technique of equatorial ionospheric fading of UHF satellite communications signals for land receiver platforms. The antenna separation for the technique to be effective must be of the order of 200 meters, and for this reason spatial diversity has been dismissed as a mitigation technique for airborne receiver platforms. If the airborne platform is over water, the sea bounce path, which typically is responsible for undesirable multipath fading, provides a signal path that is well separated in the ionosphere from the main path signal. A diversity combiner with inputs from both the usual upward looking antenna and a downward looking antenna for coupling to the sea bounce path will provide improved performance through scintillation. Results of performance improvement anlyses of the concept are presented. The concept will potentially be flight tested in the spring of 1978.

INTRODUCTION

The sea multipath spatial diversity concept is presented and discussed below. Results of a LES-8/9 performance simulation with sea multipath spatial diversity through ionospheric fading are then presented. Issues to be answered by flight testing are finally listed.

SEA MULTIPATH SPATIAL DIVERSITY CONCEPT

Spatial diversity is a recognized technique for mitigating the effects of ionospheric scintillations, however, for the technique to be effective spatial separation of the receive antennas must be on the order of at least 200 meters (Paulson and Hopkins, 1973). Antenna separations of this distance are impractical for use on airborne platforms.

It has been observed that 250 MHz satellite signals bounce off a sea surface efficiently. The sea bounce signal sometimes couples into airborne antennas strongly enough to cause sea multipath signal fading, generally considered a nuisance. Using reasonable assumptions and simple arguments, one can state that the power of a sea bounce satellite signal, originally circularly polarized, should be only 3 dB down from a main path satellite signal to an airborne platform. A horizontal polarization reflection coefficient of nearly 1 together with a vertical polarization reflection coefficient of 1/2 are reasonable values to assume. Effectively all of the horizontal polarization power plus one quarter of the vertical polarization power are consequently reflected. Ignoring earth curvature power spreading and the additional path length losses, it can be seen that slightly more than one half of the original signal power is reflected as a result; although it is, indeed, no longer circularly polarized. Thus, sea multipath fading should be expected if the antenna reception pattern favors neither the main path nor the sea bounce path signal. Typically, however, such is not the case, airborne SATCOM antennas favor signals arriving primarily from topside directions in order to avoid multipath fading.

Satellite sea bounce path signal energy arrives at an airborne platform from many different regions of the sea state. For this reason an antenna, mounted on the bottomside of the aircraft, and monitoring a satellite bounce path signal would observe a scintillating signal, because of the constructive and destructive interference of rays travelling different distances. Rician or Rayleigh amplitude fading would be, consequently, expected depending upon the roughness of the sea state. Note that the correlation of this fading over frequency is dependent upon the time delay profile of the energy received, the longer the delays the smaller the correlation bandwidth. The feature of the sea bounce path received signal of greatest significance for this paper, however, is that the sea bounce ray paths are well separated from those of a main path signal in transit through the ionosphere.

In the case of equatorial scintillations, signals received both by a sea bounce path monitoring bottomside antenna and by a main path monitoring topside antenna would exhibit amplitude fading. Because of the large ionospheric path separation, there is no correlation between the topside and bottomside fading. The sea bounce path signal provides a suitable signal for

diversity combining. This diversity source is available to airborne platforms whenever flying over large bodies of water or other highly (RF) reflective surfaces.

RESULTS OF A LES-8/9 SEA MULTIPATH SPATIAL DIVERSITY SIMULATION

The performance of the LES-8/9 UHF forward link using sea multipath spatial diversity was assessed in a detailed software simulation. Results of this simulation indicate that significant performance gains are possible with the sea multipath spatial diversity technique.

The LES-8/9 UHF forward link is a 75 bps wideband frequency hopping coded/interleaved 8-ary FSK system. The coding and full message interleaving used by the system are powerful mitigants for equatorial fading channels. Slow nonselective Rayleigh fading causes an effective power loss of only 8 dB at the 20 percent message error rate (40 character 4.4 second messages); much larger losses occur for less complex systems. Simulation results show that the use of equi-gain combining of the sea bounce path signal, however, can nevertheless restore the effective power loss of Rayleigh fading from 4 to 7 dB depending upon the assumptions involved. In detail, the actual performance gain for two different sea bounce power losses and for either frequency selective or nonselective fading were considered. A sea bounce power loss of about 3 dB is expected but a 6 dB case was also considered. Sea bounce fading is also expected to be highly frequency selective over the hopping bandwidth, but the conservative cases of nonselective sea bounce path fading was also considered. Diversity combining of a main path Rayleigh fading signal with the nonselective Rayleigh fading bounce path signals experiencing 3 dB and 6 dB of power loss gave performance improvements over the slow Rayleigh fading case with no diversity of 5 and 4 dB respectively. If the sea bounce path signal is assumed to be frequency selective the performance improvements become 7 and 6 dB for the same respective sea bounce power losses.

For the purposes of this simulation, the envelope of exponentially correlated Gaussian coherent and quadrature processes with 1 second 1/e correlation times was used to represent both the equatorial fading and the nonselective fading. For selective fading simulation the quadrature and coherent Gaussian processes were chosen independently from frequency hop to frequency hop (each hop lasts 5 ms. The doppler spread of the sea bounce signal and of the equatorial fading signal was also ignored.

An interesting fact was noted in the assessment of the simulation results. while the LES-8/9 performance loss induced by a slow nonselective Rayleigh fading channel is 8 dB, the loss induced by a highly selective Rayleigh fading channel is only 2½ dB. Thus, if the sea bounce power loss is only 3 dB, improved performance is potentially obtainable in equatorial fading by merely changing from a topside to a bottomside antenna.

Note that the performance gain results discussed in this section are for the LES-8/9 UHF forward link, a system in which strong fading mitigants exist already. Higher values of performance gain are possible for less complex systems, i.e., systems with no fading mitigants, at comparable performance levels. The LES-8/9 system performance was chosen for evaluation in preparation for a flight test of the diversity concept, discussed in the next section.

FLIGHT TEST

The Air Force Avionics Laboratory (AFAL) is fully equipped to test the sea multipath spatial diversity concept through ionospheric fading. Performance improvements predicted by the above analysis are difficult to achieve by other means without system modification. The simplicity and potential gain of the multipath technique motivate its flight testing, tentatively scheduled for April 1978.

Available to the AFAL is the NKC 135/662 test aircraft with LES-8/9 flight test instrumentation. By April 1978, a Dorne-Margolin crossed dipole will be mounted on the bottomside of this aircraft. A Motorola antenna combiner is also available for the test. AFAL has had extensive experience with flight test measurements of multipath and ionospheric fading. The tentative flight test in 1978 would be to Peru where active equatorial ionospheric scintillations have been consistently observed by others and by AFAL in the past.

The analyses of the above section opens questions which can be answered through flight testing. These questions are listed below together with their expected answers:

1. What is the actual performance improvement of the diversity technique in equatorial fading? Expected answer: 5 dB at the 20 percent message error rate.

2. What is the appropriate value for the sea bounce attenuation of LES-8/9 UHF signals? Expected answer: 4 dB ±2 dB, elevation angle dependent, value given is for 20 to 60 degree elevations.

3. To what degree is the sea bounce channel frequency selective? Expected answer: The channel decorrelates over 2 MHz at 30 degrees for nominal sea roughness. The correlation bandwidth is dependent on sea roughness, and slightly on elevation.

4. Is the sea bounce amplitude distribution Rayleigh both in ionospheric fading and

in no ionospheric fading situations? Expected answer: Yes, although it may be Rician for smooth sea no ionospheric fading conditions.

5. Is the sea bounce doppler spread significant enough to affect the results of this performance analysis? Expected answer: No.

6. Are there any other significant factors not included in the performance assessments? Expected answer: No.

CONCLUSIONS

Sea multipath spatial diversity can significantly improve performance through ionospheric scintillations. Predicted performance gains are 5 dB for the LES-8/9 system, a system that is already highly fading resistant. A flight test is planned in April 1978 to assess these predictions in equatorial fading.

REFERENCE

1. M.R. Paulson and R.U.F. Hopkins, Effects of Equatorial Scintillation Fading on SATCOM Signals, NELC/TR 1875, Naval Electronics Laboratory Center, San Diego, CA, May 1973.

ACKNOWLEGEMENT

The above work was sponsored in part by the Defense Nuclear Agency under Contract # DNA 001-76-C-0152.

A TIME DIVERSITY EXPERIMENT FOR A UHF/VHF SATELLITE CHANNEL WITH SCINTILLATION

David P. White
M.I.T. Lincoln Laboratory
Lexington, Massachusetts 02173

I. INTRODUCTION

Satellite communications in the UHF and VHF bands has at various times and in certain receiver locations on the earth's surface been subject to moderate to severe degradation as the result of ionospheric scintillation effects. This report describes the use of a particular time diversity method to provide essentially error free communications in such a degraded environment. More specifically, it is designed to protect a broadcast message at a modest data rate.

M.I.T. Lincoln Laboratory's efforts in this direction has proceeded through a number of stages. First a significant amount of the existing scintillation data (pre-1975) was reviewed and often further analyzed to derive the propagation channel parameters needed to "size" a reliable *broadcast communication system*. With the channel parameters defined within reason, a number of competitive schemes were analyzed usually with the aid of computer simulations. This stage is well documented by a colleague, Bucher [1]. Suffice it to say that the use of bit interleaving and convolutional encoding at the transmitter coupled with deinterleaving and Viterbi decoding at the receiver seemed a very promising solution. It should be reemphasized that our study is aimed at broadcast communications only. This implies that the individual receivers do not acknowledge reception of messages, good or otherwise. Moreover, finite message delays (principally due to the interleaving process) are tolerable. It is also unstated, but central to our study, that an acceptable post-processor error rate is less than 10^{-5} as contrasted to the degraded channel pre-processor error rate of 10^{-2} or higher. The error protection built into redundant human constructed text is not considered as it is in the case of a John Hopkins Applied Physics Laboratory study [2].

*This work was sponsored by the Department of the Navy.
The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

The second stage of the study consisted of the construction of proto-type error correcting equipment coupled with a series of experimental evaluations. The final testbed was designated to be the U.S. Navy Fleet Broadcast which is currently sent over the Marisat (Gapfiller) satellite system. The equipment designed had to interface on one end with the Navy's transmitter and on the other end with the standard Navy AN/SSR-1 broadcast receiver. The addition to the transmitter was essentially a series of interleavers and a convolutional encoder. The receiver additions were more substantial; the main elements being the deinterleavers, Viterbi decoder and data recording system. Upon completion of construction, an extensive hardware simulation was run with a simulated fading channel. This further refined the choices of receiver parameters and provided the basis for the field experiments. This second stage is fully documented in the report by White [3]. A number of the relevant details are also given in Part V of this report.

The third stage provides most of the content of this report. A description of the results from the Hawaii to Gapfiller to Guam link experiments is detailed and conclusions are drawn based on this field experiment along with the earlier hardware channel simulations.

II. A SYNOPSIS OF UHF SCINTILLATION MORPHOLOGY

The geographical areas experiencing moderate to severe degradation due to scintillation effects are confined primarily to the equatorial regions and to a lesser extent from the sub-auroral to the polar cap region. For the equatorial region (more specifically a band centered on the geomagnetic equator) the scintillation is largely a nighttime phenonmenon beginning shortly after sunset and diminishing after midnight. In the auroral-polar region the occurrence of scintillation also exhibits a diurnal behavior with peak scintillation periods being near local midnight and the lower amplitude scintillation periods being near local noon. The transitions from a peak scintillation period to a quiet period are less dramatic in these northern regions due partially to the less clearly defined transitions between daytime and nighttime ionospheres. Aarons, et al., (1971), [4] have

written a comprehensive review of the experimental results for both regions and more recently (1976) Aarons [5] has updated this for equatorial scintillation. Crane [6] has reviewed the application of scattering theory to weak and strong scintillation models and then has compared the functional form of the results to experimental data with a high degree of success. The analysis in his paper was the seed which led to the development of our set of experiments.

For our purposes, it is sufficient to say that the scintillation is associated with irregularities in the effective refractive index of the F region. Moreover these irregularities tend to be field aligned with axial ratios of about 50. These irregularity patterns moving transversely across the propagation path act as a diffraction grating and cause a fading rate proportional to the scale size and the horizontal velocity of the irregularities. Generally the scintillation correlates well with the appearance of "spread F" (the name originating from the phenomenon of radar pulse broadening due to multipoint wave scattering by F region irregularities). The general characteristic of equatorial scintillation is as follows. Shortly after sunset the received signal may change abruptly from a constant level to one which exhibits deep fades (fades of greater than 20 dB with a duration of about 1 sec are not uncommon). Later in the evening (around local midnight) the scintillation becomes more sporadic with lesser fade depths and finally in the early morning hours the fades may last for 10's of seconds but with fades depths closer to 5 to 6 dB.

The occurrence and degree of scintillation depends not only on latitude and time of day as mentioned above but also shows a strong dependence on longitude, earth's magnetic field activity, solar cycle and proximity to the equinoctial period. With regard to latitude, Basu, et al. [7] from OGO 6 data showed that (for Nov. 1969 at least) the probability of scintillation occurrence was greatest over the west African continent to South American sector and least in the far western Pacific. The scintillation is seasonal being greatest at the equinoxes (March 21 and September 23) and minimum at the soltices. (On the other hand, Nichols [8] shows that data from Kwajalein in the Marshall Islands showed its peak during the summer soltice of 1971.) The dependence of scintillation in the solar cycle is not well defined but there does seem to be a correlation of strong scintillation in active years.

In our experiment we chose Guam as our receiver site (magnetic latitude 11°) based on logistics and its superior telephone communication capability with respect to the transmitter complex. The transmissions were scheduled between 0900 to 1500 GMT (2000 to 0100 Local Guam time) and the equipment was deployed to cover the autumnal equinox period. The unfavorable factor is that the experiment was run in a low point in the solar cycle. Crane has stated that at 245 MHz, scintillation with fades greater than 6 dB occur less than 15% of the time at equatorial latitudes. Our experience (July 26 to December 31, 1976) indicated that the scintillation was even less frequent and this was confirmed by R. Hopkins of NOSC who was conducting spatial diversity experiments in the same time frame and at the same location.

Paulson and Hopkins [9] in their diversity experiments were able to develop error free communications with receivers spaced by about 700 meters but clearly this technique is not applicable for a small receiver platform. Crane [6] had already shown that neither polarization nor frequency diversity techniques were useful (at least for bandwidths less than 10 to 30 MHz). The experiments we ran were only concerned with the use of time diversity to mitigate against the effects of scintillation on broadcast communication.

III. Experimental Goals

Heller and Jacobs [10] have shown that convolutional encoding and Viterbi decoding in conjunction with DPSK modulation is an efficient and reliable means of communicating over a satellite channel in the absence of scintillation. However under scintillation fading conditions, the received bit stream would exhibit large groups of errors so that the message could not be decoded with the Viterbi decoder. The obvious solution was to randomize the errors by interleaving the data before sending it to the transmitter encoder and upon reception deinterleaving the data prior to presenting it to the Viterbi decoder. From prior analysis of data (Crane [6] and Bucher [1]) it was decided that it would be sufficient to protect against severe fades of duration of approximately 1 sec. The convolutional interleaver design selected takes continguous bits in and provides a separation between them of about 1.7 sec. Details can be found in White [3]. The penalty for the interleaving and deinterleaving structures implemented was the introduction of a channel delay of about 2 minutes. Both our computer and hardware simulations as well as the field experiments confirmed the choice of interleaver size as being a conservative but not an unwieldly design (the design required about 4000 bits of storage per teletype channel per soft decision bit in each of the transmitter and receiver sections).

After selection of the interleaver structure it then remained to specify the Viterbi decoder parameters. The constraint lengths were chosen to be K = 7, 8 or 9 for the code rates 1/2, 2/3 and 3/4 respectively. The relatively small number of states to be examined (2^6) meant that all the computations could all be done in serial and still allow a relatively unsophisticated controller for the decoder. The decision to use "soft" decision inputs (3 bit) in addition to two bit AGC scaling provided even further

protection with most of the hardware addition being deinterleaver memory. An evaluation was also made on the relative merits of systematic versus non-systematic codes. Systematic codes are those in which the original information is sent along with parity check bits which provide the necessary redundancy. Figure 1 shows the interleaving structure for a receiver that can protect up to three parallel (before multiplexing) data streams. It should also be obvious that

detection of the received signal is possible. If, on the other hand, the phase variation is large over a time span of many bits but the phase change between successive bits is relatively small then differentially coherent detection is preferred. For this we modified the circuitry to utilize the quadrature arm information as well. This resulted in a choice between use of a coherent versus two differentially coherent combiner for each test.

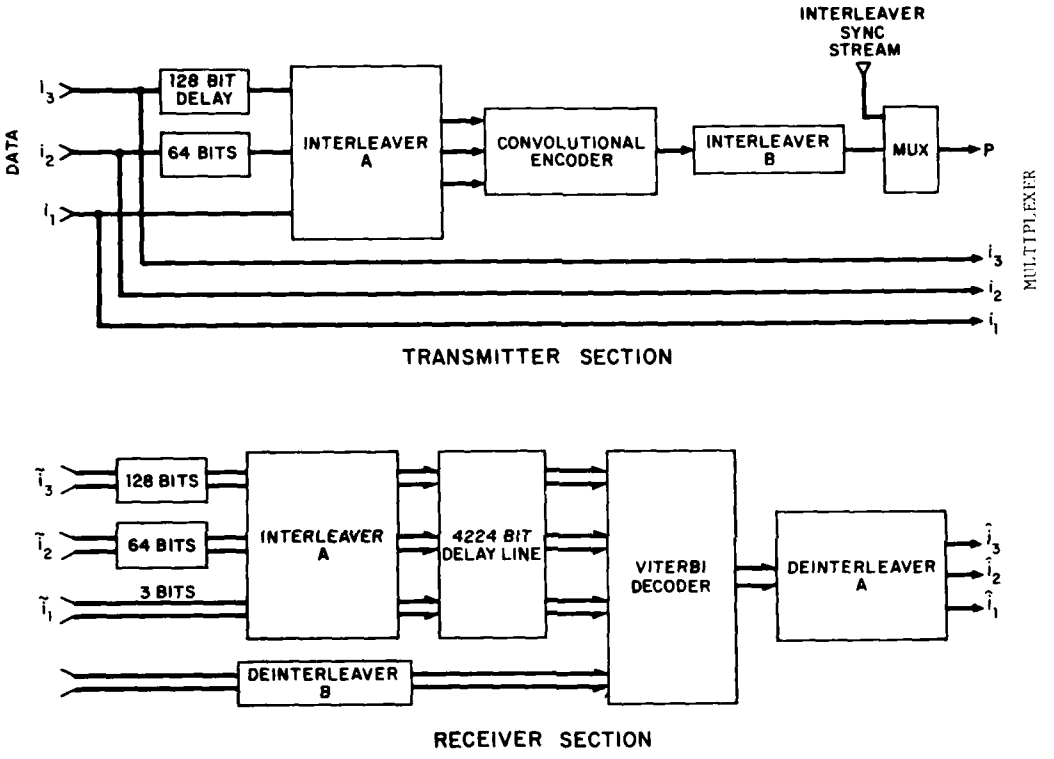


Fig. 1. Interleaver Configuration for Systematic Codes.

even an unmodified receiver (with no interleavers or error correcting hardware) can still recover the data under favorable channel conditions. The implication, of course, is that one needs only to modify those receivers that operate in the strong scintillation zones. The non-systematic codes offer even more protection for the same E_b/N_0 (energy per information bit/noise intensity) but in this case all the transmitted bits are encoded and one will always need a receiver with deinterleaving and decoder components (see Fig. 2).

There was also some question concerning the magnitude and effect of scintillation on phase variation. In the receiver used for the experiment, the Navy AN/SSR-1 (UHF Broadcast Receiver) possesses a Costas phase-lock loop of which only the in-phase arm is used in bit synchronization. This assumes coherent

Two points should be made concerning Fig. 2 (which was derived from computer simulations). First the Rayleigh fading assumption leads to a pre-processor error rate which decreases very slowly with increasing $\bar{E}_b/N_0$ ($\bar{E}_b$ is the average energy/information bit). This clearly shows the need for some type of error correction. Second with the rate 1/2 code, it is observed that for a value of $\bar{E}_b/N_0 > 12.5$ dB, the post-processing error rate will be negligible and non-measurable with any degree of statistical significance. For example, at a data rate of 75 bps, it takes almost four hours just to receive 10^6 bits and scintillation is seldom uniformly intense over such an extended duration. For the particular link used (Gapfiller to Guam) the energy per received symbol to noise ratio ($\bar{E}_s/N_0$) is about 21.4 dB and at a rate 1/2 code the energy per information bit to noise ratio ($\bar{E}_b/N_0$) is 24.4 dB. In

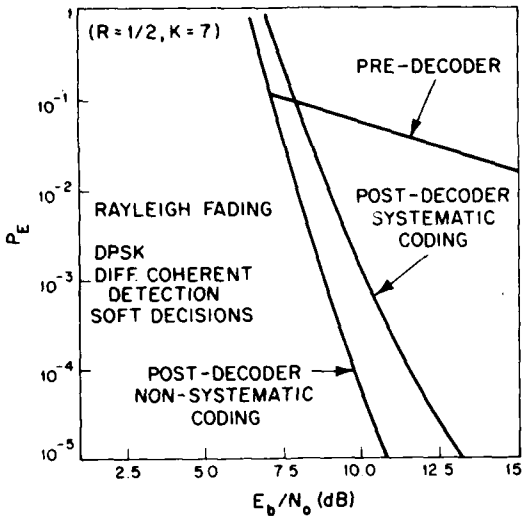


Fig. 2. Comparison of P_e for Systematic and Nonsystematic Codes.

this situation, the result of the field experiment is binary, either the system works perfectly (no post-processor errors) or there is some major problem in the hardware or some major incorrect assumption concerning the fading phenomena. It should be emphasized that extensive hardware simulations were done in the laboratory with a Rayleigh simulated channel and where there was an opportunity to measure the error rates for selectable values of E_b/N_0. Within the experimental limitations the agreement between the hardware simulation curves and the computer simulation curves was generally good. We will expand the point further in Section V.

At this point, twelve different combinations need to be evaluated (the three data rates, systematic or non-systematic codes and coherent or differentially coherent combining). The receiver configuration shown in Fig. 3 can operate in any of these modes. A key feature of the experiment is the inclusion of the data collection facility. This facility routinely records on digital magnetic tape a number of key parameters such as the following:

a) all pre- and post-decoder bit errors and accumulated frame error rates

b) decoder metrics

c) P_r/N_0 ratios, AGC values

d) dynamic phase errors from Costas loop

e) in-phase and quadrature soft decision outputs for a single 75 bps channel

All the tapes are ultimately processed on the IBM 370.

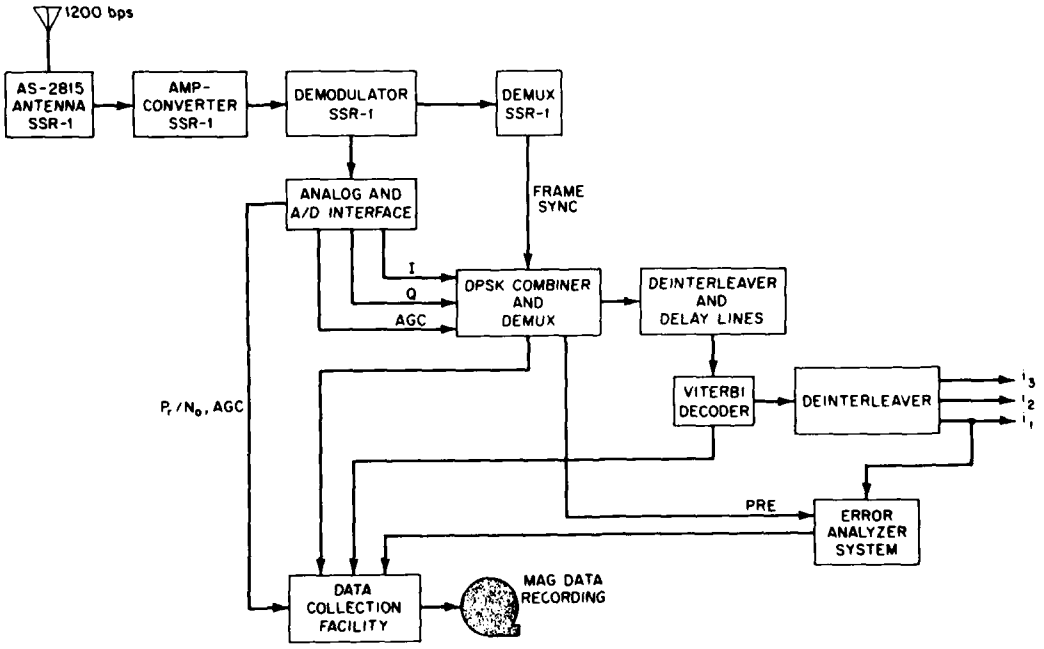


Fig. 3. Receiver and Data Collection.

IV. EXPERIMENTAL LIMITATIONS

The experimental broadcast was originally scheduled to emanate from the Navy CAMS at Hawaii (Communications Area Master Station) up to the Gapfiller (Marisat) satellite and then down to a receiver site located in a strong scintillation zone, in this case Guam. The intention was to utilize a spare transmitter multiplexer (TDM 1150) and simulate our own fleet broadcast which would then occupy the spare 25 kHz narrowband transponder channel (Narrowband α) of Gapfiller. The Fleet Broadcast format consists of fifteen independent 75 bps teletype channels which are multiplexed together (along with a 75 bps frame sync sequence) to give a 1200 bps data stream which is then DPSK modulated. Our time diversity scheme provides error correction by utilizing a number of the 75 bps streams as parity check information. For example if we want to protect one 75 bps teletype channel, the data stream could be convolutionally encoded such that for every single information bit into the encoder we get 2 bits out (this being the rate 1/2 code). These two bits would be routed in parallel (as elements of separate 75 bps streams) to the transmitter's multiplexer. If one of the streams retained its identity we would have what we earlier called systematic encoding but if both bits were encoded then the term is non-systematic. An extension to protecting two and even three independent 75 bps streams by adding an additional 75 bps stream leads to rate 2/3 and 3/4 codes (which can also be configured as systematic or non-systematic) is obvious. The essential point is that in order for the decoding to be done correctly, it is absolutely necessary that the parallel streams remain synchronized (that is there be no relative "slips" between streams) prior to the multiplexing. It is an unfortunate fact that the Navy's multiplexer with their elastic buffers on many occasions were observed to slip at least one stream relative to the others. This behavior has no effect on the normal fleet broadcast operations but is disasterous to our proposed mode of encoding. There have been a number of attempts to modify the TDM 1150 to remove this behavior and it may now be finally fixed.

Our operational period extended from July 26 to December 31, 1976. This is also the period in which the Pacific Gapfiller Satellite was being evaluated and both the CAMS EASTPAC and WESTPAC were undergoing the normal load of operational startup problems. Their primary responsibility was to maintain the fleet broadcast and other net operations first and then, if there were any assets left to satisfy our needs. It turned out that most of the time there was no spare multiplexer available for our use and we had to take whatever spare channels were available on the actual fleet broadcast. The logistics of obtaining an appropriate channel setup every night was difficult. In addition, a Navy radioman who was not familiar with the experiment patched in the transmitter encoder on Hawaii every night in a different way due to necessary operational changes. The Lincoln Laboratory representative on Guam manning the receiver complex tried to coordinate the above tasks over a much less than desirable phone link and often compromises had to be made.

On August 16, a simple 5V power supply failure occurred and the transmitter encoder drawer was shipped to Guam for repair by our representative. In order to obtain better operational control over the experiment we decided then to both transmit and receive from Guam. In October we sent a Laboratory representative to Hawaii to coordinate the encoder patching and then transmitted from Hawaii from October 9 through October 26 when further TDM problems arose. The rest of the transmissions then originated on Guam through the end of the experiment.

On Guam the problems of operational logistics were considerably simplified. However, we were often required to operate on one of the subchannels of the 500 kHz wideband channel of Gapfiller. We operated on one of the wideband channels from September 3 to September 29 and occasionally thereafter. It so happened that some of the strongest scintillation occurred while we were in the mode. It should be noted that we have introduced two more variables here. First the uplink transmissions as well as the downlink undergo some degree of scintillation and secondly the wideband channels lead to all the usual power sharing problems.

V. EXPERIMENTAL RESULTS

Of the 128 recordings, only ten of these nights showed appreciable scintillation. Appreciable scintillation is defined here as those tapes which show more than several hundred pre-decoder bit errors in a six hour recording period. These errors resulted from fades as deep as 25 dB and fade durations of approximately a second. They were also twenty-one tapes with light scintillation and consequently only a few hundred pre-decoder bit errors over the duration. (100 bit errors in one hour is an error rate of 3.7×10^{-4} which is more than acceptable for teletype of English text.)

The data for October 6, 1976 (tape 49-9: 19:00 through 14:58:00 GMT) is presented in some detail and is representative of all the heavy scintillation data recorded. The uplink originated from Guam and operations were over the narrowband satellite channel. The encoding was rate 1/2 systematic and the detection was performed coherently. Link calculations for the narrowband transponder channel yield a received power to single sided noise intensity ratio of 52.2 dB. A segment of the received signal is shown in Fig. 4. It is noted that the scintillation is terminated abruptly at about 10:17:00 which is somewhat unusual since the decay of scintillation usually is more gradual. The onset of scintillation had begun at 9:59:00 giving a scintillation period of 18 minutes. A further

10 minute scintillation period occurred beginning at 11:28:00 GMT (giving a total of only 28 minutes of heavy scintillation on the 6 hour recording period). In the figure deep fades (26 dB) with durations close to 1 sec are observable.

Detailed statistics on each of the heavy scintillation periods were collected. Crane [6] and Whitney et al. [11] have shown that the Nakagami-m probability density function [10] is a suitable amplitude fade distribution during UHF scintillation. This distribution is given by

$$p(a) = \frac{2m^m (a)^{2m-1} e^{-ma^2/\overline{a^2}}}{\Gamma(m)(\overline{a^2})^m} \qquad (1)$$

where $a = \bar{v} - v$, v is the received amplitude and $\bar{v}$ is the average of the received amplitude. The corresponding power probability density function is then equal to

$$p(P_r) = \frac{m^m (P_r/\bar{p})^{m-1} e^{-mP_r/\bar{p}}}{\Gamma(m)\bar{p}} \qquad (2)$$

where $\bar{p}$ is the average power. The variance of the average received power can be shown to be

$$\mathrm{var}\left(\frac{P_r}{\bar{p}}\right) = \frac{1}{m} \qquad (3)$$

For a value of m = 1, the above density functions reduce to the well known Rayleigh and exponential density functions.

$$p(a) = \frac{2a\, e^{-a^2/\overline{a^2}}}{\overline{a^2}} \qquad (4)$$

$$p(P_r) = \frac{1}{\bar{p}} e^{-P_r/\bar{p}} \qquad (5)$$

The sampling rate of the received signal was 50 samples/sec. The m values were routinely computed via equation 3 and indicate how close to Rayleigh the scintillation was. As an example for October 6, an analysis of a fifteen minute segment of data showed a value of m = 1.9 which indicates that the scintillation process is not far from Rayleigh distributed. In summarizing all the heavy scintillation data during the four months, m values run from m = 1 to as large as almost 3 with the lower range being more prevalent.

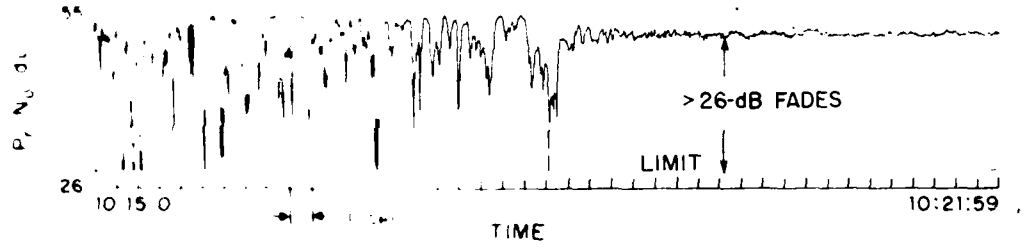


Fig. 4. Received Signal ... (Rate 1/2, Sys., Coherent Det.).

In addition to the amplitude statistics, it is also necessary to develop some statistics on the fade duration and interfade duration distributions. These statistics have a direct bearing on the amount of interleaving memory that one must employ in order to truly randomize the errors. In particular, cumulative sample probability distributions for both the fade durations and the interfade durations are taken for a series of preselected threshold levels (which usually are referenced with respect to the average $\bar{P}_r/N_0$ for the period analyzed). As pointed out by Bucher [1] the corresponding probability density function can be fairly well approximated by an exponential distribution and our recent results have confirmed this. This distribution is

$$p(T) = \frac{1}{\bar{T}} e^{-T/\bar{T}} \qquad (6)$$

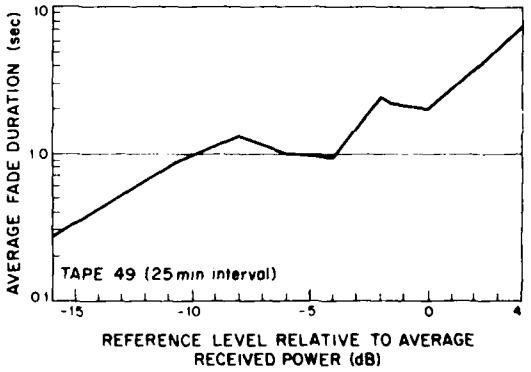


Fig. 5. Average Fade Duration vs. Reference Level.

where $\bar{T}$ can be either the average interfade duration or average fade duration. $\bar{T}$ itself is a function of the reference level. The analysis program not only plots the cummulative distribution functions and computes the fade and interfade averages at each threshold level but it also provides plots of the averages as a function of threshold level. Figure 5 shows the results for the average fade duration, again for October 6. The key feature of this figure is that even down to 12 dB below the average value of $\bar{P}_r/N_0$ the average fade duration is about 1.0 sec. The corresponding average interfade duration (curve not shown) is about 5 sec. It should be noted that for some of the heavy scintillation days the average fade (interfade) duration versus reference level curves can be well approximated by simple power low formulas (i - interfade, f - fade duration).

$$\bar{T}_i = (\bar{T}_o)_i \left(\frac{p}{p_o}\right)^{-n}$$

$$\bar{T}_f = (\bar{T}_o)_f \left(\frac{p}{p_o}\right)^{\gamma}$$

where $\bar{T}_o$ is the average duration at the average power level p_o. For example a linear approximation to Fig. 5 would yield a value of $\gamma \approx .48$. Lastly, power spectral densities and autocorrelation functions on the received signal power fluctuations are available and presented in detail in a more comprehensive report in preparation.

The most important aspect of this particular scintillation event is summarized in the following table.

October 6, 1976	$\bar{E}_b/N_0$=22dB	m = 1.9
PreDecoder Error Rate	Simulated 4×10^{-3}	Measured 5×10^{-3}
Post Decoder Error Rate	$< 10^{-6}$	0

Clearly the result is that after decoding, the message came out error free even though the raw channel error rate was fairly high. If we look at all our scintillation data over all four months we come to essentially the same conclusion. For the days of heavy scintillation in a recording period we may have between 2500 to 7000 errors per teletype channel prior to decoding and 0 errors after decoding. Another example (September 20) shows pre- and post-decoding error rates of 1.2×10^{-2} and 0, respectively, at an m value of 1.1. The pre-decoder error rates are measured over a period of tens of minutes to a few hours. The burst error rates prior to the deinterleaving certainly approach 1/2 during the fade intervals.

It must be acknowledged that about 50% of our scintillation data is corrupted by unidentifiable interference. The receiver site was located in close proximity to a large number of radars and communications transmitters. This fact limits the utility of the data for clarification of the scintillation process. Nevertheless, the experiment is still useful in that even with combined scintillation and RFI, the post-decoding error rate is still effectively zero.

We have analyzed about 30 hours of strong scintillation data of which about half is compromised by local RFI. The remainder of the data exhibits a behavior which is quite close to being a Rayleigh process. We have also observed differences in tests between the average fade and interfade duration (measured at a specified level relative to the average power value). This implies that there is also a corresponding increase or decrease in the horizontal speed of the F region irregularities causing the scintillation. Hopkins, in a personal correspondence, indicated that during heavy scintillation he measured velocities of the order of 40 m/sec as compared to the more usual 100 to 200 m/sec range.

The obvious limitation to the experiment was, first, that there was clearly not enough available data and, second, there was not any power control on the received signal so that one could, in fact, evaluate in actual conditions how much system margin was provided. However, it is strongly felt that the series of comprehensive hardware simulations performed prior to the field experiment answered many of the important questions. In brief, these simulations were run in the following way. The data sources were standard pseudo-random sequence generators. The data was fed into a test multiplexer, DPSK encoder and RF phase modulator (simulating the Navy's TDM-1150 multiplexer and WSC-5 transmitter). The fading channel was simulated by a voltage controlled attenuator. The control voltage resulted from first low pass filtering the outputs of two 31-bit feedback shift registers thereby generating two independent Gaussian distributions with zero means. Squaring these two outputs and taking the square root of the sum resulted in the voltage with the required Rayleigh distribution. The cut-off frequency of the low pass filter was adjusted to give a range of fading behavior similar to that observed from prior analysis of real scintillation data. The input signal was also controllable via a stepped attenuator so that the $\bar{E}_b/N_0$ ratio could be varied over the appropriate range. On the output of the receiver two data analyzers detect and accumulate error events in order to determine error rates. It should be noted that we did not attempt to introduce phase variations. Figure 6 shows one set of curves where the demodulation scheme was differentially coherent. The curve labelled "simulation 1/2 schematic" was derived from a computer simulation and was somewhat limited in that it did not account for the real behavior of the SSR-1 tracking loops. The fading associated with the computer simula-

tion would be somewhere between what we have called "fast" and "slow" in the hardware simulation. The maximum discrepancy between the computer and corresponding hardware simulations would be about 1.5 dB.

An example of the measured improvement obtained through interleaving and coding is shown in the following table. The pre-decoder error rate is 6×10^{-3} for $E_b/N_0 \simeq 22$ dB and the

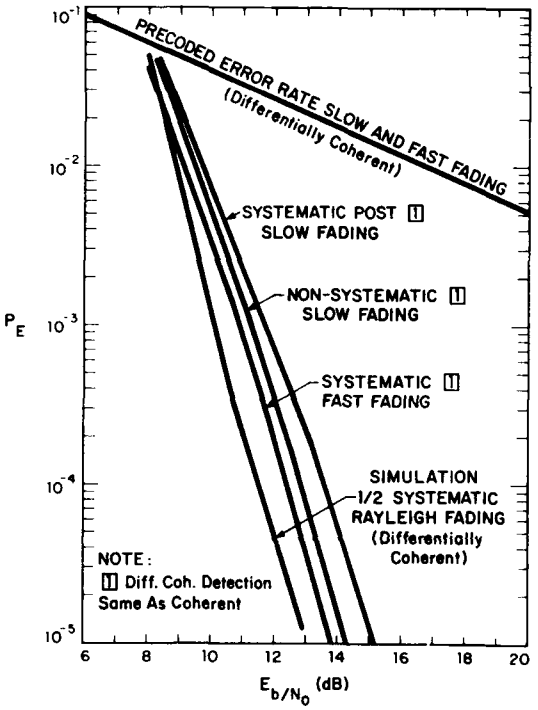


Fig. 6. Error Rate for R=1/2, DPSK Encoded and Rayleigh Fading K=7.

post-decoder error rates are as shown for coherent detection.

Table of Measured Error Rates

	R = 1/2	2/3	3/4
Systematic	$< 10^{-6}$	2×10^{-5}	6×10^{-5}
Nonsystematic	$< 10^{-6}$	$< 10^{-6}$	2×10^{-5}

It must be acknowledged that these values were obtained from the hardware simulations and that the computer simulations give error rates about a factor of 5 lower. We have not yet been able to completely resolve the discrepancy as yet. With our present system realization, the rate 3/4 (systematic) decoder is probably marginal for strong scintillation.

VI. CONCLUSIONS

From our limited field tests we can find no evidence that coherent detection is seriously degraded during heavy scintillation. Since the implementation of the coherent scheme is easier than the differentially coherent scheme we recommend the former. This observation is substantiated by the differential phase measurements made by Paulson and Hopkins [9] with spaced receivers. These measurements showed a maximum frequency deviation of about + .5 Hz which is within the tracking loop hold-in range (about a .2° shift per channel symbol).

Clearly, a good deal more of similar communication experiments need to be run during strong scintillation conditions. Most of our experiments were performed at the R = 1/2 codes. At these rates there was sufficient system margin and one could perhaps have opted for a slightly simpler error protection scheme. (For example, using 1 bit "hard" decisions in the decoder cuts the memory requirements by about 3 and loses only about 1.5 dB in margin.) On the other hand, for the rate 3/4 systematic code, one probably needs all the system margin available.

REFERENCES

1. E. A. Bucher, "UHF Satellite Communication During Scintillation," Technical Note 1975-10, Lincoln Laboratory, M.I.T. (5 August 1975) DDC AD-A014801/5.

2. Johns Hopkins/Applied Research Laboratory, "The Effect of Ionospheric Scintillation on the Fleet Broadcast of the Fleet Satellite Communication System" (December 1976) SDO-4380.6.

3. D. P. White, "A Time Diversity Coding Experiment for a UHF/VHF Satellite Channel with Scintillation: Equipment Description," Technical Note 1977-22, Lincoln Laboratory, M.I.T. (1 September 1977).

4. J. Aarons, H. E. Whitney and R. S. Allen, "Global Morphology of Ionospheric Scintillation," Proc. IEEE, Vol. 59, pp 159-172 (1971).

5. J. Aarons, "Equatorial Scintillation: A Review," Air Force Geophysical Laboratory, Report Number AFGL-TR-76-0078 (13 April 1976).

6. R. Crane, "Ionospheric Scintillation," Proc. IEEE, Vol. 65, No. 2 (February 1977).

7. S. Basu, S. Basu and B. Khan, "Model of Equatorial Scintillations from In Situ Measurements," Air Force Geophysical Laboratory, Report Number AFGL-TR-76-0080 (13 April 1976).

8. B. E. Nichols, "UHF Fading from a Synchronous Satellite Observed at Kwajalein October 1970 Through June 1972, "Technical Note 1974-19, Lincoln Laboratory, M.I.T. (24 March 1974) DDC-AD780174/9.

9. M. R. Paulson and R. U. F. Hopkins, "Spatial Diversity Characteristics of Equatorial Scintillation, "Technical Report 113, Naval Oceans Systems Center, San Diego (2 May 1977).

10. J. A. Hiller and I. M. Jacobs, "Viterbi Decoding for Satellite and Space Communications, "IEEE Trans. Communications Tech. COM-19, No. 5 (1971).

11. H. E. Whitney, "Ionospheric Scintillation Effects on VHF-UHF Communication Systems," Air Force Geophysical Laboratory, Report Number AFGL-TR-76-0078 (13 April 1976).

EFFECTS OF THE IONOSPHERE ON THE PERFORMANCE OF THE SEASAR SYNTHETIC-APERTURE RADAR

Warren D. Brown
Sandia Laboratories, Albuquerque, New Mexico

Abstract

Phase scintillation data at 1.24 GHz from the Wideband Satellite Experiment have been used to determine the expected performance degradations of SEASAR. These degradations are a result of propagation through electron density irregularities in the ionosphere. A formalism relating SEASAR performance to the Wideband phase scintillation spectrum is described. Antenna patterns and values of the pointing angle error, the quadratic phase error, the integrated sidelobe ratio, and the first sidelobe peak corresponding to selected passes of the Wideband Satellite are presented. Probability distributions of performance degradations are given for both the polar and equatorial regions. On the basis of the data from Poker Flat, Alaska, it is demonstrated that SEASAR performance problems will occur for a significant fraction of the time in the polar region.

ERROR PROBABILITIES FOR DIGITAL TRANSMISSION IN SCINTILLATION FADING CHANNELS USING GAUSS-QUADRATURE INTEGRATION

R. E. Ziemer and Youngyearl Han
E. E. Dept., Univ. of Mo.-Rolla
Rolla, MO 65401

ABSTRACT

Calculation of error probabilities for a coherent phase-shift keyed communication system operating in a transionospheric scintillation channel is accomplished by means of the Gauss-quadrature integration formula. The channel model used, patterned after Rino's work, is slowly flat fading wherein the envelope of the received signal is modeled as the envelope of correlated Gaussian quadrature random processes. The error probability for the scintillation channel is calculated using actual ionospheric scintillation data for transmission in the UHF region (30-300 MHZ).

INTRODUCTION

The scintillation of radio waves passing through the ionosphere has been an observed phenomenon for many years, first by radio astronomers and then as a result of the reception of radio signals from orbiting satellites. The amplitude distribution of ionospheric scintillation has been approximated with varying degrees of success by the Rice-Nakagami, log-normal and bivariate-Gaussian distributions [1]. Rino [2] has shown that the amplitude distribution of ionospheric scintillation is described closely as the envelope of bivariate Gaussian quadrature components and through suitable choices of parameters a good fit is obtained for all amplitudes except at the extremes of the distribution. Although similar to the Rice-Nakagami and log-normal distributions, the bivariate-normal distribution is more general because it allows the components to be correlated and to have unequal variance.

The purpose of this paper is to present a useful way of calculating the average error probability for phased-shift keyed signaling in the presence of bivariate-normal distributed flat fading. The receiver is assumed to track the phase of the received signals exactly. The Gauss-quadrature integration formula is employed in the calculation of the average error probability.

STATEMENT OF THE PROBLEM

With the correlated Gaussian process, the received signal is considered as a complex random process, $V(t)=V_x(t)+jV_y(t)$, where V_x and V_y are the quadrature components of the signal. The time-varying impulse response is given by

$$h(\tau,t)=E(t)\delta(t-\tau) \tag{1}$$

or the time-variant transfer function with flat frequency by

$$H(f,t)=E(t),\ -\infty<f<\infty \tag{2}$$

where

$$E(t)=E_o+e_c(t)+je_s(t), \tag{3}$$

In the above equations, $E_o=\overline{E(t)}$ and $e_c(t)$ and $e_s(t)$ are zero-mean, correlated Gaussian processes.

This transionospheric scintillation channel model is a good fit to the actual received signal data except for very wideband transmissions. Assuming the transmitted signal is of the form

$$X(t)=\mathrm{Re}[Be^{j\omega_o t}] \tag{4}$$

it can be easily shown that the received signal is of the form

$$Y(t)=B[E_o+e_c(t)]\cos\omega_o t-Be_s(t)\sin\omega_o t. \tag{5}$$

The quadrature components of the received signal are, respectively,

$$V_x=B[E_o+e_c(t)], \tag{6}$$

$$V_y=Be_s(t), \tag{7}$$

and the envelope is

$$A=B\sqrt{[E_o+e_c(t)]^2+e_s^2(t)}=BR(t). \tag{8}$$

For situations where the received signal envelope, A, randomly fluctuates, the overall probability of error is obtained by averaging $p(E|a)$ over A:

$$P_E=\int_{\infty}^{\infty}p(a)p(E|a)da. \tag{9}$$

$p(E|a)$ is the probability of error given A=a, $p(a)$ is the probability density function (pdf) of A, and P_E is the average probability of error. In those cases where the envelope of the signal randomly fluctuates, the lower limit of the integral becomes zero. If the received signal is assumed to be a general Gaussian process, $p(a)$ is an envelope pdf obtained from a bivariate Gaussian pdf by transforming rectangular coordinates into polar coordinates:

$$p_a(a)=\int_o^{2\pi}p_{a\theta}(a\cos\theta,a\sin\theta)ad\theta \tag{10}$$

where

$$p_{a\theta}(a,\theta)=ap_{xy}(V_x,V_y)\Big|_{\substack{V_x=a\cos\theta\\V_y=a\sin\theta}} \tag{11}$$

$$=ap(a\cos\theta,a\sin\theta),\theta\leq\theta<2\pi,o\leq a<\infty.$$

$p_{xy}(V_x,V_y)$ is a correlated bivariate Gaussian pdf.

The error probability for the phase-shift keyed signal given a, can be derived by assuming a bi-phase modulated transmitted signal of the form

$$x(t)=A\cos(\omega_o t+\cos^{-1}m\,d(t)+\theta) \tag{12}$$

where $nT\leq t\leq(n+1)T$ in which T is the signaling interval, $d(t)$ is the data sequence, and $\cos^{-1}m$ the modulation index. Synchronous detection in white, Gaussian noise backgrounds requires a correlation or matched filter detector to give minimum probability error. Here, the noise is additive, white, and Gaussian with zero mean and one-sided power spectral density N_o. In the special case m=o, assuming that the receiver tracks the phase exactly and the fading is slow so that a=constant within a T-second signaling interval, the probability of error, given a, is

$$P(E|a)=\tfrac{1}{2}\operatorname{erfc}(\sqrt{z}a) \tag{13}$$

where z is the signal-energy-to-noise-spectral density ratio. To perform the integration in Eq. (9) with the lower limit of the integral zero, the Gauss-quadrature integral formula [3] is applicable if we know the moments of $p(a)$. The moments may be used to calculate the weights and abscissas in the integration formula

$$\int_b^c f(x)w(x)dx=\sum_{k=1}^{m}W_k f(x_k)+E^1 \tag{14}$$

where $f(x)\geq 0$ in [b,c], W_k are the weights x_k are the abscissas, m is the number of weights and abscissas, and E^1 is the error of the approximation. Here, $f(x)$ can be replaced by the amplitude pdf $P(a)$ and $w(x)$ by the error probability $P(E|a)$. Also, b is zero and c is plus infinity. If we define the rth moments M_r, associated with $w(x)$ over [b, c], by the equation

$$\int_b^c x^r f(x)dx=M_r \tag{15}$$

and M_r is represented by

$$\sum_{k=1}^{m}W_k x_k^r=M_r\ ,\ r=0,1,\ldots,M-1 \tag{16}$$

it is said that the integral is approximated with degree of precision 2m-1. If we let $x_1, x_2,\ldots,x_m$ be the zeros of $g(x)$, i.e.,

$$g(x)=(x-x_1)(x-x_2)\ldots(x-x_m)=x^m+\alpha_1 x^{m-1}+\ldots+\alpha_m, \tag{17}$$

the α's can be obtained from equation (16) by transforming. Finally, the abscissas x_k and weights W_k are obtained. It is stated that there is no guarantee that the zeros of $g(x)$ will be real and distinct and that they lie in [b,c] even though $f(x)<0$. If the roots are not real and distinct, the desired formula does not exist.

The moments of $p(a)$ are given by

$$E[a^n]=\int_o^{\infty}a^n P_a(a)da=\int_o^{2\pi}g(\theta)d\theta$$

where $g(\theta)$ is the integral

$$\int_o^{\infty}p(a\cos\ ,a\sin\ )a^{n+1}da=\frac{1}{2\pi\sqrt{2D}}$$

$$\times\ B\Big[\sum_{\substack{k=0\\ \text{even}}}^{n+1}\binom{n+1}{k}\Big(\frac{\bar{v}_x E}{D}\Big)^{n+1-k}\Big(\frac{2A}{D}\Big)^{k/2}\Big[\Gamma\Big(\frac{k-1}{2}+1\Big)$$

$$+\gamma\Big(\frac{k-1}{2}+1,F^2\Big)\Big]+\sum_{\substack{k=1\\ \text{odd}}}^{n+1}\binom{n+1}{k}\Big(\frac{\bar{v}_x E}{D}\Big)^{n+1-k}$$

$$\times\ \Big(\frac{2A}{D}\Big)^{k/2}\Big[\Gamma\Big(\frac{k-1}{2}+1\Big)-\gamma\Big(\frac{k-1}{2}+1,F^2\Big)\Big]\Big] \tag{18}$$

where

$$A=\sigma_x^2\sigma_y^2-C_{xy}^2$$

$$B=\exp\frac{\bar{v}_x^2E^2-\bar{v}_x^2\sigma_y^2D}{2AD}$$

$$D=\cos^2\theta\sigma_y^2-2C_{xy}\cos\theta\sin\theta+\sin^2\theta\sigma_x^2$$

$$E=\cos\theta\sigma_y^2-C_{xy}\sin\theta$$

$$F=\frac{-\bar{v}_x E}{\sqrt{2AD}}$$

$\Gamma(q)$ is a gamma function and $\gamma(s,t)$ is the incomplete gamma function, σ_x^2 and σ_y^2 are the variances of V_x and V_y, respectively, C_{xy} is the covariance of V_x and V_y, and $\bar{V}_x$ is mean of V_x. It is assumed that the coherent signal is the phase reference; hence the mean of V_y becomes zero.

Now it remains to compute the integration formula for each signal-to-noise ratio according to Eq. (14) to obtain the error probability.

CALCULATION OF P_E

The data used are from Rino. With the values of the variances of V_x and V_y, the covariance, and the mean of V_x, moments of p(a) were calculated up to order 20. With the moments thus obtained, the weights and abscissas of the Gauss-quadrature integration formula were calculated. In finding the roots of the algebraic equation (17), an attempt was made to calculate the roots to tenth order by Graffe's root-square method. In those cases where the roots turned out to be complex they were simply disregarded.

Once the weights and abscissas are determined, the error probability is found by the equation

$$P_E=\int_0^\infty p(a)p(E|a)da\approx\sum_{k=1}^{m}\tfrac{1}{2}W_k\,\mathrm{erfc}(\sqrt{z}a_k) \qquad (19)$$

For each set of weights and abscissas the error probability was calculated for the signal-to-noise ratio, z, in the range 0 to 20 dB.

RESULTS

The technique for calculating P_E just described was tested by employing the data used by Rino [2]. The parameter values for the bivariate Gaussian approximation to the envelopes for three separate data sets are given in Table I. These values were calculated by Rino.

Table I. DATA SETS

	σ_x^2	σ_y^2	C_{xy}	θ	$\bar{v}_x$
Data Set I	0.3625	0.0074	0.0037	0.6°	0.5101
Data Set II	0.2925	0.0374	0.0174	3.9°	0.5767
Data Set III	0.1580	0.0719	0.036	18.6°	0.7336

The moments up through order 20 were then calculated using Eq. (18) and Simpson's one-third integration rule with 100 integration intervals. They are given in Table II for data sets I-III, respectively. The weights and abscissas are calculated next. These are given in Table III. The error probability is then calculated using Eq. (19). The resulting curves are shown in Figures 1-3.

CONCLUSIONS

A method has been discussed whereby the probability of error for a fading channel can be approximated by a series by knowing the moments of the envelope of the received signal. Using the Rino bivariate Gaussian model to approximate the envelope distribution of three data sets, the moments were analytically calculated from the approximating bivariate Gaussian density function. Another approach which could have been taken is to estimate the moments directly from the measured envelope time series. This could be done "on-site" through the use of a micro-computer based data acquisition system as discussed in a companion paper [4]. The calculation of confidence intervals for P_E based on the accuracy of the estimated moments, would be of importance in such a situation.

The results presented in this paper were for the case of phase-shift-keyed signaling. The application of the method to other coherent and noncoherent signaling schemes would be a straight-forward calculation. Further results are given in [5].

REFERENCES

1. Fremouw, E. J., "Modeling and Prediction of Ionospheric Scintillation", AIAA Paper No. 74-54, AIAA 12-th Aerospace Sciences Meeting, Washington, DC, Jan. 1974.

2. Hatfield, V. E. and C. L. Rino, "Non-Rician Statistics and Their Implications for Modeling Effects of Scintillation on Communication Channels", Paper 4-1, Ionospheric Effects Symposium Proceedings, Washington, DC, Jan. 1975.

3. Ziemer, R. E. and W. H. Tranter, Principles of Communications: Systems Modulation, and Noise, Houghton-Mifflin, 1976.

4. Stigall, P. D. and R. E. Ziemer, "Microcomputer Data Acquisition and Analysis for Transionospheric Communication Channels", Symposium on The Effect of the Ionosphere on Space Systems and Communications, Washington, DC, Jan. 1978.

5. Han, Youngyearl, "Calculation of the Average Error Probability for Phase-Shift Keyed Signaling in the Presence of Scintillation", MSEE Thesis, University of Missouri-Rolla, 1976.

TABLE II

MOMENTS FOR DATA SETS I - III

Order	Moments		
	Data Set I	Data Set II	Data Set III
0	0.999980927	0.999990106	0.999990463
1	0.650264092	0.69681769	0.800824761
2	0.630065560	0.662472367	0.768060029
3	0.752400778	0.762816727	0.833600342
4	1.03774561	1.00310898	0.994559586
5	1.59146214	1.45870686	1.28148223
6	2.66142464	2.30041599	1.76197243
7	4.78287220	3.88369656	2.56297588
8	9.14478493	6.9540510	3.91821384
9	18.4634857	1.31185150×10	6.26318359
10	39.1357880	2.59255066×10	1.04247742×10
11	86.6817932	5.34479523×10	1.80060272×10
12	199.851151	1.14532059×10^2	3.21809082×10
13	478.095215	2.54338699×10^2	5.93696747×10
14	1183.51831	5.83810791×10^2	1.12820801×10^2
15	3024.6027	1.38215479×10^3	2.220432693×10^2
16	7963.59766	3.36839185×10^3	4.42098633×10^2
17	21563.9102	8.43592187×10^3	9.08885010×10^2
18	59956.0820	2.167905808×10^4	1.91284497×10^3
19	170928.437	5.70896641×10^4	$4.116556\ 69 \times 10^3$
20	499029.312	1.53871250×10^5	9.04931641×10^3

TABLE III

WEIGHTS AND ABSCISSAS FOR DATA SETS I - III

I. Two-Term Expansion

	Data Set I	Data Set II	Data Set III
Weights:	0.3071531	0.3276832	0.4151988
	0.6928277	0.6723068	0.5847916
Abscissas:	1.329285	1.299296	1.223323
	0.355023	0.4031789	0.5008658

II. Three-Term Expansion

Weights:	0.06371551	0.07189083	0.1070542
	0.4672636	0.4962778	0.5987346
	0.46900170	0.4318046	0.2942014
Abscissas:	1.8249720	1.738842	1.543306
	0.9160204	0.9092638	0.8944943
	0.2344599	0.2768338	0.3400476

III. Four-Term Expansion

Weights:	0.01082798	0.01287154	0.02050125
	0.1837977	0.2080757	0.2944118
	0.4699416	0.4938852	0.5260549
	0.33541350	0.2851575	0.1590223
Abscissas:	2.231759	2.102027	1.820082
	1.381575	1.328245	1.217162
	0.6767604	0.678383	0.6950903
	0.1733095	0.207273	0.2484438

IV. Five-Term Expansion

	Data Set I	Data Set II	Data Set III
Weights:	0.158963×10^{-2}	0.171640×10^{-2}	0.169086×10^{-3}
	0.528963×10^{-1}	0.581142×10^{-1}	0.398965×10^{-1}
	0.272969	0.2958682	0.3376339
	0.421299	0.440650	0.4861735
	0.251230	0.2036411	0.1361173
Abscissas:	2.588652	2.442432	2.469329
	1.780906	1.708086	1.693333
	1.101572	1.084643	1.141331
	0.5240415	0.5408216	0.650309
	0.1372466	0.1676317	0.2301979

V. Six-Term Expansion

Weights:	0.1308124×10^{-3}	-0.26847×10^{-8}	0.204925×10^{-2}
	0.9246245×10^{-2}	0.210028×10^{-2}	0.759190×10^{-1}
	0.9811413×10^{-1}	0.6307011×10^{-1}	0.3814473
	0.3032158	0.3008584	$-0.6596536 \times 10^{-6}$
	0.3799165	0.4356013	0.4325059
	0.2093574	0.1983518	0.1080695
Abscissas:	2.981177	5.602629	2.1221986
	2.198966	2.407546	1.548269
	1.538124	1.683158	1.049717
	0.9493581	1.06779	-0.143525×10^{-1}
	0.4466255	0.5316889	0.592819
	0.1198405	0.1649917	0.204702

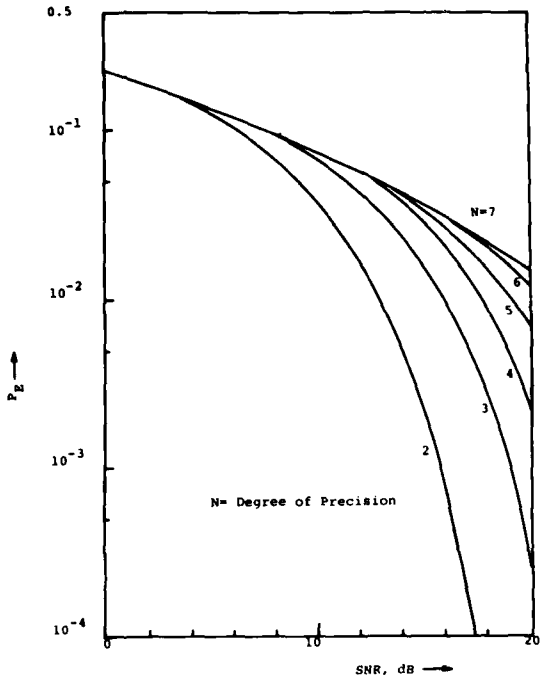


Figure 1. Gauss-Quadrature Series Approximation to P_E for PSK for Data Set I

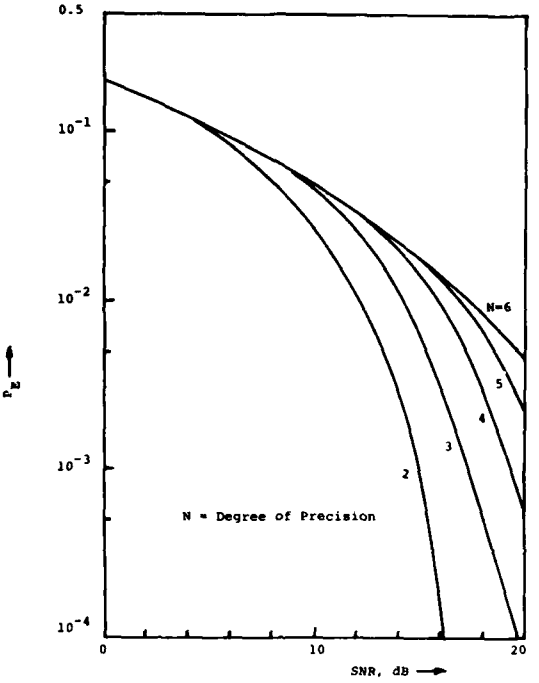


Figure 2. Gauss-Quadrature Series Approximation to P_E for PSK for Data Set II

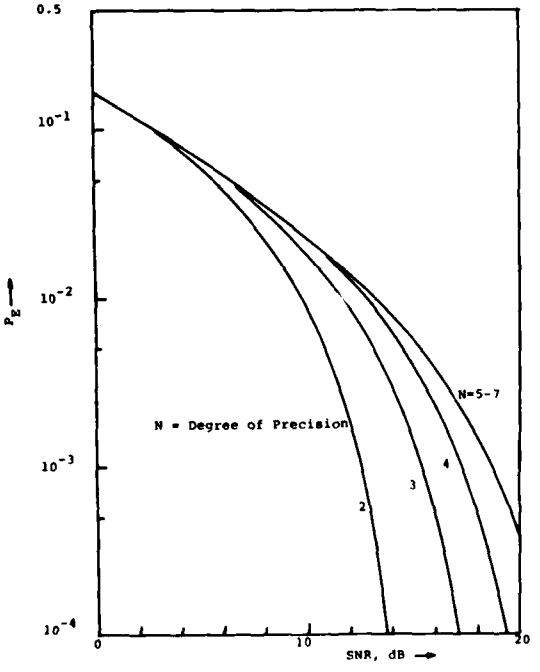


Figure 3. Gauss-Quadrature Series Approximation to P_E for PSK for Data Set III.

"A NOTE ON THE DESIGN OF FILTERS FOR A DIGITAL FADE SIMULATOR FOR THE TRANSIONOSPHERIC CHANNEL"

R. W. D. BOOTH
LINCOM CORPORATION
LOS ANGELES, CALIFORNIA

1.0 INTRODUCTION

Recent investigations of the statistical nature of the scintillations of radio waves that traverse the ionosphere indicate that the joint density function of the in-phase and quadrature components of the received signal is very nearly bivariate normal [1], [2]. Assuming that the complex amplitude of the received signal is a gaussian process then a digital fade simulator for the examination of communication links may be readily implemented such that the statistics of the complex amplitude for fixed t are bivariate normal. This has already been implemented as described in reference [3]. This note serves as an extension of this work, in particular, the design of the filters used in the fade simulator is examined and the utility of using theoretically predicted parameters in the simulator design is investigated. The properties of the fading signal of interest are the mean and variance of the received amplitude and the fade duration distribution of the received amplitude. The conclusions reached indicate that the choice of Butterworth filters results in a poor fit of the fade duration distribution for low duration (fast) fades and that a simple cascade of two one-pole filters offers improved performance. Finally it is shown that the theoretically predicted scintillation parameters produce an adequate fit of the fade duration distribution serving as another modest verification of the theory.

2.0 FADING MODEL

It is assumed that the received wave is a CW signal. For modulated signals it is assumed that the signal is narrowband, thus, the small perturbations of the scintillating medium produce minor disturbances of the modulation and major disturbances of the carrier phase and amplitude. For this case the CW results may be applied. The received signal v(t) can be written in complex notation [1].

$$v(t) = \mathrm{Re}\,(E \cdot e^{j\omega t}) \qquad (1)$$

The complex amplitude E is

$$E = b + x(t) + jy(t) \qquad (2)$$

where b is the amplitude of the nonscintillating component and x(t) and y(t) are the in-phase and quadrature components of the scintillating voltages. It is assumed that x(t) and y(t) are zero mean, that is,

$$\langle x(t)\rangle^1 = 0 \qquad (3)$$

and $$\langle y(t)\rangle = 0. \qquad (4)$$

The following second moments of E may be conveniently utilized.

$$R_o = \langle E \cdot E^* \rangle - b^2 = \sigma_x^2 + \sigma_y^2 \qquad (5)$$

and $$B_o = \langle E \cdot E \rangle - b^2 = \sigma_x^2 - \sigma_y^2 + 2j\sigma_{xy}^2. \qquad (6)$$

Note that B_o is a complex quantity. Specification of b, R_o, and B_o is enough to statistically describe the complex random variable E.

The correlations R_o and B_o may be computed from physical properties of the scattering medium. These parameters are a function of the dimensionless quantity $Z = \lambda z/(4\pi\alpha^2)$ where λ is the wavelength of the incident radiation, $z = z_1 \cdot z_2/(z_1 + z_2)$ where z_1 and z_2 are the distances from the scattering screen to the signal source and receiver respectively and α is a scale size parameter peculiar to the scattering medium. The limiting values for R_o and B_o have been computed for the cases $Z \to \infty$ and $Z \to 0$ [4].

The weak scatter limit corresponds to $Z \to 0$. For this case

$$\lim_{Z\to 0} \left|\frac{B_o}{R_o}\right| = 1$$

and angle $B_o = \pi$. Thus as $Z \to 0$ $B_o = -R_o$. By inspection of equations (5) and (6) it is clear that $R_o \to \sigma_y^2$ and $B_o \to -\sigma_y^2$, therefore,

$$\lim_{Z\to 0} \sigma_{xy}^2 = 0$$

and $$\lim_{Z\to 0} \sigma_x^2 = 0.$$

Thus in the weak scatter limit all of the scattered power is in the quadrature component. For this case very little fading is observed although large bursts or enhancements occur. The

1 < > denotes expectation

strong scatter limit corresponds to $Z \to \infty$. This occurs when the scale size α is small enough so that many scattering centers contribute to the phase fluctuations. For this case

$$\lim_{Z\to\infty} \frac{|B_o|}{R_o} = \left(1 - \frac{1}{2^{\mu+1}}\right) \cdot \sec \mu\pi/2$$

and angle $B_o = \mu\pi/2$ where $\mu = -\gamma + .5$. For the Kolomogorov spectrum $\gamma = 4/3$, hence, the angle $B_o = -5/6 \cdot \pi/2$ and $|B_o|/R_o = .42$. The angle of B_o is nearly $-\pi/2$, thus, the real part is nearly zero. From equation (6), $\sigma_x^2 = \sigma_y^2$ and $\sigma_{xy}^2 \triangleq -.21 \cdot R_o$. The correlation coefficient $\rho \triangleq \sigma_{xy}^2/\sigma_x\sigma_y$ is approximately -.42. Thus the scintillating components x and y are negatively correlated in the strong scatter limit. It is pointed out [4] that this does not correspond to Rician statistics as one might expect. These parameters and assumptions are enough to specify the statistics of x(t) and y(t) for fixed t. This is not quite enough to completely characterize x(t) and y(t) since some assumptions need to be made concerning the temporal properties of x(t) and y(t).

It will be assumed that x(t) and y(t) are gaussian random processes, thus, the joint density of [x(t), x(t + τ), y(t), y(t +τ)] is a gaussian function. Note that the assumption that x(t)(for fixed t) is a gaussian random variable is not enough to guarantee that x(t) is a gaussian process. With the gaussian process assumption it remains to specify the following correlations $\langle x(t)\, x(t+\tau)\rangle = R_x(\tau)$, $\langle y(t)\, y(t+\tau)\rangle = R_y(\tau)$, and $\langle x(t)\, y(t+\tau)\rangle = R_{xy}(\tau)$. One is faced with the possibility of evaluating three different correlation functions. It will be assumed that these three correlations are all the same within a scale factor.

$$R_x(\tau) = \sigma_x^2 \cdot \rho(\tau) \qquad (7)$$

$$R_y(\tau) = \sigma_y^2 \cdot \rho(\tau) \qquad (8)$$

$$R_{xy}(\tau) = \sigma_{xy}^2 \cdot \rho(\tau) \qquad (9)$$

$$\rho(0) = 1. \qquad (10)$$

For $\tau = 0$ the correlations reduce to the previously derived variances and cross correlation for the random variables x(t), y(t).

It remains to specify $\rho(\tau)$ or equivalently its fourier transform $S(\omega)$. Fortunately a useful estimate of the spectrum $S(\omega)$ may be obtained from spectra of the fluctuations in the intensity of the received signal. Rufenach [5] has reported some data indicating that the spectrum of the intensity is low pass in nature. The spectrum is essentially flat out to a cutoff frequency and then rolls of as F^{-3} approximately. The cutoff frequency is determined by velocity considerations, i.e., the velocity of the scintillating screen and the velocities of the signal source (nongeostationary satellites) and receiver. The high frequency roll off is determined by the spectrum of the irregularities in the screen. The spectrum $S(\omega)$ can be related to this empirically estimated spectrum of the intensity in the following manner.

The intensity $I(t) \triangleq E \cdot E^*$

where $E = (b + x(t) + jy(t))$.

Thus $I(t) = (b + x + jy) \cdot (b + x - jy)$

$$= (b + x(t))^2 + y^2(t). \qquad (11)$$

The correlation of the intensity is

$$R_I(\tau) \triangleq \langle I(t)\, I(t+\tau)\rangle = \langle [(b + x(t))^2 + y^2(t)] \cdot [(b + x(t+\tau))^2 + y^2(t+\tau)]\rangle. \qquad (12)$$

This correlation is readily evaluated since it is assumed that x(t), x(t + τ), y(t), y(t + τ) are jointly gaussian and Price's theorem may be used to evaluate the higher order moments. Thus

$$R_I(\tau) = (b^2 + \sigma_x^2 + \sigma_y^2)^2 + 4b^2\sigma_x^2 \cdot \rho(\tau) + 2(\sigma_x^4 + 2\sigma_{xy}^4 + \sigma_y^4)\, \rho^2(\tau). \qquad (13)$$

The spectrum is

$$S_I(\omega) = (b^2 + \sigma_x^2 + \sigma_y^2)^2 \cdot \delta(\omega) + 4b^2\sigma_x^2 \cdot S(\omega) \quad 2(\sigma_x^4 + 2\sigma_{xy}^4 + \sigma_y^4) \cdot 1/2\pi \cdot \quad S(\omega) * S(\omega).^1 \qquad (14)$$

The first term is an impulse at $\omega = 0$ representing the total power in the received waveform. The second term is proportional to the spectrum of the gaussian processes and the third term is proportional to the convolution of the spectrum. For weak to moderate scintillating conditions $b^2 >> (\sigma_x^2 + \sigma_y^2)$ i.e. the direct power dominates the scintillating power. As long as the weak scatter limit is not approached i.e. $\sigma_x^2 \to 0$ then $4b^2\sigma_x^2 >> 2(\sigma_x^4 + 2\sigma_{xy}^4 + \sigma_y^4)$ and the spectrum of the intensity is approximately

$$S_I(\omega) \sim 4b^2\sigma_x^2\, S(\omega) \qquad (15)$$

where the impulse has been neglected. Thus the spectrum of the processes x(t) and y(t) should be approximately the same as the intensity spectrum. $S(\omega)$ should then be low pass in nature and roll of as f^{-3}. An interesting feature of the intensity spectrum (14) is the presence of $S(\omega) * S(\omega)$. As the scintillations become more intense, i.e., as $4b^2\sigma_x^2 \sim 2(\sigma_x^4 + 2\sigma_{xy}^4 + \sigma_y^4)$ then the convolution term becomes more important. Since convolving the spectrum tends to widen it, the bandwidth of the intensity fluctuations will increase under heavy scintillation conditions.

3.0 USE OF THE SCINTILLATION INDEX S_4

A measure of the depth of the fades or the strength of the scintillation is the scintillation index. There are various types of scintillation indices, some empirical, some theoretical. A common measure is S_4 which is a measure of the RMS intensity [6].

$$S_4^2 \triangleq \langle (I(t) - \langle I(t)\rangle)^2\rangle / \langle I(t)\rangle^2 \qquad (16)$$

This can be rewritten as

$$S_4^2 = (\langle I^2(t)\rangle - \langle I(t)\rangle^2)/\langle I(t)\rangle^2 \qquad (17)$$

$$= (R_I(0) - \langle I\rangle^2)/\langle I\rangle^2 \qquad (18)$$

---

1 * denotes convolution

From expression (11)

$$\langle I(t)\rangle = \langle (b + x(t))^2 + y^2(t)\rangle$$
$$= b^2 + \sigma_x^2 + \sigma_y^2 = \text{total power.} \quad (19)$$

Using (19) and (13) then

$$S_4^2 = [4 \cdot b^2 \cdot \sigma_x^2 + 2(\sigma_x^4 + 2\sigma_{xy}^4 + \sigma_y^4)]/(b^2 + \sigma_x^2 + \sigma_y^2)^2. \quad (20)$$

This may be normalized by choosing that the total scintillating power be unity, i.e., $\sigma_x^2 + \sigma_y^2 = R_o = 1$. There are several special cases of S_4^2.

a. Strong Scatter Limit
$$S_4^2 = (1.776 \cdot b^2 + 1.18)/(b^2 + 1)^2 \quad (21)$$

b. Weak Scatter Limit
$$S_4^2 = 2 \cdot \sigma_y^4/(b^2 + 1)^2 \quad (22)$$

c. Rician Statistics
$$S_4^2 = (2b^2 + 1)/(b^2 + 1)^2 \quad (23)$$

d. Rayleigh Limit $S_4^2 = 1$ (24)

A knowledge of scintillation index S_4^2 plus an estimate of the type of scintillation encountered can be used as a first estimate of the signal strength b^2 or equivalently $b^2/(\sigma_x^2 + \sigma_y^2)$. Apparently it may be assumed that strong scintillations have occurred if $S_4 \geq .4$ and moderate to weak scintillations have occurred if $S_4 < .4$ [7].

4.0 DETERMINATION OF GAINS AND FILTER SPECTRUM FOR ANALOG AND DIGITAL SIMULATORS

An implementation of a fade simulator is illustrated in figure 1. The philosophy is to generate the two correlated gaussian processes, x(t) and y(t), add b to x(t) and use the two processes y(t) and x(t) + b to modulate quadrature and in-phase versions of the input signal. The inputs to the scintillation generator are the two statistically independent gaussian random processes $N_x(t)$ and $N_y(t)$. These may be either analog or digital signals. Consider the analog case first. $N_x(t)$ and $N_y(t)$ are then gaussian white noise processes. Let the impulse responses of the two filters be $h_x(t)$ and $h_y(t)$ respectively. Thus by inspection of Figure 1

$$x(t) = G_x \int_{-\infty}^{t} N_x(\tau) \cdot h_x(t - \tau)d\tau \quad (35)$$

and
$$y(t) = G_y \int_{-\infty}^{t} [G_1 \cdot N_x(\tau) + G_2 \cdot N_y(\tau)] \cdot h_y(t - \tau)d\tau. \quad (36)$$

The spectrum of x(t) is $S_x(\omega) = G_x^2 \cdot |H_x(j\omega)|^2$ and the spectrum of y(t) is $S_y(\omega) = G_y^2 \cdot (G_1^2 + G_2^2) \cdot |H_y(j\omega)|^2$. The cross spectrum is $S_{xy}(\omega) = G_x G_y \cdot G_1 \cdot H_x(j\omega) H_y^*(j\omega)$. With the assumption that the autocorrelation and cross correlation functions are the same within a scale factor then the spectrum of x(t) is the same as the spectrum of y(t) within a scale factor. The scale factors can be absorbed into G_x and G_y so it may be assumed that $|H_x(j\omega)|^2 = |H_y(j\omega)|^2 \triangleq |H(j\omega)|^2$. Furthermore, $H_x(j\omega)\, H_y^*(j\omega) = |H(j\omega)|^2$.

The variance of x(t) is $\sigma_x^2 = G_x^2 \cdot \int_{-\infty}^{\infty} |H(j\omega)|^2 d\omega$. Let

$$G^2 \triangleq \int_{-\infty}^{\infty} |H(j\omega)|^2 d\omega \quad (27)$$

then $\sigma_x^2 = G_x^2 \cdot G^2$. The parameter G^2 is peculiar to the filter $H(j\omega)$ and may be computed. Similar computations may be made for the variance of y(t) and the cross correlation ie $\sigma_y^2 = G_y^2(G_1^2 + G_2^2) \cdot G^2$ and $\sigma_{xy}^2 = G_x \cdot G_y \cdot G_1 \cdot G^2$. First fix $G_1^2 + G_2^2 = 1$, then $\sigma_y^2 = G_y^2 \cdot G^2$. The gains G_x and G_y are then set by the variances. Let

$$G_x = \sigma_x/G \quad (28)$$

and
$$G_y = \sigma_y/G. \quad (29)$$

The cross correlation is then $\sigma_{xy}^2 = G_1 \cdot \sigma_x \cdot \sigma_y$

or
$$G_1 = \sigma_{xy}^2/(\sigma_x \cdot \sigma_y) \triangleq \rho. \quad (30)$$

Thus the gain G_1 is just the correlation coefficient of the processes x(t) and y(t). The remaining gain G_2 is

$$G_2 = \sqrt{1 - G_1^2} = \sqrt{1 - \rho^2}. \quad (31)$$

Since $|\rho| \leq 1$ this results in a consistent specification of G_1 and G_2. Thus the network of gains in figure 1 is determined from the correlations σ_x^2, σ_y^2 and ρ.

The spectrum of the filters remains to be determined. The spectrum $|H(j\omega)|^2$ should closely match the spectrum of the intensity as derived in Section 2.0, that is, $|H(j\omega)|^2 \approx S(\omega)$. Since $S(\omega) \approx S_I(\omega) \alpha\, f^{-3}$ in the rolloff region then $|H(j\omega)|^2 \sim f^{-3}$. This is the main problem with the design of filters for the simulator. Filter designs based on the ratio of polynomials do not lead to spectra that roll off with odd integer exponents. A first order Butterworth filter rolls off as f^{-2} and a second order filter rolls off as f^{-4}, thus the best filter choice is something in between the two filters. Comparison of simulated data with measured data indicates that the first order filter is too rich in high frequency components leading to a fade duration distribution with too many short duration fades. On the other hand, a second order Butterworth filter leads to a fade duration distribution with not enough rapid fades [3]. In order to obtain an adequate fit of the fade duration distribution it is necessary to design a filter with spectral shape that approximates the f^{-3} property.

There is one property of the spectrum that may be exploited. The spectrum rolls off as f^{-3} only up to a certain point and then rolls off faster. This upper frequency cutoff corresponds to the inner scale size. Fluctuations smaller than this are not supported by the mixing effects of the medium. This may be approximated by designing filters that eventually roll off as f^{-4}, thus, there should be two more poles than zeroes in the filter transfer function. The cascade of two one-pole filters satisfies this property without resorting to second order Butterworth filters. For an analog implementation of the

fade simulator this is an easily designed and fabricated filter. The setting of the two cutoff frequencies can be arrived at empirically. If an intensity spectrum is available, the first cutoff frequency may be set near the 3 dB point for the intensity spectrum. The high frequency cutoff is about 20 times higher than this 3 dB point. This particular choice for the high frequency cutoff was deduced from comparisons of measured data [3] with digitally simulated data.

A digital implementation of the fade simulator essentially parallels that of the analog fade simulator. The input noise processes $N_x(t)$ and $N_y(t)$ are now statistically independent samples of two independent gaussian random processes. The means are zero and the variance of the samples is one. The sampling interval is T where T is chosen such that the sampling rate $f_s = 1/T$ is several times higher than the upper cutoff frequency of the filters. The gains of the amplifiers G_x, G_y, G_1, and G_2 are determined from expressions (19), (20), (21) and (22) for the digital case also. The only difference is the computation of the normalization constant G^2. The integral (27) is replaced with

$$G^2 = \int_{-\pi/T}^{\pi/T} |H(j\omega)|^2 d\omega \qquad (32)$$

where $|H(j\omega)|^2$ is the magnitude squared of the transfer function of the digital filter. This computation is readily performed in the discrete time domain i.e.

$$G^2 = \sum_{j=0}^{\infty} h^2(j) \qquad (33)$$

where h(j) is the discrete time response of the digital filter.

The design of filters for a digital fade simulator presents alternate possibilities. The first possibility is to use FFT techniques. The incoming digital noise sequences may be fourier transformed using an FFT. This FFT is then multiplied by a spectral function or mask such that the magnitude squared rolls off as f^{-3}. The result of this multiplication is then inverse transformed by the FFT and used as one or the other sequences x(t) or y(t). There are two problems with this. The first is that spectral modifications occur due to the basic nature of the FFT process, i.e., only finite lengths of samples may be transformed. The second problem is that a phase function must be assumed for the mask. The independent specification of both the phase and gain of the filter design although a particular phase function may be initially estimated and then used in an iterative procedure (utilizing Hilbert Transforms) to specify both the phase and gain of the filter. The phase function has the interesting property that it is approximately 135^0 in the region of the f^{-3} rolloff. The other possibility for the choice of digital filter design is to use digital versions of analog filters. This straightforward approach eliminates the problems of finite duration FFT's, storage of masks, and the question of the realizability of the filter. It also leads to a faster algorithm that consumes very little storage and computation time on smaller machines that may be dedicated to simulating and monitoring actual hardware systems.

The approach adopted was to use a bilinear z transform of the cascade of two one pole filters. This rather crude approximation worked surprising well on the 250 MHz data reported in reference [3].

The Laplace transform of two one-pole filters is

$$H(s) = \frac{a}{s+a} \cdot \frac{b}{s+b} \qquad (34)$$

where $a = 2\pi F_a$ (35)

and $b = 2\pi F_b$.

The parameters are the two cutoff frequencies of the filter. In general F_a is the first cutoff (-3 dB point) and F_b is the higher frequency cutoff of the filter. The bilinear z transform consists of substituting

$$S = \frac{2}{T} \cdot \frac{(1-z^{-1})}{(1+z^{-1})} \qquad (37)$$

where T is the sampling interval. The transform H(z)is just

$$H(z) = \frac{\alpha\gamma \cdot (1+2z^{-1}+z^{-2})}{\beta \cdot \Delta - (\beta+\Delta)z^{-1}+z^{-2}} \qquad (38)$$

where $\alpha = (aT/2)/(1-aT/2)$, $\beta = (1+aT/2)/(1-aT/2)$, $\gamma = (bT/2)/(1-bT/2)$ and $\Delta = (1+bT/2)/(1-bT/2)$. The corresponding difference equation is

$$\beta\cdot\Delta\cdot f(i)-(\beta+\Delta)f(i-1)+f(i-2)=\alpha\cdot\gamma\cdot(g(i)+2g(i-1)+g(i-2)). \qquad (39)$$

The input is g(i) and the output is f(i). The storage requirements are minimal; two (1x3) vectors plus constants. Note that T must be small enough so that $1-aT/2 > 0$ and $1-bT/2 > 0$.

It remains to compute G^2. Instead of computing G^2 according to equation (33) the following computation may be performed.

$$G^2(n) = \sum_{j=0}^{n} h^2(j) \qquad (40)$$

and $G^2 = \lim_{n\to\infty} G^2(n)$.

The calculation $G^2(n)$ is useful for estimating the time until the filter transients have died down. Expression (24) is valid only for filter outputs that may be considered stationary, i.e., filters that were started at $t = -\infty$. For $a \neq b$ or equivalently $F_a \neq F_b$, then

$$H(z) = A \cdot \frac{\alpha(1+z^{-1})}{(\beta-z^{-1})} - B \cdot \gamma\frac{(1+z^{-1})}{(1-z^{-1})} \qquad (41)$$

where $A = \gamma(1+\beta)/(\Delta-\beta)$ (42)

and $B = \alpha(1+\Delta)/(\Delta-\beta)$. (43)

This is just the sum of two first order filter responses, thus,

$$h(j)=0 \qquad j<0 \qquad (44)$$

$$h(j)= A\cdot\alpha/\beta-B\cdot\gamma/\Delta \qquad j=0 \qquad (45)$$

$$h(j)= A\cdot\alpha\cdot(\beta+1)\cdot\beta^{-(j+1)} - B\cdot\gamma\cdot(\Delta+1)\cdot\Delta^{-(j+1)} \qquad j>0 \qquad (46)$$

using (44), (45) and (46) in expression (40) then

$$G^2(n)= (A\cdot\alpha/\beta-B\cdot\gamma/\Delta)^2 + (A\cdot\alpha\cdot(\beta+1)1\beta)^2 \cdot (1-\beta^{2n})/(\beta^2-1) + (B\cdot\gamma\cdot(\Delta+1)/\Delta)^2 \cdot (1-\Delta^{2n})/(\Delta^2-1)-(2\cdot A\alpha\cdot B\cdot\gamma/\beta\cdot\Delta)\cdot (\beta+1)(\Delta+1)\cdot(1-(\beta\Delta)^{2n})/(\Delta\beta-1). \qquad (47)$$

taking the limit as $n\to\infty$ then

$$G^2= (A\cdot\alpha/\beta-B\cdot\gamma/\Delta)^2 + (A\cdot\alpha\cdot(\beta+1)/\beta)^2/(\beta^2-1) +(B.\gamma\cdot(\Delta+1)/\Delta)^2/(\Delta^2-1)-(2\cdot A\cdot\alpha\cdot B\cdot\gamma/\beta\cdot\Delta) (B+1)(\Delta+1)/(\Delta\beta-1) \qquad (48)$$

thus the scaling factor G^2 may be directly computed from the filter parameters. The settling time for the filter may be estimated by assuming that b>>a, i.e., that the high frequency cutoff is much higher than the -3dB point. Since the simulation results seem to bear this out then the expressions for $G^2(n)$ may be simplified.

$$G^2(n)\approx (\alpha/\beta)^2 + (\alpha(\beta+1)/\beta)^2(1-\beta^{2n})/(\beta^2-1). \qquad (49)$$

This expression attains 90% of its limiting value as $n\to\infty$ for $n\geq N_0 \triangleq \cdot5\cdot(\log 10\beta)^{-1}$.

This completes the specification of the values for the digital fade simulator. The variances and cross correlation are set by expressions (28), (29), (30), (31), the normalization constant G^2 is computed from the filter constants using expression (48), and the cutoff frequencies may be determined by examining intensity spectra or comparison of simulations of fade duration distributions with measured fade duration distributions. The ratio $b^2/(\sigma_x^2 + \sigma_y^2)$ may be estimated from scintillation index data as noted in section 3.0.

5.0 SOME RESULTS AND CONCLUSIONS

The utility of the design procedures presented in the previous sections was verified by the construction of a digital fade simulator similar to figure 1. The RF portion consisting of the mixers, phase shifter and power splitters were eliminated and the statistics of the signal amplitude were monitored by directly observing the components of E, i.e., x(t) + b and y(t). In particular, the properties of the signal amplitude were observed. This amplitude is

$$A(t) = [(b + x(t))^2 + y^2(t)]^{\frac{1}{2}}. \qquad (50)$$

The property of A(t) of interest is the fade duration distribution of A(t). This distribution is the total number of times the amplitude A(t) was less than a particular level given that it stayed below that level for a particular length of time or longer. For these distributions the particular levels were chosen to be 1.0, .80, .60 and .40 times the mean value of the received signal amplitude.

The experiment proceeds as follows. The sampling period chosen was .02 seconds for a sampling rate of 25 Hz. The filter parameters were initialized and then the processes x(t) and y(t) were generated and used to compute a discrete time series of signal amplitude A(t). The simulation is first allowed to run without collecting data in order to allow the filter transients to settle. Using the computation for N_0 in the previous section, T = .02 seconds and F_a = .04Hz then the transient period consisted of about 300 noise samples. After the filter transients have died down, 40960 samples of the amplitude A(t) were computed. This corresponds to 819.2 seconds of data or about 13.6 minutes of data. This allows comparison with measured data since actual measurements are only stationary over 10-15 minute intervals. The samples were used to compute the fade duration distribution. The case of one-pole Butterworth filters is illustrated in figure 2 for 250 MHz data from TACSAT I [3]. For this case the value of σ_x^2 = .9, σ_{xy}^2 = 0, and b^2 = 5.58. The cutoff frequency of the filter is .04 Hz. This corresponds to the results reported by Deckelman and Ziemer in their figure 6 [3]. Notice that the distribution is overly rich in short duration fades for this one pole case. The choice of σ_x^2 = .9 is not a theoretically predicted value and in spite of the reasonable fit for the amplitude distribution the phase distribution can be expected to be severely distorted. Since the scintillation index was quite large ($\geq$ .4) the strong scatter limit values for σ_x^2, σ_y^2 and σ_{xy}^2 were used. The filter cutoff frequencies were set at .04 Hz and .7 Hz. The value of b^2 was 1.6. The fade duration distribution was computed for this case and compared with the measured data. The results are plotted in figure 3 and compared with the measured data points (small circles). The fit is quite good, the rapid fades are more in line with the measured data and the deep fades are as good as the previous case. For this case the compiled value of the scintillation index was S_4 = .738. This implies a value of b^2 = 1.8 which is reasonably close to the actual value of 1.6.

It should be noted that many choices of the scintillation parameters can give adequate fits of the amplitude data. For fits of both amplitude and phase more careful choices of the constants must be made. From the results prescribed here it appears that the theoretical values of the scintillation parameters plus the use of the cascade of two one-pole filters to shape the spectrum can generate fade duration distributions that well approximate the data.

REFERENCES

1. Fremouw, E. J. and C. L. Rino, "Modeling of Transionospheric Radio Propagation," Final Quarterly Technical Report, August 1975, SRI.

2. Rino, C. L., R. C. Livingston, and H. E. Whitney, "Some New Results on the Statistics of Radio Wave Scintillation 1, Empirical Evidence for Gaussian Statistics," Journal of Geophysical Research, Vol. 81, No. 13, May 1, 1976.

3. Deckelman, W. F. and R. E. Ziemer, "Computer Modeling of the Statistical Properties of Trans-ionospheric Scintillation Channels," IEEE Transactions on Communications, April 1975.

4. Rino, C. L., "Some New Results on the Statistics of Radio Wave Scintillation 2, Scattering from a Random Ensemble of Locally Homogeneous Patches," Journal of Geophysical Research, Vol. 81, No. 13, May 1, 1976.

5. Rufenach, C. L., "Power-Law Wavenumber Spectrum Deduced from Ionospheric Scintillation Observations," Journal of Geophysical Research, September 1, 1972.

6. Briggs, B. H. and I. A. Parkin, "On the Variation of Radio Star and Satellite Scintillations with Zenith Angle," Journal of Atmospheric and Terrestrial Physics, 1963, Vol. 25.

7. Crane, R. K., "Morphology of Ionospheric Scintillation," 1974 National Telemetry Conference, NTC 74-285.

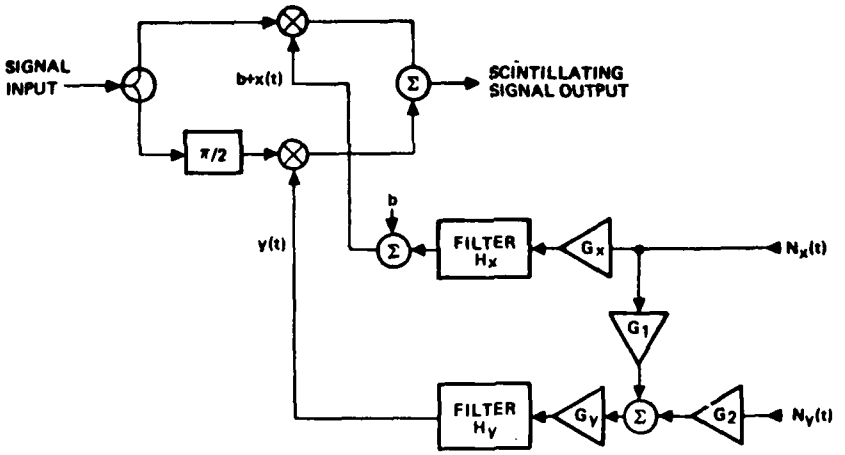


Figure 1

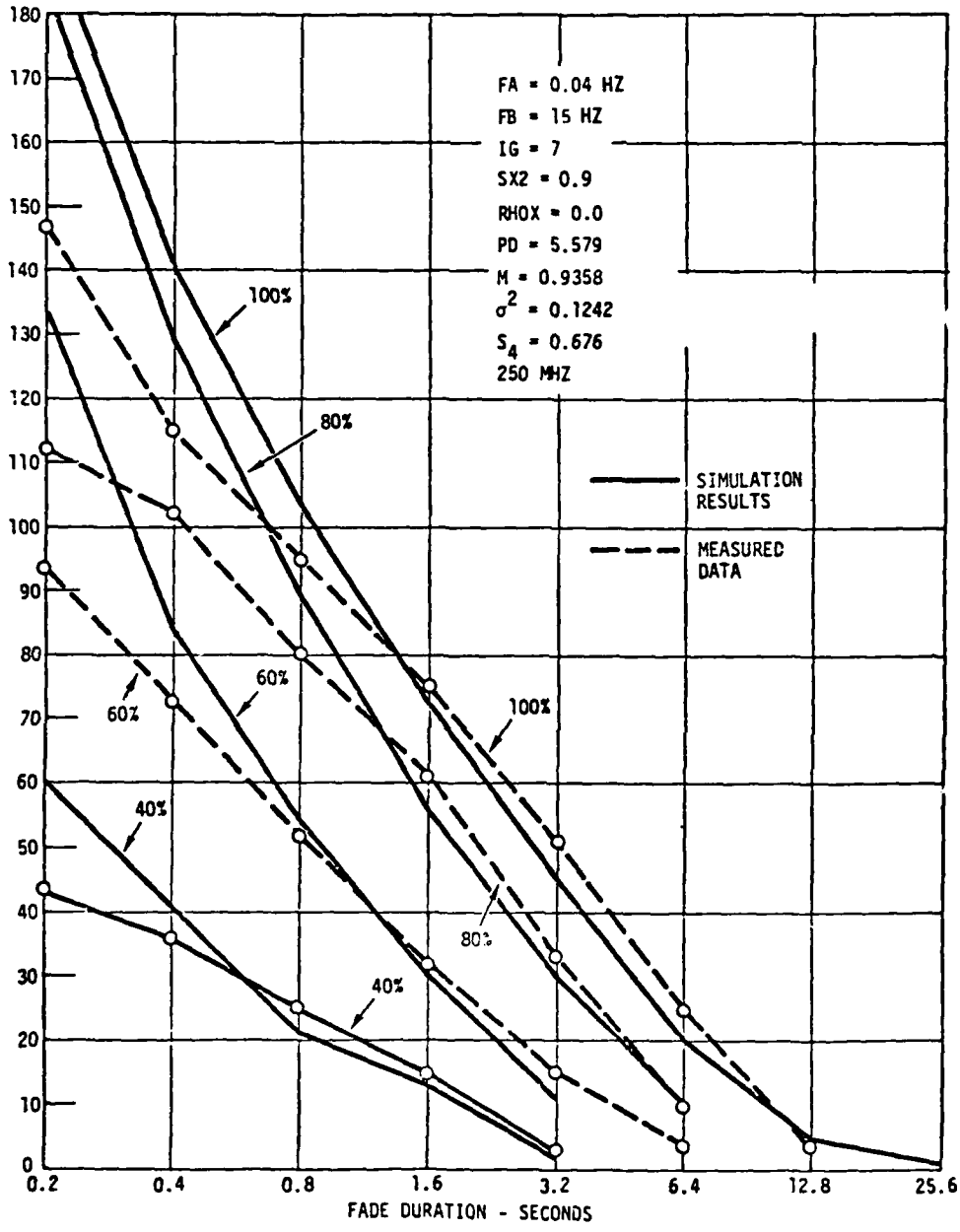


Figure 2. Total Fades Vs. Fade Duration in 819.2 Seconds (One Pole Filter)

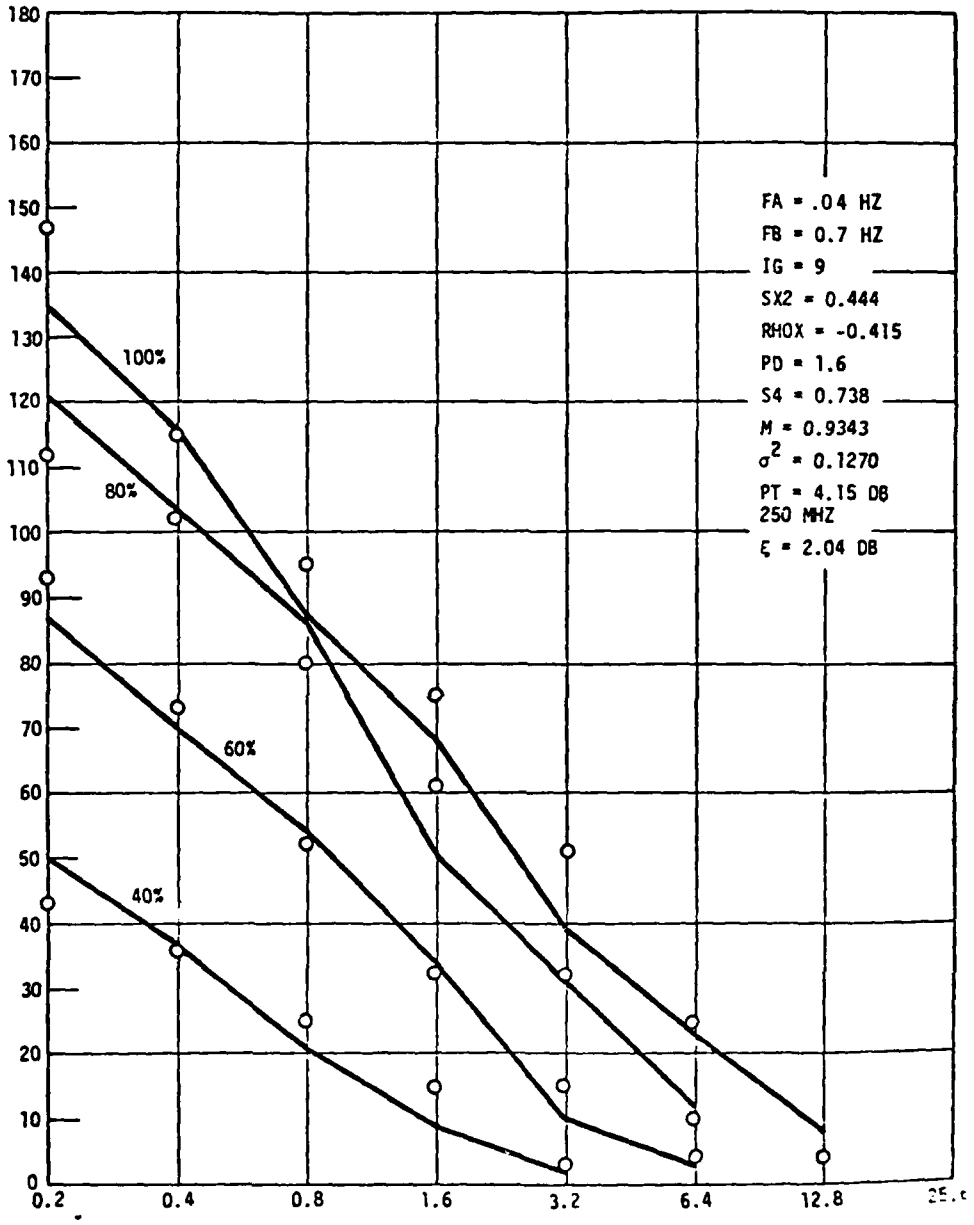


Figure 3. Total Fades Vs. Fade Duration in 819.2 Seconds
(Cascade of Two One-Pole Filters)

AN ANALYSIS AND SIMULATION OF A SCINTILLATING TIME-DIVISION-MULTIPLEXED TELETYPE CHANNEL

C. C. Kilgus, W H. Goodnight, L. G. Smith,
H. K. Stilwell, D. D. Stott J. M. Whisnant,
T. Wyatt

Applied Physics Laboratory
The Johns Hopkins Universtiy
Howard County, Maryland

INTRODUCTION

The effect of ionospheric scintillation on a geo-synchronous communications satellite UHF downlink has been considered.

The study included a survey of the existing scintillation data. The Naval Electronics Laboratory Center (NELC) and the Air Force Geophysical Laboratory (AFGL) generously provided data applicable to the effort.

The measure of channel performance used in the analysis was the average bit error probability, the probability averaged over many fades. An upper bound on the average error probability was derived and plotted against the undisturbed SNR for each scintillation group.

A simulation was then constructed to find the average bit error rate that produced readable messages in the scintillating channel. Tapes of measured scintillation data provided by AFGL were read to obtain bit by bit deviations from the undisturbed SNR. The probability of error for each bit was calculated and a random number table read to determine if the bit was to be corrupted. The resulting messages were printed and the readability graded.

The conclusions resulting from this investigation are the following:

1. The Air Force Geophysical Laboratory has established a scintillation data base that allows average bit error rate to be calculated as a function of user latitude, local time the magnetic index, and the undisturbed link SNR.

2. The average bit error rate, that is, the bit error averaged over many fades, is a good measure of message readability during scintillation. Plain text messages with an average bit error rate of 1×10^{-3} are readable with few exceptions.

3 On a channel with a 12.1 dB undisturbed SNR, the analysis indicates that an acceptable 1×10^{-3} average bit error rate is maintained more than 95% of the time at high latitudes and 100% of the time at mid-latitudes. However, at low latitudes between the hours of 1800 and 0200 local time the channel reliability will be low and additional SNR or bit stream redundancy must be incorporated to insure reliabile transmissions.

4. The analysis indicates that increasing the undisturbed SNR to 24 dB appears to provide immunity to scintillation effects at all latitudes and times.

SIGNAL DESCRIPTION

The UHF communications system considered uses PSK (NRZ-M) modulation in the channel. The SNR in the absence of scintillation is 12.1 dB, providing a bit error rate less than 1×10^{-5}.

The system teletype uses a five level code and requires a start and stop bit for each character; hence, seven bits are transmitted for each character. The total channel is composed of fifteen (15) message channels and one (1) sync channel, bit by bit multiplexed into a single bit string.

After the messages are multiplexed the bit string is converted to the

NRZ-M code. In the NRZ-M code a one bit is represented by a change in level and a zero bit by no change.

IONOSPHERIC SCINTILLATION DATA

A large body of scintillation data has been collected by many organizations. Since the study concerns only synchronous communications satellites, a data base of scintillation observations of synchronous satellites was obtained. The frequency range of interest is about 240-320 MHz; it was decided to limit the data to that which lies in the range of 200 to 400 MHz. The data must appear in a form suitable for bit error rate analysis; one suitable form is therefore amplitude probability distributions.

An excellent source of data was found, the Air Force Geophysical Laboratory, whose methods are described in more detail below. Their data, taken in all three latitude regions, were used to obtain the bit error rates determined in this study. The sites at which the AFGL measurements were made are in Greenland, Massachusetts, and Peru.

Air Force Geophysical Laboratory Data

The Air Force Geophysical Laboratory has adopted a standard scintillation index for their ionospheric studies:

$$SI = \frac{P_{max} - P_{min}}{P_{max} + P_{min}} \times 100$$

where P_{max} is the third peak down from the maximum and P_{min} is the third minimum up from the lowest excursion in the given sample period. A 15-minute sample interval was chosen in order to achieve stationarity of the signal statistics over the measurement period. In general, scintillations have periods that range from about one per second to one per minute. A 15-minute interval will therefore include many periods of the scintillations, while a longer sample interval would be subject to possible changes in signal statistics.

The SI parameter does not describe the fading characteristics of the signal in sufficient detail for the engineer interested in propagation effects on system design. To determine the fading margin necessary to overcome scintillation, the time percentage that the signal fades below various levels is also required.

At AFGL the values of the scintillation index were divided into six scintillation groups. A range of scintillation indices were included within each group. For the data used in this study, the groups and ranges of SI with corresponding changes in signal level are shown below:

Group	SI %	$P_{max} - P_{min}$ dB	m
0	<23	<2.0	165
1	24-43	2.0-3.7	32
2	44-60	3.7-5.7	14.5
3	61-82	5.7-9.6	7.3
4	83-91	9.6-13.1	4.0
5	92-99	≥13.1	2.3

The AFGL data were processed by recording the output of their receivers on magnetic tape and then digitizing at 6 samples per second. The digitized data were then divided into 15-minute segments and processed by computer. Amplitude probability distributions were obtained for each 15-minute segment and the results were sorted according to scintillation group. The total data then yield a corresponding median cumulative distribution function or model distribution for each scintillation group.

It was determined that an individual 15-minute distribution can be described by a Nakagami-m probability distribution. Although the Nakagami-m distribution has no theoretical basis for ionospheric scintillation, it provides a reasonable approximation to the measured data. The shape of the Nakagami-m distribution is completely specified by the single parameter m which may be any positive number ½ or greater. The probability density function of the Nakagami-m distribution is given by

$$P(P_r) = \frac{m^m P_r^{m-1}}{\Gamma(m) P_o^m} e^{-m(P_r/P_o)}$$

where P_r is the received power level and P_o is the average received power. Since the individual distributions for each scintillation group can be described by Nakagami distributions, the median for group is also a Nakagami distribution. The model distributions are used to determine a cumulative distribution function (cdf) which corresponds to the frequency of occurrence of the scintillation groups for a given period. The AFGL data lists the percent occurrence of the 15-minute scintillation index values for each group and sorts the values according to local time. To determine a cdf the percent occurrence of each scintillation group is multiplied by its probability at each dB level as given by the Nakagami

distribution and the values are summed to obtain the result.

EVALUATION OF SCINTILLATION EFFECTS ON THE CHANNEL

The evaluation consisted of two efforts:

1. The average bit error rate that will occur for each scintillation group was calculated as a function of the undisturbed channel SNR. The AFGL data were then used to describe the scintillation group (and therefore average bit error rate) occurrence as a function of latitude, local time, and magnetic index.

2. The average bit error rate required for readable messages was determined by computer simulation using AFGL tapes of signal levels during scintillation.

The combination of these two efforts allows the channel reliability to be determined as a function of latitude, local time, and magnetic index.

Average Error Probability in the Channel

The primary effect of scintillation is to degrade the error rate due to variation in signal level. The effect of scintillation-induced phase perturbations is considered to be inconsequential for the relatively low data rates employed at UHF.

The received signal level in a channel without fading is deterministic. The error probability can be calculated given one parameter, the SNR at the receiver detector. The bit errors are uniformly distributed in the bit stream. In a scintillating channel the received signal is a random variable with a Nakagami m distribution. Two parameters are required - the undisturbed SNR and the m parameter that describes the fading. The errors occur in bursts when the signal level is near a minimum. Average error rates can still be calculated, that is, the error probability over many cycles of the fading.

A closed form bound on the average probability of error, valid for integer values of m, was derived in the study:

$$P(\epsilon) < 1 - \frac{1}{\sqrt{\pi}} \left(\frac{\gamma_0}{m}\right)^{\frac{1}{2}} \left(\frac{m}{m+\gamma_0}\right)^{m-\frac{1}{2}} \times \sum_{k=0}^{m-1} \frac{\Gamma(m-k-\frac{1}{2})}{\Gamma(m-k)} \left(\frac{m+\gamma_0}{m}\right)^{k}$$

where γ_0 is the undisturbed SNR. This bound is plotted in Figure 1.

Channel Average Bit Error Rate

The undisturbed SNR in the link is 12.1 dB. The following error rate for each scintillation group can be determined from Fig. 1.

Scintillation Group	m	Average $P(\epsilon)$
0	165	$<1 \times 10^{-5}$
1	32	$<1 \times 10^{-5}$
2	14.5	$<1 \times 10^{-5}$
3	7.3	5×10^{-5}
4	4.0	4.5×10^{-4}
5	2.3	$\geq 3.5 \times 10^{-3}$

An average bit error rate of 1×10^{-3} was shown in the simulation to produce acceptable readability of message traffic; and for the purposes of this section is assumed to be the threshold of acceptable performance.

High, Mid and Low Latitude - AFGL data allows the channel performance to be determined at high, mid and low latitudes. The low latitude data (Figure 2) show Group 0 data, better than 1×10^{-5} ber, dominating from 0700 to 1800 but severe Group 5 scintillation with a ber greater than 3.5×10^{-3} frequently occurring at other times.

SIMULATED EFFECT OF SCINTILLATION ON MESSAGE TRAFFIC

A detailed description of the simulation is included in the study report.

Results of the Simulation

No Scintillation. A message was chosen and run through the channel with $m = \infty$ (no scintillation) and various SNR.

With a 5.2 dB SNR the average PSK (NRZ-M) bit error rate is 1×10^{-2}. Typical messages are unreadable; this performance level seems unsatisfactory. Even in the absence of scintillation the text contains long bursts of errors because of overtyping.

With a 5.9 dB SNR an effective error rate of 5.4×10^{-3} was produced. This begins the region of subjective judgment, in cases where lines were overtyped or synchronization lost the messages are unreadable, otherwise they are garbled but sometimes readable (Message 1).

With 6.3 dB SNR an effective error rate of 3.5 x 10^{-3} was produced. These messages are generally readable except for frequent cases of overtyping (Message 2).

With a 7.4 dB SNR the effective bit error rate is 1 x 10^{-3} and the messages are with few exceptions readable (Message 3); this performance level seems satisfactory.

Group 4 Scintillation (m = 3.93). An 11 dB undisturbed SNR is necessary to maintain a 1 x 10^{-3} average bit error rate in the face of group 4 scintillation. The messages readability is excellent except for rare cases of overtyping.

Group 5 Scintillation (m = 1.75). Three cases were run: a 17.5 dB undisturbed SNR corresponding to a 1 x 10^{-3} average bit error rate; a 12.1 dB undisturbed SNR corresponding to the channel studied; and a 24 dB undisturbed SNR.

The messages with 1 x 10^{-3} error rate had excellent readability (Message 4).

The messages with 24 dB SNR were essentially completely error free.

The messages with 12.1 dB SNR were generally not readable (Message 5).

STUDY REPORT

"The Effect of Ionospheric Scintillation on the Fleet Broadcast of the Fleet Satellite Communication Systems", Applied Physics Laboratory Report SDO-4380.6, December 1976

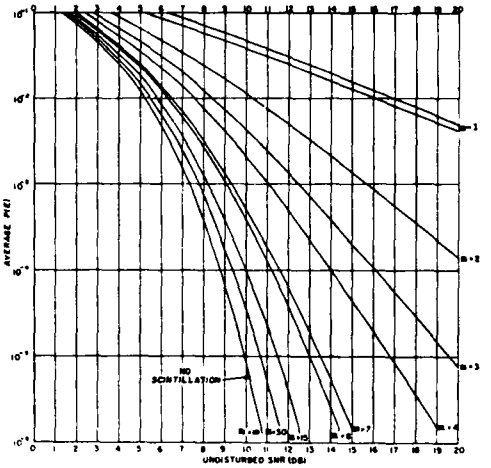


Fig. 1 Bit Error Rate in the Scintillating Channel

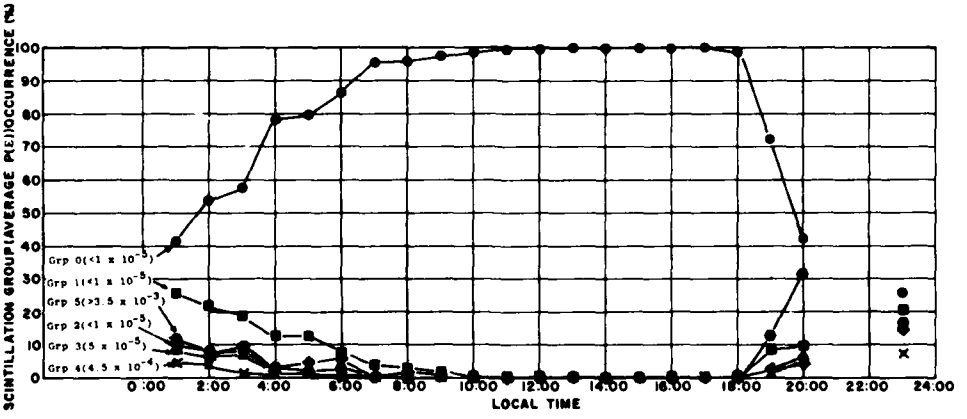


Fig. 2 Low Latitude Scintillation Group and Channel Average P(ε) Occurrence PDE(MRS-H) No RFI Undisturbed SNR = 12.1 dB K = 4-6 Winter

Message 1

1. RECENT SHIP OBSERVATIONS HAVE SHOWN THAT EXCELLENT BROADCAST EADABILITY HAS BEEN OBTAINED BEYOND THE THEORETICAL MAXIMUM BWMITS OF GAPFILLER EARTHMCOVERALENDEPICTED FIGURE 2-1,IG 2-3 OF REF A. [illegible] THIS POINT WILG ENABLE A REFINEMENT OF ACTUAL FRINGE ZGNA KOVER G, ANJ PROVE VALUABLE IN FORECASTING ACT AL RECEPTIOC ZONES.

Message 2

1. RECENTLLTIP OBSERVATIONS HAVE SHOWN THAT EXCELLENT BROADCAST [illegible] OF GAPFILLER EARTH COVERAGE DEPICTED FIGURE 2-1,PG 2-3 OF REF A. ALTHOUGH RELIABLE COPY SHOULD NOT, THEORETICALLYWN BE EXPECTED BEYOND THE FIVE DEGREE LOOK ANXB, OBSERVATIONS OF ECEPTION BEYOND THIS POINT WILL ENABLE A REFINEMENT OF ACTUAL FRINGE ZONE COVERAGE, AND PROVE VALUABLE IN FORECASTING ACTUAL RECEPTION ZONES.

Message 3

1. RECENT SHIP OBSERVATIONS HAVE SHOWN THAT EXCELLENT BROADCAST READABILITY HAS BEEN OBTAINED BEYOND THE THEORETICAL MAXIMUM LIMITS OF GAPFILLER EARTH COVERAGE DEPICTED FIGURE 2-1,PG 2-3 OF REF A. ALTHOUGH RELIABLE COPY SHOULD NOT, THEORETICALLY, BE EXPECTED BEYOND THE FIVE DEGREE LOOK ANGLE, OBSERVATIONS OF RECEPTION BEYOND THIS POINT WILL ENABLE A REFINEMENT OF ACTUAL FRINGE ZONE COVERAGE, AND PROVE VALUABLE IN FORECASTING ACTUAL RECEPTION ZONES.

Message 4

1. RECENT SHIP OBSERVATIONS HAVE SHOWN THAT EXCELLENT BROADCAST READABILITY HAS BEEN OBTAINED BEYOND THE THEORETICAL MAXIMUM LIMITS OF GAPFILLER EARTH COVERAGE DEPICTED FIGURE 2-1,PG 2-3 OF REF A. ALTHOUGH RELIABLE COPY SHOULD NOT, THEORETICALLY, BE EXPECTED BEYOND THE FIVE DEGREE LOOK ANGLE, OBSERVATIONS OF RECEPTION BEYOND THIS POINT WILL ENABLE A REFINEMENT OF ACTUAL FRINGE ZONE COVERAGE, AND PROVE VALUABLE IN FORECASTING ACTUAL RECEPTION ZONES.

Message 5

1. RECENT SHIP OBSERVATIONS HAVE SHOWN THAT EXCELLENT BROADCAST READABILITY HAS BEEN OBTAINED BEYOND THE THEORETICAL MAXIMUM LIMITS UQEEC JBDEAL [illegible] FIGURER[illegible]GK2-3 OF REF A. ALTHOUGH RELIABLE COPY SHOULD NOT, THEORETICALLY, BE EXPECTED BEYOCTHIS POINT WILL ENABLE A REFINEMENT OF ACTUAL FRINGE ZONE COVERAGE, AND PROVE VALUABLE IN FORECASTING ACTUAL RECEPTION ZONES.

A REVIEW OF RECENT STUDIES OF EQUATORIAL F-REGION IRREGULARITIES AND THEIR IMPACT ON SCINTILLATION MODELING

Sunanda Basu*
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

M.C. Kelley
School of Electrical Engineering, Cornell University
Ithaca, N.Y. 14853

Abstract Our understanding of the various plasma instabilities that cause nighttime equatorial F-region irregularities and their effects upon radio wave propagation has grown enormously since the last IES Symposium three years ago. This has been achieved through the combined use of ground-based, aircraft and in-situ measurements in the radio and optical domains. Analytical and computer simulation investigations have kept pace with the experimental results. In this review we shall outline the present state of both experiment and theory. A realistic modeling of the irregular medium must precede effective communications channel modeling. We shall thus endeavor to bring to the attention of the systems engineer those aspects of scintillation modeling most affected by the new understanding of equatorial spread-F.

1. INTRODUCTION

Four decades after the discovery of the phenomenon called 'equatorial spread-F' (Booker and Wells, 1938), the general problem of nighttime equatorial F-region irregularities has remained a fascinating one to theoretical and experimental geophysicists. This problem has also been one of serious concern to communications engineers as it is well known that these irregularities cause amplitude scintillation which can degrade the performance of satellite communication links. More recently, it has become clear that naturally occurring phase scintillation can impair the performance of satellite surveillance systems that use synthetic aperture processing to achieve high range resolution.

* NRC Resident Research Associate

A variety of different techniques, such as, in-situ rocket and satellite measurements, radar and airglow observations as well as numerical and analytical investigations have, on one hand, enlarged vastly the scope of the problem and, on the other, contributed greatly to our understanding of equatorial F-region irregularities. It is also interesting to note how our improved understanding of the characteristics of equatorial irregularities has influenced the development of more realistic models of equatorial scintillations. For instance, soon after in-situ measurements revealed that the irregular medium is best represented by a power-law type irregularity power spectrum (Dyson et al., 1974) over a wide range of scale lengths (~10 km to 10 m) rather than by a gaussian spectrum characterized by a single scale size (Briggs and Parkin, 1963), modeling of ionospheric scintillation was tailored to reflect this change (Rufenach, 1975; Costa and Kelley, 1977). It is also being recognized that saturated amplitude scintillations (>20 dB) in the VHF/UHF band and moderate fadings (4-8 dB) in the GHz band over the equatorial region can only be modelled in terms of large amplitude irregularities distributed in thick layers having a power-law type irregularity power spectrum with large, ill-defined outer scale sizes (Basu and Basu, 1976). These large scale size irregularities can contribute substantially to phase scintillation even under conditions of weak amplitude scatter (Fremouw et al., 1978) as we shall be discussing later.

In our last review on the subject presented at the 5th International Symposium of Equatorial Aeronomy in August, 1976 and recently published

(Basu and Kelley, 1977, hereafter to be referred to as Paper I), we discussed the latest information, available at the time, on equatorial irregularities obtained by the various techniques and presented a model of equatorial scintillations consistent with these observations. Since then considerable new information has become available as a result of the Air Force Geophysics Laboratory (AFGL) equatorial irregularity campaigns in Peru in October, 1976 and March, 1977, and the analysis by Stanford Research Institute (SRI) of the data obtained with DNA Wideband satellite which was launched in May, 1976. In addition, considerable theoretical work on plasma instability and fluid turbulence has been conducted during the past year. Most of the papers resulting from such current experimental and theoretical efforts are under various stages of publication and the present review will be an appropriate forum to bring such material to the attention of the ionospheric community.

As in Paper I, we shall limit the scope of the present review to cover only the irregularity generation mechanisms and their effects on both amplitude and phase scintillation as well as VHF radar backscatter. Worldwide equatorial scintillation morphology has been adequately reviewed by Aarons (1977). Since Paper I has recently been published, we intend to deal very briefly, in Section 2, with the topics mentioned in the earlier paper. The two major experimental efforts by AFGL and SRI are described in Sections 3 and 4 and current theoretical efforts are discussed in Section 5. A brief summary is given in Section 6.

2. BRIEF REVIEW

In Paper I we pointed out that for scintillation effects in the equatorial region we are primarily interested in the properties of the disturbed equatorial ionosphere near or above the peak in electron density. This is due to the fact that scintillations are controlled by the absolute fluctuations in electron density. Thus even though equatorial irregularities originate below the F peak and result in very intense fluctuations in electron density there, the scintillation effects are modest. This has been discussed in detail by Costa and Kelley (1976) who calculated the scintillation effects of the observed plasma density profiles shown in Figure 1 and concluded that only modest scintillation would result at VHF frequencies with negligible effect at GHz frequencies. On the other hand, using topside in-situ irregularity data from Ogo-6 shown in Figure 2, Basu and Basu (1976) showed that saturated VHF scintillation and moderate GHz scintillation can be explained on the basis of large amplitude irregularities in an environment of high electron density (McClure and Hanson, 1973) distributed in thick layers as observed by radar studies (Basu et al., 1977a). Nonetheless we need to briefly review the bottomside processes since they are the origin for the high altitude disturbances.

Several mechanisms have been proposed as the initial source for the bottomside structures. The Rayleigh-Taylor gravitational instability has been invoked as an explanation for observed irregular structure of the equatorial F-region ionosphere (Dungey, 1956; Haerendel, 1974; Hudson and Kennel, 1975). The growth rate is small, however, and it is possible that some other seed mechanism might actually initiate the bottomside structure. Beer (1974), and more recently, Booker and Ferguson (1977) have strongly argued that the seat of the primary instability is in the neutral medium. Martyn (1959) and Cole (1974) have suggested that neutral winds in the E region play a role in creating F region structure. The drift instability (Hudson and Kennel, 1975) and partial reflections from the steep unperturbed gradient (Balsley and Farley, 1975) have also been invoked to explain weak bottomside spread F at short wavelengths.

Whatever the initial process, there is strong evidence that regions of low plasma density rise into the high density topside region. Evidence for this comes from rocket (Kelley et al., 1976; Morse et al., 1977), radar studies (Woodman and La Hoz, 1976), satellite (McClure et al., 1977) and indirectly from airglow observations (Weber et al., 1977). As an example consider the rocket data in Figure 1 where a "hole" in plasma density is indicated just below the F peak. Less than two minutes after the rocket passed through this low density region, a ground based 50 MHz radar detected a patch of 3 m irregularities rising rapidly from the same region. The Jicamarca radar maps showing the existence of intense regions of 3 m backscatter high above the F peak and the airglow evidence will be discussed in Section 3. Satellite data in Figure 3 taken from McClure et al., (1977) shows low density topside "holes" or

"bubbles" in which the plasma drift velocity is upward and westward with respect to the background medium. Moreover these depletions have ion composition typical of low altitude regions, that is, significant numbers of molecular and metallic ions (Hanson and Sanatani, 1973; McClure et al., 1977).

Scannapieco and Ossakow (1976) have performed computer simulations of a Rayleigh-Taylor unstable ionosphere in the collision dominated case. They showed that an initial 3 km wavelength finite amplitude sinusoidal perturbation on the bottomside eventually formed a bubble-like structure which rose into the topside such as was seen by the experimental techniques mentioned earlier. The development time of 10^4 sec determined by Scannapieco and Ossakow was considerably longer than that observed but factors not included in the simulation could act to speed up the process. First, the observed zero order vertical gradient is considerably steeper than that used in the simulation. Second, the observed bottomside structures (Kelley et al., 1976; Morse et al., 1977; see also Figure 16) are of larger amplitude than the initial perturbation used in the computer simulation and were already highly nonlinear steepened structures. Attempts at treating analytically the nonlinear Rayleigh-Taylor instability in the collisionless regime have been presented by Chaturvedi and Kaw (1975) and Hudson (1977). Further recent analytical treatment of the problem will be presented in Section 5.

With this brief review of experimental and theoretical efforts made to understand the nature of equatorial irregularities we will present various recent coordinated scintillation and irregularity studies which give us a better idea of scintillation modeling in the equatorial region.

3. AFGL EQUATORIAL IRREGULARITY CAMPAIGNS

An intensive study of nighttime electron density irregularities in the equatorial ionosphere was performed in October, 1976 and March, 1977 by conducting simultaneous radar and scintillation measurements near the magnetic dip equator in Peru. The 50 MHz radar observations were made at Jicamarca and scintillation measurements were performed at the nearby ground stations of Ancon and Huancayo by receiving VHF transmissions from geostationary and orbiting satellites. In addition, the AFGL aircraft was employed to make on board scintillation, ionosonde and airglow measurements to provide spatial configuration of irregularity patches as well as to determine the existence of density depletions indicated by in-situ measurements. The multi-station scintillation measurements were used to study the localized origin of large scale irregularity patches, their drift speed, spatial extent and lifetime in the equatorial ionosphere. The simultaneous radar and scintillation observations also provided information on the relationship between the meter and kilometer sized irregularities giving rise to the radar backscatter and VHF-UHF scintillations respectively. Much new information became available as a result of this major effort and many publications and presentations have been made based on these correlated data sets. These are summarized below with specific references.

1) There is great variability of irregularity occurrence from one night to another. On certain nights, such as on October 16-17, 1976 a single irregularity patch may evolve at a particular location and then drift eastwards for a period dictated by its lifetime as discussed by Aarons et al., 1977. Figure 2 of the paper by Basu et al., 1978 in this volume gives the irregularity configuration for this night probably the simplest that was seen during the campaign periods. On some other nights, a series of irregularity patches with a large scale quasi-periodicity is observed. A good example of such a configuration was obtained on October 19-20 as shown in Figure 4 taken from the detailed report on the October campaign done by Basu and Aarons (1977). Figure 5 also taken from the report shows the respective positions for the observations. The important point to note from the radar map kindly provided by J.P. McClure is that a relatively thin layer of bottomside irregularities appeared at 1950 LT which by 2015 developed into a plume structure extending several hundred km into the topside. A second plume developed at 2140 LT. The drifting plume structures caused severe scintillations (>20 dB) at the ground stations with periods of 4 - 6 dB scintillation caused by the bottomside structures observed in between the plumes. It is interesting to note that HF forward scatter experiments in the equatorial region had earlier obtained evidence

of irregularity patches existing in quasi-periodic patterns (Röttger, 1976). On yet other nights there were no irregularities within the 900 km E-W coverage provided by the Jicamarca radar and various satellites as shown in Figure 5.

2) Although the irregularity patches occur after local sunset, these are found to evolve either to the west or the east of any specific station signifying that local conditions rather than local time dictate the generation of irregularities (Basu and Aarons, 1977).

3) The large scale irregularity patches are found to consistently drift eastwards between 1900 - 2400 LT with a speed ranging between 90-140 m/sec. Combining the temporal variation of scintillation with the drift speed, the E-W dimensions of the patches are found to range typically between 200-400 km, although some are larger (Basu and Aarons, 1977).

4) Another important aspect of the campaign was the determination of the relationship between the relatively large scale irregularities that cause scintillation ($\sim$ km to 100 m) and the small scale irregularities that cause 50 MHz backscatter (3m). Woodman and Basu (1977) using the simultaneous backscatter and scintillation data obtained from nearly common ionospheric volume on October 29, 1976 shown in Figure 6 found that both these types co-existed in the developing phase of the irregularities. However, they showed in a quantitative manner that in view of the 8 dB scintillation at 360 MHz, backscatter echo strength of approximately 80 dB above incoherent scatter levels is expected if the commonly observed monotonic power law spectrum (Dyson et al., 1974) extends upto 3m. There is thus a discrepancy of 4 orders of magnitude between the observed 40 dB enhancement and that computed. To bring computations and observations into agreement they postulated a Gaussian type of cut-off near the O^+ ion gyroradius which is on the order of 3m in the topside ionosphere. The existence of such a cut-off can be considered to be an effective inner scale of turbulence. This point is discussed further in Section 5. Recently, Booker and Ferguson (1977) have postulated an inner scale near the ionic gyroradius to explain spread-F signatures on equatorial ionograms in the HF band and Booker and Miller (1977) have discussed the importance of the inner scale in scintillation modeling. The existence of an inner scale has interesting practical implications for pulse propagation within the ionosphere. For instance if the ionic gyroradius is considered to be the effective inner scale rather than the Debye length, computations show that pulse broadening due to scattering in the 100 - 300 MHz will be 2 orders of magnitude smaller (Yeh and Liu, 1977).

As a result of further co-located scintillation and radar observations during the March, 1977 campaign, Basu et al., (1977b) have come to the conclusion that km and m scales coexist only during the developing phase in the early evening hours whereas later at night it is possible to have equally large km scale irregularities (as monitored by scintillation observations) without any associated backscatter. It is important to note that even L-band scintillations are possible without appreciable backscatter leading Basu et al., (1977b) to speculate that the cut-off scale length at this time is probably of the order of few tens of m. Further details and diagrams are given in paper by Basu et al., (1978) in this volume.

5) The AFGL aircraft scintillation observations have helped separate spatial and temporal behaviour of irregularities while the on-board ionosonde and optical imaging system have found evidence of electron density depletions in the bottomside F-region. Figure 7 using data kindly made available by J. Buchau shows the locus of the subionospheric point as the aircraft flew between ground stations on Oct 19-20, 1976. The thin lines signify the absence of scintillations while the thick lines signify their presence. The top panel shows that during 2247 UT (October 19) to 0015 UT (October 20), when the ground stations at Ancon and Huancayo did not record any scintillations as may be observed from Figure 4, the aircraft did not detect any irregularities in the entire latitude range of 11°S to 13°S and longitude interval of 72°W to 75°W. The bottom panel shows that from 0015 UT onwards the aircraft detected three irregularity patches with distinct boundaries indicating spatially localized irregularity generation after UT midnight (i.e., 1900 LT). The imaging system provided all-sky pictures of the 6300 Å OI airglow emission, which results from dissociative recombination of O_2^+ in the F region. Initial observations show the existence of north-south aligned regions of airglow depletion (Weber

et al., 1977). An isolated airglow depletion band observed on March 17, 1977 is modelled in Figure 8 as a troughlike bottomside electron density depletion which explains the observed airglow and ionosonde features. Further details regarding the association of the airglow depletion with scintillations and 3m backscatter are provided in the paper by Buchau et al., (1978) in this volume.

6) The scintillation data obtained from the various ground stations, in particular, spaced receiver measurements made at Ancon, and the data obtained from the AFGL aircraft as well as the Air Force Avionics Laboratory (AFAL) aircraft have been used to determine various diversity techniques to mitigate the effects of intense amplitude scintillation at 250 MHz (Whitney et al., 1977). In particular it was found that the high degree of decorrelation over a baseline of 366 m associated with rapid, severe intensity fluctuations ($S_4 \sim 1$) make space diversity techniques useful. Further details are given in paper by Whitney (1978) in this volume.

4. WIDEBAND SATELLITE OBSERVATIONS

The Wideband satellite was launched into a sun-synchronous near polar orbit on May 22, 1976 carrying a multifrequency coherent radio beacon on-board. The mutually coherent signals, which range from VHF to S-band are being recorded by SRI at ground stations in Poker Flat, Alaska and two equatorial stations, namely Ancon, Peru and Kwajalein in the Pacific sector. The second equatorial station at Kwajalein was set up to determine longitudinal differences in equatorial scintillation first pointed out by Basu et al., (1976) on the basis of in-situ irregularity data. We shall primarily discuss the equatorial data but point out certain basic differences with auroral scintillation structure that have been reported by the SRI group. The following information has been taken from Rino et al., (1977) and Fremouw et al., (1978).

1) The most important aspect of these observations has been the realization that ionospheric radio-wave propagation is dominated by large slowly varying phase scintillations. These can be large even in the absence of significant amplitude scintillations. The latter statement is particularly true for the auroral zone. Indeed, the measurement of simultaneous phase and amplitude scintillation has shown that computations of phase deviation based on the weak scatter theory using the observed amplitude scintillation would grossly underestimate the actual phase scintillation level. Based on these observations the SRI group has developed a multiplicative two-component model to characterize the joint first-order statistics of amplitude and phase. This model is discussed in detail by Fremouw et al., (1976). Briefly, the idea is to separate two components of the total (complex) scintiliating signal by filtering. Because of the power-law nature of the irregularity spectrum the low frequency cut-off had to be somewhat arbitrarily defined. Thus using a double detrending process it was possible to separate a 'focus' component having fluctuations with periods between 2.5 and 10 seconds and a 'scatter' component having fluctuation periods smaller than 2.5 secs. If one considers the scan velocity of the satellite in the F-layer of 3 km/sec then the focus component can be considered to be caused by refractive focusing and defocusing by irregularities between 30 km and 7.5 km acting as lenses. The scatter component is caused by smaller irregularities comparable to and smaller than the first Fresnel zone through the diffraction process. The focus component gives rise to the large slow phase fluctuations while the scatter component causes the fast intensity fluctuations. Figure 9 taken from Fremouw et al., (1978) contains one VHF data set showing these two components obtained from each of the three latitudinal regions (initially the Wideband observations were started at Stanford, the equipment being moved to Kwajalein in October, 1976) where SRI had ground stations. It is interesting to note that for the same value of the S_4 index (a measure of the scatter component), the standard deviation of phase (a measure of the focus component) varies substantially at different locations being the smallest at the equator. Later this has been found to be a consistent feature (Fremouw, 1977) and may be of importance for irregularity generation mechanisms. For communications purposes it is important to note that in the equatorial region the scatter component displays a very Rician characteristic while that obtained at midlatitude is highly nonRician. Whitney and Basu (1977) have reported earlier a difference in the slope of the intensity spectrum between an equatorial and auroral station.

2) The SRI group has categorized scintillations qualitatively into seven groups ranging from 'extremely quiet' to 'extremely active' categories. It is interesting to note that equatorial stations provide the majority of cases in both these extreme situations as compared to the auroral station where a 'modestly active' behavior is most frequent. This equatorial scintillation behavior is probably caused by the great day-to-day variability in irregularity occurrence discussed in Section 3.

3) During extremely active conditions, scintillations are found to persist in the gigahertz range such as shown in Figure 10. The observations of GHz scintillations with a low orbiting satellite proved that weak irregularities all the way upto the plasmasphere are not necessary as suggested by Booker (1975). Indeed, Booker and Miller (1977) have recently point out that weak irregularities persisting over the entire plasma envelope cannot explain GHz scintillations for which large amplitude irregularities at the F-region peak is necessary. Basu and Basu (1976) had earlier proposed such a model on the basis of in-situ and radar data.

4) The multifrequency Wideband observations have shown that the standard deviation of phase in general shows an f^{-1} behavior even under conditions of strong diffractive scatter. Occasional violation of the law has been observed in conjunction with strong scintillations in the GHz range. Figure 11 shows the frequency dependence of phase scintillation during two 20-second periods of the pass shown in Figure 10. The L-band point has been corrected for the S-band phase perturbations which were seen to be present at the beginning of the pass. In the less disturbed interval the f^{-1} law is generally followed whereas in the more disturbed sample the VHF departs significantly and even the UHF points show some scatter. At such times the slope P of the phase spectrum also flattens with values of $P > -2$ being observed at VHF and UHF.

5) The frequency dependence of intensity scintillations observed during the same periods is shown in Figure 12. The fully developed or saturated nature of the intensity scintillations at VHF and the five UHF frequencies analyzed during the more disturbed period is apparent in the departure of these six data points from a $f^{-1.5}$ dependence (Rufenach, 1974) that holds between L band and S band. Such a reduction in frequency dependence of intensity scintillations in the multiple scatter regime had been discussed earlier by Whitney and Basu (1977) and Mullen et al., (1977). Figure 12 also shows that under extremely active conditions it is possible to have intensity decorrelation across the UHF comb of frequencies. In this regard it should be noted that during the March, 1977 AFGL campaign it was found that most Wideband passes, even those showing GHz scintillation, were not associated with 3m radar backscatter (Basu et al., 1977b). This point has been discussed in Section 3. Thus whether intensity decorrelation during the developing phase of equatorial irregularities will be still greater is not yet known.

6) The SRI group on the basis of their observations at Kwajalein (9°N geog; 8°N dip) and Ancon (11°S geog; 2°N dip) have indicated that there is a local summer maximum of equatorial scintillations. On the other hand, Ghana observations (5°N geog; 8°S dip) over a five year period show a local summer minimum (Koster, 1976) in Figure 13. We are of the opinion that there is probably a longitudinal control over equatorial scintillations with differences being observed between the Atlantic and Pacific sectors as discussed by Basu et al., (1976) rather than a seasonal control. The reason for such a longitudinal variation is yet unclear.

7) There is strong indication from simultaneous scintillation and total electron content (TEC) measurements at Ghana over a 42 month period (excluding the June solstice) that scintillations in the premidnight period are associated with a decrease in TEC (Koster, 1976). The largest TEC depletion is noted at 2100 LT a time that is closely associated with the development of plumes on radar maps. This agrees well with current theoretical ideas and in-situ observations of equatorial irregularities being associated with density depletions. It is rather interesting to note that one instance of such a "hole" with 90% TEC depletion was observed in Kwajalein in August, 1977 (Fremouw and Lansinger, 1977). The fact that such a behavior has been reported only once during a year's observation probably indicates the close relationship between density depletions and the developing stage of equatorial irregularities which the Wideband satellite because of orbital constraints is in general unable to

sample.

5. RECENT THEORETICAL RESULTS

Theoretical efforts have kept pace with the multi-technique experimental efforts described above. The successful attempt to numerically simulate the rising bubbles have been described in Section 2. These computer simulations are strongly supported by an analytic treatment due to Ott (1977) who, assuming the experimental validity of the bubble concept, calculated their shape and velocity as a function of altitude and size. At low altitudes he found that the bubbles should be cylindrical (circular cross sections and uniform along $\underline{B}$) and that they rise with the velocity

$$u = \frac{n_0 - n_b}{n_0 + n_b} \cdot \frac{g}{\nu_{in}} \qquad (1)$$

where n_b is the density inside the bubble, n_0 is the density outside, g is the gravitational constant and ν_{in} the ion neutral collision frequency. At high altitudes where ion neutral collisions can be neglected, Ott showed that the equations governing the motion are identical to the equations of two dimensional ordinary fluids (for example, a liquid constrained between two planes with separation distance much less than the size of the planes). Considerable theoretical and experimental work has already been done on such fluids and can thus be directly applied to the present problem. The observed bubble shape is shown in Figure 14 and the velocity is given by

$$u \simeq .5\,(Rg)^{1/2} \qquad (2)$$

where R is the radius of the circular 'cap'. (It should be noted that all statements concerning velocity fields or velocity turbulence can be replaced by discussions of electric fields or electrostatic turbulence since $\underline{V} = \underline{E} \times \underline{B}/B^2$).

From our previous discussions we know that an outstanding feature of equatorial spread F on the topside is the presence of large scale ($\sim$ 10 km scale) regions of low plasma density with very sharp gradients at the edges. The source of energy for this interchange of low and high density plasma is the excess gravitational potential energy which resides in the F region plasma supported against gravity by the magnetic field. From a plasma physics standpoint the zero upward density gradient is disrupted by the electrostatic fields in the instability process. The second most important feature of equatorial spread F is the wide range of wavelengths exhibited by spread F irregularities as discussed in Sections 3 and 4. For instance, the 3 m scalelength is 3-4 orders of magnitude smaller than the primary process described above. Two processes have been recently suggested which may well acount both for the wide range of wavelengths as well as the extended spatial extent of spread F.

Costa and Kelley (1978a,b) have considered the steep gradients in plasma density associated with bottomside irregularities (see Figure 16) and presumably also with topside bubbles, and have shown that drift waves should grow rapidly on the observed gradients. The normalized growth rate is plotted in Figure 15 as a function of the parameter $b = k_\perp^2 \rho_i^2/2$ where $k_\perp$ is the wavenumber perpendicular to $\underline{B}$ and ρ_i is the ion gyroradius. For reference, b at the Jicamarca backscatter wavelength is 100. The growth rate peaks at $k_\perp \rho_i = 1.5$ which corresponds to wavelengths of about 20 m. The growth rate is linearly proportional to the gradient scale length, L, at the bubble edge and is 1 sec^{-1} for L = 125 m under typical ionospheric conditions. These drift waves should not only contribute to backscatter but play a role in the eventual merging of bubbles into the background (via destruction of the steep gradient).

A second and perhaps more important process from the standpoint of scintillations is the generation of a thick irregularity layer as suggested by Kelley and Ott (1978). Again use was made of Ott's results that in the collisionless case the equations of motion are identical to a two dimensional fluid. In the fluid case, the upward buoyancy force is balanced by emission of vortices in the background fluid. This injects velocity turbulence into the background fluid in a wake which trails behind the bubble. In the plasma case each vortex corresponds to a line charge surrounded by a circular $\underline{E}\underline{x}\underline{B}$ drift of the plasma. Since there is a background density gradient this vortex will create density irregularities with the same scale size as the velocity turbulence.

Numerous theoretical studies of turbulence in two dimensional fluids (Kraichnan, 1967; Lilly, 1969) have shown that velocity injected into such a fluid at some characteristic wave number k_0 will cascade to wavenumbers both higher and lower than k_0. In the space domain this 'inverse cascade' to lower k implies that the velocity

(and density) turbulence will spread into the largest volume available in the system. Evidence for this dual cascade in two dimensional turbulence has been found in the Earth's atmosphere (Kao and Wendell, 1970) and magnetosphere (Kelley and Kintner, 1978).

Kelley and Ott (1978) suggest that the vortices emitted by upwelling bubbles not only form the wake but tend to fill the between bubble regions (i.e., in an E-W direction) with velocity and density irregularities via this cascade to lower k values. Cascade to higher k also occurs with the ultimate spectral cutoff occurring at some large wavenumber where ion viscosity plays a role. This dissipation is most effective near the ion gyroradius and hence it seems likely that the 3 m scale (Jicamarca backscatter) lies in the dissipative subrange. Such a 'cutoff' in the amplitude of irregularities at short wavelengths was postulated by Woodman and Basu (1977) as discussed in Section 3.

In the intermediate range of wavenumbers $L_s^{-1} < k < L_d^{-1}$, where L_s is the stirring length (bubble size) and L_d is the dissipation scale, Kelley and Ott (1978) predicted a k^{-3} spectrum for velocity or electric field fluctuations and a k^{-1} spectrum for density irregularities. This seems in contradiction to numerous reports of one dimensional k^{-2} density spectrum associated with equatorial spread F. However, Costa and Kelley (1978a) have shown that at least in the case of bottomside non-linear structures, the k^{-2} spectrum is due to the steep edges in density encountered. Several examples of the steep edges are clear in the data presented in the top panel of Figure 16, obtained during the passage of a sounding rocket through bottomside spread F. The relative density ($\delta n/n$) is plotted for an 8 second interval ($\sim$ 16 km along trajectory) . As an illustration of the steep edge effect, Costa and Kelley (1978a) performed a fourier transform of the data in panel 1, then randomized the phase in each fourier component and reassembled the data in the time domain. Two such random phase runs were made and are plotted in the lower two panels. The power spectrum of all three sets of data is identical and varies as f^{-2} over most of the spectrum.

It is interesting to note that R. F. Woodman (private communication, 1976) had first suggested a phase coherence test as a way of distinguishing between the dominance of steep structures or turbulence in equatorial spread -F. More recently Chaturvedi and Ossakow (1977) have considered the non-linear development of bottomside spread -F. They conclude that steepened structures should evolve in time and in fact a k^{-2} spectrum would result due to the steep edges. They also suggested a phase coherence test to distinguish between such process and a turbulence like mechanism proposed earlier by Chaturvedi and Kaw (1976). The results of Costa and Kelley (1978a) described above seem to agree with the sharp edge hypothesis.

6. SUMMARY

The following picture evolves from the discussion above. Bottomside spread-F occurs in a widespread region (at least $\pm 10°$ latitude) centered on the magnetic equator in the post sunset hours. Some of the resulting low density regions rise into the topside leaving being a trail of counter rotating vortices which mix the background density gradient and create irregularities in density. Since the whole flux tubes partake in this upwelling a two dimensional bubble originating anywhere in this latitude range will have effects in the topside. The dual cascade of velocity turbulence tends to fill the E-W region between bubbles with both velocity and density irregularities. The density gradients in the bubbles themselves will eventually dissipate possibly via the "universal" drift wave instability.

Radio waves incident upon the equatorial ionosphere under these conditions are thus subject to at least three different types of scintillating environment: 1. Just after sunset during the development phase, the bottomside irregularitiies cause modest amplitude scintillation effects due to the low absolute fluctuations in electron density. It is not known whether large phase fluctuations accompany such weak amplitude scintillations as the Wideband satellite does not have equatorial crossings at this time. 2. In the fully developed stage bubbles are continually emitted into the topside over a wide range of latitudes. These bubbles continually stir the medium at wavelengths comparable to the bubble dimensions ($\sim$10km). Two dimensional dual cascade spreads the velocity and density irregularities to larger and smaller k values. The cascade to smaller k combined with the vertical wake formation tends to fill a large volume with density irregularities. Incident radio waves thus are subject to sharp discontinuities in electron

density in the bubbles as well as very thick (in altitude) regions of power law type density irregularities. Very intense amplitude scintillations have been observed during such times as discussed in Section 3 and it is expected that large phase fluctuations will also be present. 3. In the decay phase (later local time) the velocity turbulence injected by the bubbles as well as the bubble gradients themselves decay via dissipation at large k. In this phase the dissipation scale moves to smaller k values (Kelley and Ott, 1978) so the smallest scale structures disappear first. Thus significant kilometer scale structures will still exist in the medium and strongly affect scintillation even though backscatter measurements become less intense. This is probably the phase of irregularity evolution most often sampled by the Wideband satellite as discussed by Basu et al., (1977b). However these authors have pointed out that they are as yet unable to conclude in a definitive manner, whether these non-backscatter associated irregularities are drifting 'fossil' structures of irregularities generated after sunset or freshly generated structures appearing later in the evening having little or no power at 3m scalelength.

It has to be pointed out that two-dimensional models of irregularity bubbles do not explain why density depletions are observed to rise more frequently than enhancements fall (Hudson, 1977). G. Haerendel (private communication, 1977) has noted that density enhancements may be tied to the E-region, whereas density depletions are free to interchange and drift up. Alternatively, McClure et al., (1977) have suggested that density enhancements may spread out over a broader E-W extent than depletions, thereby producing a smaller relative enhancement. More extensive experiments are required during which both velocity and density irregularities are measured simultaneously as a function of k to solve the three-dimensional problem. Electric field measurements of the velocity fields are easier to perform on technical grounds and are equivalent since $\underline{V} = \underline{E} \times \underline{B}/B^2$.

From the scintillation point of view, it would be highly desirable to have a geostationary satellite with the capabilities of the present Wideband one to monitor the focus and scatter components during the developing phase of irregularities. Such measurements in conjunction with the Jicamarca radar observations would provide valuable information on spectral characterisitics of equatorial irregularities. The finding of Costa and Kelley (1978a) and Chaturvedi and Ossakow (1977) regarding steepened bottomside structures and the associated k^{-2} spectrum is an important one. It is thus quite crucial to determine whether steepened structures also dominate topside irregularity behavior. This may have far-reaching implications in scintillation modeling.

ACKNOWLEDGMENTS

Numerous people have contributed to the development of the ideas expounded in this paper and it is hoped that the reference list adequately recognizes their contribution. We also thank the various authors who made their diagrams available to us for reproduction. We wish to thank Santimay Basu for many helpful comments and a careful reading of the manuscript. The AFGL equatorial irregularity campaigns were conducted under the guidance of J. Aarons to whom Sunanda Basu is indebted for many constructive suggestions. She extends her thanks to the National Research Council for the award of an Associateship. M.C. Kelley acknowledges helpful discussions with E. Costa and D.T. Farley. The work at Cornell University was sponsored by ONR under N00014-75-C-0780.

REFERENCES

Aarons, J., Equatorial scintillations: A review, IEEE Trans. Ant. & Propagat. AP 25, 729, 1977.

Aarons, J., J. Buchau, Santimay Basu and J.P. McClure. The localized origin of equatorial F-region irregularity patches, J. Geophys. Res., in press, 1977.

Balsley, B.B. and D.T. Farley, Partial reflections: A source of weak VHF equatorial spread-F echoes, J. Geophys. Res, 80, 4735, 1975.

Basu, Sunanda and Santimay Basu, Correlated maesurements of scintillations and in -situ F-region irregularites from Ogo-6, Geophys. Res. Lett., 3, 681, 1976.

Basu, Sunanda, Santimay Basu and B.K. Khan, Model of equatorial scintillations from in -situ measurements, Radio Sci., 11, 821, 1976.

Basu, Sunanda and M.C. Kelley, Review of equatorial scintillation phenomena in the light of recent developments in the theory and measurement of equatorial irregularities, J. Atmos. Terr. Phys., 39 (9), 1977.

Basu, Sunanda, J. Aarons, J.P. McClure, C. LaHoz, A. Bushby and R.F. Woodman, Preliminary comparisons of VHF radar maps of F-region irregularities with scintillations in the equatorial region, J. Atmos. Terr. Phys., 39(9), 1977a.

Basu, Santimay and J. Aarons, Equatorial irregualrity campaigns: Part I, Correlated scintillation and radar backscatter measurements in October, 1976 Air Force Geophysics Laboratory Technical Report, under preparation, 1977.

Basu, Santimay, Sunanda Basu, J. Aarons, J.P. McClure and M.D. Cousins, On the coexistence of km and m scale irregularities in the nighttime equatorial F-region, Submitted to J. Geophys. Res., 1977b.

Basu, Santimay, H. Whitney, J. Aarons, J.P. McClure, Large and small scale properties of nighttime equatorial irregularities from scintillations and radar backscatter measurements, presented at the Symposium on "The Effect of the Ionosphere on Space and Terrestrial Systems", Washington, D.C., Jan. 24-26, 1978.

Beer, T., On the dynamics of equatorial spread-F, Aust. J. Phys., 27, 391, 1974.

Booker, H.G. and H.W. Wells, Scattering of radio waves by the F-region of the ionosphere, Terres. Magn., 43, 249, 1938.

Booker, H.G., The role of the magnetosphere in satellite and radiostar scintillation, J. Atmos, Terr. Phys., 37, 1089, 1975.

Booker, H.G. and J.A. Ferguson, A theoretical model for equatorial ionospheric spread-F echoes in the HF and VHF bands, Submitted to J. Atmos. Terr. Phys., 1977.

Booker, H.G. and D.C. Miller, A theoretical model for ionospheric scintillation caused by weak irregularities of ionization density, Submitted to Radio Science, 1977.

Briggs,B. H. and I. A. Parkin, On the variation of radio star and satellite scintillation with zenith angle, J. Atmos. Terr. Phys., 25, 339, 1963.

Buchau, J., E.J. Weber and J.P. McClure, Radio and optical diagnostics applied to an isolated equatorial scintillation event, presented at the Symposium on "The Effect of the Ionosphere on Space and Terrestrial Systems", Washington, D.C., Jan. 24-26, 1978.

Chaturvedi, P.K. and P.K. Kaw, Steady state finite amplitude Rayleigh Taylor modes in Spread-F, Geophys. Res. Lett., 2, 381, 1975.

Chaturvedi, P.K. and P.K. Kaw, An interpretation for the power spectrum of spread-F irregularities, J. Geophys. Res., 81, 3257, 1976.

Chaturvedi, P.K. and S.L. Ossakow, Nonlinear theory of the collisional Rayleigh-Taylor instability in equatorial spread-F, NRL Report 3604, Naval Research Laboratory, Washington, D.C. August, 1977.

Cole, K.D., Energetics of and a source of energy for equatorial spread-F events, J. Atmos. Terr. Phys., 36, 1099, 1974.

Costa, E. and M.C. Kelley, Calculations of equatorial scintillations at VHF and gigahertz frequencies based on a new model of the disturbed equatorial ionosphere, Geophys. Res. Lett., 3, 677, 1976.

Costa, E. and M.C. Kelley, Ionospheric scintillations based on in situ irregularity spectra, Radio Sci., 12, 797, 1977.

Costa, E. and M.C. Kelley, Evidence for and development of a 2-step theory for equatorial spread-F, submitted to J. Geophys. Res., 1978a.

Costa, E. and M.C. Kelley, Linear theory for collisionless drift mode waves with wavelengths near the ion gyro-radius, Submitted to J. Geophys. Res., 1978b.

Dungey, J.W., Convective diffusion in the equatorial F-region, J. Atmos. Terr. Phys., 9, 304, 1956.

Dyson, P.L., J.P. McClure and W.B. Hanson, In-situ measurements of the spectral characteristics of ionospheric irregularities, J. Geophys. Res., 79, 1497, 1974.

Fremouw, E.J., C.L. Rino and R.C. Livingston, "A two-component model for scintillation," in Proc. Symp. Cospar Satellite Beacon Group, June 1-4, 1976, Boston University, Boston, MA., 1976.

Fremouw, E.J., Examination of data from the Wideband Satellite experiment, Report PD-N-77-118-R01, Physical Dynamics, Inc., Seattle, Washington 98109, August 1, 1977.

Fremouw, E.J. and J.M. Lansinger, Examination of data from the Wideband Satellite experiment, Report PD-NW-118-R02, Physical Dynamics, Inc., Seattle, Washington 98109, October 15, 1977.

Fremouw, E.J., B.C. Fair, R.T. Tsunoda, R.A. Long and A.A. Burns, Design considerations for a microwave scintillation experiment, Report DNA 4092F, Stanford Research Institue, Menlo Park, CA 94025, March 15, 1977.

Fremouw, E.J., R.L. Leadabrand, R.C. Livingston, M.D. Cousins, C.L. Rino, B.C. Fair and R.A. Long, Early results from the DNA Wideband satellite experiment-complex signal scintillation, Radio Sci., 13 (1), 1978.

Haerendel, G., Theory of equatorial spread-F, preprint, Max-Planck-Institut fur Physik und Astrophysik, Garching, West Germany, 1974.

Hanson, W.B. and S. Sanatani, Large Ni gradients below the equatorial F peak, J. Geophys. Res., 78, 1167, 1973.

Hudson, M.K. and C.F. Kennel, Linear theory of equatorial spread-F, J. Geophys. Res., 80, 4581, 1975.

Hudson, M.K., Spread-F bubbles: Nonlinear Rayleigh-Taylor mode in two dimensions, preprint, University of California, Berkeley, CA 94720, January, 1977.

Kao, S.K. and L.L. Wendell, The kinetic energy of the large scale atmosphere motion in wavenumber frequency space, J. Atmos. Sci., 27, 359, 1970.

Kelley, M.C., G. Haerendel, H. Kappler, A. Valenzuela, B.B. Balsley, D.A. Carter, W.L. Ecklund, C.W. Carlson, B. Hausler and R. Torbert, Evidence for a Rayleigh-Taylor type instability and upwelling of depleted density regions during equatorial spread-F, Geophys. Res. Lett., 3, 448, 1976.

Kelley, M.C. and P.M. Kintner, Evidence for two-dimensional inertial turbulence in a cosmic scale low β plasma, Astrophys. J., in press, 1978.

Kelley, M.C. and E. Ott, On the role of two-dimensional turbulence in equatorial spread-F, submitted to J. Geophys. Res., 1978.

Koster, J.R., Study of the equatorial ionosphere, Report AFGL-TR77-0165, University of Ghana, Legon, Ghana, November, 1976.

Kraichnan, R.H., Inertial ranges in two-dimensional turbulence, Phys. Fluids, 10, 1417, 1967.

Lilly, D.K., Numerical simulation of two-dimensional turbulence, Phys. Fluids, Supplement II, II-240, 1969.

Martyn, D.F., Large-scale movements of ionization in the ionosphere, J. Geophys. Res., 64, 2178, 1959.

McClure, J.P. and W.B. Hanson, A catalog of ionospheric F-region irregularity behavior based on Ogo-6 retarding potential analyzer data, J. Geophys. Res., 78, 7431, 1973.

McClure, J.P., W.B. Hanson and J.H. Hoffman, Plasma bubbles and irregularities in the equatorial ionosphere, J. Geophys. Res., 82, 2650, 1977.

Morse, F.A., B.C. Edgar, H.C. Koons, C.J. Rice, W.J. Heikkila, J.H. Hoffman, B.A. Tinsley, J.D. Winningham, A.B. Christensen, R.F. Woodman, J. Pomalaza and N.R. Teixeira, Equion: An ionospheric irregularity experiment, J. Geophys. Res., 82, 578, 1977.

Mullen, J.P., H.E. Whitney, Santimay Basu, A. Bushby, J. Lanat and J. Pantoja, Statistics of VHF and L band scintillation at Huancayo, Peru, J. Atmos. Terr. Phys., 39(9), 1977.

Ott, E., Theory of Rayleigh-Taylor bubbles in the equatorial ionosphere, submitted to J. Geophys. Res., 1977.

Rino, C.L., E.J. Fremouw, R.C. Livingston, M.D. Cousins and B.C. Fair, Wideband satellite observations, Report DNA4399F, SRI International, Menlo Park, Calif. 94025, June, 1977.

Röttger, J., The macroscale structure of equatorial irregularities, J. Atmos. Terr. Phys., 38, 97, 1976.

Rufenach, C.L., Wavelength dependence of radio scintillation: Ionosphere and interplanetary irregularities, J. Geophys. Res., 79, 1562, 1974.

Rufenach, C.L., Ionospheric scintillation by a random phase screen: Spectral approach, Radio Sci., 10, 155, 1975.

Scannapieco, A.J. and S.L. Ossakow, Nonlinear equatorial spread-F, Geophys. Res. Lett., 3, 451, 1976.

Weber, E.J., J. Buchau, R.H. Eather and S.B. Mende, North/south aligned equatorial airglow depletions, J. Geophys. Res. (in press), 1977.

Whitney, H.E. and Santimay Basu, The effect of ionospheric scintillation on VHF/UHF satellite communications, Radio Sci., 12, 123, 1977.

Whitney, H.E., J.P. Mullen, J. Buchau, E.J. Weber and A. Johnson, Report on Peru scintillation tests, October 1976 and March, 1977, AFGL Technical Report, under preparation, 1977.

Whitney, H.E., Spaced receiver measurements of intense equatorial scintillations, presented at the Symposium on "The Effect of the Ionosphere on Space and Terrestrial Systems", Washington, D.C., Jan. 24-26, 1978.

Woodman, R.F. and C. LaHoz, Radar observations of F-region equatorial irregularities, J. Geophys. Res., 81, 5447, 1976.

Woodman, R.F. and Sunanda Basu, Comparison between in-situ spectral measurements of equatorial F-region irregularities and backscatter observations at 3m wavelength (abstract), EOS Trans. AGU, 58, 49, 1977, submitted to Geophys. Res. Lett., 1977.

Yeh, K.C. and C.H. Liu, Pulse delay and pulse distortion by random scattering in the ionosphere, presented at AGARD Meeting, Cambridge, MA, Oct. 3-7, 1977.

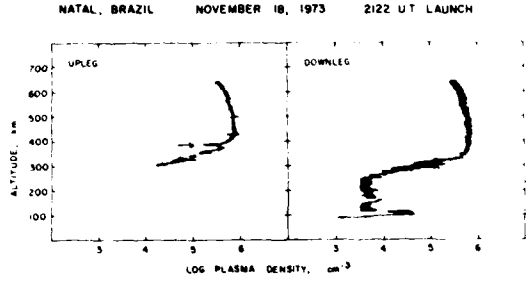


Fig. 1. Upleg and downleg plasma density profiles obtained on a sounding rocket during equatorial spread-F. The arrow shows the location of a deep depression in plasma density (after Kelley et al., 1976).

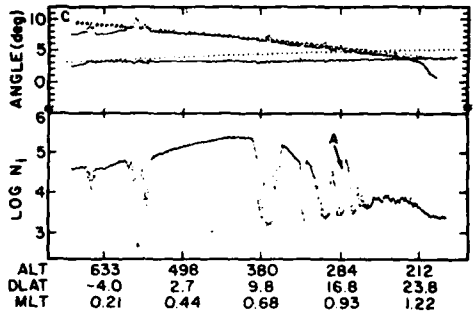


Fig. 3. Ion drift meter data from AE-C orbit 2282. The observed pitch and yaw angles are shown in the upper and lower curves of the top panel and the total density on the bottom. Positive angles correspond to ions moving up or left with respect to spacecraft (after McClure et al., 1977).

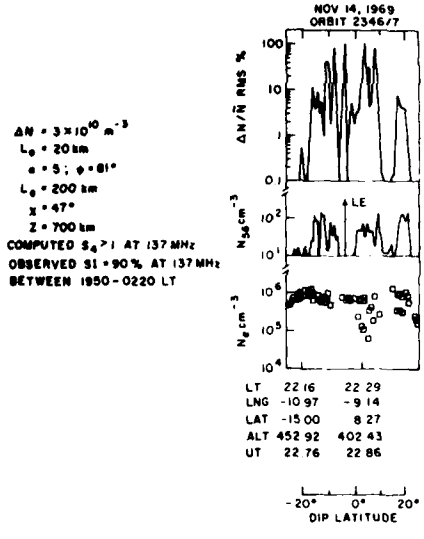


Fig. 2. Ogo-6 orbit in Legon, Ghana longitude sector during a strong VHF scintillation event. The scintillation model parameters are also indicated (after Basu and Basu, 1976).

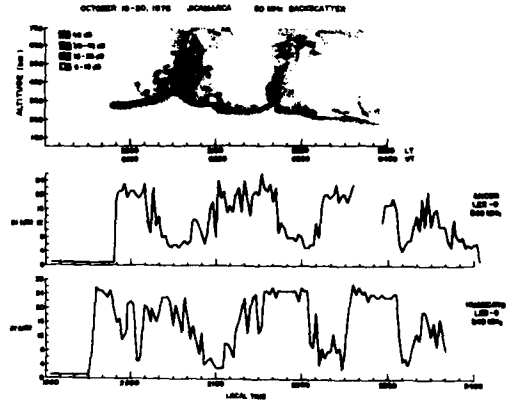


Fig. 4. Temporal variation of the 50 MHz radar backscatter observed at Jicamarca and scintillation index, SI(dB), of the 249 MHz transmissions from LES-9 satellite recorded at Ancon and Huancayo on October 19-20, 1976 (after Basu and Aarons, 1977).

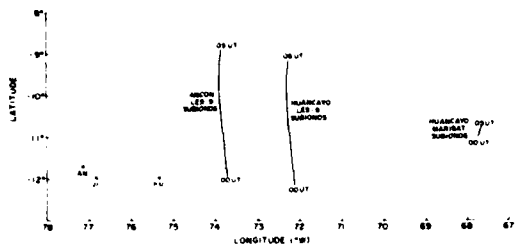


Fig. 5. The geometry of radar backscatter and scintillation measurements during AFGL equatorial campaign in October, 1976. The ground locations for various observatories and subionospheric (400 km) locations for LES-9 are shown (after Basu and Aarons, 1977).

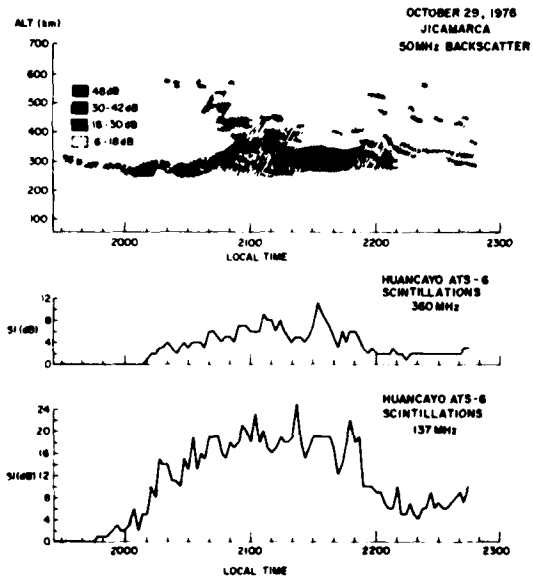


Fig. 6. Comparison of backscattered power map from Jicamarca and ATS-6 scintillation data corresponding to a nearly common ionospheric volume on October 29-30, 1976 (after Woodman and Basu, 1977).

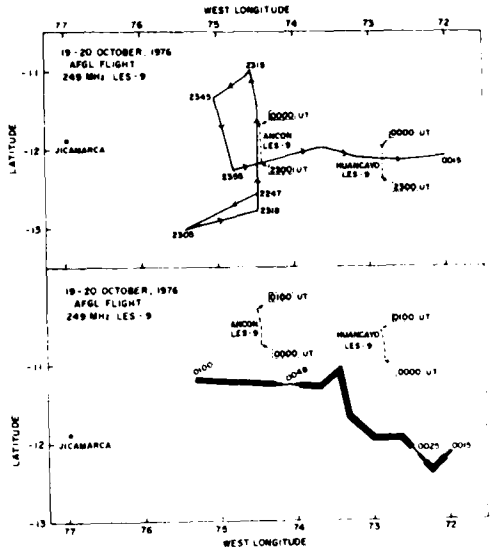


Fig. 7. Subionospheric flight path of the AFGL aircraft making 249 MHz scintillation observations with LES-9 on October 19-20, 1976. Thin and thick lines indicate respectively the absence or presence of scintillations (after Basu and Aarons, 1977).

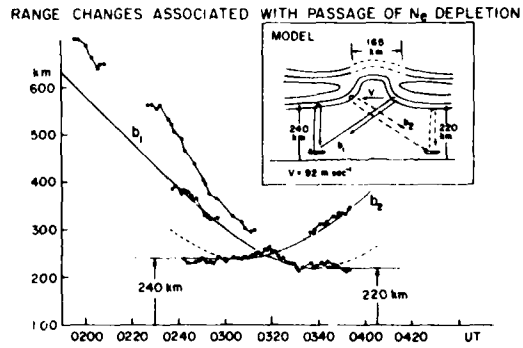


Fig. 8. Model of an eastward drifting, bottomside N_e depletion based on observed ionosonde and airglow parameters. The measured ranges of oblique returns and the virtual height of overhead F-region are compared with range/height changes expected from the passage of the model bottomside structure over the ionosonde (after Weber et al., 1977).

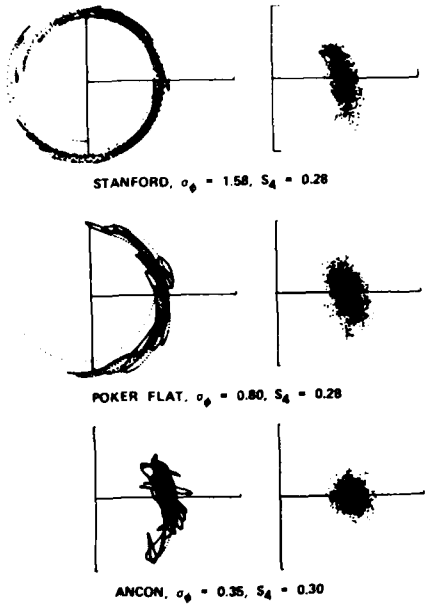


Fig. 9. Scatter diagrams for the focus (left) and scatter (right) components isolated from segments of VHF records obtained at a midlatitude (top), auroral (center) and equatorial (bottom) station (after Fremouw et al., 1978).

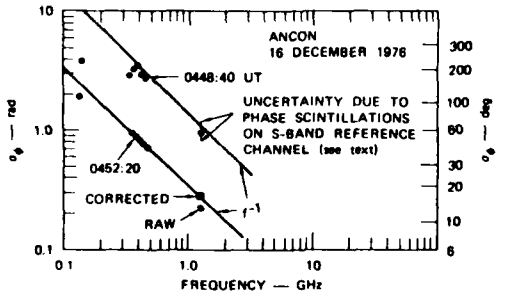


Fig. 11. Frequency dependence of phase-scintillation index during two 20-second periods of pass shown in Figure 10, compared with an f^{-1} dependence arbitrarily passed through the 413 MHz data point (after Fremouw et al., 1978).

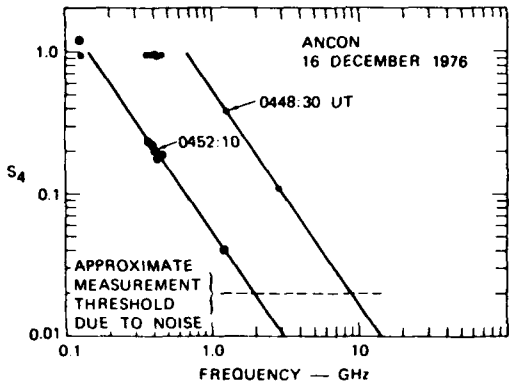


Fig. 12. Frequency dependence of intensity-scintillation index during two 20-second periods of the pass shown in Figure 10, compared with an $f^{-1.5}$ dependence (after Fremouw et al., 1978).

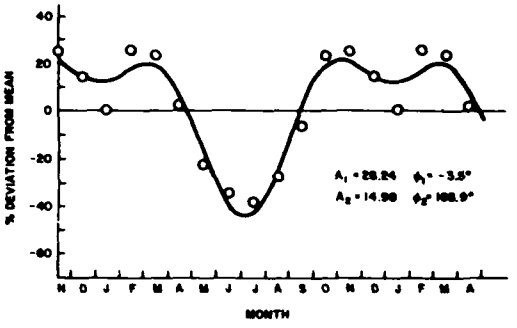


Fig. 13. Seasonal variation of 136 MHz scintillation from ATS-3 at Legon, Ghana during the period Sept. 71 to Dec. 75. Note that the annual component (with a minimum in the June solstice) has twice the amplitude of the semi-annual component (after Koster, 1976).

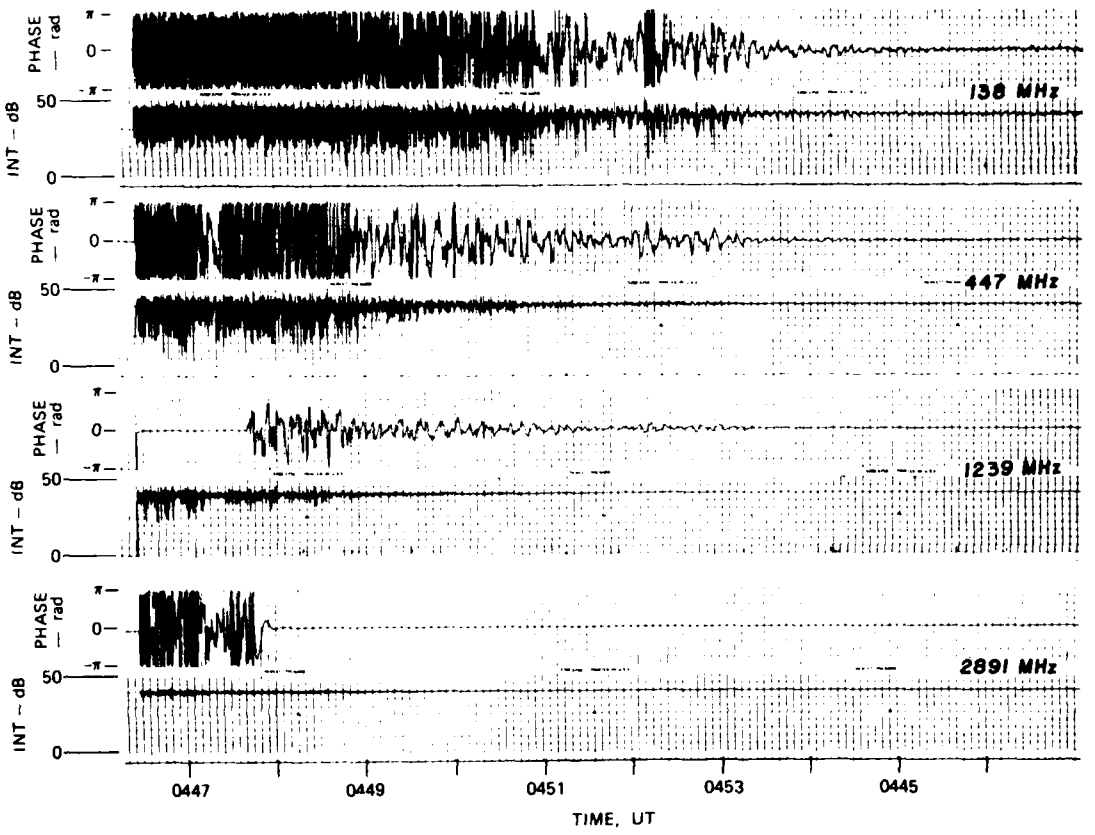


Fig. 10. Example of an equatorial Wideband pass at Ancon on December 16, 1976 showing phase and intensity scintillations at VHF, UHF, L band and S band. Note change of phase reference from L band to S band just before 0443 UT (after Fremouw et al., 1977).

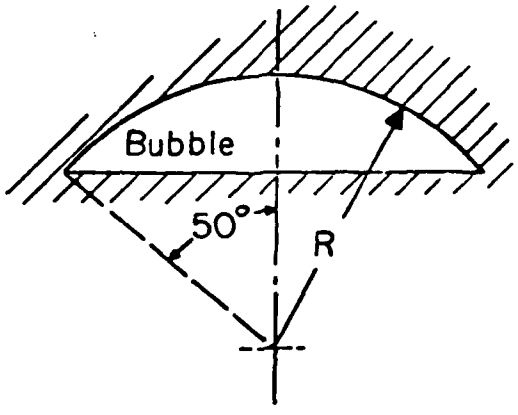


Fig. 14. Shape of bubbles in fluids (after Ott, 1977).

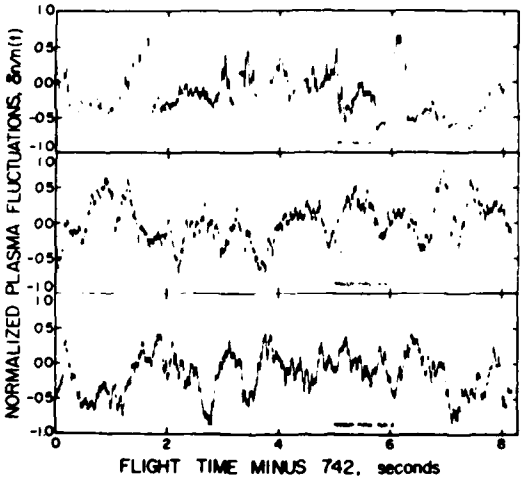


Fig. 16. The upper plot is a detrended data stream obtained on the bottomside of the F-peak during equatorial spread-F. The lower two plots are reconstructed time domain samples using the same data, but with the addition of a random phase in the FFT before reconstruction (after Costa and Kelley, 1978a).

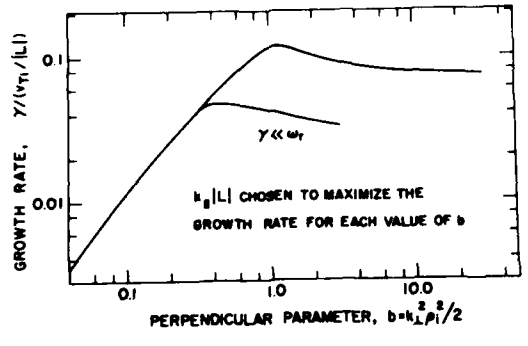


Fig. 15. Growth rate γ as function of the parameter b. The curve labeled $\gamma << \omega_r$ is obtained from the 'small growth rate' expression while the unlabeled curve is obtained from the 'generalized' expressions (after Costa and Kelley, 1978b).

LARGE AND SMALL SCALE PROPERTIES OF NIGHTTIME EQUATORIAL IRREGULARITIES FROM SCINTILLATIONS AND RADAR BACKSCATTER MEASUREMENTS

Santimay Basu
Emmanuel College
Boston, MA 02115

H. Whitney and J. Aarons
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

J.P. McClure
University of Texas at Dallas
Richardson, TX 74080

INTRODUCTION

The nighttime F-region irregularities of electron density at equatorial latitudes have been studied for several decades by observing the spread-F signature in ionograms and scintillations in transionospheric communication links (Booker and Wells, 1938; Aarons et al., 1971; Skinner and Kelleher, 1971). In recent years, these studies have been supplemented by a variety of new and powerful experimental techniques, such as, in-situ measurements by rockets and satellites, HF forward scatter and VHF radar backscatter measurements (Farley et al., 1970; Dyson et al., 1974; Kelley et al., 1976; Röttger, 1976; Woodman and LaHoz, 1976; McClure et al., 1977). These experimental investigations have been supported by computer simulations and analytic work on the formation of equatorial irregularities (Scannapieco and Ossakow, 1976; Haerendel, 1974; Hudson and Kennel, 1975).

Basu and Kelley (1977) have reviewed the recent advances in our knowledge of equatorial irregularities but point out that some of the basic questions remain unresolved. For example, it has not been established if the generation of nighttime irregularities is related to the passage of the terminator (or sunset line) or to the existence of suitable conditions in certain spatial locations. The nighttime equatorial irregularities also exhibit a marked day-to-day variability; a night indicating the presence of strong irregularities may be followed by another without any irregularities within the detectable limit. The question of co-existence of large and small scale irregularities also remains unexplored. Nevertheless, it has important implications in regard to the irregularity power spectrum and hence of importance in communications channel modelling and the development of theories of irregularity generation.

In order to explore some of these unresolved questions, we performed simultaneous VHF radar backscatter and scintillation measurements near the magnetic equator during October, 1976 and March, 1977 (Basu and Aarons, 1977). The 50 MHz radar backscatter measurements were performed at Jicamarca, Peru (11.97°S, 76.86°W) and scintillation measurements were made at two nearby ground stations Ancon (11.7°S, 77.15°W) and Huancayo (11.97°S, 75.34°W). The VHF-UHF scintillations are sensitive to irregularities within the scale length range of about one kilometer to a hundred meter whereas the 50 MHz radar backscatter is caused by 3 meter irregularities. Thus simultaneous radar and scintillation measurements allowed a study of irregularity development over a three decade range of scale lengths. The scintillation measurements made at the two stations with various geostationary satellites made it possible to track the irregularity patches and thereby determine their drift and lifetime.

RESULTS AND DISCUSSIONS

Figure 1 shows the results of simultaneous 50 MHz radar backscatter observations performed at Jicamarca and scintillation measurements made at Huancayo with the 137 and 360 MHz transmissions from ATS-6 satellite on October 29. 1976. During the October period, the ATS-6 satellite was being moved to its new location and on the night of October 29, 1976, the 400 km intersection point from the satellite to Huancayo was only 30 km to the east of the ionospheric volume illuminated by the radar beam. The top panel of Figure 1 exhibits the backscattered power map. This map has been redrawn from the original digital power map which is difficult to reproduce. The radar map shows the temporal variation of the range of backscattering region and the strength of the echo power in decibels above the approximate maximum incoherent scatter level. A description of the digital power mapping technique has been outlined in Woodman and LaHoz (1976). The second and third panels represent the temporal variation of scintillation index, SI(dB), as recorded at 360 and 137 MHz respectively. It may be noted that the radar backscatter and scintillation measurements are very well correlated. The onset of scintillations at 20 LT is coincident with the radar detecting moderately strong (18-30 dB) irregularities in a thin layer. Later, the intensity of 137 MHz scintillations attained saturated levels when the 3 m irregularity layer thickened. The measurements indicate that, in the early evening period, the 3 meter irregularities detected by the radar co-existed with large scale irregularities (1 km - 100 m) causing scintillations over a nearly common ionospheric volume. Woodman and Basu (1977) have made a quantitative comparison of the backscattered power level and 360 MHz scintillation level indicated in Figure 1 and found that a monotonic power law form of the 3-dimensional irregularity power spectrum with spectral index 4 cannot explain the backscatter observations. The in-situ irregularity data suggested such a spectra between scale lengths of several kilometers to tens of meters. These authors have indicated that a sharper roll-off of the irregularity power spectrum at irregularity scale lengths of several meters due to finite ion gyro-radius can explain the observations.

Figure 2 shows the radar and scintillation measurements performed on October 16-17, 1976. The top panel shows the 50 MHz radar backscattered power map. The second panel shows the temporal variation of scintillation recorded at Ancon by receiving the 249 MHz transmissions from LES-9 satellite. The 400-km intersection of Ancon scintillation observations was located at a distance of 350 km to the east of Jicamarca radar observations. The third panel illustrates the results of 249 MHz scintillation measurements made at Huancayo. The 400 km intersection of Huancayo measurements was located 175 km to the east of the ionospheric location probed by the Ancon measurements. In contrast to the measurements presented in Figure 1, the different panels of Figure 2 represent observations at different locations. From Figure 2, it may be observed that the Jicamarca radar exploring the westernmost location first detected the 3 meter irregularities as early as 19 LT but an extended irregularity structure termed as 'plume' evolved later at 1950 LT. The panel 2 of the figure shows that the onset of scintillations at Ancon occurred at about 2050 LT. Considering that the plume structure drifted towards east and traversed 350 km distance to cause the onset of Ancon scintillation, an eastward drift speed of 97 m/sec is obtained. From panel 3, it is found that the sharp onset of strong scintillations at Huancayo was delayed from that at Ancon by about 25 minutes. This delay is again consistent with a drift speed of about 116 m/sec. Thus the results illustrated in Figure 2 shows that an irregularity structure evolved first in the west possibly due to the existence of suitable background conditions and the structure drifted eastwards at a speed of approximately 100 m/sec to cause delayed onsets of scintillations on two propagation paths (Aarons et al., 1977). Considering the onset time of plume structure at Jicamarca and decay of scintillations at Huancayo a total lifetime of at least 3 hours for the irregularity structure is obtained.

The results of October 16, 1976 and October 29, 1976, indicated that large scale irregularities (~1 km - .1 km) co-existed with 3 m irregularities. The above fact did not, however, hold in all cases. It was found that scintillation measurements detected larger number of irregularity patches than did the radar. In Figure 3, we illustrate the radar maps, in schematic form, and results of scintillation measurements performed at Ancon and Huancayo, on seven nights in

October, 1976. It may be noted that the irregularity patches detected in scintillation experiments were more preponderant than 3 m plume structures. Based on the October measurements, a definite conclusion regarding the above fact could not be made because, except for October 29, 1976, the radar and scintillation experiments were probing different ionospheric volumes with a minimum separation of 350 km. The problem of co-existence of large and small scale irregularities can only be investigated if the measurements refer to a common volume.

During March, 1977, scintillation measurements with a host of geostationary satellites were so organized that ionospheric locations to the west, the east and in close proximity to the volume illuminated by the radar backscatter observations could be probed. As in October, the ground stations at Ancon and Huancayo performed scintillation measurements. The geostationary satellites LES-8 (abbreviated as L-8), ATS-3 (A-3), GOES-1 (G), LES-9 (L-9) and MARISAT were used. The 400 km subionospheric positions of all these measurements are shown in Figure 4, each location being specified by the station name (A for Ancon, H for Huancayo) followed by the abbreviated name of the satellite. Owing to finite orbital inclination, some of these vary considerably with time. The ground stations have been indicated by dotted circles. It may be noted that scintillation measurements performed at Ancon with the GOES-1 satellite probed an ionospheric location which was only 20 km to the east of the magnetic field line passing through the radar illuminated volume.

We shall illustrate only one night's observation (March 20, 1977) to indicate the nature of co-existence of small and large scale irregularities. Figure 5 shows the results of all scintillation measurements on this night, the panels from top to bottom being arranged in the order of increasing east longitude. The exact subionospheric locations corresponding to these different panels can be obtained from Figure 4. It is found that the onset of scintillations on this night occurred between 00 UT - 01 UT. When referred to the local time of these various subionospheric locations, the onset seems to have occurred within 4 minutes of 19h 40m LT. Thus unlike the results of October 16-17, 1976, when the irregularities evolved first in the west and were localized spatially, the irregularities seem to be under temporal control on March 20, 1977. It is to be noted that panels 2-5 of Figure 5 exhibit two discrete irregularity structures to which we shall have occasion to discuss again. Figures 6 and 7 represent the results of simultaneous radar observations at Jicamarca and scintillation measurements at Ancon with the GOES-1 satellite which, as mentioned earlier, refer to approximately a common ionospheric volume. From Figure 6 we find that during the early evening hours, the onset of scintillations occurred precisely when a plume evolved in the radar map. However, later (after 2045 LT) when the radar backscatter became weak, the scintillations maintained a high level. This indicates that the relative spectral powers of large scale and meter scale irregularities varied considerably during a two hour period (20 LT - 22 LT). In Figure 7 we show that the scintillation experiment detected another irregularity structure on this night between 23 LT - 01 LT (bottom panel) but virtually no radar backscatter was detected during this time (top panel). Referring to Figure 5, it may be noted that this second irregularity structure having negligible spectral power at 3 m was detected in panels 2 through 5, corresponding to the longitude interval of 78°W - 74°W. The reason for the absence of any signature of this event in panel 1 (79°W position) may be due to the evolution of the irregularity to the east of 79°W location. The significant fact regarding the second irregularity structure evolving an hour before midnight is a cut-off in the irregularity power spectrum at meter wavelengths. By performing power spectral analysis of amplitude scintillations recorded in the post sunset and near midnight period, we find no difference in the spectral form over the frequency interval of 0.2 - 2 Hz. The spaced receiver scintillation measurements made on this night at Ancon with LES-9 indicated that the eastward drift speed was about 75 m/sec at 1900 LT which increased to about 125 m/sec around 2300 LT. Considering an approximate drift speed of 100 m/sec, the frequency interval of 0.2 - 2 Hz may be translated to irregularity scale lengths of 2 km - 200 m. Thus the irregularity power spectrum between the scale lengths 2 km - 200 m remained unaltered between the post sunset and near midnight periods although their associations with 3 m irregularities was markedly different.

CONCLUSIONS

We have shown that on certain nights the equatorial irregularities are generated in particular spatial locations. On other nights, the irregularity generation is predominantly controlled by local time. The measurements of October, 1976 and March, 1977 indicate that between 1900 - 0100 LT the irregularity patches drift eastwards at speeds ranging from 85 - 140 m/sec. By tracking the irregularity patches with multisatellite scintillation measurements, it is found that the patches have typical lifetimes of several hours. Thus the temporal variation of scintillations recorded at an equatorial station, is dictated by spatial or temporal localization of irregularity generation, subsequent drift and lifetime of irregularity patches.

The equatorial irregularities that evolve in the post-sunset period are found to have a wide scale length range between several kilometers to a few meters. Such a scale length range can give rise to both scintillations and radar backscatter. The irregularity structures that arise around midnight are found to have little spectral powers at meter scale lengths. Such irregularity structures give rise to scintillation but no radar backscatter. For both these types, the irregularity power spectrum between a few kilometers to several hundred meters remain unaltered. Thus the major difference between these two types of irregularity power spectrum lies at scale lengths of tens of meters.

REFERENCES

Aarons, J., H.E. Whitney, R.S. Allen, Global morphology of ionospheric scintillation, Proc. I.E.E.E., 59, 159-172, 1971.

Aarons, J., J. Buchau, Santimay Basu, and J.P. McClure, The localized origin of equatorial F-region irregularity patches, to appear in J. Geophys. Res., 1977.

Basu, Sunanda, and M.C. Kelley, Review of equatorial scintillation phenomena in the light of recent developments in the theory and measurement of equatorial irregularities, to appear in J. Atmos. Terr. Phys., 1977.

Basu, Santimay and J. Aarons, Equatorial irregularity campaigns: Part I. Correlated scintillation and radar backscatter measurements in October, 1976, Air Force Geophysics Laboratory Scientific Report, 1977 (in press).

Booker, H.G., and H.W. Wells, Scattering of radio waves by the F-region of the ionosphere, J. Terres. Magn., 43, 249-256, 1938.

Dyson, P.L., J.P. McClure, W.B. Hanson, In-situ measurements of the spectral characteristics of F-region ionospheric irregularities, J. Geophys. Res., 79, 1495-1502, 1974.

Farley, D.T., B.B. Balsley, R.F. Woodman, and J.P. McClure, Equatorial spread-F: Implications of VHF radar observations, J. Geophys. Res., 75, 7199-7216. 1970.

Haerendel, G., Theory of equatorial spread-F, Preprint, Max-Planck-Institut für Physik und Astrophysik, Garching, West Germany, 1974.

Hanson, W.B. and S. Sanatani, Large N_i gradients below the equatorial F-peak, J. Geophys. Res., 78, 1167-1173, 1973.

Hudson, M.K., and C.F. Kennel, Linear theory of equatorial spread-F, J. Geophys. Res., 80, 4581-4590, 1975.

McClure, J.P., W.B. Hanson, and J.H. Hoffman, Plasma bubbles and irregularities in the equatorial ionosphere, J. Geophys. Res., 82, 2650-2656, 1977.

Röttger, J., The macroscale structure of equatorial irregularities, J. Atmos. Terr. Phys., 38, 97-101, 1976.

Skinner, N.J., and R.F. Kelleher, Studies of F-region irregularities at Nairobi, I - From spread-F on ionograms 1964-1970, Ann. Geophys. 27, 181-194, 1971.

Scannapieco, A.J., and S.L. Ossakow, Nonlinear equatorial spread-F, Geophys. Res. Lett., 3, 451-454, 1976.

Woodman, R.F., and C. LaHoz, Radar observations of F-region equatorial irregularities, J. Geophys. Res., 81, 5447-5466, 1976.

Acknowledgement. The work at the University of Texas at Dallas was partially supported by NSF grant ATM 76-17351.

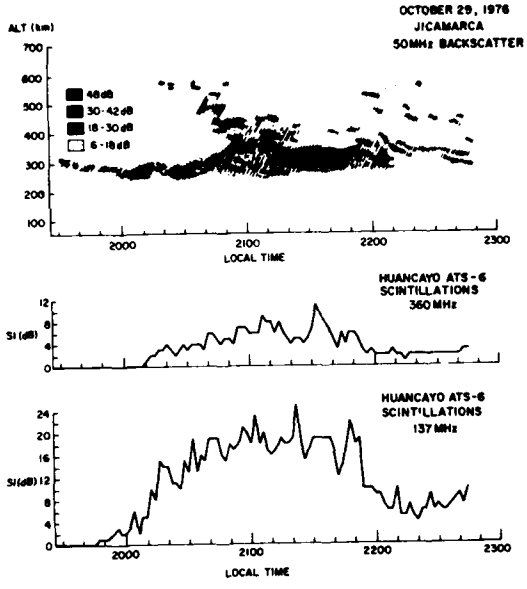


Figure 1. Radar (50 MHz) backscattered power maps of F-region irregularities obtained at Jicamarca and results of simultaneous scintillation measurements performed at Huancayo on October 29, 1976, with 137 and 360 MHz transmissions from the ATS-6 satellite. The east-west separation between the ionospheric locations explored by the radar and scintillation experiments was only 30 km at an ionospheric height of 400 km.

Woodman, R.F., and Sunanda Basu, Comparison between in-situ spectral measurements of equatorial F-region irregularities and backscatter observations at 3 m wavelengths (abstract), EOS Trans. AGU, 58, 449, 1977.

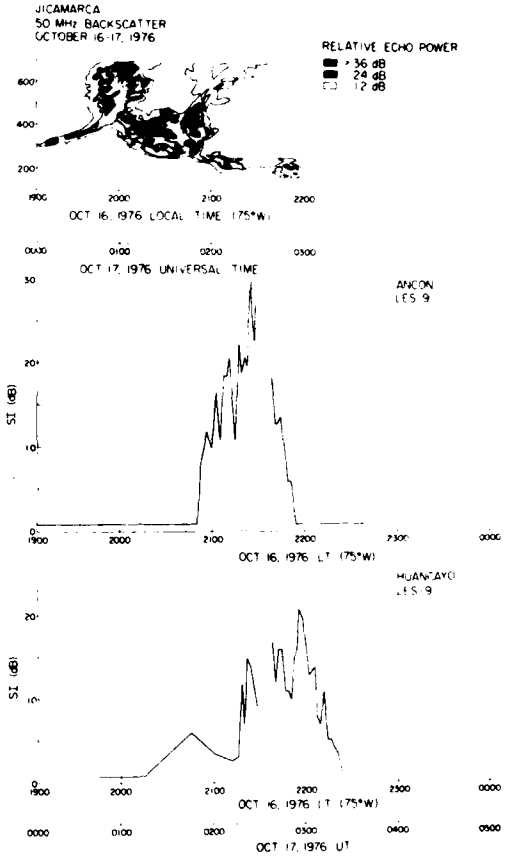


Figure 2. Results of simultaneous radar backscatter and 249 MHz scintillation observations performed on October 16-17, 1976. The ionospheric volumes pertaining to the results in the top and middle panels were separated by 350 km whereas those pertaining to the middle and the lower panels by 175 km.

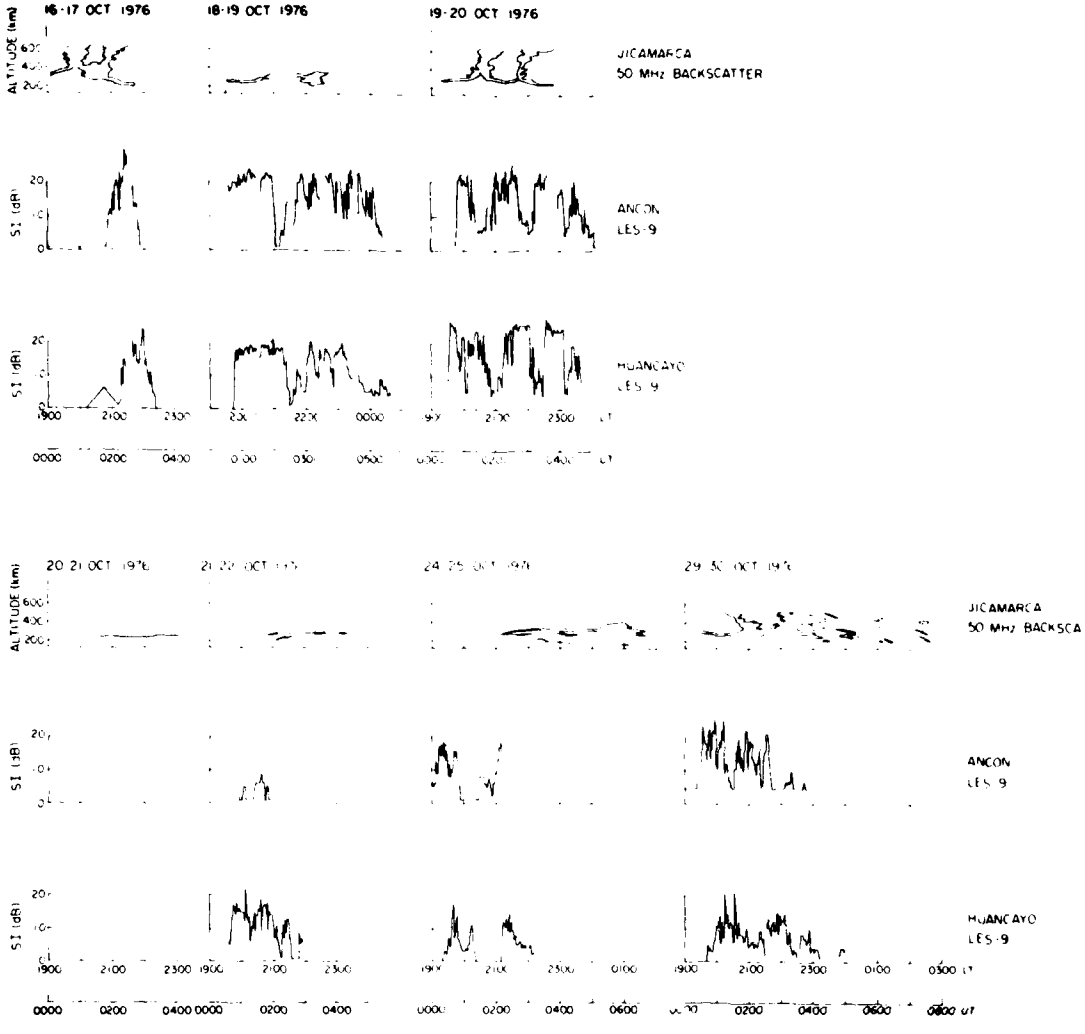


Figure 3. Comparison of radar power maps, in schematic form, and results of simultaneous scintillation measurements made at Ancon and Huancayo on seven nights in October, 1976.

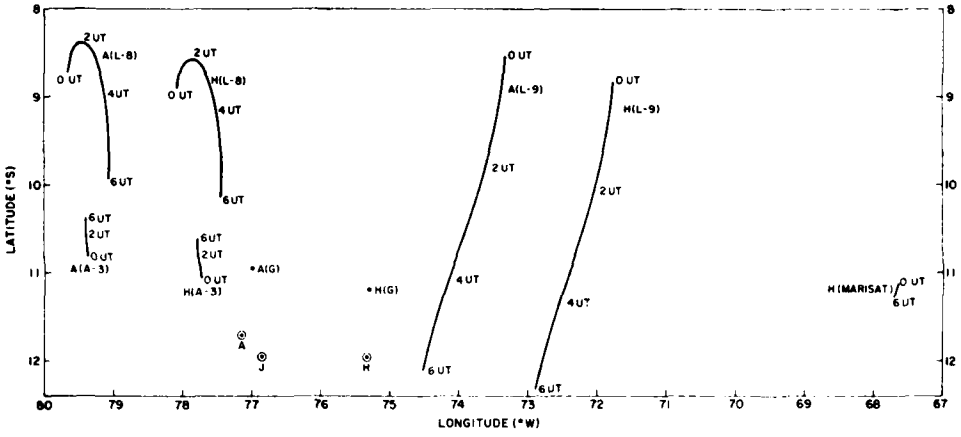


Figure 4. The subionospheric (400 km) locations of scintillation measurements made during March, 1977, each location being specified by the station name ('A' for Ancon and 'H' for Huancayo) followed by the abbreviated name of the geostationary satellite. For satellites with finite orbital inclination, the universal time (UT) is indicated alongside the variable subionospheric position.

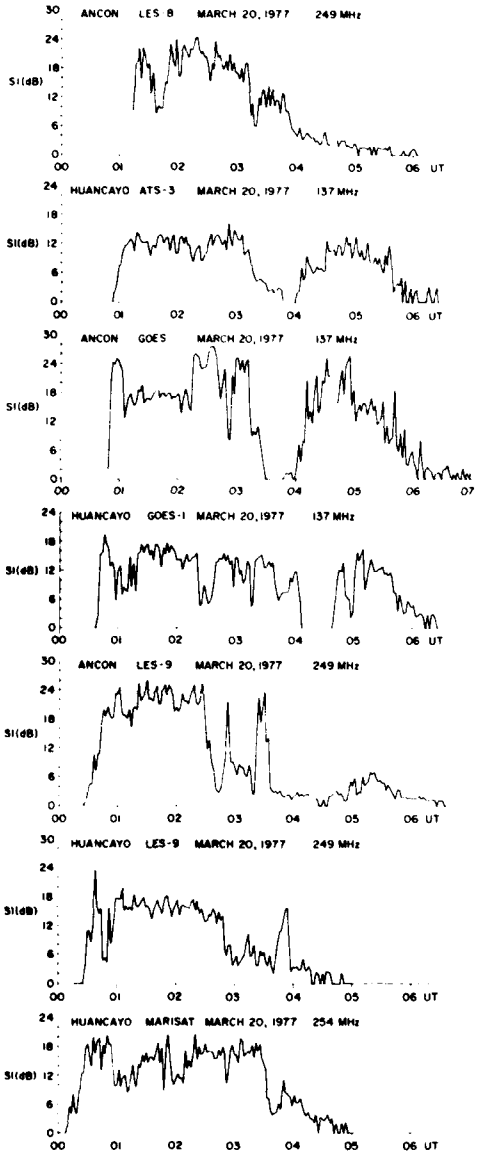


Figure 5. The results of simultaneous scintillation measurements performed on March 20, 1977. The subionospheric (400 km) location corresponding to each panel may be obtained from Figure 4.

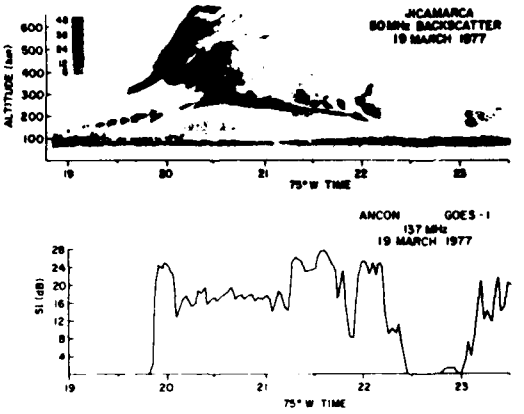


Figure 6. Comparison of radar power map and simultaneous 137 MHz scintillation measurements over a nearly common ionospheric volume during the development phase of irregularity generation on March 20, 1977.

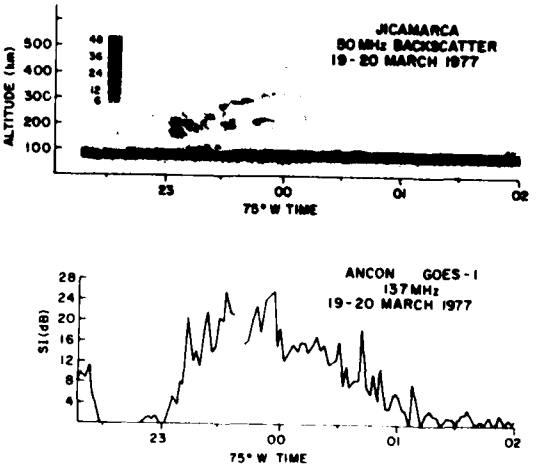


Figure 7. Comparison of radar power map and simultaneous 137 MHz scintillation measurements obtained on March 20, 1977, in the late phase of irregularity generation

Radio and Optical Diagnostics Applied to an Isolated Equatorial Scintillation Event

By

J. Buchau and E.J. Weber
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

And

J.P. McClure
The University of Texas at Dallas
Richardson, TX 74080

ABSTRACT

During equatorial scintillation studies carried out in March 1977 a new all-sky spectrophotometric imaging system and an ionospheric sounder, both onboard a USAF research aircraft and the Jicamarca Backscatter Radar were operated to monitor the development, structure and motion of F region irregularities leading to scintillations on satellite signals. The imaging system provided all-sky pictures of the 6300 Å OI airglow emission, which results from dissociative recombination of O_2^+ in the F region. This technique provided minute by minute images of the large scale structure of the bottomside F-layer. Initial observations show the existence of north-south aligned regions of airglow depletion. These bands often extend more than 1200 km in the north-south direction and from a few km to several hundred km in the east-west direction. They drift eastward at ~100 m sec^{-1}. Their presence is detected by the ionospheric sounder as oblique echoes or spread F and they correlate with scintillation events.

An isolated airglow depletion band was observed during a flight on 17 March 1977 and is described here in detail. The depletion drifted east at 92 m sec^{-1} and was observed for three hours during which time it moved from the western to the eastern observational horizon. The ionospheric sounder showed strong electron density gradients associated with the eastern and western edges of the airglow depletion. Substantial raising of the bottom height of the F-layer and strong spread F characteristics were associated with the passage of the depletion over the aircraft. The Jicamarca Backscatter radar observed a patch of irregularities in the 175 to 700 km height range, drifting eastward over the radar, in good correlation with the passage of the airglow depletion.

250 MHz signals from the LES-9 satellite exhibit 6-14 dB scintillations during passage of the disturbance through the ray path. Geometrical considerations suggest that irregularities between 275 and 800 km, collocated with the airglow depletion, are responsible for the observed scintillations.

A troughlike bottomside electron density depletion, extending for more than 1200 km in N/S direction with an E/W width varying between 100 and 200 km and with a base height of the F-layer around 300 km, collocated and moving with the 6300 Å depletion, explains both the observed 6300 Å airglow structure and the sounder observations. The backscatter and scintillation measurements indicate the simultaneous presence of F-region irregularities of scale sizes ranging from 3 m to several km. We suggest that the at times rapidly rising and at times stationary strong electron density biteouts (plasma bubbles) observed by satellites in the equatorial ionosphere are the topside signature of disturbances such as the one we describe.

INTRODUCTION

An intensive experimental campaign was conducted in March 1977 in Peru to investigate the development, dynamic behavior and lifetime of equatorial nighttime F-layer irregularities and to determine their relation to scintillations observed on trans-ionospherically propagated rf-signals. Participating in this campaign were two USAF research aircraft conducting communication and environmental studies, Air Force Geophysics Laboratory (AFGL) ground stations for satellite reception and the large VHF radar at Jicamarca.

Out of a total of seven flights, we have selected one, flown on 17 March 1977, during

which an intense isolated scintillation event was observed on a 250 MHz signal transmitted from the LES 9 satellite. This event permitted the detailed correlation of various airborne and ground based geophysical and propagation measurements for a unique description of the F-layer irregularities responsible for the observed scintillations. From these investigations we obtained both a three-dimensional model of the irregularity region and a description of its temporal behavior. Our results are representative of other strong equatorial scintillation events.

EXPERIMENTS

Figure 1 shows the flight track flown on 17 March 1977. The objective was to have the ray path from the LES 9 satellite to the AFGL Airborne Ionospheric Observatory intersect the Jicamarca magnetic meridian at a height of 400 km. During the flight the LES 9 satellite, stationed nominally at 40°W, moved from 17°40'N at 00:40 UT to 14°55'S at 06:10 UT (subsatellite points) and the elevation/azimuth angles to the satellite changed from 29°/52° East of North to 51°/101° East of North respectively. The intersection of the satellite-to-aircraft ray path with the Jicamarca magnetic meridional plane was at all times within 250 km of the ground station. Considering the field alignment of F-layer irregularities, it was assumed that the occurrence of backscatter over Jicamarca and the simultaneous observation of scintillations

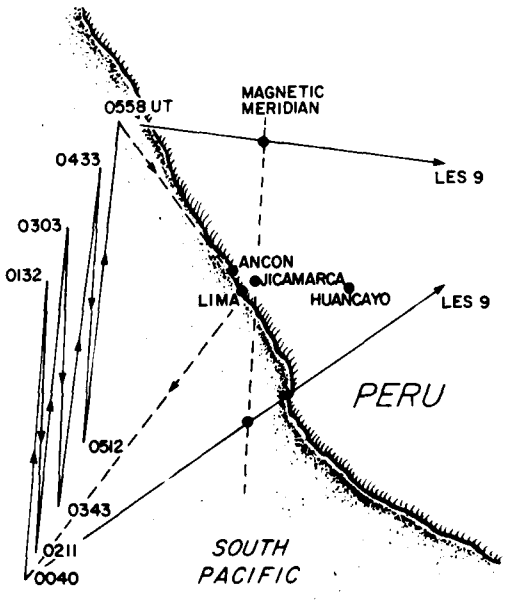


Figure 1. Flight track of 17 March 1977. Arrows labelled LES 9 are ground projections of ray paths to the satellite at the beginning and at the end of the measurements.

on the satellite-to-aircraft path were the result of the formation or arrival of F-layer irregularities belonging to the same event, at the Jicamarca meridian.

The aircraft carried satellite beacon radio receivers, vertically looking spectrometers and a new all-sky spectrophotometric imaging system (Mende and Eather, 1976; Weber et al., 1977). The ionospheric sounder used for these studies was a modified Granger 3905-I sounder (Gowell and Whidden, 1968), recording three ionograms every five minutes. The satellite receive system consisted of a Dorne-Margolin VHF/UHF wide band circularly polarized antenna, a Spectrum International preamplifier converter and a modified Collins R-390 HF receiver used as an IF amplifier. The AGC was recorded on a strip chart recorder and magnetic tape, as a measure of field strength with a post-detection bandwidth of 20 Hz. For the purpose of this study, the scintillation index, SI, in dB (Whitney et al., 1969) was determined by measuring the dB difference between the third peak down from the maximum and the third fade up from the minimum in each 2 minute time interval.

Significant airborne measurements were made with the all-sky photometer, a new spectrophotometric imaging system added to the instrumentation of the Airborne Ionospheric Observatory. The instrument is a wide field of view (155°), narrow spectral bandwidth TV system designed to operate in a time exposure mode. All-sky images of the equatorial airglow were made through 6300 Å and 5577 Å narrow band (30 Å) interference filters, using alternate 2.5 sec exposures to produce an image at each wavelength every 30 seconds. The resulting TV frames were then recorded on video tape and by photographing a TV monitor. An example of an all-sky photometer 6300 Å image is shown in the left half of Figure 2. The grid lines are magnetic meridians in 1° increments, assuming a 250 km emission height. The 6300 Å airglow emission originates in the F-region primarily as a result of dissociative recombination of molecular ionized oxygen. This emission feature is a sensitive indicator of F-region height and electron density changes; decreased intensity is associated with regions of low density or increased height of F-layer ionization. The bright airglow filling the portion of the sky from overhead to the western horizon and a second bright region extending from 2° east of the aircraft to the eastern horizon are indicative of substantial F-layer ionization below 300 km. The dark field aligned band or airglow depletion between these two bright regions is a phenomenon which was routinely observed during these equatorial flights and is the result of decreased ionization below 300 km. The ground projection on the right side of Figure 2 gives an indication of the size of the field of view, and the dimensions of this airglow depletion.

Care must be exercised in the interpretation of features near the edge of the field of

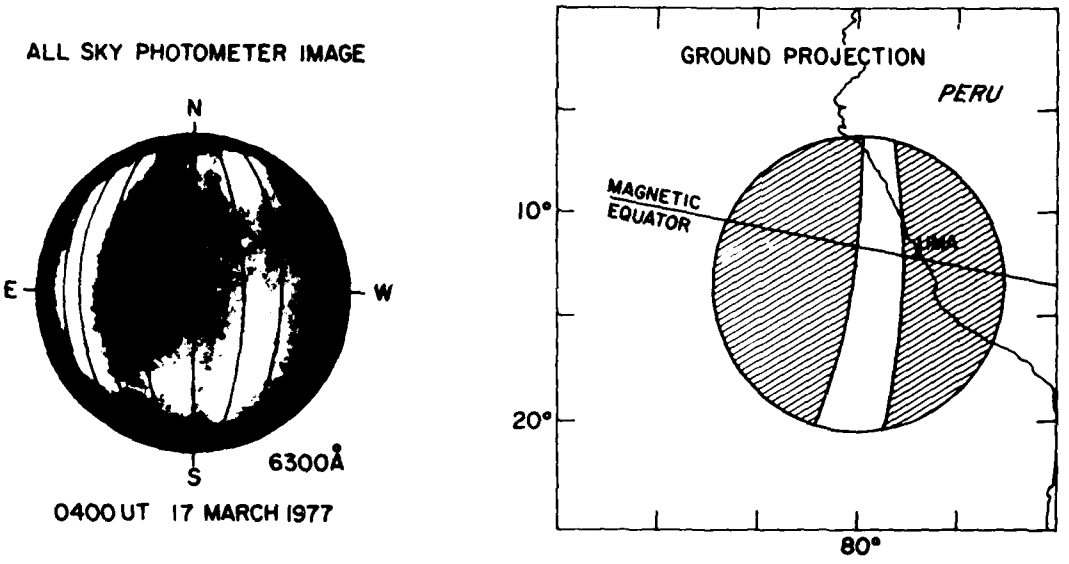


Figure 2. Example of an All-Sky Photometer Image and Its Ground Projection Assuming a 250 km Emission Height. The superimposed grid indicates the projection of corrected geomagnetic longitudes, at one degree intervals.

view. The van Rhijn effect and vignetting act in opposition, but vingetting exceeds van Rhijn enhancement at the edges, often resulting in a perceptable dark band around the image for the weak airglow features under consideration. The narrow north-south striations about ½° to the west of zenith are the ionosonde antennas which stretch above the all-sky lens.

The Jicamarca 50 MHz radar has been used for the study of equatorial irregularities for several years. The observed "plumes" resulting from backscatter due to 3 meter irregularities have been described in detail and tentatively modelled elsewhere (Woodman and La Hoz, 1976). These plumes are thought to result from the development and upward propagation of low electron density bubbles in the equatorial F-region. During these experiments the Jicamarca radar was operated to record the development and structure of the irregularities for subsequent correlation with the airborne and ground based data observations.

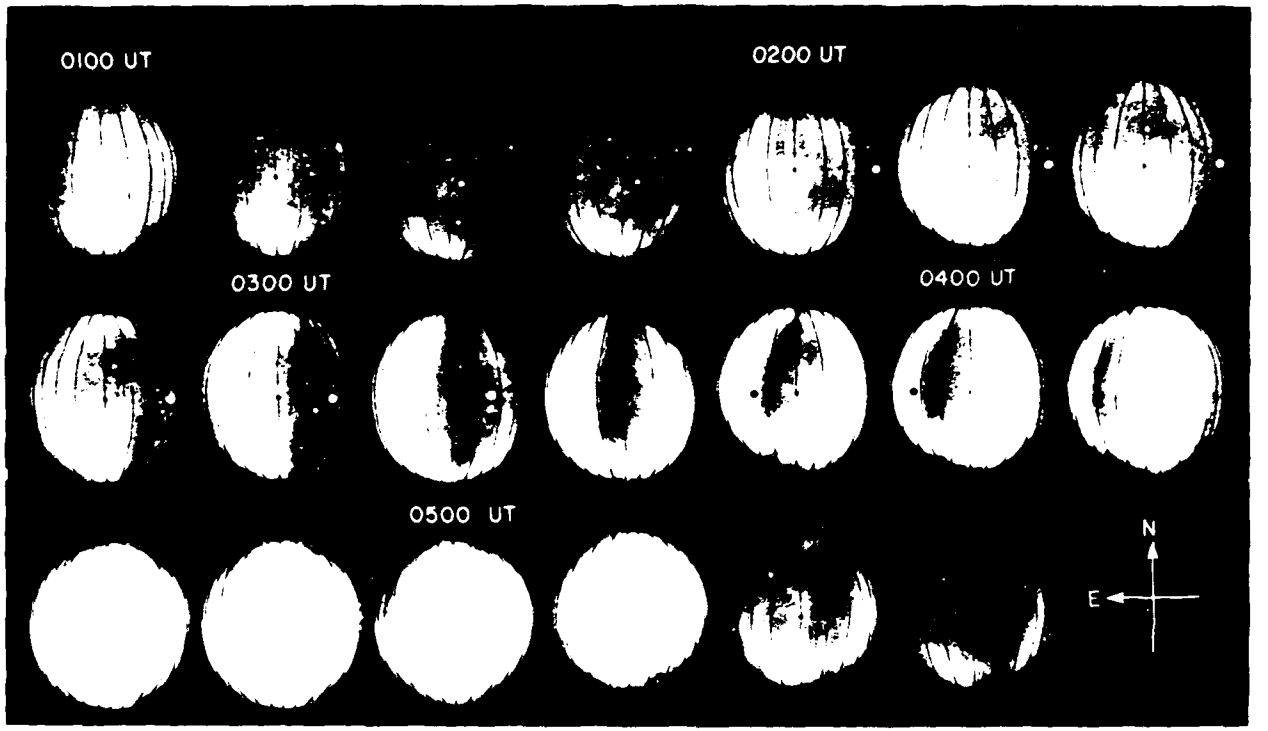
As discussed by Basu et al. (1977), VHF scintillations are the result of F-region irregularities with scale sizes on the order of 1 km, while the radar measures irregularities of 3 m scale size. Based on other measurements and theoretical calculations, Basu et al. (and references therein) established that the smaller irregularities are sometimes but not always observed in conjunction with the larger ones. Comparison of the radar maps with scintillation and airglow data will permit us to further elucidate the relation of the larger and smaller-scale irregularities.

OBSERVATIONS ON 17 MARCH 1977

All-Sky Photometer

Figure 3 presents a series of 6300 Å images (photographs of the tape recorded video frames) obtained during the flight on 17 March 1977, at 15 minute intervals between 0100 and 0545 UT. For the discussions in this paper, only the 6300 Å images were considered. All images have been reoriented with magnetic north to the top as shown in Figure 1.

The images between 0100 UT and 0200 UT show a low-level, unstructured glow with some enhancement towards the South, probably enhanced emission from the maximum of the Appleton anomaly. The Milky Way is visible in the 0100-0145 UT images as a slight enhancement aligned in the southeast-northwest direction. The 0215 UT image shows a prominent depletion in the 6300 Å airglow in the form of a dark band which extends from south to north along much of the western horizon. The formation of this dark band can


0100 UT
0200 UT
0300 UT
0400 UT
0500 UT
N
E

be seen as early as 0200 UT. Within the next 2.5 hours, this band travels across the sky, leaving the instrumental field of view on the eastern horizon by approximately 0445 UT. Generally the images show that the eastern or leading edge of the airglow depletion is closely aligned in the magnetic North/South direction, (best seen in the 0330 UT image). The leading edge displays a sharp intensity gradient in the east-west direction while the western edge of the depletion region shows a somewhat more gradual, structured transition to the adjacent bright airglow region. The width of the depletion when directly overhead at 0330 UT is approximately 150-200 km. In the north-south direction, these regions extend across the entire field of view to include a horizontal distance of more than 1200 km, assuming a 250 km emission height.

Unstructured airglow covers (most of) the observable sky until 0515 UT and then rapidly falls in intensity, leaving only minor enhancements towards the southern and western horizons.

Drifts

The Corrected Geomagnetic (C.G.) longitudes of the eastern and western edges of the airglow depletion, measured on an east-west great circle through the aircraft zenith, are shown in Figure 4. The depletion drifted eastward with a relatively constant velocity of approximately 92 m/sec, while maintaining on almost constant east-west size of approximately 165 km. Eastward drifts from 50 to 100 m/sec were observed for similar depletions on other evenings.

Ionosonde Measurements

Figure 5 presents results of the ionospheric soundings conducted simultaneously onboard the aircraft. The virtual height of the observed F layer and the virtual range of oblique F-layer echoes are shown as a function of time. The oblique echoes are first observed at a virtual range of 800 km at 0157 UT, almost coincident with the initial observation of the airglow depletion on the western horizon. These oblique returns decrease in range, consistent with the approach of a reflecting or scattering region, and merge with the overhead F-layer by 0317 UT.

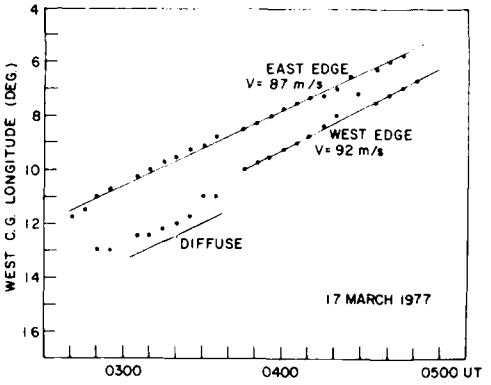


Figure 4. The Corrected Geomagnetic Longitudes of the Eastern and Western Edges of the Airglow Depletion Shown in Figure 3.

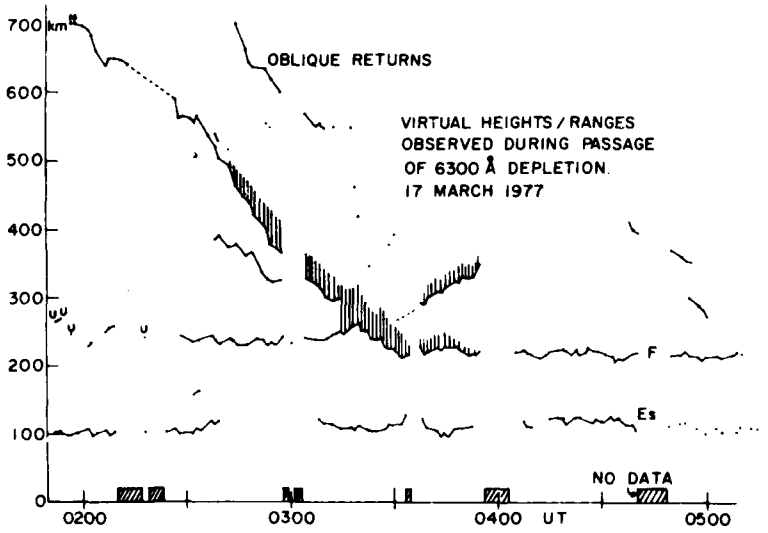


Figure 5. Virtual Heights/Ranges of Ionosonde Returns Observed During Passage of 6300 Å Depletion. The shaded areas represent range spread on backscatter returns or overhead spread F conditions.

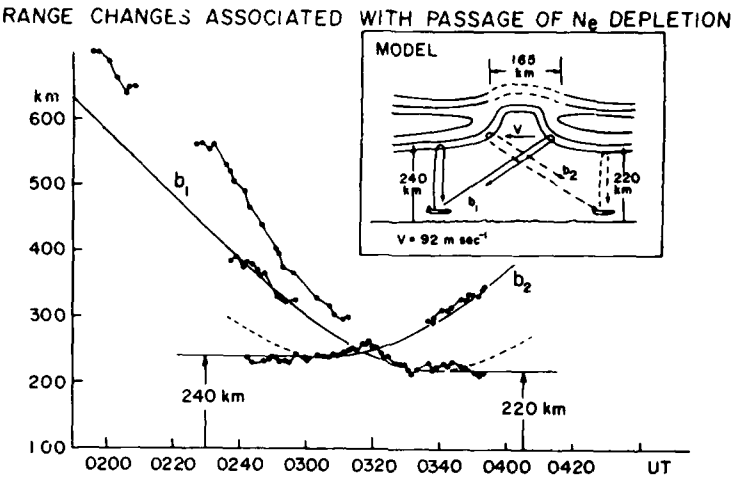


Figure 6. Model of an eastward drifting, bottomside Ne depletion or corrugation based on observed ionosonde and airglow parameters. The measured ranges of oblique returns and the virtual height of the overhead F-region are compared with range/height changes expected from the passage of the model bottomside structure over the ionosonde.

By 0335 UT oblique echoes are seen to separate from the lowest F-region trace and to increase in range to about 350 km by 0354 UT. After this time they are obscured by other returns and cannot be further identified. Even though the omnidirectional sounder antenna does not permit determination of the angle of arrival of the oblique echoes, the coinciding time histories of the airglow depletion movement and the backscatter range change suggest that the sounder observes the motion of ionospheric scattering regions associated with the motion of the depletion. To verify the identity of airglow depletion edges and backscatter regions, the location of the backscattering irregularities was superimposed on the all-sky images. Assuming a height of 250 km for the scattering region, ranges of the approaching backscatter front were converted to ground distances. The estimated locations of these approaching scattering regions are shown as white dots in the respective airglow images in Figure 3, to the west of zenith. Ground ranges derived from the receding backscatter branch, observed after the overhead passage of the leading edge of the depletion, were entered as black dots to the east of the zenith of the respective images. As Figure 3 shows, best seen in the 0245 to 0315 UT images, the approaching backscatter is tracking the trailing (western) edge of the depletion, while the leading (eastern) edge is tracked by the receding echoes observed after 0330 UT.

A model of an ionization depletion in the bottomside of the ionosphere, shown in the upper right of Figure 6, produces a sequence of approaching and receding echo traces (branch b_1 and b_2) as well as a variation in h'F, which closely resemble the observations. With the aircraft located to the east of the structure, returns are received vertically and via ray path b_1; after passage of the structure to the east of the aircraft, returns are received vertically and ia ray path b_2. As mentioned above the width of the structure has been taken as 165 km and the velocity as 92 m sec^{-1}. The time of passage of the eastern edge through the aircraft zenith was determined as 0308 UT. The virtual heights of the F-layer before and after the passage of the depletion were taken from Figure 3 as 240 km and 220 km respectively. Figure 6 shows the results of the model computations with the relevant sections of the h' plot from Figure 3 superimposed.

The fit is generally good, but additional strong returns associated with the trailing edge of the depletion at ranges larger than derived from the simple model, and approaching at a greater speed suggest a more complex structure of the trailing edge than assumed for the model. Examination of the 16 mm all-sky photometer photographic record, having 30 second time resolution, reveals diffuse structure having turbulent motion at the western edge of the depletion, in contrast to the smooth and stable features observed at the eastern edge. The east-west asymmetry is also visible in Figure 3. The 0300 UT image shows diffuse and patchy structure at the trailing edge compared to the well defined leading edge. In-situ satellite observations of presumably similar depletion regions (McClure et al., 1977) show similar variations in the structure of the edges of the depletions. The structural differences may be responsible for the differences in the ionosonde backscatter associated with the leading and trailing edge.

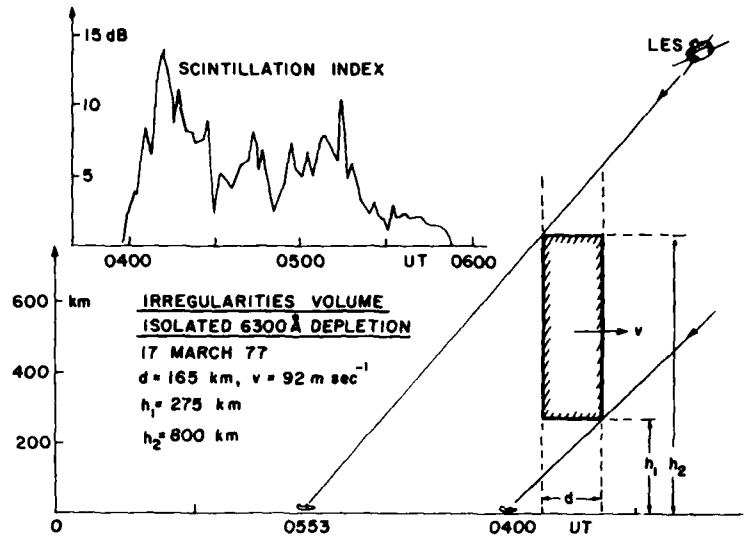


Figure 7. Isolated Scintillation Event Observed on 17 March 1977 and Estimated Irregularities Volume Cross-Section.

After the approaching trace merged with the overhead trace, strong unstructured spread F developed, and persisted for the period of passage of the depletion. Throughout the evening, the F-layer had come down, from initially 275 km (0000 UT) to 230 km (0300 UT) just prior to the overhead arrival of the leading edge of the depletion. The layer rapidly moved upwards by 35 km, reaching a maximum virtual height (h') of 265 km by 0319 UT, the time of strongest spread F. After this, the layer again rapidly moved down to 215 Km (0332 UT) and fluctuated around this level for the remainder of the observations.

Determination of foF2 between 0100 and 0500 UT is somewhat uncertain because of spread conditions and high nighttime HF noise levels. Sporadic ionograms from this time interval showing clear traces indicate that foF2 fluctuated between 8.5 and 9.5 MHz. After 0500 UT the spread and noise conditions improved and a clear decrease of the foF2 from 9.2 MHz (0503 UT) to 8.0 MHz (0523 UT) and finally to 5.8 MHz (0549 UT) is observed, which follows the decrease of the overall brightness of the 0500 UT to 0545 UT all-sky photometer images. Since h'F does not change appreciably (from 212 km at 0503 to 226 km at 0549 UT) this change in airglow level is directly attributable to the Ne decay.

Scintillations

Three clearly separated scintillation events were observed in the aircraft during this flight lasting from 0104 UT to 0242 UT, from 0215 UT to 0317 UT and the last one from 0400 UT to 0553 UT. Here we are concerned only with the last event, since it occurred during the time when the previously described depletion moved through the LES 9 ray path. The eastward drift in effect moved the ray path from lower to higher altitudes through the region. The scintillations observed during this last event are shown in the upper left of Figure 7. The scintillations start with SI showing a strong 14 dB peak; after that they fluctuate for about one hour between 5 and 10 dB and decay towards the end to weak 2 dB scintillations.

If we assume, that the irregularities responsible for the observed scintillations reside in a volume coincident with and extending vertically above the airglow depletion shown by the all-sky photometer, we can determine a minimum and a maximum height for the volume containing the irregularities. Using the width of 165 km and the eastward velocity of 92 m sec^{-1} determined from the all-sky photometer measurements, the minimum height can be estimated as 275 km and the maximum height as 800 km. The geometry of the ray paths at the start and the end of the isolated scintillation event and the cross-section of the volume defined by these rays and the depletion is shown in the lower right of Figure 7. It has to be pointed out, however, that neither the all-sky photometer nor the sounder measurements allow any inference about the ionospheric structure above 300 km.

Jicamarca Backscatter Measurements

The 50 MHz backscatter measurements (Figure 8) show the time history of the development or drift of 3 m irregularities above Jicamarca. The picture can be understood either as the time history of an eastward

drifting irregularity region observed from a fixed location or, assuming a rigid ionosphere, as an east-west cross-section of such a region.

Some irregularities are seen in the lower F-region from the beginning of the observations until 0300 UT. Starting at 0357 UT the first echoes from an extended region of irregularities are observed at 500 to 600 km height. This disturbance eventually involves the whole F-region between 175 and 670 km. Irregularities in the F-region below 500 km disappear between 0430 UT (500 km level) and 0450 UT (200 km level), while some very weak irregularities above 500 km are observed until 0535 UT. The relatively uniform diffuse background appearing at the first digital level (0 to 6 dB above threshold) was caused by a computer malfunction, and does not indicate the existence of a diffuse background of weak irregularities.

To assess the relation between the airglow depletion and the approximate volume of irregularities resulting in scintillations on the one hand, and the observed backscatter on the other, the volume cross-section shown in Figure 7 is superimposed in dashed lines on the backscatter power map (Figure 8). Using a speed of 90 m sec^{-1}, the arrival time of the airglow depletion at the Jicamarca meridian was estimated to be 0410 UT. At a velocity of 90 m/sec the 165 km wide airglow depletion passed over Jicamarca in exactly 30 minutes. This determines the horizontal extent of the volume cross-indicated in Figure 8. As the figure shows, the estimated volume is in general agreement with the observed backscatter cloud. There is a discrepancy between the shape of the simple rectangle based on the local airglow and scintillation observations and the envelope of the 3-m irregularities shown in the figure. This may be indicative of

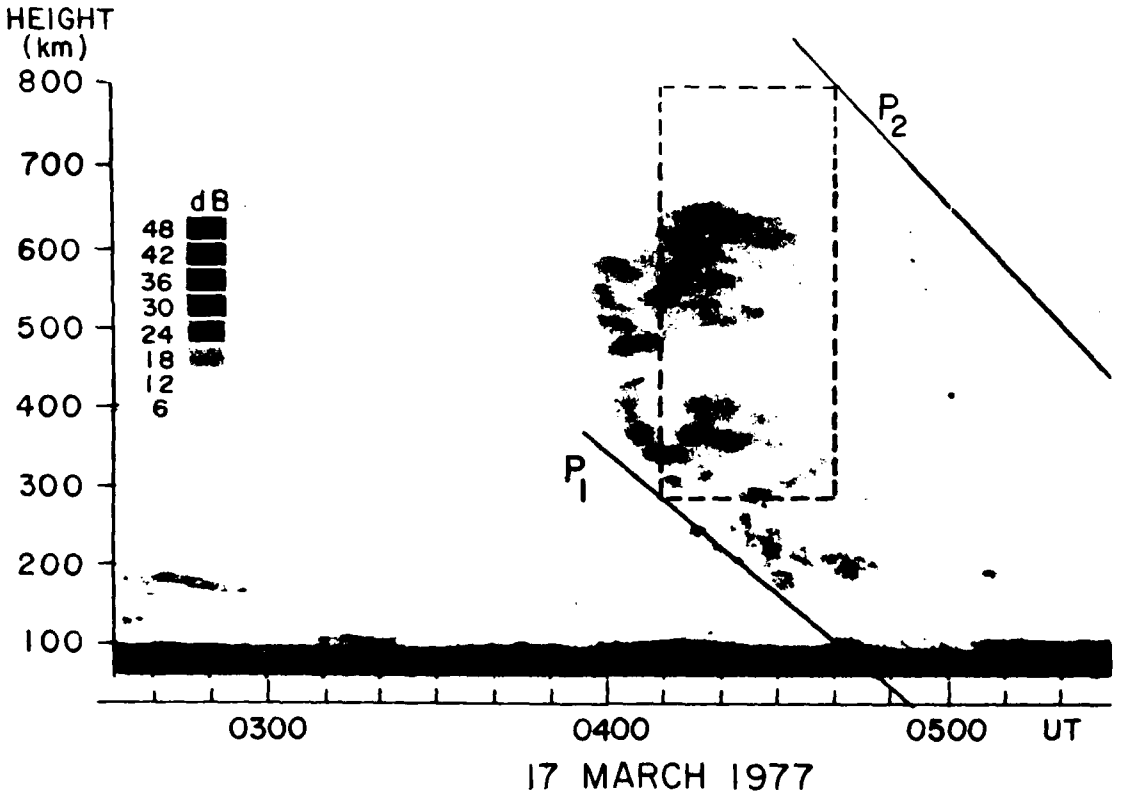


Figure 8. Range time-intensity map of isolated F-region disturbance passing over the Jicamarca radar. The dashed lines represent the irregularity volume determined from airglow and scintillation measurements. P_1 and P_2 denote the trans-ionospheric ray path, through the disturbed region at the beginning and end of the associated scintillation event.

1) changes with altitude of the eastward drift of the ionosphere and/or 2) changes with time and altitude of the upward and relative westward drift velocity of plasma depletion regions (plasma bubbles) with respect to the background ionosphere, as discussed by McClure et al. (1977). Airglow observations, made near the intersection with the 250 km altitude surface, of these middle and upper F-region geomagnetic field lines which cross over Jicamarca, would be useful for comparison with the detailed shape of the region of echoes above Jicamarca.

In Figure 8 we have also shown the position of the ray path from the aircraft to the LES 9 satellite at two selected Universal Times. P1 is the ray path at 0400 UT, the start, and P2 the ray path at 0553 UT, the end of scintillation event. The figure now shows that the crossing of 3m irregularities into the ray path at heights between 175 and 275 km coincides with the onset of the scintillations. The strong 14 db scintillation event coincides approximately with the movement of intense irregularities in the 300-400 km height range into the ray path. And finally, the weak scintillations observed after 0520 UT (see Figure 6) appear to result from the trailing irregularities associated with the weak backscatter seen for about one hour at heights between 500 and 600 km, after the major cloud has passed.

CONCLUSIONS

A new technique of monochromatic, all-sky imaging used onboard an aircraft has revealed the existence of magnetically North-South aligned regions of airglow depletions with E/W dimensions of 50 to 200 km and a North/South extent larger than the 1200 km field of view of the all-sky photometer.

Airglow images, ionosonde and 50 MHz backscatter data have been analyzed to describe the airglow depletion and its relation to the F-region plumes observed at Jicamarca, using a well defined airglow depletion observed on 17 March 1977.

It is concluded that the airglow depletions, the oblique echo/range variations and h'F changes seen by the ionosonde and the 50 MHz backscatter plumes, can all be understood as different aspects of the same phenomenon, a volume of reduced electron density, containing meter to kilometer size irregularities, extending from the lower to the upper F-region causing scintillations on ray paths transiting this volume.

ACKNOWLEDGEMENTS

The authors wish to thank the Jicamarca Radio Observatory personnel for their support in taking and processing the radar information. The success of the airborne missions was due to engineering support provided by R.W. Gowell, J.B. Waaramaa and J. W.F. Lloyd, logistics support by R. Carnevale, and airborne support from the 4950th Test Wing, Wright-Patterson AFB, OH.

This research was supported, in part, by the Air Force Laboratory Independent Research Fund of the Air Force Geophysics Laboratory, Air Force Systems Command.

REFERENCES

Basu, S., J. Aarons, J.P. McClure, C. LaHoz, A. Bushby and R.F. Woodman, Preliminary Comparison of VHF Radar Maps of F-region Irregularities with Scintillations in the Equatorial Region, to appear in J. Atmos. Terr. Phys., Sept. 1977.

Gowell, R.W. and R.W. Whidden, Ionospheric Sounders in Aircraft, AFCRL-68-0369, Instrumentation Papers, No. 144, 1968.

McClure, J.P., W.B. Hanson and J.H. Hoffman, Plasma Bubbles and Irregularities in the Equatorial Ionosphere, J. Geophys. Res., 82, 2650, 1977.

Mende, S.B. and R.H. Eather, Monochromatic All Sky Observations and Auroral Precipitation Patterns, J. Geophys. Res., 81, 3771, 1976.

Weber, E.J., J. Buchau, R.H. Eather and S.B. Mende, North/South Aligned Equatorial Airglow Depletions, J. Geophys. Res., (in press) 1977.

Whitney, H.E., C. Malkk and J. Aarons, A Proposed Index for Measuring Ionospheric Scintillations, Planet. Space Sci., 17, 1069, 1969.

Woodman, R.F. and C. LaHoz, Radar Observations of F Region Equatorial Irregularities, J. Geophys. Res., 81, 5447, 1976.

Theoretical and Numerical Simulation Predictions of the Equatorial Spread F Environment

S. L. Ossakow, S. T. Zalesak, B. E. McDonald, and
P. K. Chaturvedi*

Naval Research Laboratory, Washington, D.C. 20375
*University of Maryland, College Park, Maryland 20742

ABSTRACT

The nonlinear evolution of the collisional Rayleigh-Taylor instability in the nighttime equatorial F region ionosphere has been investigated by using a two-dimensional (perpendicular to the magnetic field) numerical simulation. The peak of the F region is taken to be at 430 km. Irregularities (plasma density depletions) produced by the instability on the bottomside grow and nonlinearly bubble through to the topside with vertical rise speeds $\sim$ 160 m/sec and 85% depletions. The one-dimensional (altitude) power spectral density for the irregularities is $\propto k^{-2.3}$.

I. INTRODUCTION

Equatorial Spread F is now generally believed to be initiated by the collisional Rayleigh-Taylor instability occuring in the evening bottomside equatorial F region [Balsley et al., 1972; Haerendel, 1974; Scannapieco and Ossakow, 1976]. Much of the recent experimental evidence supports this postulate [Kelley et al., 1976; Woodman and LaHoz, 1976; McClure et al., 1977]. Associated with the generation of this instability on the bottomside is the formation of rising plasma density depletions (or bubbles). These bubbles have been observed experimentally [Kelley et al., 1976; Woodman and LaHoz, 1976; McClure et al., 1977] and by numerical simulation studies [Scannapieco and Ossakow, 1976]. McClure et al. [1977] have observed the vertical rise speed of these bubbles to range from several meters/sec (and even zero) to several hundred meters/sec. Costa and Kelley [1976] have attempted to show that these bubbles or plasma density depletions have properties that could cause intense VHF scintillations as well as scintillations in the gigahertz range. McDonald et al. [1976] have performed scintillation calculations (using a parabolic equation solver) for 1 meter (300 MHz) waves propagating through the actual bubble regions provided by the numerical simulation output of Scannapieco and Ossakow [1976]. Significant S_4 indices (as a function of time) resulted from these calculations.

These previous studies show the importance of being able to predict the nighttime equatorial Spread F/scintillation environment. It is well known that this environment has deleterious effects on C^3 systems. In the present paper we extend our previous studies [Scannapieco and Ossakow, 1976] by raising the altitude of the F region peak to 430 km [Kelley et al., 1976] and find important new simulation results (compared with our previous ones) associated with bubble information.

II. THEORY AND NUMERICAL SIMULATION RESULTS

We now wish to present the two dimensional set of equations which will be used to describe the nonlinear evolution of the collisional Rayleigh-Taylor instability for equatorial nighttime F region conditions. The ambient geomagnetic field, B, is taken to be constant and in the z direction, the y axis is vertically upward (gravity is in the negative y direction), and the x-axis points westward. The plasma two-fluid equations describing the system are

$$\frac{\partial n_\alpha}{\partial t} + \nabla\cdot(n_\alpha \underline{V}_\alpha) = -\nu_R(n_\alpha - n_{\alpha o}) \tag{1}$$

$$\underline{V}_e = \frac{c}{B}\underline{E} \times \hat{z}, \quad \hat{z} \equiv \frac{\underline{B}}{|B|} \tag{2}$$

$$\underline{V}_i = \left(\frac{\underline{g}}{\Omega_i} + \frac{c}{B}\underline{E}\right)\times \hat{z} + \left(\frac{\underline{g}}{\Omega_i} + \frac{c}{B}\underline{E}\right)\frac{\nu_{in}}{\Omega_i} \tag{3}$$

$$\nabla\cdot\underline{j} = 0, \quad \underline{j} = e(n_i\underline{V}_i - n_e\underline{V}_e) \tag{4}$$

where the subscript α denotes species (e is electron, i is ion), n is density; $\underline{V}$ is velocity; ν_R is recombination rate

($n_{\alpha o}$ is the equilibrium value such that in equilibrium the right hand side of (1) is zero); $\underline{E}$ the electric field; g is gravity, e is electronic charge; ν_{in} is ion-neutral collision frequency; c is speed of light; and $\Omega_i = eB/m_ic$ is the ion gyrofrequency. Eqns. (2) and (3) are derived from the electron and ion momentum equations. In particular, for the ions we have neglected inertial effects ($\partial/\partial t + V_i \cdot \nabla$ terms) and taken $\nu_{in}/\Omega_i << 1$ in the momentum equation and this results in eqn. (3). The temperature and neutral wind have been set to zero. The recombination rate, ν_R, represents the rate limiting charge exchange, $O^+ + O_2 \rightarrow O_2^+ + O$, and ion-molecule reaction, $O^+ + N_2 \rightarrow NO^+ + N$, rates [McFarland, et al., 1973].

Taking eqns. (1) - (4), assuming $n_e = n_i = n$, and setting $\underline{E} = -\nabla\phi$, we obtain

$$\frac{\partial n}{\partial t} - \frac{c}{B}(\nabla\phi_1 x\hat{z})\cdot\nabla n = -\nu_R(n-n_o) \qquad (5)$$

$$\nabla\cdot(\nu_{in} n\nabla\phi_1) = \frac{B}{c}(\underline{g} \times \hat{z})\cdot\nabla n \qquad (6)$$

where $\phi = \phi_o + \phi_1$, $\nabla\phi_o = m_i g/e$ and we have neglected any other ambient zero order electric fields. Eqns. (5) and (6) are the basic two dimensional (x,y; ∇ refers to these directions) set which we will solve numerically. In our simulations n_o, ν_{in}, and ν_R have realistic altitude (y) dependences. Linear perturbation theory [ϕ_1, $\delta n \propto \exp(ikx + \gamma t)$, $n \equiv n_o + \delta n$) applied to eqns. (5) and (6) yields

$$\gamma = -\frac{g}{\nu_{in}} \cdot \frac{\nabla n_o}{n_o} - \nu_R \qquad (7)$$

This exhibits positive values (growth) on the bottomside of the F region peak and negative values (damping) on the topside.

The numerical simulation was performed over a two-dimensional mesh corresponding to an altitude (y) range between 332 km and 532 km and an East-West (x) extent of 8 km (see Fig. 1). The mesh was such that there were 102 points in the y direction and 42 in the x direction with periodic boundary conditions (note: this corresponds to $\Delta y = 2$km and $\Delta x = 200$ meters). The background (zero order) ionospheric electron density, n_o, employed is depicted in Fig. 1 (The F peak is at 434 km). Superimposed on this at t=0 was an initial cosinelike perturbation, $\delta n/n_o \sim 5\%$ in the x direction with a wavelength ~ 3 km with a maximum in the center of the x direction. For the n_o used in the simulation the shortest bottomside density gradient scale lengths, $L = n_o(\partial n_o/\partial y)^{-1}$, are ~ 10 km. The largest linear growth rate, $\gamma = 6.07 \times 10^{-3}$sec^{-1} and occurs at an altitude of 378 km (at this altitude $\nu_R = 3.2 \times 10^{-5}$sec^{-1}). Unlike our previous simulation [Scannapieco and Ossakow, 1976], here the entire bottomside is linearly unstable. That is, between 332 km and 432 km the recombination term in eqn. (7) does not balance the gradient term (in this range ν_R goes from 1.3×10^{-4} to 6.4×10^{-6}sec^{-1} while the growth rate goes from 2.5×10^{-3} to 1.0×10^{-4}sec^{-1}; note exactly at the F peak $\partial n_o/\partial y = 0$ and there is damping, i.e., $\gamma = -\nu_R$).

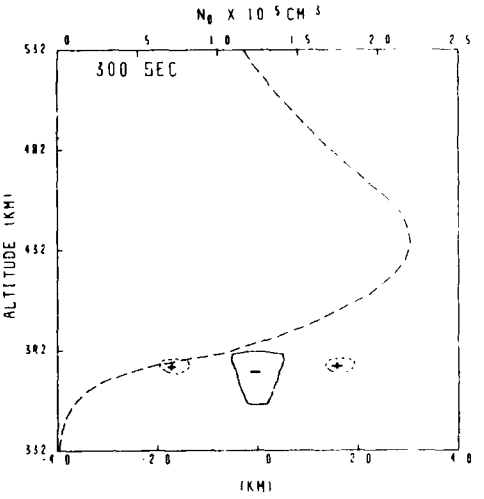


Figure (1). Contour plot of $\delta n/n_o$ at t = 300 sec. The plus and minus signs indicate enhancement and depletions over ambient electron number density. The large dashed curve represents a plot of ambient electron number density (values on upper horizontal axis), n_o, as a function of altitude. The vertical axis (y) represents altitude, the lower horizontal axis (x) East-West range, and the ambient magnetic field is out of the figure (z). The contour level in this figure is -16% and +19%.

Figures 1-4 exhibit isodensity contour plots of $\delta n/n_o$ at t = 300,700, 1000 and 1400 secs and each is overlayed with a plot of n_o (dashed curve) as a function of altitude. Figure 1 exhibits the early phase of growth of the collisional Rayleigh-Taylor instability and shows the formation of plasma enhancement (+) and depletions (-). The depletion contour represents a 16% depletion; whereas, the enhancement contour represents a 19% enhancement (note: Here and in subsequent figures the contour plotting is such that the first (outer) depletion contour n/n_o is $2^{-1/4}$ and each subsequent one is $2^{-1/2}$ times the previous one. Then one subtracts 1.0 to find $\delta n/n_o$. For the

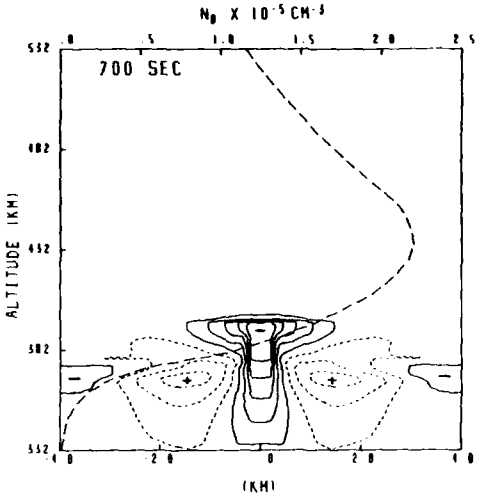


Figure (2). Same as figure 1 except t = 700 sec. The depletion contours in the central region range from -16% (outermost) to -79% (innermost) with a maximum of -84% within the bubble. The enhancement contours range from 19% (outermost) to 138% (innermost) with a maximum of 187% inside the innermost contour.

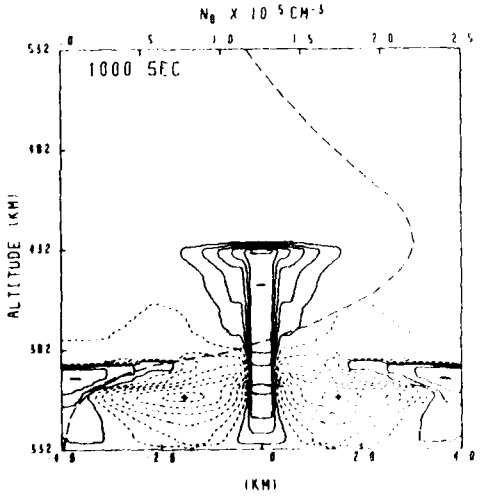


Figure (3). Same as figure 1 except t = 1000 sec. The central region depletion contours range from -16% to -85% with a maximum depletion inside the innermost contour of -85.2%. The enhancement contours range from 19% to 572% with a maximum of 674% inside the innermost contour.

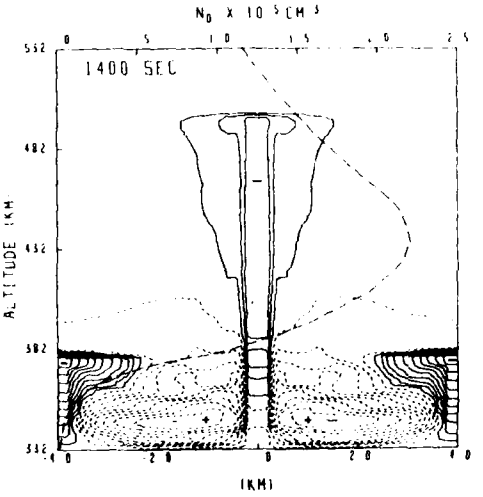


Figure (4). Same as figure 1 except t = 1400 sec. The central depletion contours (bubble) now range from -16% to -58% with a maximum inside of -70%. The depletions off to the side have an innermost contour of -96%. The enhancement contours range from 19% to 2600%.

enhancements, the first (outer) contour is $2^{1/4}$ and the subsequent ones are $2^{1/2}$ times the previous ones. Then, one substracts 1.0 to find $\delta n/n_o$ for enhancements). The greatest depletion ins he depletion contour in Fig. 1 is 26% at this time; whereas, the largest enhancement, inside the enhancement contour, is 23%. Figure 2 depicts a clear density depletion (bubble) in the center of the x direction rising toward the F peak. At this time (700 sec), the innermost depletion contour, in the center where the bubble is forming, is 79% (with the maximum depletion within the bubble being 84%). Depletions in the wings in the x direction (16% contour) are also in evidence. The innermost enhancement contour represents a 138% enhancement (the maximum enhancement inside this contour is 187%).

Figure 3 depicts the bubble rise at 1000 sec. as it crosses the F region peak. At this time the innermost contour in the bubble corresponds to an 85% depletion (maximum depletion 85.2% inside this contour). We

notice that the depletions in the wings show an innermost contour of 58% depletion. Also, we note that the center bubble has not completely broken from the lower altitude region and is forming a long trailing tail. This is unlike the previous numerical simulation [see Fig. 3 of Scannapieco and Ossakow, 1976] performed for the altitude regime 250 km to 450 km (within this region recombination effects did not allow growth between 250 and 270 km, i.e., this was a damped region). The enhanced regions in Fig. 3 exhibit 572% as the innermost contour (a maximum of 674% inside of this contour). In this figure we also note that the enhancements encompass a large region in the x direction though confined in altitudes $\leq$ 390 km (basically enhancements move downward while depletions move upward).

Figure 4 shows the evolution of depletions and enhancements at 1400 sec. The top of the main bubble (depletion) in the central x region is at an altitude $\sim$ 500 km which is beyond the F peak (434 km). Moreover, it has a long trail connecting it to an altitude of 357 km. The widest part of the top outermost contour (16% depletion) of the bubble is $\sim$ 3 km; whereas, the innermost contour is $\sim$ 0.5 km wide. Consequently, we have a long narrow bubble which extends $\sim$ 150 km in altitude and a few km in East-West dimension. Woodman and LaHoz [1976], on occasion, have observed radar backscatter "plumes" to extend over 100 km in altitude. They have identified these plumes with depleted plasma regions. The innermost contour in the central bubble region in Fig. 4 represents a 58% depletion (with a maximum depletion within this contour of 70%). The innermost contour in the depleted wings represents a 96% depletion (maximum depletion within this contour is 97%). The innermost contour in the enhancement represents a 2600% enhancement (maximum enhancement within this contour is 2640%). Also from Fig. 3 to Fig. 4 the top most part of the bubble has moved $\sim$ 65 km in 400 sec which represents a rise velocity $\sim$ 160 m/sec (compared with $\sim$ 10 m/sec in Scannapieco and Ossakow, 1976).

Figure 5 displays contours of constant induced potential, ϕ_1, from eqn. (6) over the computational mesh at 1400 sec. This helps to interpret Fig. 4. We note that the more isolated high altitude part of the central bubble is acted on by an induced electric field that points from West to East. This causes the bubble to rise with a velocity $-(c/B)\,\nabla\phi_1 \times \hat{z}$. However, the lower portion of the mesh is acted on by an induced electric field which points from East to West (also much weaker than the field pointing from West to East) and causes the enhancements and depletions to move down by $-\nabla\phi_1 \times \hat{z}$. This causes the depletions in the wings as well as the central portion of the bubble to be

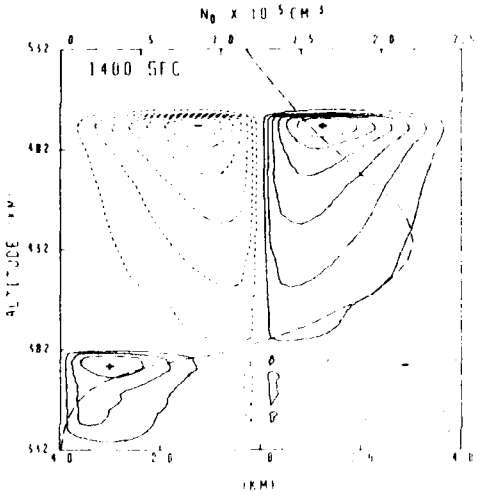


Figure (5). Contours of constant induced potential, ϕ_1, over the mesh at t = 1400 sec. Plus and minus denote positive and negative values, with values decreasing in magnitude as one goes from the innermost to the outermost contours.

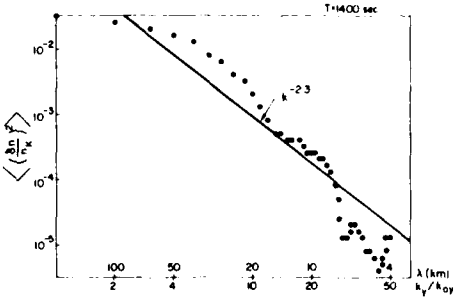


Figure (6). Log-log plot of irregularity amplitude squared vs the mode number (and wavelength) in the altitude (y) direction. This is the one-dimensional power spectrum. The points represent results from the numerical simulation. The solid straight line is a least squares fit of a power law to all of the points.

captured by the enhancements. Basically isolated depletions and enhancements should move up and down, respectively, at the equator. However, depletions surrounded by

large enhancements mitigates this view. Also, the central bubble does not break as easily at low altitudes as our previous numerical simulation [Scannapieco and Ossakow, 1976] because the plasma at the present altitudes are more incompressible (ν_R was more effective in the previous case), i.e., in eqn. (5) $\partial n/\partial t - (c/B)\ (\nabla\phi_1 \times \hat{z}) \cdot \nabla n \approx 0$.

Figure 6 depicts the power spectral density (fluctuation density squared vs wavenumber) in the vertical direction at 1400 sec. This plot was obtained by two-dimensionally Fourier analyzing $\delta n/n_o$ in Fig. 4 and then integrating over k_x. The numerical simulation points were then least squares fitted with a power law power spectrum. This procedure yields

$$(\delta n/n_o)^2(k_y) \propto k_y^{-2.3}$$

III. SUMMARY AND CONCLUDING REMARKS

We have performed a numerical simulation of the nonlinear collisional Rayleigh-Taylor instability for nighttime equatorial Spread F conditions. Ion-neutral collision frequencies, recombination effects and background electron densities with realistic altitude profiles have been utilized. The F region peak altitude was 434 km. We find that a bubble (plasma density depletion) forms on the bottomside of the F region which then rises through the F peak by induced $\underline{E} \times \underline{B}$ motion. This produces irregularities on the topside where linear analysis would predict no irregularities.

As the bubble moves through the F peak (at 1000 sec) the rise velocity is $\sim$ 160 m/sec and the fractional depletion within the bubble, $\delta n/n_o \sim 85\%$. These values are much larger than the bubble speed $\sim$ 10m/sec and depletions, $\delta n/n_o \sim 40\%$ (In this case the bubble went through the peak in 8000 sec) found by Scannapieco and Ossakow [1976]. The only difference between the two numerical simulations is that in the present case the altitude of the F region peak is 434 km, whereas, in the previous simulation it was 350 km. This 80 km rise of the peak has made a significant difference. The greater bubble velocity can be understood as follows. Ossakow and Chaturvedi [1977] have shown that the bubble rise velocity is given by $V_B = (g/\nu_{in})\ f\ (\delta n/n_o)$, where $f\ (\delta n/n_o)$ is an increasing function of the fractional depletion $\delta n/n_o$ and depends on bubble shape. Raising the F peak altitude decreases ν_{in}. Furthermore in the present simulation $\delta n/n_o$ is also larger, both factors tend to increase V_B. Furthermore, $\delta n/n_o$ can be predicted to be larger, at least by linear theory, because $\delta n \propto \exp \gamma t$, where $\gamma = g/\nu_{in}L - \nu_R$. Raising the ionosphere (height of F peak) increases the growth rate, γ.

At 1400 sec, the bubble has an altitude extension $\sim$ 150 km (with the top of the bubble at 500 km altitude). The maximum depletion within the bubble is $\sim$ 70%, with other depletions $\sim$ 97% in the wings away from the main bubble remaining well below the peak, surrounded by large enhancements. Fourier analyzing these irregularities produces a one-dimensional power spectral density, in the altitude direction, $\propto k^{-2.3}$.

These numerical simulations correlate well with existing experimental (radar backscatter and rocket and satellite in situ measurements) data. The results we have presented indicate the state of the art in obtaining a predictive capability for equatorial Spread F/scintillation phenomena. The predictive capability is one of understanding the physical phenomena, deriving equations which describe the mechanisms, performing numerical simulations of those equations, and then producing the irregularity environment.

ACKNOWLEDGEMENT

This work was supported by the Defense Nuclear Agency.

REFERENCES

Balsley, B. B., G. Haerendel and R. A. Greenwald, Equatorial Spread F: Recent observations and a new interpretation, J. Geophys. Res., 77, 5625, 1972.

Costa, E., and M. C. Kelley, Calculations of equatorial scintillations at VHF and gigahertz frequencies based on a new model of the disturbed equatorial ionosphere, Geophys. Res. Letts., 3, 677, 1976.

Haerendel, G., Theory of Equatorial Spread F, preprint, Max-Planck Institute fur Physik und Astrophysik, 1974.

Kelley, M. C., G. Haerendel, H. Kappler, A. Valenzuela, B. B. Balsley, D. Carter W. Ecklund, C. W. Carlson, B. Hausler, and R. Torbert, Evidence for a Rayleigh-Taylor type instability and upwelling of depleted density regions during equatorial Spread F, Geophys. Res. Lett. 3, 448, 1976.

McClure, J. P., W. B. Hanson, and J.H. Hoffman Plasma bubbles and irregularities in the equatorial ionosphere, J. Geophys. Res., 82, 2650, 1977.

McDonald, B. E., S. L. Ossakow, and A. J. Scannapieco, A nonlinear model for equatorial Spread F irregularity and scintillation calculations, in

Symposium Proceedings of Geophysical Use of Satellite Beacon Observations, edited by M. Mendillo, Boston University, June 1976.

McFarland, M., D. L. Albritton, F. C. Fehsenfeld, E. E. Ferguson, and A. L. Schmeltekopf, Flow-drift technique for ion mobility and ion-molecule reaction rate constant measurements. 2. Positive ion reaction of N^+, O^+ and N_2^+ with O_2 and O^+ with N_2 from thermal to ∿ 2eV, J. Chem. Phys., 59, 6620, 1973.

Ossakow, S. L., and P. K. Chaturvedi, Morphological studies of rising equatorial Spread F bubbles, J. Geophys. Res. (submitted September 1977).

Scannapieco, A. J., and S. L. Ossakow, Nonlinear equatorial Spread F, Geophys. Res. Letts., 3, 451, 1976.

Woodman, R. F. and C. LaHoz, Radar observations of F-region equatorial irregularities, J. Geophys. Res. 81, 5447, 1976.

CHATANIKA MODEL OF THE HIGH-LATITUDE IONOSPHERE

R. Vondrak, R. Tsunoda, E. Hatfield, P. Perreault and G. Smith
Radio Physics Laboratory, SRI International
Menlo Park, CA 94025

ABSTRACT

Electron density measurements made with the incoherent-scatter radar at Chatanika, Alaska have been used to obtain a synoptic model of the high-latitude ionosphere. This Chatanika model is a modification of the ionospheric model developed by the USAF (RADC/ET, Hanscom AFB) for use in ray-tracing codes for HF propagation prediction. The Chatanika radar has been used to map out the two-dimensional (altitude, latitude) spatial variation of ionization. A region extending over a geomagnetic latitudinal range of 62^{o} to 68^{o} in the E-region and 56^{o} to 73^{o} in the F-region can be surveyed with a spatial resolution of about 10 km and a temporal resolution of approximately 15 minutes. Characteristic features that are commonly observed include the trough, intense E layers corresponding to the diffuse and discrete auroras, and the daytime E and F layers. Of particular interest is the valley between the E and F layers because it may serve as an HF duct for transauroral propagation. The spatial location of all these features and their dependence on magnetic activity, local time, and season are being investigated by using a number of 24-hr observations made during the last two years. In addition, an auroral absorption model that includes a magnetic-activity dependence has been developed from previously compiled riometer data. Both the ionospheric morphology model and the auroral absorption model are incorporated into a subroutine package compatible with existing raytracing codes.

MODELLING OF LOW LATITUDE IONOSPHERE

Y.V. Somayajulu and A.B. Ghosh

Space Research Section, Radio Science Division
National Physical Laboratory, New Delhi-110012

Abstract

Ionospheric models are in demand for forecasting ionospheric characteristics for radio system applications and for evaluating and correcting for refraction effects in position fixing. In order to correct for the refraction effects, one has essentially to utilize a ray- tracing technique which requires a knowledge of the electron density distribution along the ray path. The technique of building up electron density profiles described in this paper is essentially based on the method proposed by Somayajulu et al (1965). The region around hmF2 built by assuming an alpha-Chapman layer extending to 1.5 scale height above hmF2. The bottomside profile from one scale height below hmF2 down to E-region is represented by an exponential fit and the D-region by Somayajulu-Mitra model(1977). The topside region above (hmF2+1.5H) is divided into two exponentially decaying regions, the lower one corresponding to heavy ion(O^+) as the dominant constituent and the upper one corresponding to lighter ions (He^+ or H^+) as the dominant constituent. The two constraints put on the development of topside electron density profiles are electron content obtained from satellite beacon experiment and electron density at 1000 km obtained by satellite probes.

1. Introduction

Ionospheric models are needed for forecasting ionospheric characteristics for radio system applications and for evaluating and correcting for refraction effects in position fixing. With increasing demand on the accuracy of position fixing, there is need first to evaluate the errors introduced by ionospheric refraction and then to develop correction techniques so that from measured values of range or range rate and elevation angles, the true values may be computed very quickly. Such methods employ the three-dimensional ray-tracing technique which requires the specification of the refractive index profiles in the ionosphere which is essentially one of specifying the electron density distribution along the raypath. The primary ionospheric input could then be either real time data, such as foF2, h'F and/or the total electron content. If real time data are not readily available the predicted values based on ionospheric models will have to be used. In a practical situation the inputs may be a combination of real time data and predicted values. The problem then is one of generating a family of model electron density profiles with sufficient latitudinal and longitudinal resolution and with local time dependence introduced. Several models have been developed to characterize the true height dependence of electron density in the ionosphere. Models based on theoretical considerations of ionospheric processes such as photoionization processes, thermal distribution, electric currents etc., in the ionosphere, are useful for evaluating the effects of perturbations and in reproducing, transient events. One such model is the Penn State MK-I model[1]. It is primarily a daytime model of N-h resulting from Solar production and atmospheric loss. The shape of the electron density profile is found from the solution of the equations of continuity. Such models, although more accurate, have the disadvantage of requiring large computers and hence time consuming.

On the other hand empirical or statistical models are simpler, less time consuming but have lower accuracy. In a practical situation like forecasting, position fixing by satellites or tracking where time is a constraint and other limitations like computer capacity prevail and where a high degree of accuracy is not needed, a semi-empirical model like the present one would be suitable. It is important, however, that the ionospheric models represent the properties of the ionosphere as accurately as possible as functions of geophysical indices, with some statistical description of their variability. Some empirical models are available. However, these models are more suited for midlatitude ionosphere. This paper is concerned with development of a modelling technique for building electron density profiles for low latitude regions such as the Indian subcontinent where the electron density profiles have strong latitude dependence, particularly, during daytime.

2. Currently Available Empirical Models

2.1 AFCRL Model

It makes use of the Penn State model near the E and F1 regions with a smoothing to fit F2 profile. The thickness parameter is specified internally with reference to neutral temperature model as a function of space, time and solar flux. The shape of F2 region is specified externally either by Chapman, parabolic or Sin^2X.[2]

2.2 BENT Model

Bent et al[3,4] developed electron density models based on the analysis of many tens of thousands of topside and bottomside profiles. It essentially makes use of the bottomside profile from ionosonde data with biparabolic layer at the F2 peak. The topside region is expressed exponentially with three arbitrary height ranges having decay constants K_L, K_M and K_U.

2.3 IRI Model

The IRI model[5] is based mainly on empirical data. For the topside ionosphere, analytical expressions are used based on topside profiles from Millstone Hill and Arecibo, middle and low latitude stations respectively. Exponential relation is used for the bottomside F2 region. An empirical formula is used for the electron density corresponding to foF1. A parabolic supplement is added to the F2 profile for the F1 region. For a continuous transition to the upper limit of the valley region, a concave parabola was introduced. An analytical expression for the valley shape to approach the typical profiles recorded by incoherent scatter radar from Malvern is used for midlatitudes. At low latitudes near noon no valley is usually noticed. The maximum electron density of layer E was taken from NOAA-ITS mapping program. A parabolic transition is then used from NmE to NmD for daytime but an exponential with parabolic argument is used for nighttime. The D region profiles are based upon Mechtly's compilation[6].

3. Present Modelling Technique

The method described in this paper is based on the one proposed by Somayajulu et al[7] for building electron density profiles upto 1000 km by making use of the bottomside ionograms and the electron content measurements. Certain improvements are incorporated in the present paper.

The technique of building the electron density profiles(Fig.1) is as follows:

3.1 Bottomside Profile

3.1.1 Height and Electron Density of F2 Region Peak

The first step is to choose appropriate values of hmF2 and NmF2 for a given set of solar activity conditions to construct the bottomside profile. The bottomside profile is developed with the help of height of the F2 region peak hmF2 and its peak electron density NmF2. The height of the F2 region peak, hmF2 is obtained either from ionograms or from the predicted values of M(3000)F2 published in standard ionospheric data. The M(3000)F2 factor itself is derived from the ionograms. Using M(3000)F2, hmF2 is obtained by using the Appleton-Beynon formula

$$hmF2 = 1346.92 - 526.40M + 59.82M^2 \quad ..(1)$$

where M = M(3000)F2 factor

$$= \frac{MUF(3000)F2}{foF2}$$; foF2 the F2 region critical frequency for a range of latitudes and longitudes and over several solar cycles are available from published ionospheric data.

Making use of foF2, the peak electron density N_mF_2 is obtained as

$$N_mF_2 = 1.24 \times 10^{-2} (foF2)^2 \text{ -- } (el/m^3) \quad .. \quad (2)$$

where foF2 is in Hz.

3.1.2 Electron Density Distribution Around F2 Region Peak

Martyn[8] concluded theoretically that part of the profile above F2 peak should be alpha-Chapman. Nisbet and Bowhill's [9] results also support this conclusion. Based on the rocket data, Seddon[10] proposed a model for quiet ionosphere according to which F2 region can be represented by an alpha-Chapman layer during daytime from one scale height (Hm) below hmF2 to about 1.5 Hm above the peak and during nighttime from two scale height below hmF2 to 1.5 scale height above hmF2. Thus, we have,

$$N = N_mF_2 \cdot Ch(z) \text{ where } \begin{cases} -1 \leq Z \leq 1.5 \text{ for day} \\ -2 \leq Z \leq 1.5 \text{ for night} \end{cases} \quad (3)$$

where $Z = \frac{h - hmF2}{Hm}$ and Ch(z) is the Chapman function given by $\exp(1/2(1-z-e^{-z}))$. The parameter Hm is taken from CIRA[11,12].

3.1.3 D and E Region Profiles

From one scale height below hmF2 down to E region, the profile for daytime is represented by

$$N = Nm(E) \cdot e^{(h-110)/S} \quad .. \quad (4)$$

where $Nm(E) \leq N \leq 0.7\ NmF2$, $110 \leq h \leq (hmF2 - Hm)$.

The parameter hoE is taken to be 110 km. The parameters Nm(E), and S are functions of local time, latitude and solar activity. From the analysis of past data from several ionospheric observations, the values of NmE, and S can be predicted for a given set of conditions[13].

3.1.4 D Region Profile upto Base of E Region

The D region profile upto the base of E region is built by using model based on electron density profiles derived from a series of rocket flights from Thumba[14] covering a complete range of X from 6^0 to 86^0.

3.1.5 Comparison of Bottomside Profile Using Present Model with Actual Profiles Reduced from C^4 Ionograms

The exponential fit given by equation 4 between 110 km to the height (hmF2-Hm) agrees closely with the true height profile derived from actual ionograms. Therefore, an exponential profile from 110 km. to the height (hmF2-Hm) would be a valid representation of bottomside region. This would not require the time consuming reduction of the ionograms to true height profiles.

3.1.6 Electron Content upto 1.5 Hm above hmF2

The value of electron content from the ground upto a height of 1.5Hm above hmF2 is then obtained by integrating Equation (3).

3.1.7 Scale Height Near the F2 Peak

Scale height near the F2 peak is taken from CIRA 1972 for the neutral constituents according to Nisbet[1]. The scale heights obtained from ionograms at Delhi are found to show a reasonably good agreement with those taken from CIRA 1972 as shown in Fig.2

3.2 Topside Profile

In order to derive the electron densities in the topside ionosphere the usual assumption is made that diffusive equilibrium exists for the major ionic constituent above the reference altitude which in our case is hmF2 + 1.5Hm. The major ionic constituent in the topside region is O^+ upto at least 1000km during the day for maximum solar activity conditions. However, for low solar activity conditions the transition from O^+ to lighter ionic constituent, usually H^+, occurs below 1000 km. At night, for all solar activity conditions the transition altitude lies below 1000 km. In the topside ionosphere, the electron temperature(Te) the ion temperature(Ti) and the mean ionic mass vary with height, thereby making scale height vary with height. So the computation of electron density distribution becomes quite complex and cumbersome. Hence a simpler approach is adopted which yields reasonable electron density distribution in the topside ionosphere.

The electron density in the topside above the reference height is assumed to fall off exponentially as

$$N = No\ e^{-kh} \quad ... \quad ... \quad (5)$$

where h is the height measured from the reference height hmF2+1.5Hm and No is the corresponding electron density (No=0.7NmF2).

In order to take into account the transition from heavy-to-light ionic constituent in the topside, two decay constants KL and Ku are used in equation (5) to derive the electron density distribution. Thus the electron density profile is built from (hmF2+1.5km) to the transition height using an exponent KL and using Ku above this height to 1000 km in equation(5). The values of the exponents KL and Ku are chosen such that the topside electron content obtained from the model profile matches the average electron content data appropriate for the location and period. Thus the decay constants take into account the overall effect of the values of Te and Ti and the mean molecular masses of the mixture of O^+ and light ion at the respective altitudes for the corresponding solar activity, season and diurnal hour. In addition, use is made of the electron density data at 1000 km observed by satellite probes [15-18] for matching the model profile at 1000 km altitude. Electron content contours [19-21] used in building the profiles are shown in figures 3&4.

3.2.1. Choice of the Transition Height

The transition height in the topside ionosphere, i.e., where O^+ and H^+ species have equal number density, depend upon the ionic temperature and ratio of $n(H^+)$ and $n(O^+)$ [22,23]. If $no(H^+)$ and $no(O^+)$ represent ion densities at the reference level then

$$\frac{n(H^+)}{n(O^+)} = \frac{no(H^+)}{no(O+)} \cdot \exp g_o \int \frac{(mi(O^+)-mi(H^+)}{KTi}\, dz' \quad ..(6)$$

where Z'=geopotential altitude and go= acceleration due to gravity at the reference height. If Ti is constant with altitude, the transition height is given by [24].

$$Z' = Zo + \frac{(1+\frac{Te}{Ti})XHo(0)}{\frac{1+\frac{1}{To}\frac{no(H^+)}{no(O^+)}}{1+\frac{no(H^+)}{no(O+)}}} \cdot \ln \frac{no(O^+)}{no(H^+)} \quad ..(7)$$

For high solar activity conditions, the transition altitude during daytime lies above 1000 km. For low solar activity conditions however this lies below 1000 km. Using the electron temperatures from Mahajan[25] and ion temperatures from Risal Singh et al [26] and number densities of electrons and hydrogen ions from Bhatnagar et al [27], transition heights are evaluated for all the stations over India. It is found that the transition height during solar minimum daytime conditions varies from 700-850 km. Therefore a constant transition height of 750 km is used in the present profile building. However the change in transition altitude is ±10%. The corresponding change in decay constant in the upper region is very small since the $n(H^+)$ value from 750 km to 850 km is nearly the same as well as the electron temperature more or less constant at that altitude range.

3.2.2 Matching the NmaxE , Bottomside and Topside of alpha-Chapman Layer

The shapes near the NmaxE and bottomside and topside of the alpha-Chapman layer and exponential topside profiles are matched by using a cubic interpolation of electron density as a function of height between profile points.

4. Results and Discussion

4.1 Computed Profiles

Electron density profiles upto 1000 km. altitude are generated for low and high solar activity conditions, to illustrate the profile building technique described in the previous sections. These profiles have been built for the locations of Delhi, Ahmedabad, Calcutta, Hyderabad and Kodaikanal covering a geomagnetic latitude range of 20°N to the equator. These are shown in figs. 5 to 8. In figs. 9 and 10 the computed profiles are shown alongwith electron density profiles derived from Alouette [28, 29] ionograms for similar solar activity conditions and latitudes. Due to non-availability of published profiles for Indian longitude range, profiles for Western hemispheric longitudes are used. For this reason the NmF2 values are seen to differ widely. Therefore, the Alouette profiles are matched at the reference altitude to No in order to compare the shapes of the topside profiles. It should be noted that there are day-to-day irregular variations in the ionosphere in NmF2, hmF2 as well as the electron and ion temperatures. Therefore one would expect a considerable scatter in the shape of the topside profiles. This is shown by the shaded area. It should also be remembered that the model profiles generated are average quiet time profiles.

In view of these considerations, the agreement between the model profiles and the Alouette profiles is considered to be good. The K-values defining the model profiles in the topside are tabulated in tables II to V.

Table I

Station	Geomagnetic Latitude	Geographic Latitude
Delhi	19°11'N	28°38'N
Ahmedabad	14°01'N	23°01'N
Calcutta	12°15'N	22°58'N
Hyderabad	07°39'N	17°21'N
Kodaikanal	0°44'N	10°14'N

Table II

K Values for Solar Minimum Summer Daytime

Station	K_L (cm)	K_U (cm)
Delhi	8.23×10^{-8}	3.42×10^{-8}
Ahmedabad	7.34×10^{-8}	3.39×10^{-8}
Hyderabad	6.58×10^{-8}	3.36×10^{-8}
Kodaikanal	7.30×10^{-8}	3.4×10^{-8}

Table III

K Values for Solar Minimum Winter Daytime

Station	K_L (cm)	K_U (cm)
Delhi	9.32×10^{-8}	3.45×10^{-8}
Ahmedabad	8.10×10^{-8}	3.4×10^{-8}
Hyderabad	7.86×10^{-8}	3.38×10^{-8}
Kodaikanal	7.44×10^{-8}	3.35×10^{-8}

Table IV

K Values for Solar Maximum Summer Daytime

Station	K(cm)
Delhi	6.41×10^{-8}
Ahmedabad	6.35×10^{-8}
Calcutta	7.73×10^{-8}
Hyderabad	$\cdot19\times10^{-8}$
Kodaikanal	$\cdot\cdot\times10^{-8}$

Table V

K Values for Solar Maximum Winter Daytime

Station	K(cm)
Delhi	6.76×10^{-8}
Ahmedabad	6.67×10^{-8}
Calcutta	6.54×10^{-8}
Hyderabad	6.47×10^{-8}
Kodaikanal	6.25×10^{-8}

4.2 Comparison with Bent's Model

The principal difference between the present model and Bent's model is in the topside profile. Bent divides the topside exponential profile into three arbitrary height ranges and adjusts the decay constant K to match the observed profile. In our work, the topside region is divided, on physical grounds, into two altitude ranges, the lower one corresponding to heavy ion (O^+) as the dominant constituent and the upper one corresponding to lighter ions as the dominant constituents. The transition height between the lower and upper region is chosen based on the experimental data on the transition altitudes from O^+ to lighter ions as discussed earlier. In addition to foF2 data, we also have used electron content data as a constraint on the profile.

Another important difference is that in Bent's model, he divides the whole latitude range into 3 groups, 0° to 30°, 30° to 60° and 60° to 90°. It is well known that in the 0° to 30° latitude range, there is a strong latitudinal dependence of ionosphere due to the presence of equatorial anomaly. In our model, the latitudinal dependence at low latitudes is explicitly introduced in the profile building.

4.3 Comparison with International Reference Ionosphere(IRI) Model

The IRI model gives only 18°N (geomag.) latitude profile of Arecibo as representative of low latitude profiles. Delhi latitude being 19°11'N (geomag.) is compared with above profile. It is found that NmF2 values used in IRI model for the particular solar activity conditions are substantially lower than those typically observed at Delhi. Therefore, for comparing the shape of the profiles, the NmF2 given in IRI model are matched with the Delhi values and the resulting profiles are compared in Figs. 11-12. It may

be seen that there is no significant difference in the bottomside profile except near NmE peak. We have taken the foE at 110 km to fix electron density at 110 km. The discrepancy may be due to the layer structure of E-region and perhaps can be reduced if a true height profile from ionogram is utilized. The main difference comes in the topside region. The shapes in both the models are somewhat similar except that the actual values of IRI model are substantially lower. Thus in Indian Subcontinent as well as for low latitude regions it is more appropriate to use our model than the IRI model.

5. Limitations of the Model Profiles

The electron density distribution in the topside ionosphere is principally controlled by two factors:
(1) the ion-and electron temperatures and their gradients;
(2) the relative abundances of light-to-heavy ionic constituents.

of these two the latter is more important. The major ionic constituent at the reference altitude is taken to be O^+ and the light ionic constituent is assumed to be H^+ throughout the range. The presence of helium is deliberately ignored because when He^+ becomes the dominant constituent it appears only in a very thin layer(in comparison to H^+ scale height). Thus neglecting the role of helium ions has very little effect on topside electron density profiles under most conditions and results primarily in reducing the sharpness of the O^+ to H^+ transition [30,31]. The relative abundance ratio $\eta(=n(H^+/n^{(O+)})$ can vary widely between 10^{-4} to 10^{-2}. Using the relationship: $n(O^+)/n(H^+)=(9/8)(n(^O)/n(H))$ the relative abundance ratio n is estimated from the neutral O and H densities at the lowest altitude available in CIRA(1972) model. The error involved in this procedure cannot be estimated easily but is considered to be reasonably small.

The relative abundance ratio η, the ion temperature Ti and the ratio Te/Ti determine the heavy-to-light ion transition altitude(see eqn.7). In our modelling Ti is assumed to be constant with altitude. Temperature gradients of Ti at night and during low solar activity conditions are considered to change the assumed transition altitude only marginally. However, during daytime under high solar activity Conditions the change in the transition height is considered to produce about 10 to 15% changes in electron densities above 750 km. For some applications analytical representation of the entire profile is desirable[32].

References

1. Nisbet,J.S(1971),Radio Sci.,6, 437.
2. Cookingham,R.E(1972),Modelling the bottomside electron density profile, AFCRL-72-0340.
3. Bent,R.B., LLewellyn, S.K., Nesterzvk, G. and Schmid, P.E(1975), Proc. Symp. Effects of Ionosphere on Space Systems and Communications, USA,559.
4. Bent, R.B., Llewellyn, S.K. and Schmid, P.E(1972), Space Research XII, 1208.
5. Rawer, K., Ramakrishna, S. and Bilitza, D(1975), IPM-Scientific Report W.B.2, Freiburg(FRG).
6. Mechtly, E.A. and Bilitza, D(1974), International Report 1 PW, Freiburg.
7. Somayajulu, Y.V., Tyagi, T.R. and Bhatnagar, V.P(1965),Space Res.V, 641.
8. Martyn,D.F(1956),Austr.J.Phys9,161.
9. Nisbet,J.S.,Bowhill,S.A(1960),JGR,65,3609.
10. Seddon, J.C(1963),JGR, 68, 1339.
11. CIRA(1965).
12. CIRA(1972).
13. Ionospheric Data, RSD, N.P.L.
14. Somayajulu, Y.V., Mitra, A.P(1977), under preparation.
15. Brace, L.H., Reddy, B.M., Mayr,H.G. (1967), JGR, 73, 265.
16. Chan, K.L., Colin, L(1969), Proc. IEEE, 57, 990.
17. Matuuara, N. and Ondoh, T(1969), Proc. IEEE, 57, 1150.
18. Jelly, D.H. and Petrie, L.E(1969), Proc. IEEE, 57, 1005.
19. Somayajulu, Y.V., Tyagi, T.R., Ghosh, A.B.(1972), Proc. Symp. Future Applications of Beacon Satellite, Graz, 289.
20. Tyagi, T.R., Ghosh, A.B., Mitra, A.P. and Somayajulu, Y.V(1972), Space Res., XII, 1195.
21. Ghosh, A.B(1976), Ph.D. thesis, University of Delhi, Delhi.
22. Bauer, S.J(1966),JGR, 71, 1508.
23. Bauer, S.J(1963), Nature, 197, 36.
24. Goel, M.K., Rao, B.C.N., Chandra,S. and Maier, E.J(1976), JATP, 38,389.
25. Mahajan, K.K(1977), JATP, 31, 637.
26. Risal Singh and Rao, B.C.N(1974), IJRSP, 3, 273.
27. Bhatnagar, V.P. and Mitra, A.P(1966), Proc. IQSY Symp., 282.
28. Alouette I observations recorded at Hawaii, NASA SP 3038(1967).
29. Alouette-II observations over Japan 2, Radio Research Laboratories, Tokyo, Japan(1971).
30. Carlson, H.C. and Gordon, W.E(1966), J. Geophy. R. 71, 5573.
31. Titheridge, J.E(1972), Planet. Space Sci., 20 , 353.
32. Booker, H.G.(1977),JATP,39,619-623.

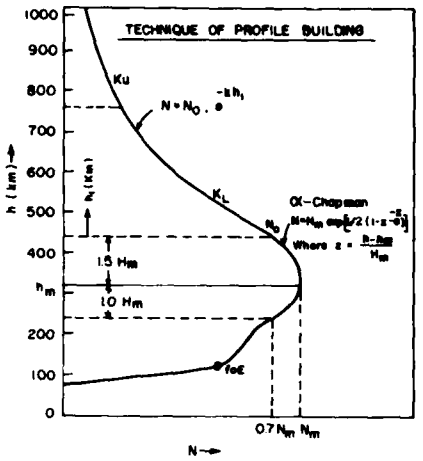


Fig.1-Present technique of profile building

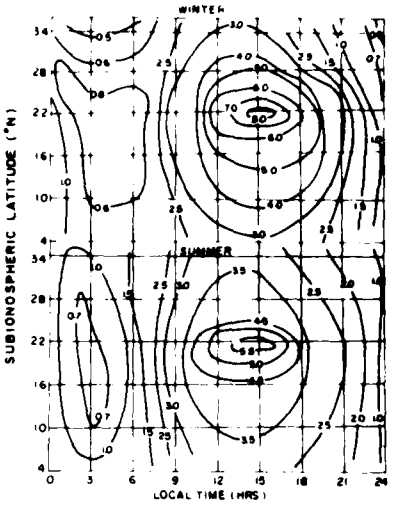


Fig.3 Electron content contours over the Indian Subcontinent for high solar activity conditions.

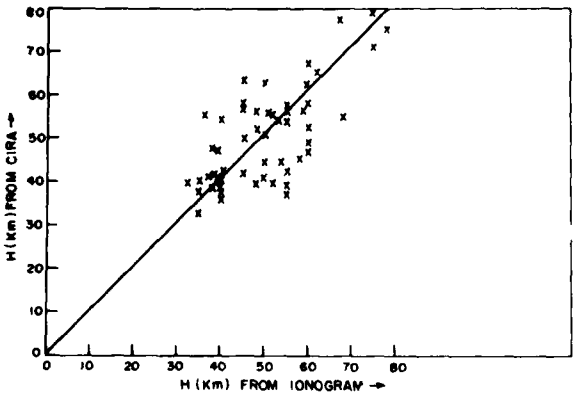


Fig.2 Comparison of H from CIRA with H reduced from bottomside ionogram.

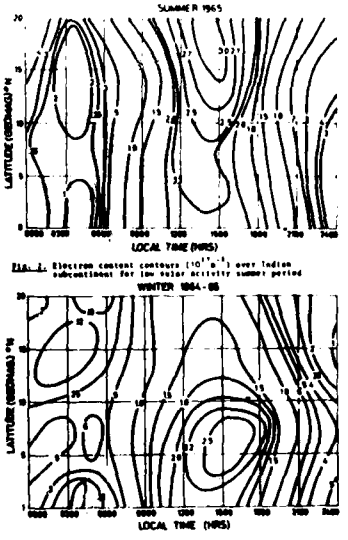


Fig.4 Electron content contours over Indian Subcontinent for low solar activity conditions.

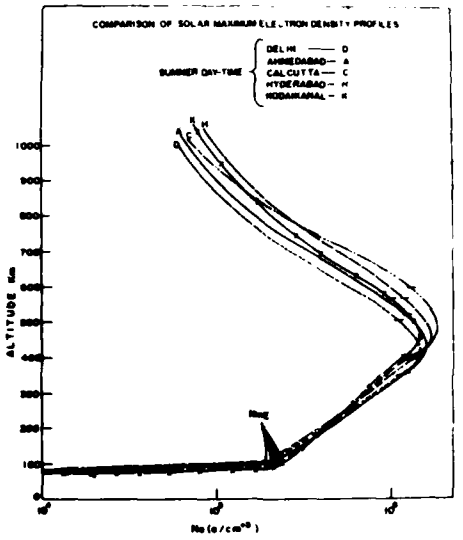


Fig.5 Electron density profiles for solar maximum Summer Daytime conditions

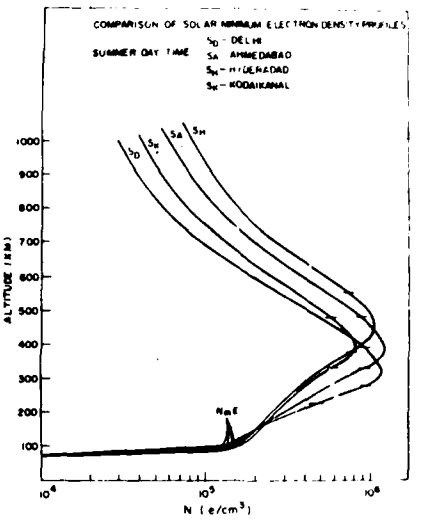


Fig.7 Electron density profiles for solar minimum summer daytime conditions

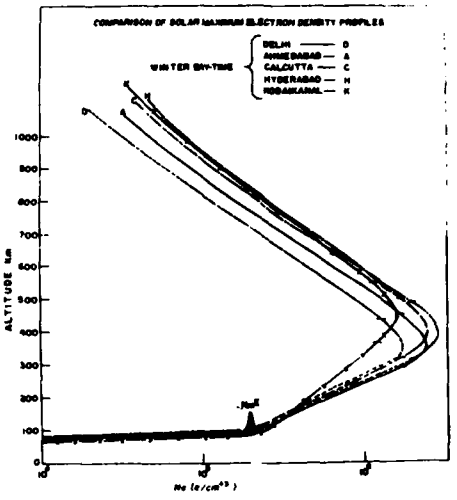


Fig.6 Electron density profiles for solar maximum winter daytime conditions.

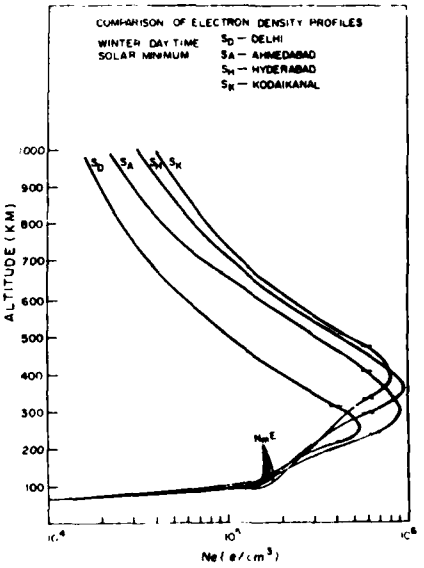


Fig.8 Electron density profiles for solar minimum winter daytime conditions

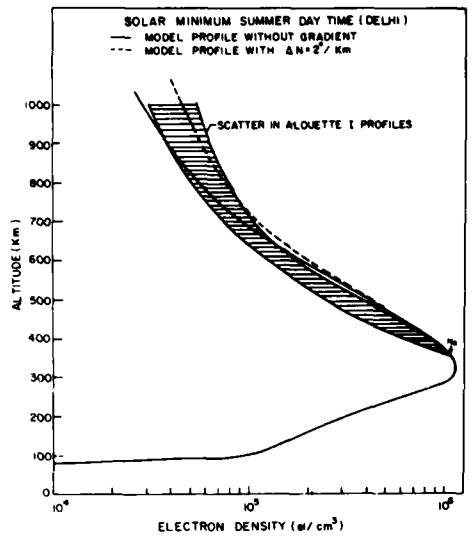


Fig.9 Comparison of present model with gradient based model and Alouette I profiles for Delhi

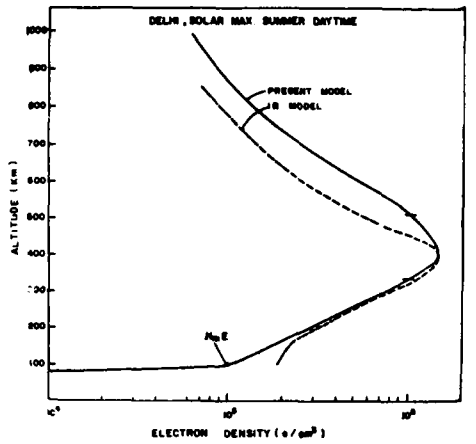


Fig.11 Comparison of present model with IRI model for solar maximum summer daytime

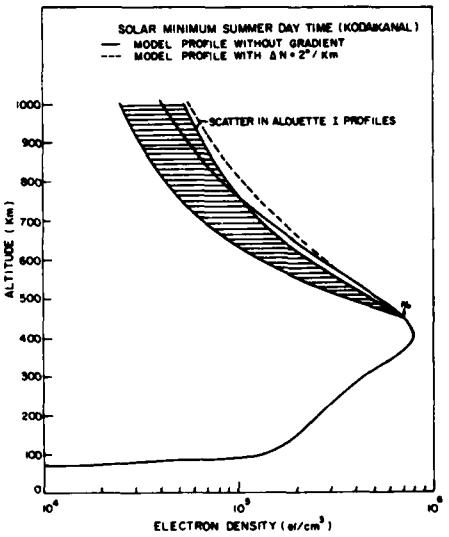


Fig.10 Comparison of present model with gradient based model and Alouette I profiles for Kodaikanal

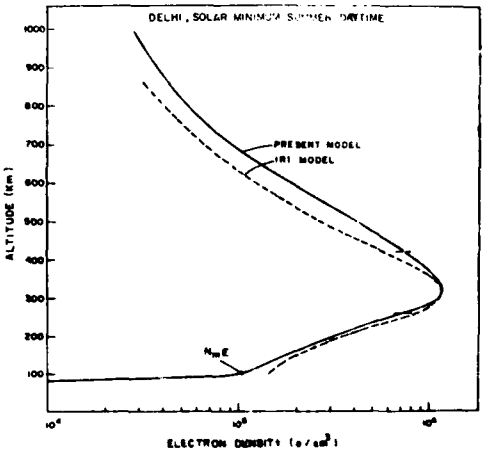


Fig.12 Comparison of present model with IRI model for solar minimum summer daytime

TOPSIDE IONOSPHERIC TROUGH MORPHOLOGY AT MID- AND HIGH-LATITUDES

M. Ahmed
Regis College
Weston, MA 02193

and

R.C. Sagalyn
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

ABSTRACT

The main trough in the topside ionosphere has been studied using the thermal positive ion and electron densities measured over a three year period (1969-1972) by means of spherical electrostatic analyzers aboard the ISIS I and INJUN V satellites in the 560-3600 km altitude range. The trough is found to be a persistent feature at night with an occurrence frequency of approximately 95%. The occurrence frequency decreases to approximately 60% near the dawn-dusk meridian and to approximately 48% near local noon. At altitudes below about 1500 km during quiet to moderate conditions ($Kp \leq 3$) the trough equatorward boundary is found at $L = 3.5 \pm 0.5$ near midnight and $L = 12.5 \pm 1.0$ near local noon. The trough equatorward location on the nightside is in good agreement with published plasmapause locations. Near local noon the trough occurs at the equatorward edge of the magnetospheric cleft and is at significantly higher L values than those reported for the dayside plasmapause. The seasonal variation of the trough location at a given local time is negligible except near sunrise.

With increasing altitude between 1500 and 3600 km, the equatorial boundary of the trough moves to continually lower latitudes during the night hours. The equatorward trough wall becomes a dominant feature of the trough often extending from 15^{o} to 20^{o} in width during quiet magnetic periods. The poleward edge of the trough becomes less well marked with increasing altitude often being defined only by a sharp spike in ionization extending over a few degrees within the auroral zone. The amplitude of the dayside high latitude trough reduces gradually with increasing altitude. However, a second region of dayside plasma depletion is observed between L = 2 and 6 approximately 50% of the time. The equatorward wall of this depletion region probably represents partial flux tube filling in the outer plasmasphere.

INTRODUCTION

Mid-latitude troughs or depressions in the F layer ionization were first reported by Muldrew (1965) and Sharp (1966). Further studies of trough characteristics have been carried out by a number of workers including Miller and Brace (1969), Rycroft and Thomas (1970), Tulunay and Sayers (1971), Taylor et al. (1975) and Grebowsky et al. (1976). Statistical studies of the relation of the trough to the plasmasphere (Thomas and Andrews, 1968, Tulunay and Sayers, 1971, and Grebowsky et al. 1976), have shown that mid-latitude charge density depletions occur near the same L shells as the average plasmapause position on the nightside. Rycroft and Thomas (1970) and Tulunay and Sayers (1971) established that the trough position varies with Kp in a manner similar to the plasmapause dependence upon

magnetic activity. Taylor et al. (1974) have shown that the trough exists in the individual ion species H^+, He^+ and O^+.

Studies of the diurnal, seasonal and altitude variation of the trough characteristics have led to confusing results (Wrenn and Raitt, (1975). Miller (1974), for example, reported that dayside troughs persisted above 2500 km while below they were seldom detectable. There are several reasons for the differences in trough results. These include use of a limited data base, widely varying altitude of the measurements, varying spatial resolutions of the measurement from a few km to nearly 1000 km, difficulties in detecting dayside features due to photoelectrons as well as varying criteria used to define the trough location. Tulunay and Sayers (1971) for example, used the trough minimum, while Brace and Theis (1974) used a density level of 10^3 electrons/cm^3 along the equatorial trough wall.

The present study seeks to contribute new knowledge of the trough characteristics as a function of local time, season and altitude. The trough occurrence frequency is also presented as a function of local time. The study is based on the examination of thirty-seven months of ISIS I positive ion and INJUN V electron probe data. The relation of the trough to the plasmapause is examined at all local times.

THE EXPERIMENT

ISIS I thermal charged particle measurements were obtained with a spherical electrostatic analyzer that measured the thermal ion density, the ion energy distribution from 0 to 50 eV, the satellite potential, and the ratio of the ion mass to the ion temperature. The sensor mounted on a 96 cm boom, consists of three concentric spherical electrodes with radii of 1.90, 2.54 and 3.18 cm respectively. The operation of the probe is based on the motion of charged particles in a central force field (Sagalyn, et al., 1963, Sagalyn and Smiddy, 1967, Smiddy and Stuart, 1969, Whitteker et al., 1972). The sensitivity range of the instrument is 10-10^6 ions cm^{-3}, in sunlight, however, photoelectron currents limit the lower sensitivity to about 700 ions cm^{-3}. Ion densities were sampled 60 times per second, corresponding to a spatial resolution of 150 m. The ratio of mass to temperature was sampled once per minute, and the energy distribution was samples once every 2 min. The results in this report are based on the ISIS I ion density mode of operation and also on measurements made with a two electrode spherical probe flown on INJUN V with a spatial resolution of 2 km.

DATA BASE

The ISIS I satellite served as an ideal platform for the study of diurnal and seasonal morphology of the trough, since its orbital plane precessed nearly one hour of local time per week, providing a complete 24 hour coverage in about 3 months. Data obtained within 45 days of the equinox and solstice periods were utilized for the seasonal study.

The statistical portion of the study reported here utilizes thermal positive ion density measurements on about 12,000 ISIS I orbits between Febraury 1969 and March 1972. The measurements have a spatial resolution of 150 meters and are well suited for the study of trough characteristics. In addition, thermal ion and electron densities measured by similar instruments aboard the INJUN V satellite on about 3000 orbits between November 1968 and November 1970 were used to improve the data base within a few hours of local noon. The data cover the altitude range 550-3600 km and all local times and seasons.

For the diurnal studies of trough characteristics about 20 to 25 clearly identifiable troughs were available in each hour of local time in each season. Seasonal variations were deduced utilizing data collected in three successive years. Northern hemisphere data obtained during magnetically quiescent and slightly disturbed conditions i.e. Kp $\leq$ 3 were used in the statistical portion of this study.

RESULTS BELOW 1500 km

a. Definition of Trough Parameters

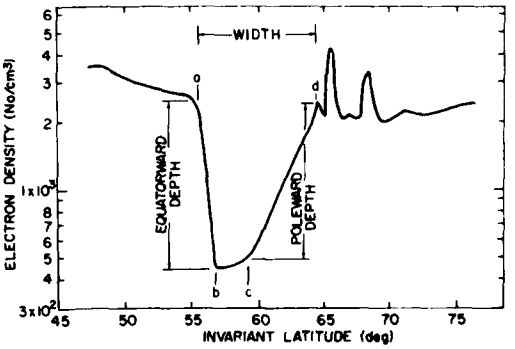
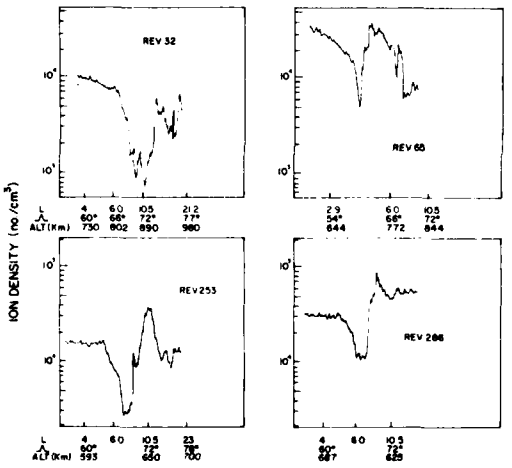


Fig. 1. Example of low altitude trough (<1500 km) where points 'a' and 'b' denote the location of top and base of the equatorial edge respectively. Points 'c' and 'd' denote the location of the base and top of the poleward edge.

A representative low altitude trough is shown in Fig. 1. Point 'a' denotes the location of the trough equatorward edge. It is the point marking the onset of a major decrease in ionization density with increasing latitude.

'b' denotes the base of the equatorward wall, 'c' the base of the trough poleward wall. The top of the poleward edge, point 'd' is the intersection of lines drawn along the trough poleward wall and through the first ionization maximum in the precipitation region. The ratio of the densities at points a/b and d/c represent the trough amplitudes at the equatorward and poleward edges respectively.

While there is considerable variation in the trough amplitude, width and gradient from orbit to orbit as illustrated in Figs. 2, 3 and 4, points 'a' through 'd' can usually be readily measured for altitudes below about 1500 km. The widths of the equatorward and poleward walls typically range from 1^{o} to 4^{o}. The high latitude precipitation region extends over several degrees and is highly structured The trough width and amplitude are consistently smaller on the dayside than on the nightside as illustrated in Figs. 4 and 2, respectively. The equatorward edge of the trough wall point 'a' of Fig. 1, was taken as the trough location for the study of results below 1500 km.



MLT. = 22 hrs $\pm$ 30 min

Fig. 2. Nighttime low altitude (<1000 km) trough examples.

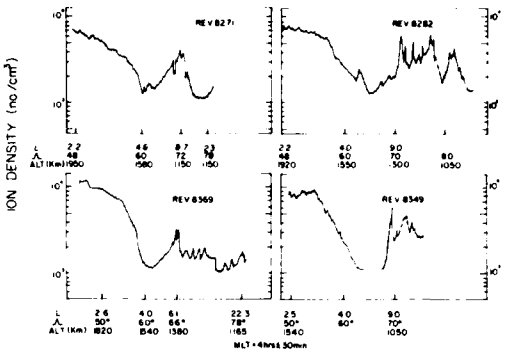


Fig. 3. Examples of troughs between 1000 and 1600 km within 4 hours of midnight.

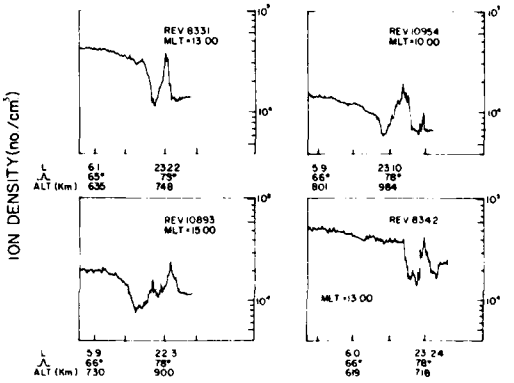


Fig. 4. Low altitude (<1000 km) daytime trough examples within 3 hours of noon.

b. Trough Occurrence Frequency

The trough occurrence frequency as a function of local time is given in Fig. 5 for altitudes below 1500 km.

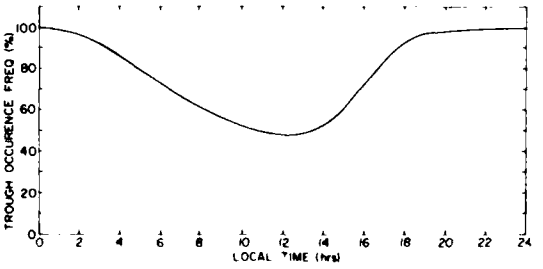


Fig. 5. The trough occurrence frequency versus local time for altitudes $\leq$1500 km.

For each hour of local time, data from at least 3 months were utilized. For example, the results at 19:00 LT includes data from March, August, September 1969, January and February 1970 giving a total of 65 observations. The results are based on data from all seasons and include about 830 clearly identifiable trough observations.

In the nighttime hours, 19:00 to 05:00 LT the trough occurrence frequency is very high, approximately 96% as seen in Fig. 5. Near 05:00 LT a sharp decrease in the occurrence frequency is observed. After 08:00 LT there is a gradual decrease in occurrence frequency with time reaching a minimum value of approximately 48% near local noon. At about 13:00 LT the occurrence frequency begins to increase reaching the maximum nighttime value at 19:00 LT. It is seen that the afternoon increase in the occurrence frequency is somewhat steeper than the morning decrease.

The results of Fig. 5 represent the first quantitative determination of trough occurrence frequency over a 24 hour period. These results are in agreement with the findings of Tulunay and Sayers (1971) which show a broad maximum during the night hours and a minimum within $\pm$ 3 hours of local noon in the northern winter hemisphere. Taylor et al. (1970) and Tulunay (1973) have reported the existence of persistent daytime troughs. Miller (1974) found no daytime troughs below 2500 km from the ISIS I electron probe measurements. Miller's ISIS I sensor had a spatial resolution of 9^{o} and hence could not resolve the high latitude troughs with widths typically ranging from 2^{o} to 6^{o}.

Brinton et al (1969 and 1970) have shown that the ion composition in the topside ionosphere at mid and high latitudes consists predominantly of O^{+} on the dayside. The midday occurrence frequency of 48% shows that the production of O^{+} by solar UV radiation frequently dominates over ion depletion mechanisms.

c. Trough Location

Fig. 6 shows the mean location of the trough equatorward edge (point 'a' of Fig. 1) versus local time for each of the four seasons. Each of these four profiles is based upon at least 400 well defined troughs. It is seen that the summer profile is symmetrical within $\pm$ 4 hours of midnight and is located between L = 3.3 and 4.0. In the morning between 04:00 and 07:00 LT, there is a sharp poleward movement with an increase of 3 L units. In the dawn-noon sector, the summer profile gradually moves to higher L values reaching a maximum of L = 12.5 $\pm$ 1.0 around 11:30 LT. Within 4 hours of local noon, the trough is found to be located at the equatorward edge of the cusp precipitation region. In the afternoon sector, the trough moves gradually equatorward to a value of L = 5 near dusk. The summer trough location profile is nearly elliptical in shape except for the LT period, 04:00 to 07:00 LT.

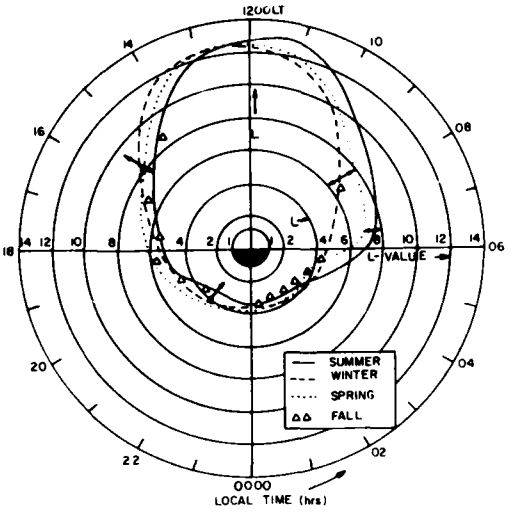


Fig. 6. Diurnal variation of trough location below 1500 km for the four seasons. The bars are the standard deviation of the measurements.

The winter trough location moves gradually inward from L = 5.5 to 3.5 between 18:00 and 04:00 LT. This equatorward movement at night is consistent with the characteristic nighttime movement of the plasmapause in the equatorial plane (Carpenter, 1966). In the 04:00 to 07:00 LT sector, the winter profile shows a gradual movement toward higher L values until 09:00 LT after which a rapid poleward movement is observed locating the trough at its peak value of about L = 12.5 at noon. The rapid movement after 09:00 LT represents the effect of the onset of sunrise in the winter hemisphere. In the noon-dusk sector, the winter trough profile moves gradually from L = 12.5 to 5.5. The winter profile is also elliptical in form.

The spring and fall trough profiles shown in Fig. 6 are similar in form to the summer and winter profiles and generally lie within the envelope defined by the summer and winter locations throughout a 24 hour period. The profiles for all the four seasons approach their highest L values near noon.

The standard deviations of the measured trough locations are shown as bars in Fig. 6; they were deduced using all trough data available in a specific hour of LT in a given season. They are of the order of $\pm$ 1 L unit

during the daytime and ± 0.4 L near midnight.

The standard deviation calculations show that there is no significant seasonal variation in the trough location except in the 04:00 to 09:00 LT sector, where the effect of the onset of sunrise with season is large.

d. Comparison with Other Trough Measurements

Since the seasonal effects are small, the mean trough location over all seasons was calculated for a comparison with published trough data (Fig. 7). The trough location deduced by Muldrew (1965) from f_oF_2 observations from the Alouette 1 Topside Sounder from September 1962 to March 1963 is also shown in Fig. 7.

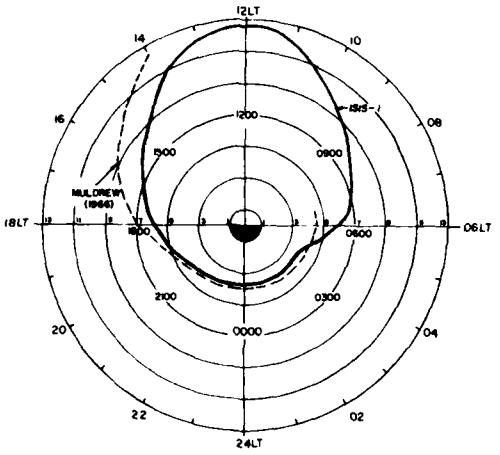


Fig. 7. Comparison of the diurnal variation of the trough location averaged over all seasons with the trough measurements of Muldrew (1965) (<1500 km).

Between 14:00 and 18:00 LT the two profiles have similar shapes although Muldrew's curve is located 1 to 2 L units higher. From 18:00 to 05:00 LT, the two profiles follow one another very closely. It should be noted that there is no "evening bulge" in the trough location in the two results as has been observed in the equatorial plasmapause location by Carpenter (1966), Chappell (1972) and others. The results are also in agreement with the low altitude trough studies of Tulunay and Sayers (1971) and Tulunay and Grebowski (1975) on the nightside. The last two investigations were confined to regions of L < 7 and hence the high latitude daytime troughs would be outside the range of these investigations.

RESULTS ABOVE 1500 km

a. Nightside

The region of significant nightside ionization depletion extends to lower latitudes with increasing altitude. Above 2000 km the equatorial wall of the trough frequently extends 15°-20° in latitude with highly variable shapes. Fig. 8 shows some typical examples of trough density variations. It is seen that there may be a uniform decrease of density with increasing latitude or two or more changes in slope. The high latitude boundary becomes more difficult to define. The poleward edge of the depletion region is marked by one or several sharp spikes in the ionization, each having a width of 1°-2°. The nightside high altitude trough is found on over 95% of the orbits examined.

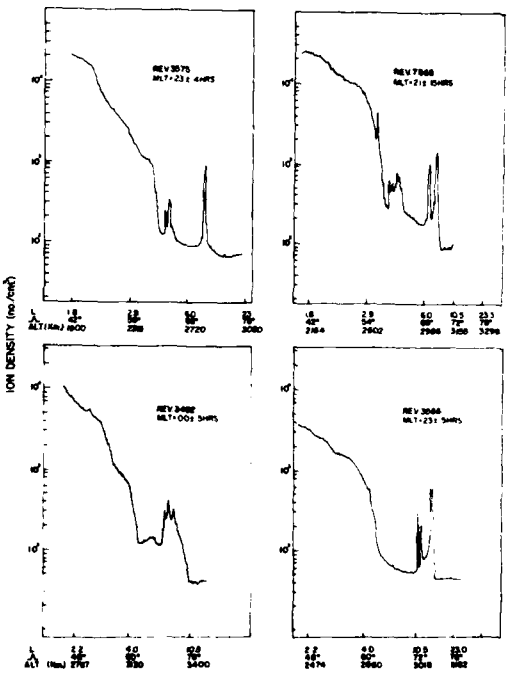


Fig. 8. Examples of nighttime trough measurements at altitudes greater than 2300km.

b. Dayside

Examples of thermal ion density observations near local noon at altitudes greater than 1500 km are given in Figs. 9 and 10. The amplitude of the high latitude trough (hereafter referred to as T_1), which was approximately 3 for altitudes less than 1500 (Fig. 4) gradually reduces to values less than 2 for altitudes greater than 2000 km (Fig. 9).

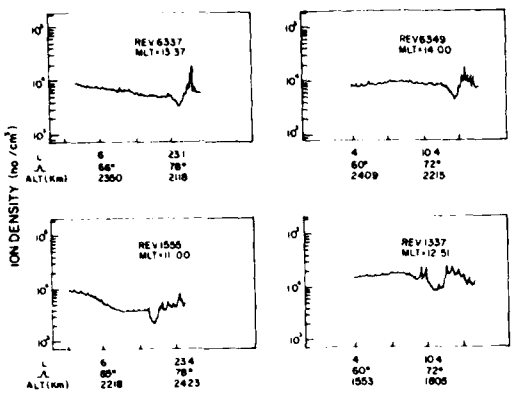


Fig. 9. Daytime high altitude trough examples of T_1 - the high latitude trough, within 2 hours of noon. The low latitude gradient T_2 is also seen on Rev. 1555.

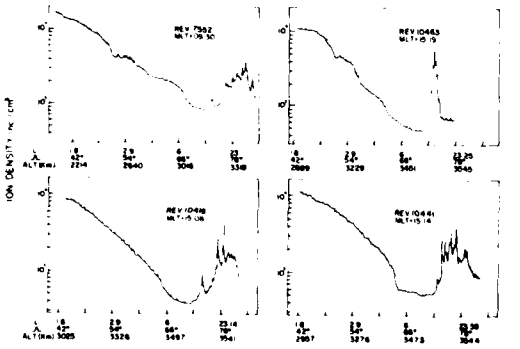


Fig. 10. Daytime trough examples between 2300 and 3500 km showing T_2, the low - mid latitude trough.

The frequent appearance of the smaller amplitude high latitude trough T_1 at altitudes above 1500 is consistent with the findings of Münch et al., (1977) based on INJUN V results. Simultaneous comparison of the trough densities with INJUN particles and electric field measurements showed that T_1 is located at the equatorial edge of the cusp where there are found sharp electric field reversals which indicate shifts in the plasma convection velocity. Increases in the electron temperature by factors of 2 to 3 and a shift from isotropic high energy particles to lower energy anisotropic magnetosheath-like electrons and protons were also observed.

Schunk et al. (1975) have shown that the existence of horizontal electric fields and elevated electron and ion temperatures produce depletions of O^+, a major ion at high latitudes, by increasing the ion loss rate, increasing the ion scale height and by changes in chemical composition. The present results together with the INJUN measurements reported by Münch et al. (1977) strongly support the conclusion of Schunk et al. (1975) that electric fields and enhanced charged particle temperatures make a significant contribution to the formation of the dayside high latitude trough.

Significant daytime gradients are also found to develop between L = 2 and 6 with increasing altitude (Fig. 10). These lower latitude gradients or troughs which we will refer to as T_2, are observed on 50% of the dayside orbits. As on the nightside, the poleward edge of the daytime depletion region is usually marked by spikes in ionization. The transition from T_1 only to T_1 and T_2 type dayside structures occurs gradually in a transition region between 1300 and 2000 km. An example of a fully developed T_1 and T_2 type ionospheric structure is given in Fig. 11. Here the equatorward walls of T_1 and T_2 are found at invariant latitudes of 78.5 and 54°, respectively.

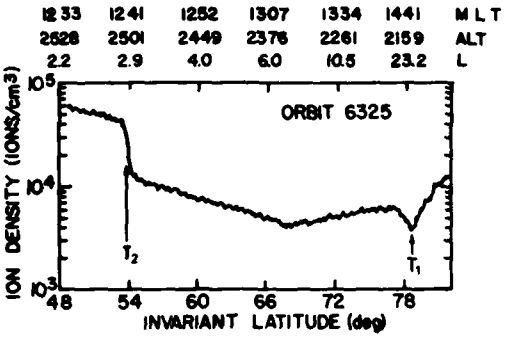


Fig. 11. An observation of both high and low latitude troughs, T_1 and T_2 respectively. Altitude 2100 to 2500 km.

The statistically derived location of the top and base of the "equatorward trough wall" of T_2, vs local time for altitudes above 2000 km is shown in Fig. 12. It is seen that on the nightside the top of the trough wall is consistently located close to L = 2.0 (invariant latitude ≈ 45°) and its base is found at L = 4.4 ($\Lambda \approx 61^\circ$). On the dayside the base of the ion depletion region, T_2, is found at slightly higher latitudes. At local noon it is located at L = 6.5 ($\Lambda = 67^\circ$). The equatorward wall is the dominant feature of the high altitude ionization depletion region.

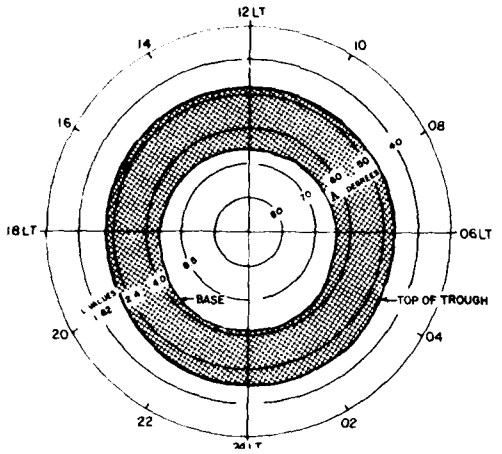


Fig. 12. Mean location of the top and base of the high altitude trough wall (T_2) versus local time for altitudes between 2000 and 3600 km.

As illustrated in Figs. 8 and 10 the ion density frequently changes by two orders of magnitude between the top and the base of the wall. This is much greater than the change in density below 1000 km, either day or night (Figs. 4 and 2). The high latitude trough wall T_2, of Fig. 12 corresponds to the plasmapause reported by Brace and Theis (1974) from ISIS I electron probe measurements above 2500 km. They identified the location of an ionization level of 10^3 electrons/cm^3 along the trough wall as the plasmapause, and obtained a nearly circular plasmapause boundary at about L = 4. As would be expected from the Brace and Theis plasmapause criteria their results lie in the shaded area of Fig. 12. T_2 also corresponds to the high altitude ISIS I dayside trough reported by Miller (1974). The T_2 type of ionization gradients were also observed by an electron sensor aboard the polar orbiting satellite OV3-1 in the 3000-5700 km altitude region between L = 2 and 6 (Bewersdorff and Sagalyn, 1971 and 1972).

We consider that the trough wall of Fig. 12 between about L = 2 and 6.5 is the result of partial filling and/or erosion of the outer plasmasphere. The combined effect of magnetic substorms (Park and Banks, 1974 and 1975) and of diurnal plasma depletions and replenishments prevent the outer shells of the plasmasphere from reaching saturation density.

DISCUSSION: TROUGH LOCATION AND THE PLASMAPAUSE

Nishida (1966 and 1967) first explained the formation of the plasmapause in terms of a magnetospheric convection model. Thomas and Andrews (1968), Aycroft and Thomas (1970), Tulunay (1972) and Tulunay and Grebowsky (1975) have all demonstrated a close relationship between the plasmapause and the mid-latitude trough at night. However, there are disagreements among these workers regarding the details of this association. Nishida and Thomas and Andrews, consider the equatorward trough edge to be the physically more significant part of the trough and have shown it to be correlated with the equatorial plasmapause, while Aycroft and Thomas (1970), Tulunay (1972) and Tulunay and Grebowsky (1975) have shown the trough minimum location to be correlated with the plasmapause. Tulunay and Hughes (1973) have shown that the location of the trough minimum is partially influenced by the location and movement of the auroral precipitation region at night and hence recommend the use of the more precisely measurable trough equatorward edge.

In Fig. 13 the trough location identified as the low latitude edge of the equatorward trough wall, for altitudes below 1500 km is compared with the plasmapause position deduced by Chappell et al. (1972), Carpenter (1966) and Taylor et al. (1970). At night the trough and plasmapause variations are very similar although the ISIS I trough locations are slightly equatorward of the plasmapause position. Thomas and Andrews (1968) obtained similar results. Since the trough equatorward wall is typically 1^o to 4^o wide, the equatorial trough base is located approximately 0.3 to 1 L unit poleward of the results shown. These results strongly suggest that the mid-point of the trough wall is an excellent indicator of the plasmapause during the night hours at low altitudes.

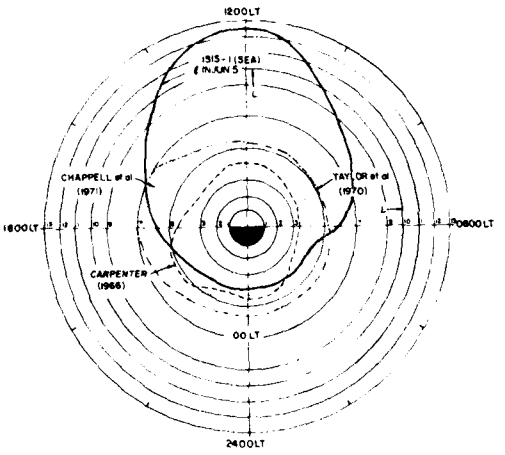


Fig. 13. Comparison of the mean location of the low altitude trough (<1500 km) versus local time with the plasmapause measurements of Carpenter (1966), Taylor et al. (1970) and Chappell et al. (1972).

The daytime trough location below 1500 km deduced from ISIS I and INJUN V measurements shown in Fig. 13 is considerably different from the reported plasmapause profiles. The ISIS I low altitude troughs are found at substantially higher latitudes ($L = 12.5 \pm 1^{o}$), just equatorward of the cusp precipitation region. The trough moves gradually to lower latitudes toward dusk while the plasmapause position lies between L = 4 and 7 from dawn to 1600 LT. The three sets of plasmapause locations in Fig. 13 do not show marked asymmetry between night and day while it is quite pronounced in the ISIS I trough location. The results of Tulunay (1973) and Tulunay and Grebowsky (1975) indicate a similar assymetry.

Examination of the location of the high altitude equatorward trough wall, T_2, (Fig. 12) shows that it is the base which may be reasonably compared with the plasmapause locations shown in Fig. 13. The base located at L = 4.4 near midnight and at L = 6.5 at noon is in good agreement with the plasmapause measurements. However, it should be noted in agreement with the finding of Brace and Theis (1974) that the trough shows no afternoon bulge.

Recent measurements by Gringauz and Bezrukikh (1976) using ion traps aboard Prognoz and Prognoz 2 satellites led them to conclude "a considerably higher latitude for the plasmapause at noon compared to that at midnight as a typical feature of the quiet magnetosphere". Their noon-midnight measurements of May 8 and August 24, 1972, for example, show a difference of approx. 3.5 L in the noon-midnight plasmapause location. On the average, however, they find the plasmapause location to be 1.5 L higher at noon than at midnight. Lemaire's (1976) theoretical studies of the steady state plasmapause position deduced from McIlwain's (1974) E3H convection electric field model for magnetically quiet conditions (Kp = 1 to 2) clearly show a noon-midnight asymmetry with the plasmapause at higher L values at noon than at midnight. Carpenter and Seely (1976) from recent quiet time Whistler drift path observations noted significant noon-midnight asymmetry in the plasmapause location contrary to earlier measurements (Carpenter, 1966). The whistler, satellite measurements and theoretical deductions of Wolf (1974) and Lemaire (1976) provide strong evidence for a noon-midnight asymmetry of the plasmapause, with the noon plasmapause located at higher L values. The ISIS I high altitude measurements of T_2 show that the base of the high altitude trough wall is in good agreement with these recent plasmapause results.

SUMMARY AND CONCLUSIONS

Examination of thermal ion and electron troughs in the topside ionosphere obtained by means of instruments flown on the ISIS-I and INJUN V satellites over a 3 year period under conditions when $Kp \leq 3$ has led to the following results:

1. The trough occurrence frequency is over 90% within 3 hours of midnight, it decreases to 60% near dawn and dusk, and reaches a minimum value of 48% near local noon.

2. At altitudes below 1500 km on the nightside, the midpoint of the equatorial trough wall at about L = 3.8 is found to be in good agreement with reported plasmapause positions. At altitudes between 1500 and 3500 km, the top of the equatorward trough wall moves to continually lower latitudes. At these higher altitudes, the equatorward wall becomes the dominant feature of the ionization distribution often extending 15^{o} to 20^{o} in latitude. The poleward edge of the trough becomes less well defined at high altitudes and is marked by ionization spikes 1^{o} to 2^{o} in width.

3. On the dayside below 1500 km the mean location of the trough T_1 at noon is $L = 12.5 \pm 1.0$, much higher than the dayside plasmapause location. T_1 persists above 1500 km with decreasing amplitude above this level.

4. The diurnal variation of the trough location was not found to vary with season except near sunrise.

5. The ISIS I data and simultaneous INJUN V measurements of electron density, ion & electron temperature, horizontal electric fields and particle measurements show that the high latitude trough, T_1, is located at the equatorward edge of the cusp. It is concluded that electric fields and enhanced thermal electron and ion temperature contribute to the formation of T_1 by increasing ion loss rates, ion scale height, and producing changes in the chemical composition of the region, in agreement with the theoretical analysis of Schunk et al. (1975).

6. At all altitudes above 1500 km a second ionization depletion region or trough T_2, is observed between L = 2 and 6.

7. The base of the high altitude trough wall (T_2 on the dayside) is found to lie between L = 4 and 6 in good agreement with plasmapause locations in the equatorial plane.

REFERENCES

Bewersdorff, A.B. and R.C. Sagalyn, Evidence of an evening ionization anomaly within the plasmasphere, Space Research XI, p. 1289, Akademic-Verlag, Berlin, 1971.

Bewersdorff, A.B., and R.C. Sagalyn, Spatial and temporal variations of the thermal

plasma between 3000 and 5700 km at L = 2 to 4, J. Geophys. Res., 77, 4734, 1972.

Brace, L.H. and R.F. Theis, The behavior of the plasmapause at mid-latitudes, J. Geophys. Res. 79, 1871-1884, 1974.

Brinton, H.C., R.A. Pickett, and H.A. Taylor, Diurnal and seasonal variations of atmospheric ion composition; correlation with solar zenith angle, J. Geophys. Res., 74, 4064, 1969.

Brinton, H.C., J.M. Grebowsky, and H. Mayr, Altitude variation of ion composition in the mid-latitude trough region - Evidence for upward plasma flow, J. Geophys. REs., 76, 3738, 1971.

Carpenter, D.L., Whistler studies of the plasmapause in the magnetosphere, 1. Temporal variations in the position of the knee and some evidence on plasma motions near the knee, J. Geophys. Res., 71, 693, 1966.

Carpenter, D.L. and N.T. Seely, Cross L plasma drifts in the outer plasmasphere: Quiet time patterns and some substorm effects, J. Geophys. Res., 81, 2728-2736, 1976.

Chappell, C.R., Recent satellite measurements of the morphology and dynamics of the plasmasphere, Rev. of Geophys. Space Physics, 10, 951, 1972.

Grebowsky, J.M., N.C. Maynard, Y.K. Tulunay and L.J. Lanzorotti, Coincident observation of ionospheric troughs and the equatorial plasmapause, Planet. Space Science, 24, 1177, 1976.

Gringauz, K.L. and V.V. Bezrukikh, Asymmetry of the earth's plasmasphere in the direction noon-midnight from data of measurements of satellites Prognoz and Prognoz 2, J. Atmos. Terr. Phys., 38, 1071, 1976.

Lemaire, J., Steady state plasmapause positions deduced from McIlwain's electric field models, J. Atmos. Terr. Phys., 38, 1039, 1976.

McIlwain, C.E., Magnetospheric Physics, Edited by B.M. McCormack, Reidel Publishing Co., Dordrecht, Holland, p. 143, 1974.

Miller, N.J., and L.H. Brace, Some winter characteristics of the northern high latitude ionosphere, J. Geophys. Res., 74, 5752, 1969.

Miller, N.J., The dayside mid-latitude plasmatrough, J. Geophys. Res., 79, 3795, 1974.

Muldrew, D.B., F-layer ionization troughs deduced from Alouette data, J. Geophys. Res., 70, 2635, 1965.

Münch, J.W., H.J. Braun, G.R. Pilkington, R.C. Sagalyn, and P.J.L. Wildman, Thermal electron densities and temperatures in the dayside cusp, J.A.T.P., 39, 699, 1977.

Nishida, A., Formation of plasmapause or magnetospheric knee by the combined action of magnetospheric convection and plasma escape from the tail, J. Geophys. Res., 5699, 1966.

Nishida, A., Average structure and storm-time change of the polar topside ionosphere at sunspot minimum, J. Geophys. Res., 72, 6051, 1967.

Park, C.G., and P.M. Banks, Influence of thermal plasma flow on the mid-latitude nighttime F_2 layer; Effects of electric fields and neutral winds inside the plasmasphere, J. Geophys. Res., 79, 4661, 1974.

Park, C.G., and P.M. Banks, Influence of thermal plasma flow on the daytime F_2 layer J. Geophys. Res., 80, 2819, 1975.

Rycroft, M.J. and J.O. Thomas, The magnetospheric plasmapause and the electron density trough at the Alouette I orbit, Planet, Space Sci., 18, 65, 1970.

Sagalyn, R.C., M. Smiddy and J. Wisnia, Measurement and interpretation of ion density distributions in the daytime F regions, J. Geophys. Res., 68, 199, 1963.

Sagalyn, R.C. and M. Smiddy, Charged particle measurements by means of electrostatic probes, electron density and temperature measurements in the ionosphere. COSPAR Tech. Manuals, p. 90, Commun. Space Res., Paris, 1967.

Sagalyn, R.C., M. Smiddy and M. Ahmed, High latitude irregularities in the topside ionosphere based on ISIS-I thermal ion probe data, J. Geophys. REs., 79, 4252, 1974.

Sharp, G.W., Midlatitude trough in the night ionosphere, J. Geophys. Res., 71, 1345, 1966.

Schunk, R.W., W.J. Raitt, and P.M. Banks, Effect of electric fields on the daytime high latitude E and F regions, J. Geophys. Res., 80, 3121, 1975.

Smiddy, M. and R.C. Stuart, An analysis of the behavior of a multigrid spherical sensor in a drifting Maxwellian plasma, Rep. AFCRL-69-0013, Phys. Sci. Res. Paper 304, U.S. Air Force Cambridge Res. Labs., Bedford, MA, 1969.

Taylor, H.A., H.C. Brinton and A.R. Deshmukh, Observations of irregular structure in thermal ion distributions in the dusk side magnetosphere, J. Geophys. Res., 75, 2481, 1970.

Taylor, H.A. and G.R. Cordier, In-situ observations of irregular ionospheric structure associated with the plasmapause, Planet. Space Sci., 22, 1289, 1974.

Taylor, H.A., J.M. Grebowsky, and A.J. Chen, Ion composition irregularities and ionoshperic-plasmaspheric coupling: Observations of a high latitude ion trough, J. Atmos. Terr. Phys., 37, 613, 1975.

Thomas, J.O. and M.K. Andrews, Transpolar Ionospheric Plasma, 1. Plasmasphere Termination, J. Atmos. Terr. Phys., 37, 613, 1975.

Tulunay, Y. and J. Sayers, Characteristics of the mid-latitude trough as determined by the electron density experiment on Ariel III, J. Atmos. Terr. Phys., 33, 1737, 1971.

Tulunay, Y., Magnetically symmetric detection of the mid-latitude electron density trough by Ariel III satellite, J. Atmos. Terr. Phys. 34, 1547, 1972.

Tulunay, Y., Global electron density distributions from the Ariel III satellite at mid-latitudes during quiet magnetic periods, J. Atmos. Terr. Phys., 35, 233, 1973.

Tulunay, Y. and A.R.W. Hughes, A satellite study of the mid-latitude trough in electron density and VLF radio emissions during the magnetic storm of 25-27 May 1967, J. Atmos. Terr. Phys., 35, 1973.

Tulunay, Y. and J.M. Grebowsky, Temporal variations in the dawn and dusk mid-latitude trough positions measured (Ariel III, Ariel IV) and modelled, Ann. Geophys., t. 31, fasc. 1, 29, 1975.

Wolf, R., Calculations of magnetospheric electric fields in magnetospheric physics (Ed. B.M. McCormac, D. Reidel Pub. Col., Dordrecht, Holland), p. 167, 1974.

Wrenn, G.L. and W.J. Raitt, In-situ observations of mid-latitude ionospheric phenomena associated with the plasmapause, Ann. de Geophys., 31, 17, 1975.

ATS-6 OBSERVATIONS OF IONOSPHERIC/PROTONOSPHERIC ELECTRON CONTENT AND FLUX

L. Kersley*, H. Hajeb-Hosseinieh and K.J. Edwards
University College of Wales
Aberystwyth, U.K.

Abstract Measurements of protonospheric content obtained at Aberystwyth from observations of the ATS-6 satellite radio beacon are reported. The monthly median diurnal behavior shows protonospheric contributions of approximately 15 to 20% to the total content along the ray path by day, rising to a predawn maximum of 35% in summer and more than 40% in winter. The results are shown to be typical of those expected from other European stations and differences from earlier American measurements are explained in terms of ionospheric interactions in the conjugate hemisphere. The temporal gradients of protonospheric content provide information on the net integrated ionospheric/protonospheric plasma fluxes and the results obtained confirm the importance of plasma exchange with both local and conjugate ionospheres.

INTRODUCTION

For more than a decade observations of polarization rotation of VHF transmissions from geostationary satellites have provided measurements of the so-called total electron content along the ray path, a parameter of importance in ionospheric group delay corrections. In practice, the weighting effect of the geomagnetic field limits the measurement to the ionospheric part of the path below some arbitrary height generally taken to be about 2500 km; however at times a significant fraction of the total number of electrons may be above this level. The ATS-6 satellite radio beacon group delay transmitter has enabled measurements to be made of the actual total electron content along the path in addition to polarization measurements of what can now be termed Faraday electron content. The difference between these quantities thus provides an estimate of the protonospheric contribution to the columnar electron content. In this paper we present results of ATS-6 observations from a mid-latitude European station, Aberystwyth, U.K. (52.4°N, 4.1°W), for the solar minimum period November 1975 to July 1976 when the geostationary satellite was stationed above 35°E longitude.

EXPERIMENT

The ATS-6 radio beacon experiment has been described by DAVIES et al. (1972) and DAVIES et al. (1975). Two basic measurements are of concern here. First, the polarization rotation of a VHF carrier and secondly the phase of a modulation on the VHF transmission with respect to that of an identical modulation on a coherent UHF carrier.

(a) Polarization Rotation and Faraday Content

Measurement of the phase difference between the ordinary and extraordinary circular modes of the 140.056 MHz carrier allows the polarization rotation (Ω) to be determined. This rotation can be related to ionosphere quantities along the ray path (s) by

$$\Omega = \frac{K}{f^2} \int_s N f_L ds \qquad 1.$$

where f is the wave frequency, K a constant, N electron density and f_L longitudinal gyrofrequency. If a mean value of f_L can be determined representative of conditions along the

* At present: NRC/NAS Senior Resident Research Associate,
Air Force Geophysics Laboratory, Hanscom AFB, MA 01731, U.S.A.

ionospheric path then the above equation can be written

$$\Omega = \frac{K}{f^2 \bar{f}_L} \int_0^s F \, Nds = \frac{K}{f^2 \bar{f}_L} N_F \qquad 2.$$

where N_F is the so-called Faraday content along the ray path. Because of the weighting of the geomagnetic field N_F can be considered as a measure of the ionospheric electron content up to some arbitrary height (h_F) which model studies have shown to be generally between 2000 and 3000 km.

The variation of f_L along the Aberystwyth/ATS-6 path, plotted as a function of height, is shown in Fig. 1. It can be seen that the parameter maximizes at about 200 km and is relatively insensitive to height below about 500 km, a shape which has important consequences for the accuracy of the Faraday and protonospheric content measurements.

A fixed value of $\bar{f}_L = 0.817$ MHz corresponding to a mean ionospheric height of 420 km has been used in the present work. Thus, from Equation 2, the relationship between the measured polarization rotation (in degrees) of the 140.056 MHz carrier and the Faraday electron content (in m^{-2}) becomes

$$N_F = 4.96 \times 10^{14} \, \Omega \qquad 3.$$

(b) Modulation Phase and Total Electron Content

The other angular parameter measured by the Aberystwyth receiver is the phase of a 1 MHz modulation on the 140 MHz carrier with respect to that of an identical modulation on a coherent 360 MHz carrier. The ionospheric modulation phase delay, for ordinary wave components, is given by

$$\phi = -K \int_s \left[\frac{1}{f_2+\bar{f}_L} - \frac{1}{f_1+\bar{f}_L} - \frac{1}{f_4+\bar{f}_L} + \frac{1}{f_3+\bar{f}_L}\right] N ds \qquad 4.$$

where the suffixes refer to the respective signal frequencies, $f_1 = 140.0560$ MHz, $f_2 = 141.0564$ MHz, $f_3 = 360.1440$ MHz and $f_4 = 361.1444$ MHz.

The longitudinal gyrofrequency is always less than 0.83 MHz for the ATS-6 to Aberystwyth geometry and is thus small compared to the signal frequencies. Hence if an average value $\bar{f}_L$ is chosen for the path the above equation can be written

$$\phi = -K \left[\frac{1}{f_2+\bar{f}_L} - \frac{1}{f_1+\bar{f}_L} - \frac{1}{f_4+\bar{f}_L} + \frac{1}{f_3+\bar{f}_L}\right] \int_s N ds \qquad 5.$$

that is

$$\phi = -K \left[\frac{1}{f_2+\bar{f}_L} - \frac{1}{f_1+\bar{f}_L} - \frac{1}{f_4+\bar{f}_L} + \frac{1}{f_3+\bar{f}_L}\right] N_T \qquad 6.$$

where N_T is the total electron content along the slant path from the satellite to the receiving antennas.

Using $\bar{f}_L = 0.817$ MHz, corresponding to a mean ionospheric height of 420 km, the total electron content (in m^{-2}) can be obtained from the modulation phase delay (in degrees)

$$N_T = -4.87 \times 10^{14} \, \phi \qquad 7.$$

(c) F-factor

An additional experimental parameter which can be obtained from the ATS-6 angular data is the shape factor or F-factor which is a weighted average electron gyrofrequency along the path. This has been defined by DAVIES et al. (1975) as

$$F = \int_s N f_L \, ds \; / \int_s N \, ds \qquad 8.$$

By substitution from Equations 1 and 5 F can be related to the measured parameters by

$$F = 0.832 \, \frac{\Omega}{\phi} \qquad 9.$$

(d) Protonospheric content

It can be seen that absolute measurements of the total and Faraday contents to a sufficient accuracy will allow the electron contribution from ~2500 km to the geostationary satellite to be determined. The effective upper boundary is however the plasmapause which is usually well below the geostationary orbit height. The protonospheric (some workers prefer the term plasmaspheric) electron content (N_p) can thus be found from

$$N_p = N_T - N_F \qquad 10.$$

RESULTS

Diurnal plots of median values of N_T, N_F, N_p N_p/N_T% and F for each of the nine months under study are presented in Fig. 2. The diurnal variations of N_T and N_p are essentially similar in character indicating that the ionospheric contribution dominates the total electron content at least during daytime. By contrast, N_p shows only a small diurnal variation with minimum values in the predawn

hours around 3 x 10^{16} m^{-2} for all months and an afternoon maximum of about 4 x 10^{16} m^{-2} in winter rising to approximately 5 x 10^{16} m^{-2} in summer. The protonospheric content expressed as a percentage of the total has values by day typically in the range 15 to 20% while the predawn maximum of about 35% in summer exceeds 40% in winter. The diurnal variations of F show a predawn minimum about 0.5 MHz while the daytime value is generally in the region of 0.66 MHz.

Some insight can be gained into the physical processes responsible for the observed variations in N_p by considering the temporal rate of change of the parameter. Protonospheric content is a measure of the plasma content of a region of the high plasmasphere above ∿2500 km altitude where production and loss are unimportant and the dominant mechanism for content changes is plasma diffusion up and down the field lines in response to interaction with the underlying ionosphere. The temporal gradient of protonospheric content thus provides a near continuous estimate of the net integrated protonospheric flux responsible for the filling and draining of the region. It should be remembered that the protonospheric content measured in the ATS-6 experiment refers to the slant path to the satellite intersecting flux tubes in the range L∿1.7 to L∿4 on average at the plasmapause. The protonospheric flux estimated from these measurements is thus an integrated parameter in contrast to the more direct measurement essentially along a particular field tube given by the whistler technique (PARK, 1970).

Protonospheric flux measurements (dN_p/dt) estimated from three-hour running mean smoothed hourly values of protonospheric content are presented in Fig. 3. This shows the mean diurnal variation of protonospheric flux for December 1975, January 1976 and June, July 1976 representing winter and summer conditions respectively. In interpreting these fluxes it must be remembered that a given field line interacts with the ionosphere at its terminations in both local and conjugate hemispheres, so that the flux being measured represents the net integrated filling or draining rate of the tubes of interest. To assist in interpretation, the times of sunrise and sunset in both local and conjugate hemispheres at 350 km at the ionospheric base of the L=2 flux tube, which intersects the ATS-6 ray path at ∿5000 km altitude, are marked. In June and July a positive upwards flux is found some hours after sunrise. It should be noted that time constants of the order of hours are probably involved before ionospheric changes can be reflected in protonospheric content, allowing for the O^+-H transition barrier and the travel time of the plasma along the long flux tubes. Following the post sunrise period of upwards flux, for much of the summer day the fluxes do not differ significantly from zero, however after conjugate sunset a strong downwards net flux is observed at a time when the still sunlit ionosphere in the local hemisphere is displaying the late afternoon peak in both maximum density and content. It would thus appear that depletion at this time is essentially to the conjugate ionosphere. In December and January, upwards fluxes are found by day with a post sunset depletion. However, it is interesting to note that net upwards fluxes are also found in the early morning hours apparently associated with the much earlier sunrise in the conjugate summer hemisphere.

DISCUSSION

The protonospheric content variations obtained at Aberystwyth when the ATS-6 satellite was stationed at 35° E differ significantly from those already reported from the earlier American sector phase of ATS-6 operations. Two contrasting features of note are the generally lower magnitudes found in the American observations particularly in summer and a diurnal variation of the earlier data for winter through equinox which minimizes by day (FRITZ, 1976; SOICHER, 1976; JOHANSON and KLOBUCHAR, 1977). In interpreting these differences it must be remembered that the measured protonospheric content changes are largely consequent upon interaction with the ionosphere at the feet of the flux tubes intersecting the ray path. Fig. 4 shows the ionospheric terminations at 350 km altitude in both local and conjugate hemisphere of the field lines crossing the ray path for the Aberystyth observations while the ATS-6 satellite was at 35° E. The important field lines for protonospheric content studies are those with intersections above ∿2500 km up to the plasmapause at say L∿4. It can be inferred from Fig. 4 that because of the convergence of the ray paths towards the satellite the geographic locations of the ionospheric terminations of the relevant field lines will alter by only a relatively small amount

for mid-latitude stations covering a wide range of longitudes. Thus, satellite, rather than observing station, longitude is an important factor in the protonospheric content measurements. Hence, the present observations can be taken as representative of mid-latitude observations from the European sector while the general concurrence of the earlier observations, already referred to, from several American stations is explained. The most apparent difference between the observing situation from Europe presented here and that from America with ATS-6 at 94° W longitude lies in the geographic latitudes of the field line terminations in the conjugate hemisphere. While for the present study the ionospheric regions in the conjugate hemisphere, interacting to contribute towards the observed protonospheric changes, lie at mid-latitudes, for the earlier American work the corresponding locations are at much higher geographic latitudes close to the Antarctic continent. The anomalous behavior of the continuously sunlit Antarctic ionosphere in this longitude sector in local summer (December/January) under the influence of the neutral wind is well documented (KING et al., 1968) and this factor, together with the long winter nights and the generally higher L-shells intersecting the American sector ray paths, can account for the differences between the European and American protonospheric contents (KERSLEY and KLOBUCHAR, 1977). The observed protonospheric fluxes show features which are in broad agreement with the theoretical model studies of MURPHY et al. (1976); namely post dawn upwards fluxes and post sunrise downwards fluxes which maximize and decay. However, the theoretical work has been confined to date to equinoxial studies with assumed equatorial symmetry as opposed to interactions with both hemispheres applicable to the present observations. The magnitudes of the fluxes are generally lower than those reported from whistler studies (PARK, 1970). However, it must be remembered that the latter are obtained from effectively direct measurements of tube contents whereas the present work reports median content changes along a slant path intersecting many field lines, albeit essentially along the flux tubes for the important lower protonospheric part of the range.

CONCLUSIONS

The monthly median protonospheric content shows diurnal variations which reflect the interactions with the underlying ionosphere. The contribution from the protonosphere amounts to some 15-20% of the total content by day, rising by night to a predawn maximum of 35% in summer and more than 40% in winter. It has been shown that observations reported here are expected to be typical of those obtained from other mid-latitude European stations and differences from earlier American measurements are explained in terms of ionospheric interactions in the conjugate hemisphere. The studies of protonospheric flux support the conclusion of the importance of the conjugate region to the interpretation of the data.

Acknowledgement Thanks are expressed to Mr. J. M. Johanson for the computations involved in the preparation of Fig. 4.

REFERENCES

Davies, K., R. B. Fritz and R. N. Grubb, The ATS-F/G radio beacon experiments, J. Environ. Sci., 15, 31-35, 1972.

Davies, K., R. B. Fritz, R. N. Grubb and J. E. Jones, Some early results from the ATS-6 Radio Beacon Experiment, Radio Sci., 10, 785-799, 1975.

Fritz, R. B., ATS-6 Radio Beacon Electron Content Measurements at Boulder, July 1974-May, 1975, Report-UAG-58, World Data Center A, Boulder, U. S. A., 1976.

Johanson, J. M. and J. A. Klobuchar, A comparison of average plasmaspheric electron content at two mid-latitude stations, Paper presented at A. G. U. Spring Meeting, Washington, D. C., U.S.A., June, 1977.

Kersley, L. and J. A. Klobuchar, Comparison of protonospheric electron content measurements from the American and European sectors, submitted to Geophysical Research Letters, 1977.

King, J. W., H. Kohl, D. M. Preece and C. Seabrook, An explanation of phenomena occurring in the high latitude ionosphere at certain Universal Times, J. Atmos. Terr.

Phys., 30, 11-23, 1968.

Murphy, J. A., G. J. Bailey and R. J. Moffett, Calculated daily variations of O^+ and H^+ at mid-latitudes - I. Protonospheric replenishment and F-region behavior at sunspot minimum, J. Atmos. Terr. Phys., 38, 351-364, 1976.

Park, C. G., Whistler observations of the interchange of ionization between the ionosphere and the protonosphere, J. Geophys. Res., 75, 4249-60, 1970.

Soicher, H., Diurnal, day-to-day and seasonal variability of N_E, N_T and N_O at Fort Monmouth, New Jersey, in Proceedings of COSPAR Satellite Beacon Symposium, pp. 231-243, Boston University, U. S. A., 1976.

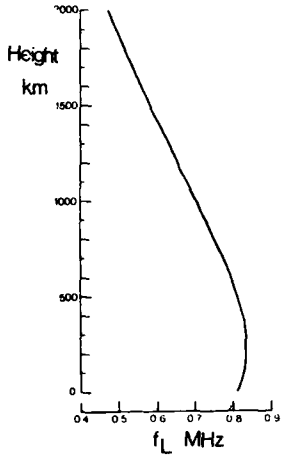
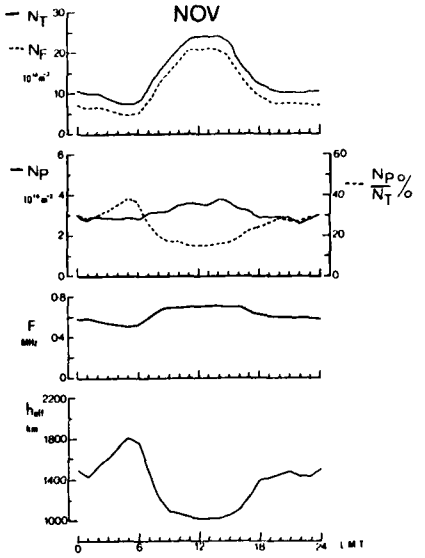
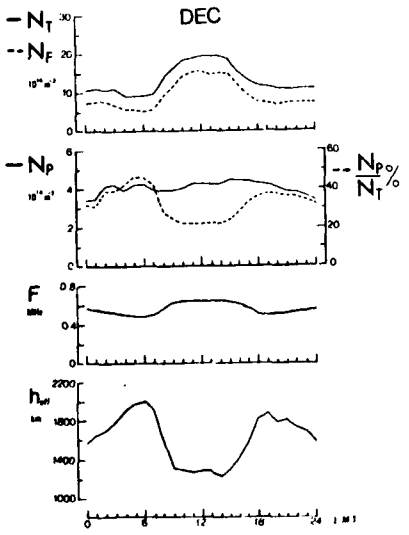


Fig. 1. Height variation of longitudinal gyrofrequency for the ATS-6 to Aberystwyth path below 2000 km.





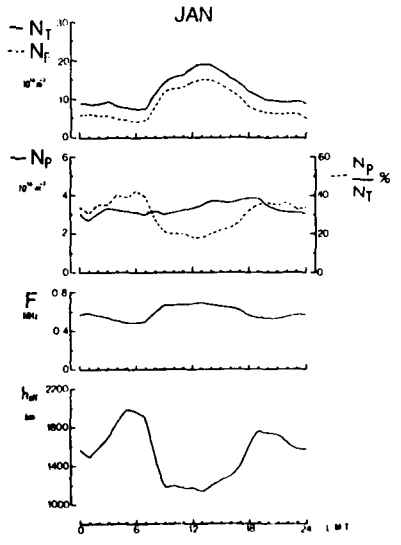


Fig. 2. Monthly median hourly values of N_T, N_F, N_p, N_p/N_T% and F.

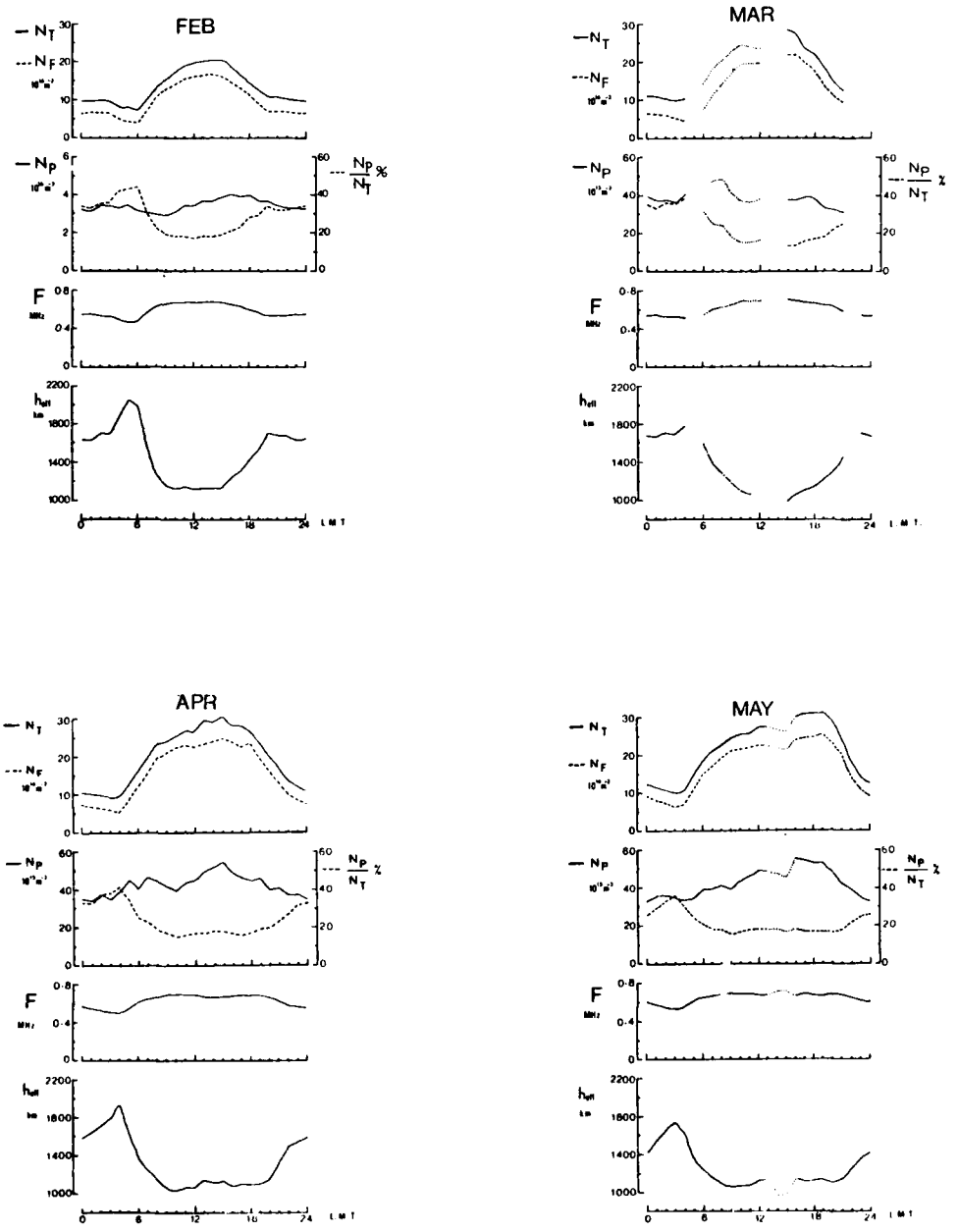


Fig. 2 (cont). Monthly median hourly values of N_T, N_F, N_p, N_p/N_T% and F.

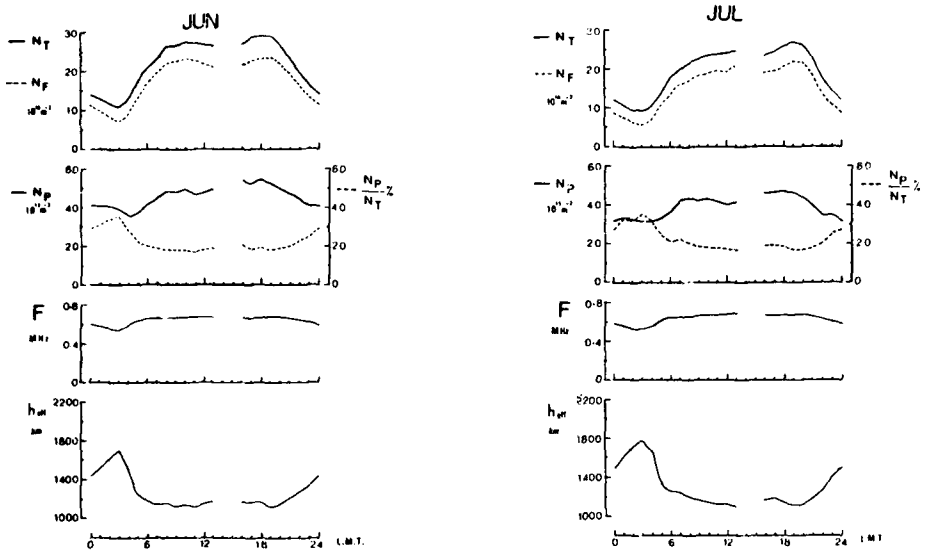


Fig. 2 (cont). Monthly median hourly values of N_T, N_F, N_p, N_p/N_T% and F.

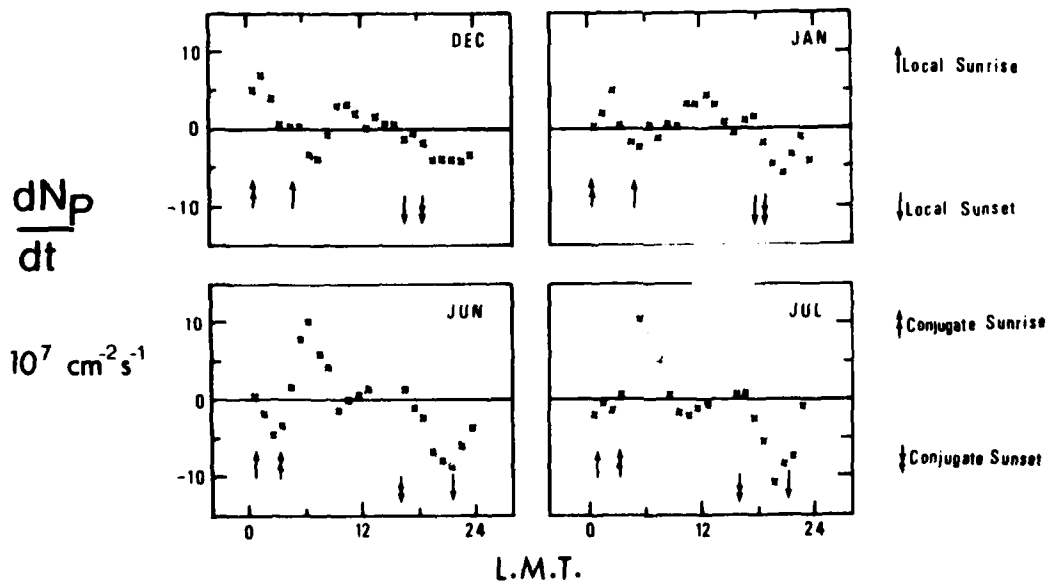


Fig. 3. Diurnal variations of integrated protonospheric flux (dN_p/dt) by month for December 1975, January, June and July, 1976.

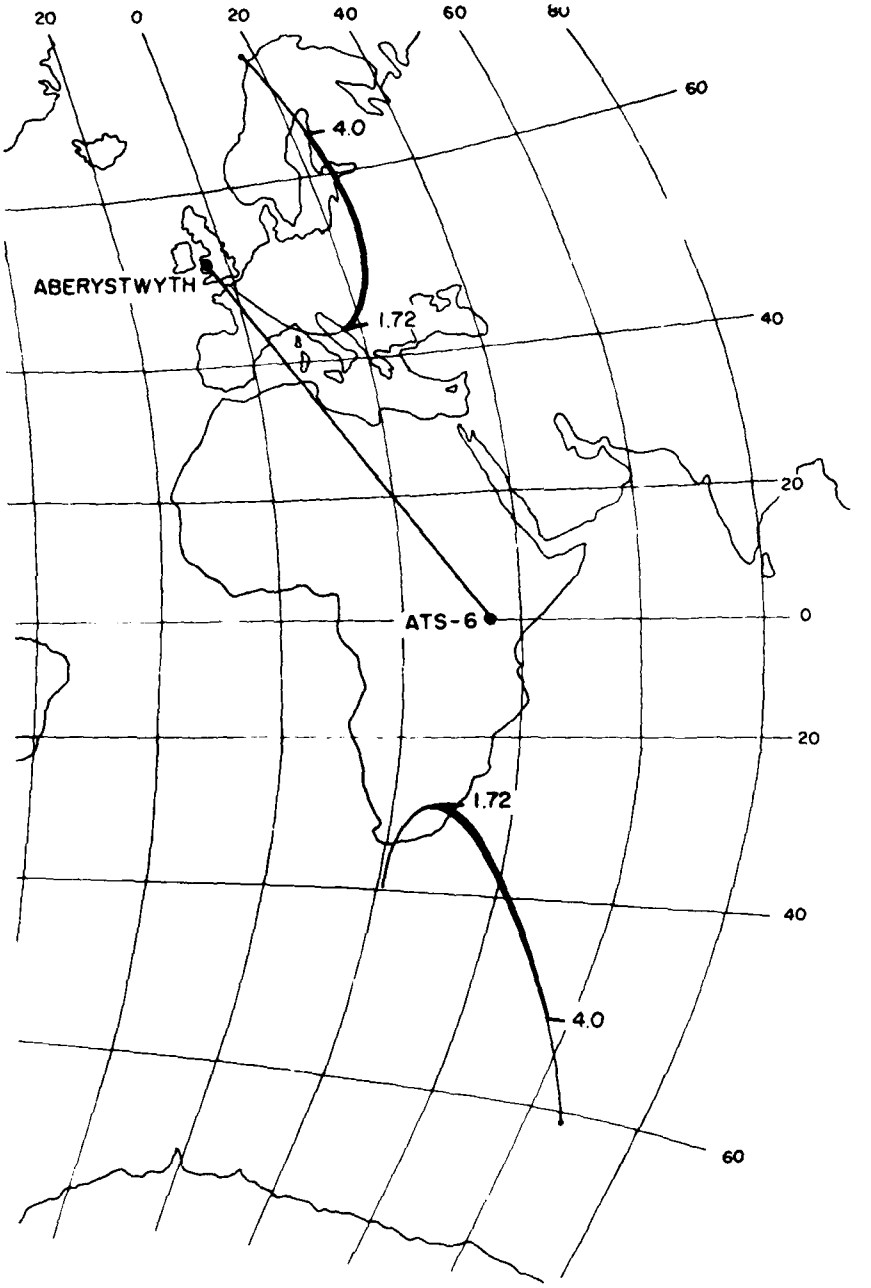


Fig. 4. Terminations at 350 km in both local and conjugate hemispheres of field lines intersecting the ATS-6 to Aberystwyth ray path. The regions of importance to measurements of N_D, corresponding to field line intersections from a height $\sim$ 2500 km to L $\sim$ 4, are marked by the heavier part of the curves.

SATELLITE MEASUREMENTS FROM WITHIN IONOSPHERIC STRUCTURES RESPONSIBLE FOR AURORAL ACCELERATION PROCESSES

R. D. Sharp, E. G. Shelley, and R. G. Johnson
Lockheed Palo Alto Research Laboratory
Palo Alto, California 94304

INTRODUCTION

Ionospheric disturbances at high latitudes associated with auroral precipitation events are known to cause radio frequency interference problems and so the physical causes of these disturbances are of direct relevance to the theme of this symposium. One class of auroral electron precipitations known as "inverted V" events are thought to be caused by large-scale ionospheric structures containing dc electric fields, with a component parallel to the geomagnetic field resulting in the downward acceleration of the precipitating electrons (Gurnett, 1972; Swift, 1975). Until recently, no direct measurements within these structures have been reported although some of the properties of the auroral acceleration process have been inferred from the characteristics of the precipitating electrons (see review by Evans, 1976).

Recent measurements from within the structures have been acquired by means of directed barium jets (Haerendel, 1976) and from the S3-3 satellite which, because of its unique orbital characteristics, is traversing the structures at altitudes of $\sim 1\ R_E$ in the polar regions. Office of Naval Research experiments on the S3-3 satellite are measuring the characteristics of both the particles and fields within the structures (Shelley et al., 1976; Mozer et al., 1977) and in association with the other spacecraft experiments (Mizera and Fennel, 1977) are providing the first detailed determinations of their geometry.

EXPERIMENT

The S3-3 satellite was launched into an elliptical polar orbit with an initial apogee of 8050 km and perigee of 260 km. The orbital inclination is 97.5°. The satellite is spinning at about 3 rpm with its spin axis perpendicular to the orbital plane. The measurements to be described in this report were acquired from an ion and electron spectrometer experiment mounted with its view direction perpendicular to the spin axis. The experiment consists of 3 ion mass spectrometers and 4 electron spectrometers similar in design to those described previously (Shelley et al., 1972; Reed et al., 1969). The energy ranges and geometric factors of the electron spectrometers are listed in Table 1. The ion mass spectrometers each acquire a 30-channel mass-per-unit-charge spectrum at a single energy-per-unit-charge every second. The energy-per-unit-charge setting is cycled through 4 values every 64 seconds, remaining on each step for 16 seconds. Thus, a 12-point energy spectrum is acquired from the 3 spectrometers every 64 seconds. These 12 measured energy-per-charge values are also listed in Table 1.

OBSERVATIONS

Initial observations from the S3-3 ion mass spectrometer experiment showed intense beams of upward-directed newly-accelerated ionospheric ions as a persistent feature of auroral field lines at altitudes of about 1 RE. Two distinct classes of ion acceleration phenomena have been detected, one involving the parallel and one the perpendicular component of the ion velocity (Shelley et al., 1976; Sharp et al., 1977).

Figure 1 shows a segment of data from the low-energy mass spectrometer acquired in the northern polar region at a local time of about 1430. The relative responses of the O^+ and H^+ ions are plotted versus time and can be compared with the pitch angle determined from the on-board magnetometer shown in the upper panel. The energy-per-unit-charge of the measured ions is also indicated. One sees sharply peaked pitch-angle distributions for both O^+ and H^+ corresponding to ions streaming up the field lines from the ionosphere. The lower panel shows the electron fluxes from the CMEB detector ($0.35 \leq E \leq 1.1$ keV). The deep minima corresponding to the atmospheric loss cone are clearly evident at the same locations as the ion peaks. The narrow ion pitch-angle distributions evident in Figure 1 (e.g., FWHM $\approx 10°$ for the proton peaks) are indicative of an acceleration mechanism which primarily energizes the parallel component of the ion velocity vector. A possible mechanism of this type is a dc electric field with a component parallel to the geomagnetic

field and, as we will see below, a detailed analysis of the ion and electron distribution functions in selected events shows that such electric fields do occur on auroral geomagnetic field lines at high altitudes.

An example of an observation of the second class of ion acceleration mechanism is shown in Fig. 2. These data were acquired in the polar cusp at an altiutde of 7600 km. One sees in Figure 2 a latitudinally extended region of upward-flowing oxygen ions with a "conical" pitch-angle distribution, i.e., having a minimum along the direction of the geomagnetic field. Pitch-angle distributions of this type are expected if the ion acceleration mechanism acts primarily on the perpendicular component of the ion velocity, and at an altitude substantially below that at which the ions are observed. The initial 90° peaked pitch-angle distribution then folds into the observed conical distribution under the action of the first adiabatic invariant. According to this model the ions shown in Figure 2 were accelerated in the altitude region of 4000-5000 km. A possible mechanism resulting in such angular distributions is acceleration by means of ion cyclotron waves.

15 SEPTEMBER 1976 EVENTS

Many of the auroral acceleration events detected by the S3-3 satellite, especially those occurring during active times, are characterized by rapid temporal variations (determinable by differences in the electric potential at altitudes below the satellite as measured by the rapidly moving electrons and the slowly moving ions) and by overlapping and complex spatial structures. In an initial attempt to determine some of the essential features of the structures, two events have been selected for detailed analysis which show evidence for a relative lack of such complicating features and which have a latitudinal scale size suitable for repeated pitch-angle scans by the relatively slowly spinning S3-3 spacecraft. One of these events at 1058 UT on 15 September 1976 has been described in detail by Cladis et al. (1977) and Cladis and Sharp (1977a,b) and, in view of the constraints on the length of this presentation, will not be further discussed here. Survey plots showing the overall characteristics of the second of these events are presented in Figure 3.

The top three panels show, on a logarithmic scale, the summed count rates from the 3 ion spectrometers from ions with the mass-per-charge ratios indicated. The instrument pitch angle is shown in the center panel and the lower four panels show the electron data (see Table 1 for energy ranges). Along the abscissa are given Universal Time (SYST) geographic longitude and latitude, altitude (km), invariant latitude and magnetic local time. An examination of Figure 2 shows an "event" with a relatively symmetric pattern of fluxes. Downward streaming ($\sim$ 100 eV) electrons with a sharply peaked pitch angle distribution form the borders. The central region is characterized by an enhanced flux level of $\sim$ 1 keV electrons with dramatically enlarged loss cones and upstreaming ions (primarily protons in this example) evidencing a parallel electric potential drop below the satellite of about 1 kilovolt in magnitude. The presence of a parallel dc electric field above the satellite in the central portion of the structure is indicated by the reduced level of the CMEA fluxes and the symmetry between the upward and downward moving soft (E $\sim$ 100 eV) electrons. In this interpretation the primary electrons have been accelerated out of the energy range of the DMEA detector by a potential difference above the spacecraft of magnitude $\phi_A > 0.24$ kV. The upward moving secondary electrons in the energy range of CMEA produced in the atmosphere by the incident primary beam, are reflected by the parallel electric field above the spacecraft giving rise to the observed symmetry. Another important indicator of a dc electric field above the spacecraft is the presence of minima at 90° in the pitch-angle distributions of the primary (CMEB) electrons. The equation for the first adiabatic invariant for these particles is:

$$\frac{B_1}{E_1 \sin^2 \alpha_1} = \frac{B_2}{(E_1 - e\phi_A) \sin^2 \alpha_2}$$

where B is the intensity of the geomagnetic field. E is the electron energy and α is their local pitch angle. The subscript 1 refers to the altitude at the upper edge of the parallel electric field structure and the subscript 2 refers to the satellite altitude. By setting $\alpha_1 = 90°$ in this equation we have a relationship between the angular width of the observed 90° minimum in the pitch-angle distribution and the vertical extent of the parallel electric field structure above the spacecraft (through B_1). The details of the calculations leading to the determination of this parameter are beyond the scope of this brief report. The results for both the event illustrated and the event analyzed by Cladis and Sharp (1977a,b) are that this vertical dimension is inferred to be greater than 1000 km.

RELEVANCE TO SPACECRAFT CHARGING

The presence of sharply peaked ion and electron pitch-angle distributions such as have been described here have implications with respect to the electrostatic charging of spacecraft systems (Johnson et al., 1977). An observed source of spacecraft malfunctions is the differential electrostatic charging of adjacent insulated surfaces leading to high-voltage discharges and accompanying material damage and electromagnetic interference (Rosen, 1975).

If an anisotropic field-aligned ion flux is incident on a spacecraft with a hole in the outer skin, then a non-conducting surface on a component inside the skin and on the same magnetic field line as the hole will become positively charged providing the hole subtends an angle

from the component surface equal to or less than the pitch-angle range over which the positive ion flux is larger than the electron flux. Assuming that the electron flux is higher than the ion flux at the larger pitch angles (which is typical) then a large negative potential could be formed on the component surface adjacent to a large positive potential. This configuration is illustrated schematically in Figure 4, and to simplify this example, the secondary electron emission from the surface is assumed to be negligible. The surface position L on the component lies along the magnetic field line through the hole in the spacecraft skin. Angles l_1 and l_2 are taken to be less than the pitch-angle range over which the ion flux is greater than the electron flux so that a positive potential will occur at position L. Position N illustrates a surface region at angles between n_1 and n_2 to the magnetic field direction where the electron flux is larger than the ion flux. At this position a negative potential will occur. At some position, M, between N and L the electron and ion fluxes will be equal and a zero potential will occur. It can be seen that the surface charging at each position on the surface is related to the pitch angles subtended at the hole in the skin and thus, in analogy to pinhole cameras for electromagnetic radiations, this will be referred to as the "pinhole camera charging effect". Although it has been illustrated for a net positive flux along the field line, an anisotropic electron flux will also produce a potential gradient across the surface in essentially the same way.

CONCLUSIONS

The ONR low-energy particle experiment on the S3-3 satellite is providing detailed observations of the ion and electron energy and pitch-angle distributions within the ionospheric structures responsible for at least some types of auroral acceleration processes.

For two such structures which have been analyzed in detail, a geometry is inferred which has many properties consistent with the oblique electrostatic shocks predicted by Swift (1975) to be the cause of auroral arcs and "inverted V" events. The observed structures have a latitudinal extent of $\sim$ 300 km, a vertical extent greater than 1000 km, and contain field-aligned electrostatic potentials of the order of several kilovolts. The highly field-aligned pitch-angle distributions of ions and electrons resulting from accelerations within these structures have particular significance with respect to differential charging phenomena on high-altitude spacecraft through the "pinhole camera effect".

ACKNOWLEDGMENTS

This research has been principally supported by the Office of Naval Research under Contract N00014-72-C-0234. Additional support has been provided by the Lockheed Independent Research Program.

REFERENCES

Cladis, J. B., and R. D. Sharp, "Electrostatic Potential Differences Along Magnetic Field Lines Inferred from Satellite Measurements of Electron and Ion Distributions," EOS, 58, 473, 1977a.

Cladis, J. B., and R. D. Sharp, "Distribution of Electrostatic Potential Along Magnetic Field Inferred from Observations of Ion and Electron Fluxes," EOS, 58, 716, 1977b.

Cladis, J. B., L. L. Newkirk, M. Walt, G. T. Davidson, and W. E. Francis, "Investigation of Ionospheric Disturbances," Report No. DNA-4225F, Defense Nuclear Agency, Washington, D. C. 20305, 28 January 1977.

Evans, D. S., "Evidence for Low-Altitude Acceleration of Auroral Particles," in Physics of the Hot Plasma in the Magnetosphere, edited by B. Hultqvist and L. Stenflo, pp. 319-340, Plenum Publishing Company, New York, New York, 1976.

Gurnett, D. A., "Electric Field and Plasma Observations in the Magnetosphere," in Critical Problems of Magnetospheric Physics, edited by E. R. Dyer, IUCSTP Secretariat, National Academy of Sciences, November 1972.

Haerendel, G., "Observations of Electrostatic Acceleration in the Magnetosphere," paper presented at the International Symposium on Solar Terrestrial Relationships, sponsored by the American Geophysical Union, Boulder, Colorado, June 1976.

Johnson, R. G., R. D. Sharp, and E. G. Shelley, "Composition of the Hot Plasma near Geosynchronous Altitude," Proceedings of the Spacecraft Charging Technology Conference, U. S. Air Force Academy, Colorado Springs, Colorado, 27-29 Oct. 1976, in press.

Mizera, P. F., and J. F. Fennel, "Signatures of Electric Fields from High and Low Altitude Particle Distributions," Geophys. Res. Letters, 4, 311, 1977.

Mozer, F. S., C. W. Carlson, M. K. Hudson, R. B. Torbert, B. Parady, J. Yatteau, and M. C. Kelley, "Observations of Paired Electrostatic Shocks in the Polar Magnetosphere," Phys. Rev. Letters, 38, 292, 1977.

Reed, R. D., E. G. Shelley, J. C. Bakke, T. C. Sanders, and J. D. McDaniel, "A Low-Energy Channel Multiplier Spectrometer for ATS-E," IEEE Trans. Nucl. Sci., NS-16, 359, 1969.

Rosen, A., "Large Discharges and Arcs on Spacecraft," Astronautics and Aeronautics, 13, 36, 1975.

Sharp, R. D., R. G. Johnson, and E. G. Shelley, "Observations of an Ionospheric Acceleration Mechanism Producing Energetic (keV) Ions Primarily Normal to the Geomagnetic Field Direction,"

J. Geophys. Res., 82, 3324, 1977.

Shelley, E. G., R. G. Johnson, and R. D. Sharp, "Satellite Observations of Energetic Heavy Ions during a Geomagnetic Storm," J. Geophys Res., 77, 6104, 1972.

Shelley, E. G., R. D. Sharp, and R. G. Johnson, Satellite Observations of an Ionospheric Acceleration Mechanism," Geophys. Res. Letters, 3, 654, 1976.

Swift, D. W., "On the Formation of Auroral Arcs and the Acceleration of Auroral Electrons," J. Geophys. Res., 80, 2096, 1975.

Table 1

DETECTOR CHARACTERISTICS

Detector	Particle	Energy (E, in keV) or Energy-Per-Unit-Charge	$G\Delta E$ (cm^2-sec-str)
CMEA	Electrons	0.07 → 0.24	1.2×10^{-6}
CMEB	Electrons	0.35 → 1.1	6.5×10^{-6}
CMEC	Electrons	1.6 → 5.0	1.9×10^{-5}
CMED	Electrons	7.3 → 24	6.5×10^{-5}
CXA1	Ions	0.50, 0.68, 0.94, 1.28	Depends on ion species $\approx 7 \times 10^{-5}$ E
CXA2	Ions	1.76, 2.4, 3.3, 4.5	
CXA3	Ions	6.2, 8.5, 11.6, 16.0	

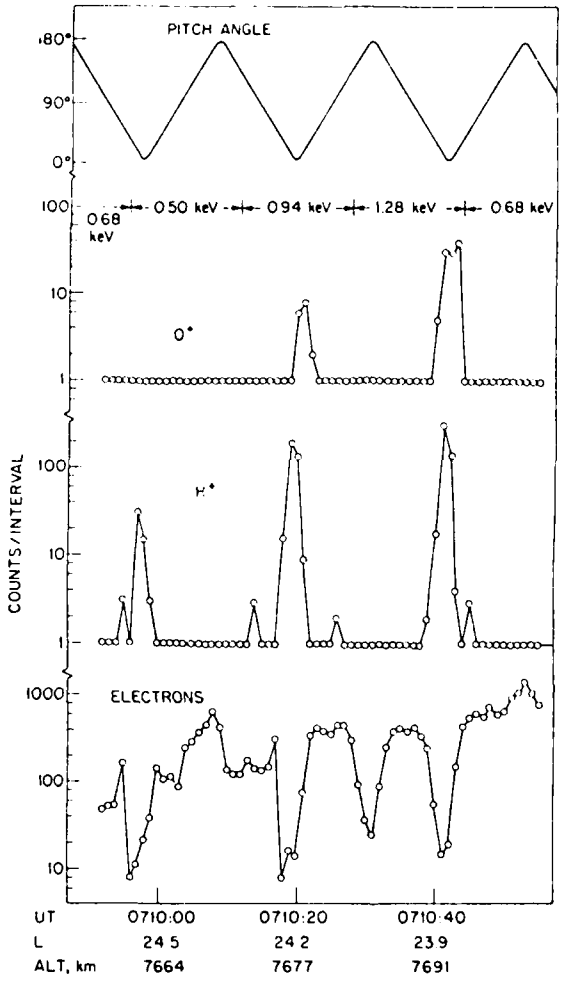


Figure 1. Data from Revolution 67 on 17 July 1976. The upper panel shows the pitch angle of the instrument axis. The two center panels show data from the mass spectrometer at the indicated energies, and the lower panel shows electron fluxes in the energy range from 0.37 to 1.28 keV.

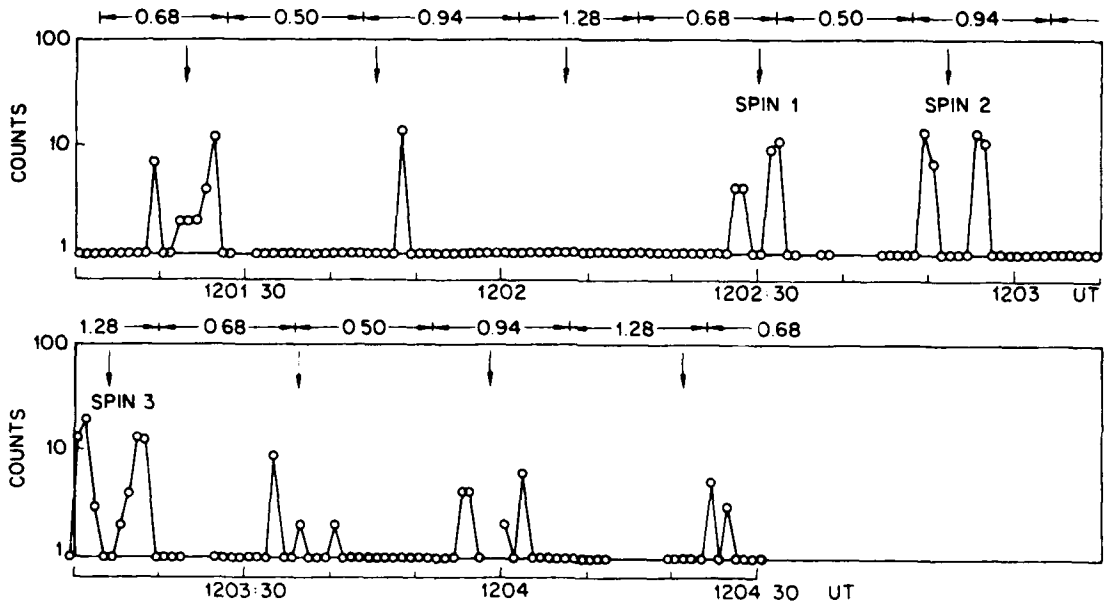


Figure 2. O^+ data from the lowest energy mass spectrometer (CXA1) at the times indicated, on 29 July 1976. The arrows indicate those times (once per spin) when the instrument was oriented nearly parallel to the magneitc field and pointing downward. The numbers above the panels are the energy-per-unit-charge setting of the spectrometer (in keV).

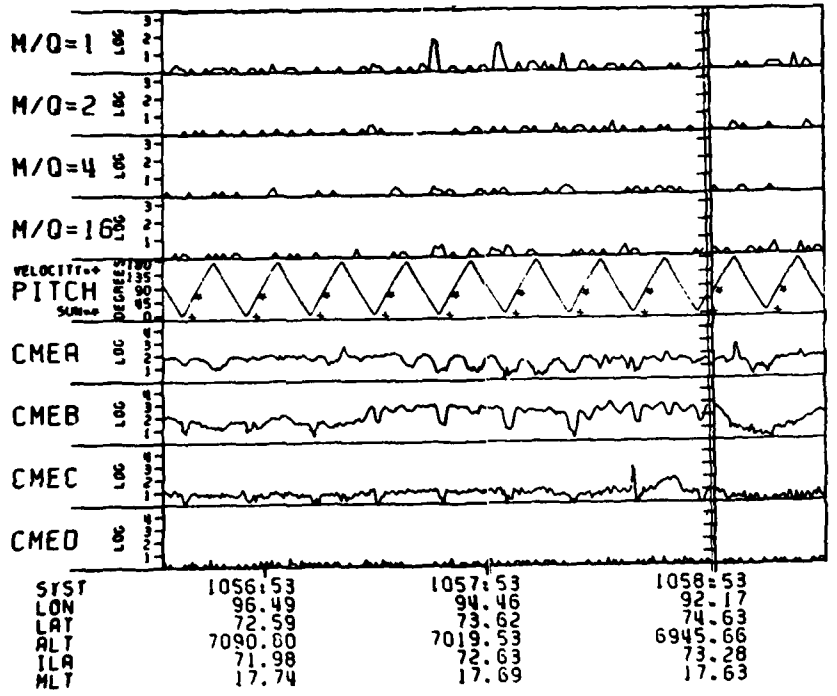


Figure 3. Survey plots of ion and electron data during an auroral acceleration event on 15 September 1976. See text for details.

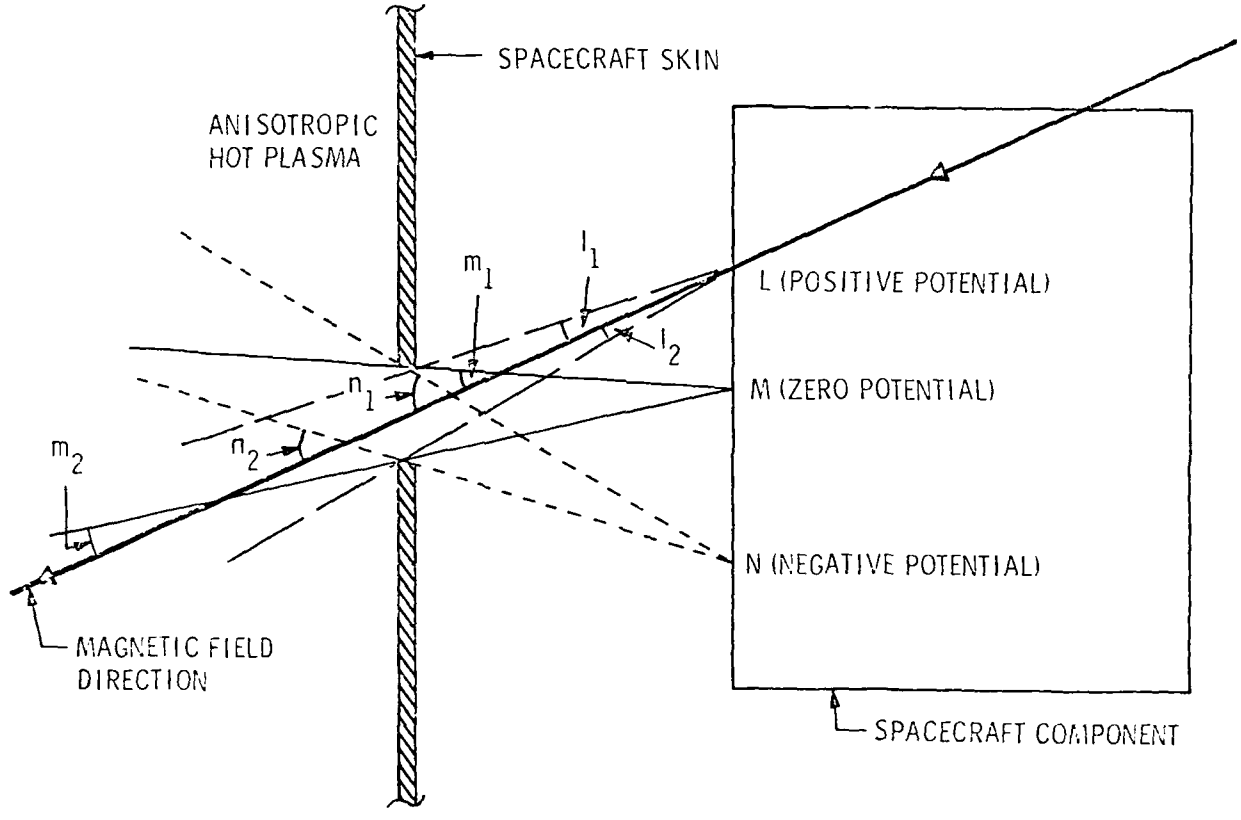


Figure 4. A schematic drawing illustrating the geometry for the pinhole camera charging effect in anisotropic hot plasmas.

EFFECTS OF AURORAL CLUTTER ON SPACE-BORNE MICROWAVE TRANSMISSIONS

H.L. Groginsky, A.H. Katz, D.B. Odom and G.D. Thome
Raytheon Company
Wayland, Massachusetts 01778

ABSTRACT

The effects of auroral clutter on space-borne microwave transmissions are likely to have a significant impact on the design, operation and performance of space-borne surveillance and navigation systems. System designers require knowledge of (1) the distribution in time and space of the occurrences of auroral clutter, (2) how the auroral clutter will affect the microwave transmission as a function of system parameters (such as transmitted frequency and beamwidth of transmissions), and (3) how auroral clutter will affect the microwave system as a function of the location of the space-borne platform with respect to the spatial distribution of auroral clutter. Global description of auroral clutter is required by space-borne systems in contrast to ground-based systems which require only a localized model. Based on available measurements, such a global description of the auroral clutter has been generated for use in the design of space-borne microwave systems. Further, this description has been used in conjunction with a geometrical model relating the position of the space-borne platform and the auroral clutter to generate maps which define where auroral clutter interactions are likely to occur.

1. INTRODUCTION

The design of space-based microwave systems requires knowledge of the backscatter clutter from ionospheric irregularities (such as auroral ionization, sporadic E, and spread F). This includes the geographical, temporal and height distribution of the ionized regions as well as its amplitude and spectral characteristics. Aurora gives rise to both scattering and dispersion of the radar signals. The former leads to an additive clutter signal in the radar while the latter behaves like multiplicative noise inducing some attenuation but principally producing an alteration of the spectrum of the radar signal. The regions of backscatter will occur principally where the radar line of sight is nearly normal to the geomagnetic field lines, while dispersion can occur wherever sufficient ionization exists. In this paper we have studied the problem of characterizing the regions in a space-borne radar's field of view producing the backscattered portion of the auroral induced clutter and the duration of these clutter echoes.

This paper represents an initial step in establishing a model of the space-borne radar auroral interaction. Its main objective is to illustrate the scope of the effort required to identify features of the backscatter clutter associated with space-borne systems. The analysis has been limited to the following conditions:

1. Irregularities which are modeled are associated with auroral clutter at E-layer heights. Further study is required to determine the effects of both F-layer and equatorial irregularities.

2. A single satellite height of 10,000 km was chosen for detail study. However, where appropriate, the effects of satellite height are described.

3. Only backscatter from auroral clutter has been investigated. Spectral dispersion due to transmission through the irregularity regions will be the subject of a future effort.

The key results of the analysis, described by this paper, are as follows:

1. Regions where auroral clutter can be observed can be very large rings, spanning up to 12 hours of longitude. The size of the rings increases with the altitude of the satellite.

2. The rings move with the satellite and thus the intersection of the rings with auroral events changes rapidly in space-time. This results in a very rapid change in the duration of the periods of observation of auroral clutter during a single satellite orbit.

3. The earth's rotation beneath the satellite's orbital plane can result in several local regions of perpendicularity that do not significantly change during an orbit. These regions in space-time will be susceptible to long duration auroral clutter dependent, of course, on the likelihood of the occurrence of auroral clutter in these regions.

This paper is organized as follows: Section 2 describes the unique geometrical factors associated with spaceborne measurements and develops the unique loci of perpendicularity. Section 3 discusses the amplitude, spatial and temporal distribution of the auroral clutter and develops the extent and duration of the perpendicular loci which might be involved in an auroral event.

2. SATELLITE OBSERVABLE ECHO REGIONS

Previous studies[1-4] have shown that ground-based radar observations of auroral clutter are geographically limited because of the aspect sensitivity of the backscatter. It is well known that auroral backscatter clutter occurs when the radar transmissions are perpendicular to the earth's magnetic field. Development of a model which predicts the effects of auroral activity on a satellite based observer is based on two factors:

1. The radar transmissions are perpendicular to the earth's magnetic field at a specified location (discussed in this section).

2. Auroral ionization occurs at the specified location.

Figure 1 shows the earth as viewed from a satellite. In the example chosen, namely a satellite in polar orbit at an altitude of 10,000 km, the radar commands a large view to the horizon ($\approx \pm 67^\circ$ latitude) and observes a large portion of the geomagnetic field. The analysis which locates the region of perpendicularity is based on a dipole approximation to the geomagnetic field at the altitude of 100 km.*

Consider any plane through the dipole axis. The geomagnetic field must lie in this plane. If one draws a line of constant altitude in this plane and then the line of sight from the satellite to any point on this locus, one may observe that the line perpendicular to the line of sight and that which lies in the defining plane is unique. In Figure 1 we have attempted to depict the variation of this line as a function of position from horizon to horizon. Clearly at the horizon, the direction of this line is perpendicular to the earth's surface and in between becomes tangent to the earth's surface at the point where the line of sight is perpendicular to the constant altitude locus. For a dipole field, it can be shown that the perpendicular line defined above makes the angle ψ with respect to the local vertical at a point defined by the geomagnetic latitude φ and longitude θ, which is given by

$$\tan \psi = \frac{\cos \alpha_m - \cos \varphi \cos \alpha_\varphi}{\sin \varphi \, (k - \cos \alpha_\varphi)} \quad (1)$$

*Cohen and Dwarkin[5] solved the ideal dipole case analytically using a method based on earlier work of Chapman.[6] The method used here allows measured earth models to be used to modify the dipole field model to determine the regions of perpendicularity.

where

$\cos \alpha_m$ = direction cosine between the earth centered magnetic dipole axis and the satellite

$\cos \alpha_\varphi$ = direction cosine between the satellite vector and a vector (from earth center) to the point in question in the dipole field (defined by φ and ℓ

$$k = \frac{R_e + h_{ion}}{R_e + h_s}$$

R_e = earth's radius

h_{ion} = height of the ionosphere

h_s = height of the satellite in compatible units

It is possible to invert (1) and solve for the geomagnetic latitude φ along a geomagnetic meridian at which the required dip ψ is found. By joining such points for different meridians it is possible to plot lines of constant required dip. Figure 2 shows the loci of constant required dip on a plot of actual geomagnetic dip for one location of the satellite. The earth surface location of the magnetic pole has been used to establish the geomagnetic meridians. Earth rotation is easily accounted for in this calculation by suitably locating the satellite. The overlay of the plots allows the matching of the required and actual dips and thereby permits a manual solution of the locations where required and actual dips match. This procedure neglects the distortions to the field induced by geomagnetic storm activity but is thought to be fairly accurate at the altitudes from which the backscatter originates. The heavy line on Figure 2 is the solution we have found. Regions of perpendicularity were determined for two satellite orbits illustrated in Figure 3. The regions of perpendicularity were calculated when subsatellite points were at the number locations.

Figures 4A and 4B show the results of the two satellite passes on a polar projection map. The heavy lines on the figures are the regions of perpendicularity within the radar coverage and are coded to permit interpretation in time sequence (see Figure 3).

The figures indicate that as the satellite traverses northward, the locus of perpendicularity appears as an arc. It then expands rapidly to a circle near the radar horizon. It travels along with the satellite as it traverses the pole, and again the circle collapses to an arc as the satellite passes to lower altitudes.

It is important to note that, in general, the region traverses the earth at the speed of the subsatellite point. In fact, if the effect is observable only between $\pm 1^\circ$ of perpendicularity, a point on the ground near the satellite orbit plane might be masked by ionospheric induced clutter for only 106 seconds.

Because of the satellite motion, e.g. for satellites below geosynchronous altitude, parts of the coverage in the easterly direction can have much longer persistence in the perpendicular condition. This effect, called the zone of convergence, is shown in Figure 4. Clearly, ionospheric disturbances at such regions can have more serious effects on a radar than elsewhere.

3. CHARACTERISTICS OF AURORAL (IONOSPHERIC) IRREGULARITIES

This section deals with the characteristics of the auroral ionospheric irregularities. This, together with the geometrical model described above, determines where and how often auroral clutter can be observed. Ground-based measurements[1-4] have demonstrated the following:

1. Aspect sensitivity appears independent of frequency with the clutter cross-section decreasing by approximately 10 dB per degree of perpendicularity.

2. Doppler spectra at 1295 MHz varies from coherent signals displaced 2.5 kHz with 6.5 kHz spectral width to spectral widths which are difficult to characterize and can be in excess of 20 kHz.

3. Cross-section versus frequency varies as a function of storm conditions by λ^7 in the 1000 to 3000 MHz region and λ^3 in the 100 to 1000 MHz regime.

3.1 Auroral Clutter Cross-Section

In order for an auroral clutter model to be useful for making system tradeoff decisions the intensity of the clutter and its variation with operating frequency must also be predictable, at least on a statistical basis. Unfortunately, at frequencies above a few hundred MHz, auroral clutter measurements are available from only a few isolated spots on the earth and for only limited periods of time. As might be expected of a phenomenon which is highly aspect sensitive and highly variable in time and space, measurements of volume scattering coefficient made at different places and at different times but at the same frequency rarely agree to within an order of magnitude and do not lead to a meaningful statistical model for system design purposes. For example at L-band (1295 MHz), observations during 1970 using the Millstone Hill radar led to a volume scattering coefficient of about 10^{-13} m^2/m^3, whereas observations with the same radar equipment the next year led to a value two orders of magnitude smaller.[2] On the other hand, an L-band volume scattering coefficient two orders of magnitude greater (10^{-11} m^2/m^3) has been reported based on measurements made in Alaska.[3]

The best way to develop a statistically meaningful model of the auroral clutter environment for system design studies of space-borne radars is to make synoptic measurements from space in the frequency band of interest. Taking 10^{-13} m^2/m^3 as a typical volume scattering coefficient at L-band and assuming that the scattering irregularities are confined to a layer 10 km thick, Table I illustrates that a scattering cross-section of order 1 m^2 can be expected from the main beam for a radar in orbit at 10,000 km. For this example, an antenna aperture of 25 m and a range resolution of 10 km has been assumed. The corresponding "spot size" of the radar beam (3 dB contour) at a range of 10,000 km is 100 km in diameter. If a volume scattering coefficient at L-band of 10^{-11} m^2/m^3 is assumed instead, a clutter cross-section of order 100 m^2 would be expected.

Table I
Clutter Cross-Section Calculation

Radar Parameters
Frequency: 1200 MHz
Aperture: 25 m
Range to Aurora: 10,000 km
Range Resolution: 10 km
Auroral Parameters
Volume Scattering Coefficient: 10^{-13} m^2/m^3
Scattering Layer Thickness: 10 km
Resulting Clutter Cross-Section: 1 m^2

The sidelobes of the hypothesized antenna may, of course, see a much larger geographic zone of irregularities than the spot size illuminated by the main beam. These extended auroral scattering zones are expected to be similar to those described in Section 3.2. The effect of the return from this larger volume scatterer will depend upon both the antenna sidelobe characteristics as well as the property of the scattering coefficient throughout this larger volume. Values for the clutter cross-section which may be in excess of 100 m^2 are possible for the viewing geometries shown in Section 3.2 during large portions of the day. An important sidelobe suppression requirement which the system designer must address can thus be characterized by the statistical model which would be developed.

3.2 Geographical and Temporal Distribution of Auroral Clutter

The times of maximum occurrence of the E-layer irregularities are shown in Figures 5A and 5B.[7] The most probable time of occurrence of this phenomenon is at night and is more likely to occur at earlier times as the latitude increases. For instance, it varies from about 1000 LMT at Winnipeg to about 1500 LMT at Resolute Bay. The lines of constant local time (Figure 5B) are not concentric with geographic latitudes, but are centered approximately on the geomagnetic pole. The geomagnetic storm activity producing these localized zones of enhanced E-layer ionization is believed to be caused by late arriving

particles (at an energy level of 1 MeV or less) that were ejected 20-40 hours earlier during the initiation of a major solar flare.[8,9]

The data shown in Figure 6 suggests a relationship between the observed E-layer plasma enhancement in the northern latitudes and a similar ionospheric enhancement observed at F-region heights in the mid-latitudes.[10] This relationship is further supported by the fact that the E-layer plasma irregularities in the mid and equatorial latitudes occur at the same time as the F-region enhancement.[11] The magnitude of the maximum increase in plasma density during storm conditions in both the E and the F-regions varies as a function of geomagnetic latitude with the intensity of the plasma disturbance observed to increase significantly in both the auroral and equatorial zones.

In Figure 6 we see that the period associated with the "storm-induced" ionization enhancement is typically in excess of eight hours. Thus, conditions leading to auroral echoes can persist for four (or more) hours on either side of the maximum occurrence "UT contours" shown in Figure 5. The echoes may then be observed on several orbits of a low altitude satellite or for a large portion of a single orbit of a medium to high altitude satellite.

Figures 5 and 6 are used to derive the geographic auroral clutter zones as a function of time. Two examples are shown in Figures 7A and 7B. In these figures, the contours of perpendicularity observed from a space-borne platform are also shown. The intersection of the clutter zones and regions of perpendicularity identify where a space-based radar can observe auroral clutter. Depending on the area illuminated by the radar the auroral clutter will either be in the mainlobe or sidelobe of the space-based radar antenna system.

The intersection which takes place between the clutter zones shown in Figure 7 and the regions of perpendicularity during a 2-hour orbit shown in Figure 4 can be used to provide a picture of the time-spatial history of the clutter during an orbit. Figure 8 shows this history for the orbits in Figure 3 for 2000 UT and 0400 UT. The occurrence of auroral clutter from one satellite pass to another is highly dependent upon both the characteristics of the orbit and the time at which the orbit occurs.

Figure 8B shows a "zone of convergence" of the contours of perpendicularity which occurs near the state of Utah. The 0400 UT clutter zone is near this "zone of convergence." Clutter which falls within this "convergence zone" will be seen for a large part of the orbit. The location of these geographic convergence areas and the position of the observed clutter zones is of considerable importance to the system designer. For example, in Figure 8A the interaction of the 2000 UT clutter zone with the perpendicularity aspect contours does not include a zone of convergence. This means that for the orbit at 2000 UT the occurrence of auroral clutter will be restricted to one or two parts of the orbit and not to a large portion of the satellite observation time as was the case in Figure 8B.

4. SUMMARY

This study has demonstrated analytically that the auroral backscatter clutter seen by a satellite radar observer can be spread over a large geographic area. For low to medium altitude satellites, a given location can be masked by auroral induced clutter for periods which are several seconds to many minutes in duration. Earth rotation and satellite motion can combine on certain orbits to yield a "zone of convergence" producing auroral induced clutter which can mask the satellite radar for nearly half an orbit (up to several hours).

As a satellite orbits the earth the regions of perpendicularity, if any, follow this motion. This is a consequence of (1) location of the platform in high earth orbit and (2) the spatial orientation of the earth's magnetic field with respect to this platform. The intersection of perpendicularity regions with auroral clutter is summarized in Figure 9 for the northern hemisphere satellite passes (see Figure 3). The longitudinal extent of the clutter for each subsatellite point

(specified in UT hours) for each satellite orbit is shown in this figure.

The orientation of the satellite radar with respect to the geomagnetic field and the auroral clutter zone changes as a function of Universal Time. The data displayed in Figures 9A and 9B show two such orientations illustrating the effect which a change in time can have on the auroral clutter observed during a given satellite pass.

The resulting auroral clutter interaction changes rapidly throughout each orbit. As the satellite moves the time and location at which auroral clutter is observed varies as a function of the time in orbit. Thus, a space-borne surveillance radar should be prepared to accept and compensate for time varying losses in sector coverage. A major concern is the occurrence of the zones of convergence which can result in auroral clutter induced loss in coverage during most of the orbit for selected geographic locations.

A procedure has been described for computing the locus of points in space from which auroral clutter can be expected for a space-borne radar. Examples have been given for the case of a radar in polar orbit at 10,000 km. It is straightforward to calculate the positions in space where the geometrical conditions for observing auroral clutter are satisfied (orthogonality between the radar beam and the earth's magnetic field). Whether or not clutter will in fact be observed from these "permitted" positions depends on the radar sensitivity and the intensity of the electron density irregularities which are present. A satisfactory statistical model is not available for the probability that irregularities of a given intensity will be present at a given point in the ionosphere. Radar observations from space are well suited for developing such a model and would provide a better basis than now exists for predicting the impact of auroral clutter on operational space-borne radar systems.

Clearly some experimental program is needed to confirm these findings. The system designer, however, must appreciate that satellite based measurements will essentially sample an observable clutter event with a rapidly moving spatial window. The satellite geometry restricts the sampling rate of these events to the orbital period with the "optimum" event sampling being achieved only in restricted latitude regions due to orbital coverage limitations.

REFERENCES

1. Chestnut, W.G., J.C. Hodges and R.L. Leadabrand, Auroral Radar Backscatter Studies from Homer, Alaska, AGARD Conf. Proc. No. 97 on Radar Propagation in the Arctic, Jan. 1972.

2. Hagfors, T., Some Properties of Radar Auroral Echoes as Observed at a Frequency of 1295 MHz, Ibid 1.

3. Chestnut, W.G., Polar Propagation Effects on VHF-UHF Radars, Ibid 1.

4. Evans, J.V., Millstone Hill Radar Propagation Study: Scientific Results, Part III, MIT Tech. Rpt. 509, 13 November 1973.

5. Cohen, M.H., M.L. Swarkin, "Perpendicular Intersections with a Dipole Field," J. of Geophysical Res., Vol. 66, No. 2, Feb. 1961, pp. 411-419.

6. Chapman, S., "The Geometry of Radio Echoes from the Aurorae," J. of Atmosph. and Terrest. Physics, Vol. 3, 1952, pp. 1-29.

7. Thomas, L., "The Temporal Distribution of Storm-Type Sporadic E in the Northern Hemisphere," Some Ionospheric Results - International Geophysical Year, Elseweir Publ. Co., 1960.

8. Obayashi, T. and Hakura, J., "Enhanced Ionization in the Polar Ionosphere," Proc. 1st Internat. Space Sci. Symp. - Space Research, North Holland Pub. Co., 1960.

9. Davies, K., "Ionospheric Radar Propagation, NBS, US Dept. of Commerce, April 1, 1969.

10. Matsushita, S., "Morphology of Ionospheric Storms," J. Geophys. Res., Vol. 64, No. 3, March 1959.

11. Smith, E.K., Worldwide Occurrence of Sporadic E, NBS Circ. 582, 1957.

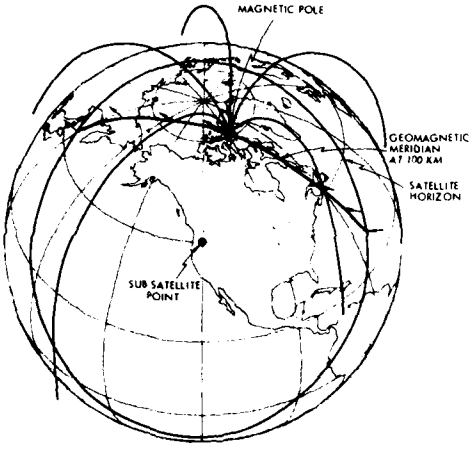


Figure 1 Geometry of 10,00 km Satellite and Earth's Magnetic Field

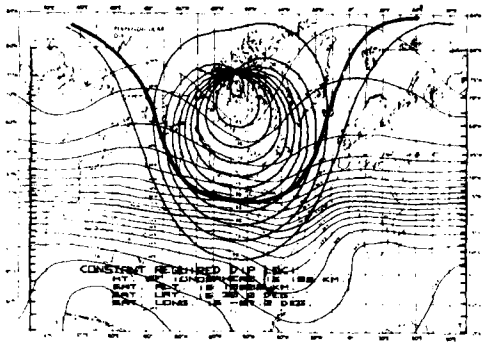


Figure 2 Loci of Constant Required Dip (to achieve perpendicularity) on Plot of Actual Geomagnetic Dip for One Satellite Location

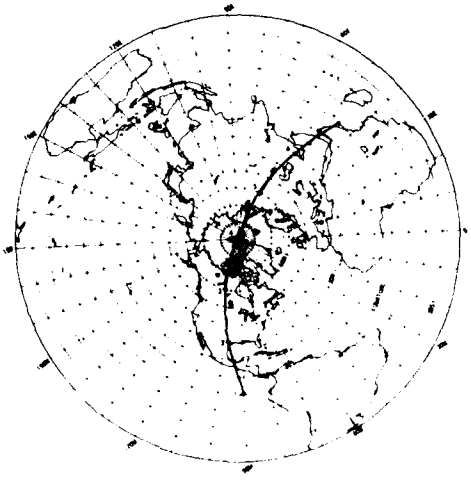


Figure 3A Satellite Orbit Used to Calculate Regions of Perpendicularity for Pass 1

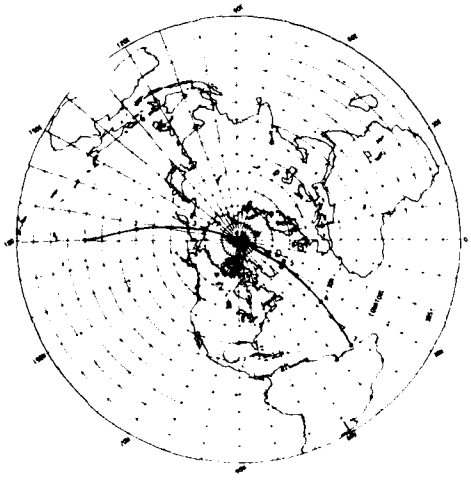


Figure 3B Satellite Orbit Used to Calculate Regions of Perpendicularity for Pass 2

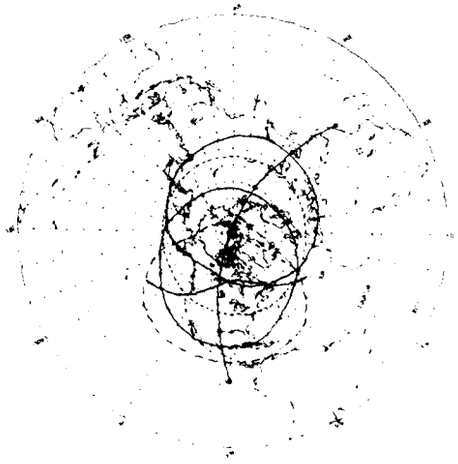


Figure 4A Regions of Perpendicularity for Satellite Orbits Shown in Figure 3A

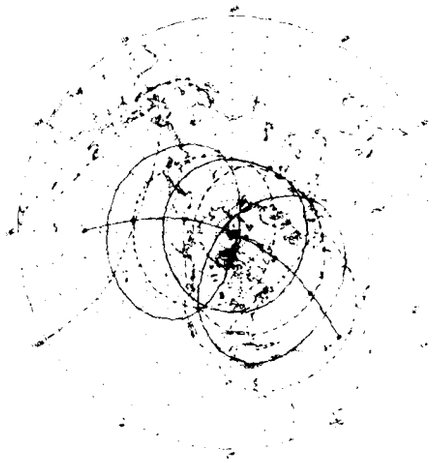


Figure 4B Regions of Perpendicularity for Satellite Orbits Shown in Figure 3B

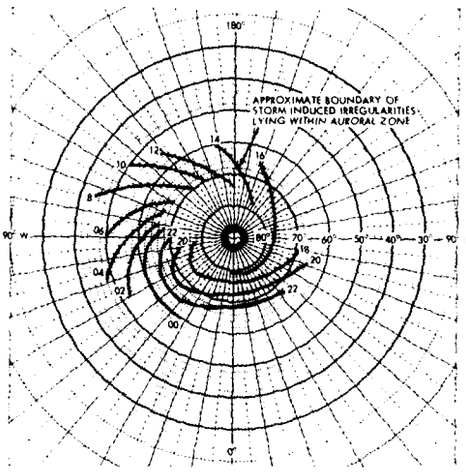


Figure 5A Universal Time of Maximum Frequency of Occurrence of Storm-Induced E Layer Plasma Irregularities

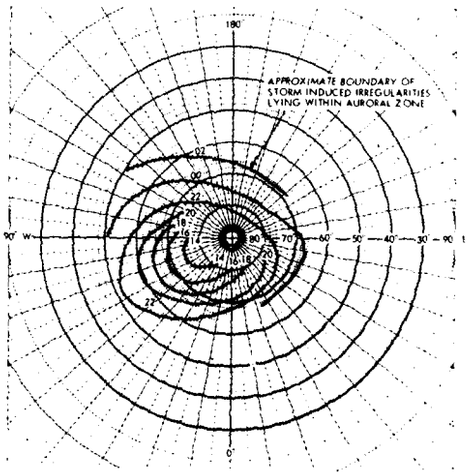


Figure 5B Local Mean Time of Maximum Frequency of Occurrence of Storm-Induced E Layer Irregularities

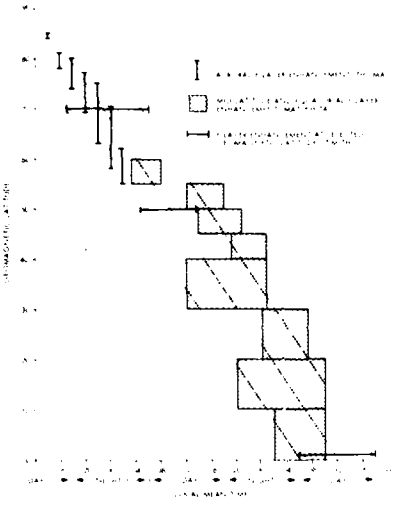


Figure 6 Observed Time of Maximum Occurrence of Storm-Induced Ionospheric Plasma Enhancement as a Function of Geomagnetic Latitude (after Thomas, 1960; Matsushita, 1959; and Smith, 1957)

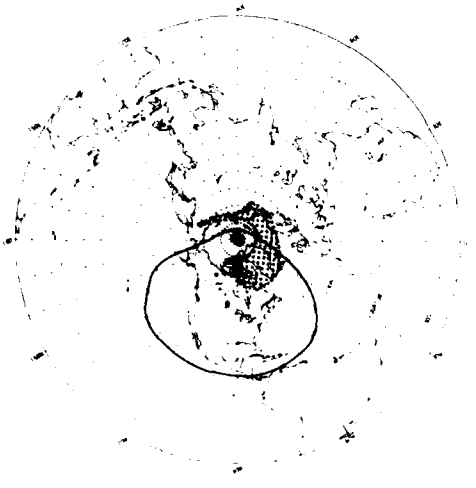


Figure 7A Auroral Clutter Zone Present at 2000 UT with Perpendicular Aspect Concour Associated with Early Portion of Satellite Pass #1

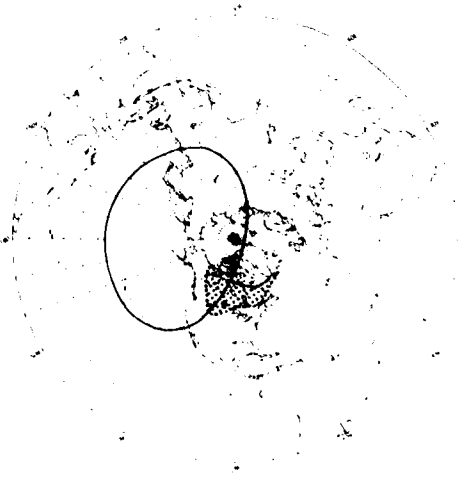


Figure 7B Auroral Clutter Zone Present at 0400 UT with Perpendicular Aspect Contour Associated with Early Portion of Satellite Pass #2

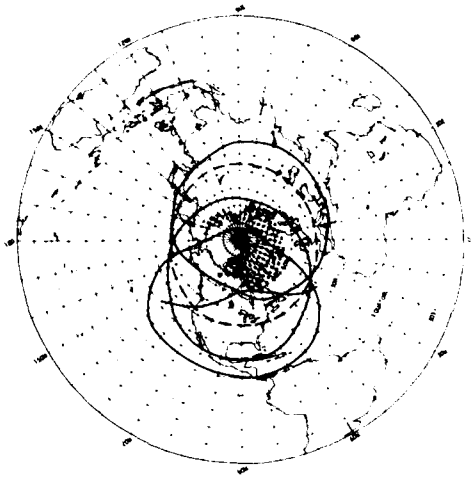


Figure 8A Auroral Clutter Zone Present at 2000 UT with Perpendicular Aspect Contours for Pass #1

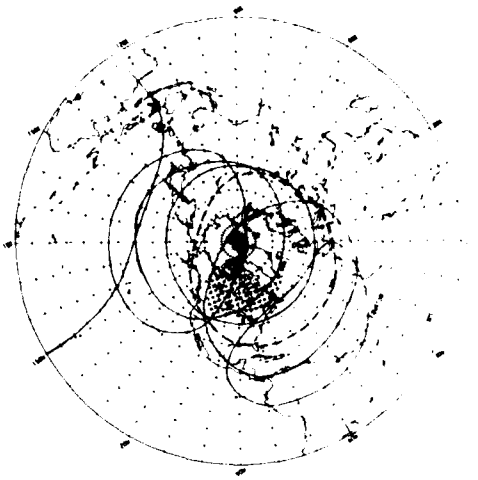


Figure 8B Auroral Clutter Zone Present at 0400 UT with Perpendicular Aspect Contours for Pass #2

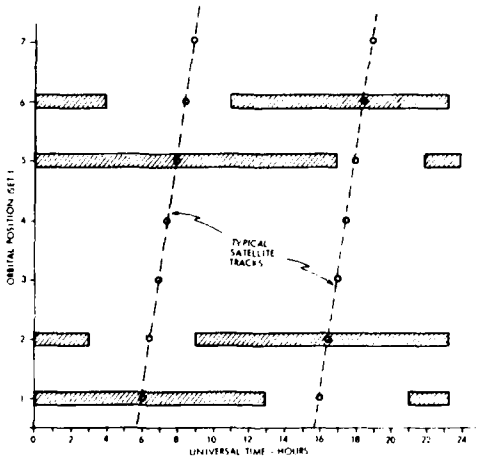


Figure 9A Auroral Clutter Occurrence for Orbital Positions of Satellite Track Shown in Figure 3A

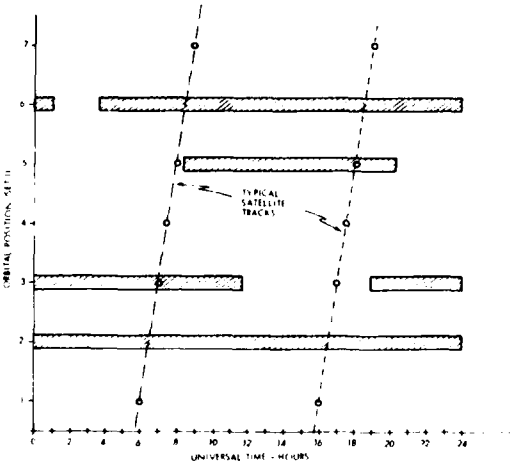


Figure 9B Auroral Clutter Occurrence for Orbital Positions of Satellite Track Shown in Figure 3B

WORLDWIDE MINIMUM ENVIRONMENTAL RADIO NOISE LEVELS (0.1 Hz to 100 GHz)

A. D. Spaulding* and G. H. Hagn**
*Office of Telecommunications, Boulder, Colorado
**SRI International, Arlington, Virginia

ABSTRACT

It is desirable to design receiving systems so that the receiver will be close to being limited by external noise. Indeed, it is cost ineffective to design radio receiving systems with noise figures significantly less than the minimum values of the environmental radio noise. The best available estimates of the minimum values of the environmental radio noise for vertically polarized antennas expected worldwide (0.1 Hz to 100 GHz) presented in this paper should facilitate cost-effective future receiver design.

PRE-DETECTION SIGNAL-TO-NOISE RATIO AND RECEIVING SYSTEM OPERATING NOISE FACTOR

The pre-detection signal-to-noise ratio(S/N) is an important telecommunications system design parameter. It is useful to refer (or translate) the noise from all sources to one point in the system for comparison with the signal power (desired signal). A unique system reference point exists: the terminals of an equivalent lossless antenna having the same characteristics (except efficiency) as the actual antenna (see CCIR Report 413). Consider the receiving system shown in Fig. 1. The output of block (a) is this unique reference point. The output of block (c) represents the actual (available) antenna terminals to which one could attach a meter or a transmission line. Let s represent the signal power and n the average noise power in watts which would be observed at the output of block (a) in an actual system (if the terminals were accessible). We can define a receiving system overall operating noise factor, f, such that $n = fkt_ob$, where

- k = Boltzmann's constant = 1.38×10^{-23} J/K,
- t_o = The reference temperature in K taken as 288K,
- b = the noise power bandwidth of the receiving system in Hz.

We can also define a system overall operating noise figure $F = 10 \log_{10} f$ in dB. The ratio s/n can be expressed:

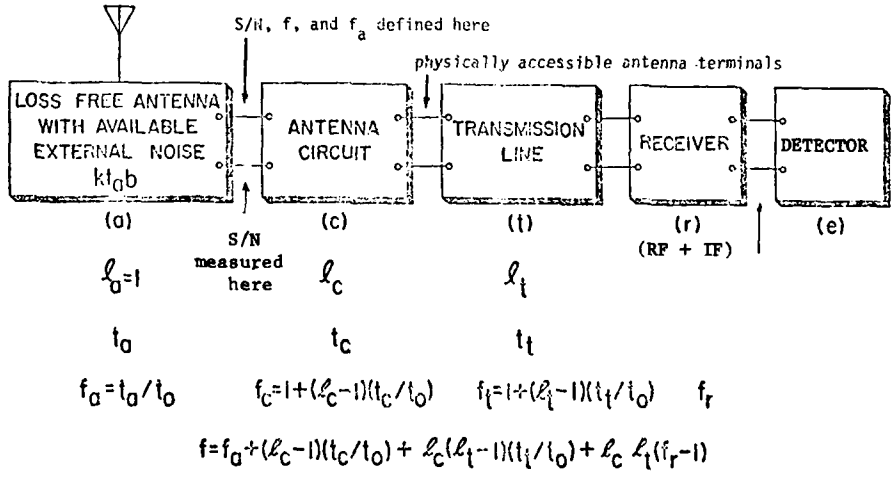


Figure 1. The receiving system and its operating noise factor, f.

$$10 \text{ LOG}_{10}(s/n) = S-N \qquad (1)$$

where

S = the desired average signal power, in dB(1W)
= 10 $\log_{10}$s, and
N = the average system noise power in dB(1W)
= 10 $\log_{10}$n.

Let us now explore the components of n in greater detail, with emphasis on environmental noise external to the system components.

For receivers free from spurious responses, the system noise factor is given by

$$f = f_a + (\ell_c - 1)\left(\frac{t_c}{t_o}\right) + \ell_c\,(\ell_t - 1)\left(\frac{t_t}{t_o}\right) + \ell_c \ell_t\,(f_r - 1) \qquad (2)$$

where,

f_a = the external noise factor defined as

$$f_a = \frac{p_n}{kt_o b}\,, \qquad (3)$$

F_a = the external noise figure defined as F_a = 10 log f_a
p_n = the available noise power from a lossless antenna (the output of block (a) in Fig. 1).
ℓ_c = the antenna circuit loss (power available from lossless antenna/power available from actual antenna)
t_c = the actual temperature, in K, of the antenna and nearby ground
ℓ_t = the transmission line loss (available input power/available output power)
t_t = the actual temperature, in K, of the transmission line
f_r = the noise factor of the receiver (F_r = 10 log f_r = noise figure in dB).

Let us now define noise factors f_c and f_t, where f_c is the noise factor associated with the antenna circuit losses,

$$f_c = 1 + (\ell_c - 1)\left(\frac{t_c}{t_o}\right), \qquad (4)$$

and f_t is the noise factor associated with the transmission line losses,

$$f_t = 1 + (\ell_t - 1)\left(\frac{t_t}{t_o}\right). \qquad (5)$$

If $t_c = t_t = t_o$, (2) becomes*

$$f = f_a - 1 + f_c f_t f_r\,. \qquad (6)$$

*Note specifically, that when $f_c = f_t = 1$ (lossless antenna and transmission line) then $F \neq F_a + F_r$.

Relation (3) can be written

$$P_n = F_a + B - 204 \text{ dB(1W)}\,, \qquad (7)$$

where P_n = 10 log p_n (p_n = available power at the output of block (a) in Fig. 1, in watts), B = 10 log b, and $-204 \approx 10 \log kt_o$. For a short ($h \ll \lambda$) grounded vertical monopole, the vertical component of the rms field strength is given by

$$E_n = F_a + 20 \log f_{MHz} + B - 95.5 \text{ dB}(1\mu V/m) \qquad (8)$$

where E_n is the field strength in bandwidth b and f_{MHz} is the center frequency in MHz. Similar expressions for E_n can be derived for other antennas (Lauber, 1977). For example, for a half-wave dipole in free space,

$$E_n = F_a + 20 \log f_{MHz} + B - 89.9 \text{ dB}(1\mu V/m). \qquad (9)$$

The external noise factor is also commonly expressed as a temperature, t_a, where, by definition of f_a

$$f_a = \frac{t_a}{t_o}\,, \qquad (10)$$

and t_o is the reference temperature in K and t_a is the antenna temperature due to external noise.

More detailed definitions and discussions (including the case with spurious responses) are contained in CCIR Report 413 (1966), and supplementary discussions on natural and man-made noise are provided in CCIR Reports 322 (1963) and 258 (1976) respectively.

RELATIONSHIPS AMONG F_a, NOISE POWER, SPECTRAL DENSITY AND NOISE POWER BANDWIDTH

Note that f_a is a dimensionless quantity, being the ratio of two powers. The quantity f_a, however, gives, numerically, the available power spectral density in terms of kt_o and the available power in terms of $kt_o b$. The relationship between the noise power, P_n, the noise power spectral density, P_{sd}, and the noise power bandwidth, b, are summarized in Fig. 2 (from Spaulding, 1976 and Hagn, 1978). When F_a is known, then P_n or P_{sd} can be determined by following the steps indicated in the figure. For example, if the minimum value of F_a = 40 dB and b = 10 kHz, then the minimum value of noise power available from the equivalent lossless antenna is P_n = -164 dB(1W). If ℓ_t = 3, then the noise power available from the receiving antenna transmission line is -172 dB(1W).

ESTIMATES OF MINIMUM ENVIRONMENTAL NOISE LEVELS

The best available estimates of the minimum expected values of F_a along with other external noise levels of interest are summarized in this section as a function of frequency. Fig. 3 covers the frequency range 0.1 Hz to 10 kHz. The solid curve is the minimum expected values of F_a at the earth's surface based on measurements

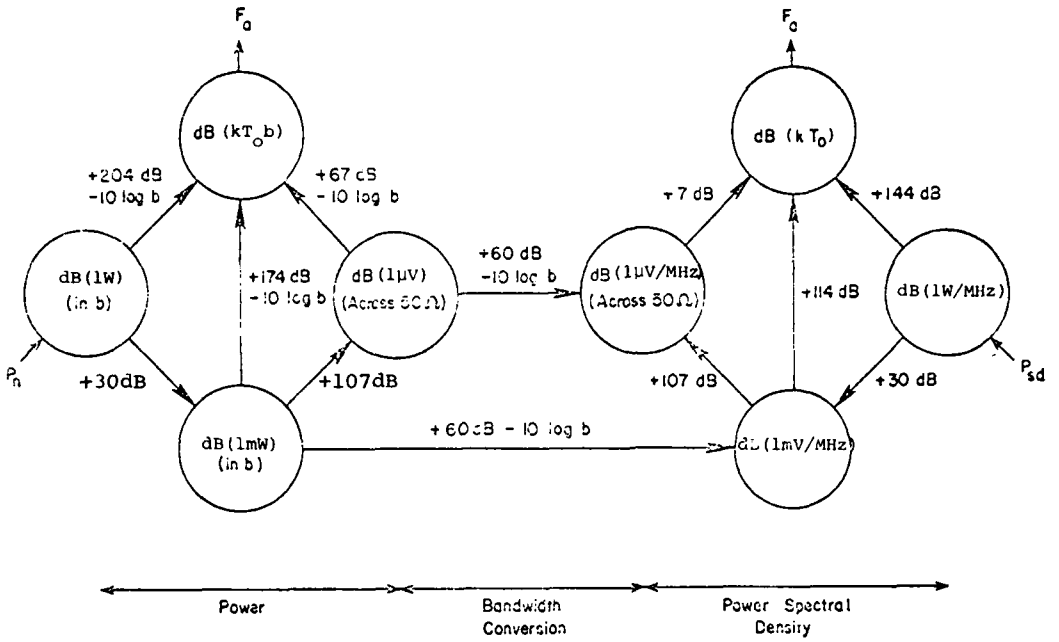


Figure 2. Relationships between power, power spectral density, and bandwidth (rsms detector).

(taking into account all seasons and times of day for the entire earth) and the dashed curve gives the maximum expected values. Note that in this frequency range there is very little seasonal, diurnal, or geographic variation. The larger variability in the 100-10,000 Hz range is due to the variability of the earth-ionosphere waveguide cutoff.

Fig. 4 covers the frequency range 10^4 - 10^8 Hz, i.e., 10 kHz - 100 MHz. The minimum expected noise is shown via the solid curves and other noises that could be of interest as dashed curves. For atmospheric noise, ($f>10^4$ Hz) the minimum values expected are taken to be those values exceeded 99.5% of the time and the maximum values are those exceeded 0.5% of the time. For the atmospheric noise curves, all times of day, seasons and the entire earth's surface have been taken into account. More precise details (geographic and time variations) can be obtained from CCIR Report 322 (1963). The man-made noise (quiet receiving site) is that noise measured at carefully selected, quiet sites worldwide as given in CCIR Report 322. The atmospheric noise below this man-made noise level was, of course, not measured and the levels shown are based on theoretical considerations [CCIR, 1963, and references therein], and engineering judgment (Crichlow, 1966). Also shown is the median expected business area man-made noise. Further details concerning man-made noise and its variation can be obtained from CCIR Report 258 (1976), Spaulding and Disney (1974) and references therein, and Hagn and Shepherd (1974).

On Fig. 5, the frequency range 10^8 - 10^{11} is covered, i.e., 100 MHz - 100 GHz. Again, the minimum noise is given by solid curves while some other noises of interest are again given by dashed curves.

The majority of the results shown on the three figures are for omni-directional vertically polarized antennas (except as noted on the figures). The average value of F_a for directional antennas will be the same if we assume random direction. Studies have indicated that at HF (for example), for atmospheric noise from lightning, there can be as much as 10 dB variation (5 dB above to 5 dB below the average F_a value shown) with direction for very narrow-beam antennas.

For galactic noise, the average value (over the entire sky) is given by the solid curve labled galactic noise (Figures 4 and 5). Measurements indicate a ± 2 dB variation about this curve. The minimum galactic noise (narrow-beam antenna towards galactic pole) is 3 dB below the solid galactic noise curve shown on Figure 5. The maximum galactic noise for narrow-beam antennas is shown via a dashed curve on Figure 5.

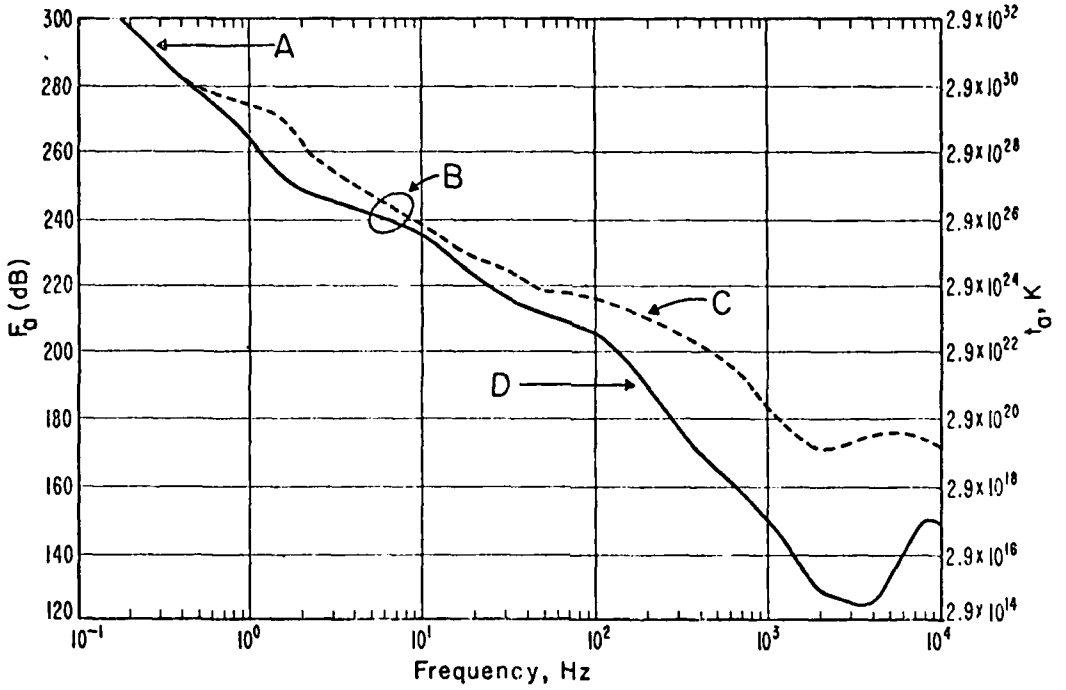


Figure 3. F_a, minimum and maximum, versus frequency (0.1 to 10^4 Hz).

A: Micropulsations
B: Atmospheric noise from lightning
C: Maximum expected
D: Minimum expected

EXAMPLE DETERMINATION OF REQUIRED RECEIVER NOISE FIGURE

We now want to consider a simple example to show how to determine the required receiver noise figure. At 10 kHz, for example, the minimum external noise is F_a = 145 dB (see Fig. 4). If we assume $t_c = t_t = t_o$, and $\ell_c = \ell_t = 1$ (that is, no antenna or transmission line losses), then

$$f = f_a - 1 + f_r \ . \qquad (11)$$

We can take f_r to be that value which will increase F by only 1 dB. This gives us a noise figure, F_r, of 140 dB, or an overall noise figure, F, of 147 dB. Any smaller noise figure, F_r, no matter how small, cannot decrease F below 146 dB. Consider now that $\ell_c = \ell_t = 100$, i.e., 20 dB antenna losses and 20 dB transmission losses. Then,

$$f = f_a - 1 + 10000\ f_r \ . \qquad (12)$$

In order to raise the F no more than 1 dB (to 147 dB) for the above situation, F_r can only be as large as 100 dB.

CONCLUDING REMARKS

Throughout this paper we have considered noise as a source of system degradation and interference as the degradation produced (see Hagn, 1977 for a discussion of the basic definitions and rationale). This is consistent with current internationally accepted usage (CCIR, 1974). Finally, it should be noted that many of the external noises (e.g., many forms of atmospheric and man-made noise) are impulsive in nature. This means that the external noise can still limit performance of a communications system, even though the receiver noise is made as high as possible so as not to increase the overall operating noise factor f. That is, system performance depends not only on the level of the interfering noise power (signal-to-noise ratio), but also on the detailed statistical characteristics of the noise.

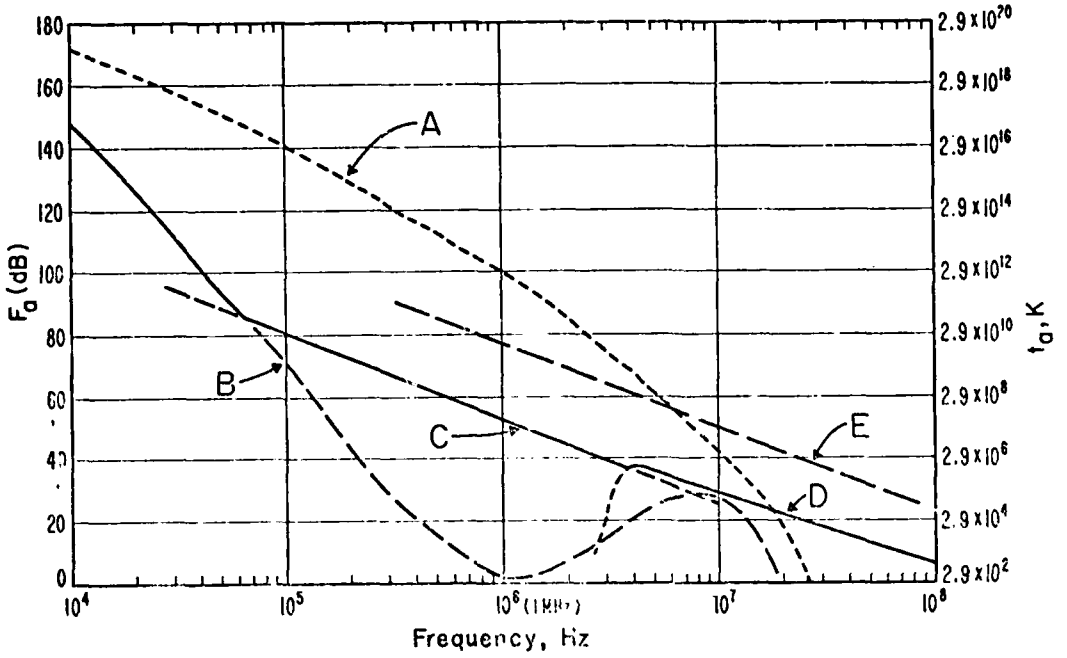


Figure 4. F_a versus frequency (10^4 to 10^8 Hz)

A: Atmospheric noise from lightning, value exceeded 0.5% of time
B: Atmospheric noise from lightning, value exceeded 99.5% of time
C: Man-made noise, quiet receiving site
D: Galactic noise
E: Median business-area man-made noise

REFERENCES

CCIR, "World Distribution and Characteristics of Atmospheric Radio Noise," Report 322, International Radio Consultative Committee, International Telecommunication Union, Geneva, 1963.

CCIR, "Operating Noise-Threshold of a Radio Receiving System," Report 413, International Radio Consultative Committee, International Telecommunication Union, Geneva, 1966.

CCIR, "Man-Made Noise," Report 258-2, International Radio Consultative Committee, International Telecommunication Union, Geneva, 1976.

CCIR, "Definitions of Interference," Report 529, International Radio Consultative Committee, International Telecommunication Union, Geneva, 1974.

W. Q. Crichlow, Private Communication, 1966.

G. H. Hagn and R. A. Shepherd, "Man-Made Electromagnetic Noise from Incidental Radiators: A Summary," Conference Proceedings No. 159, NATO/AGARD Meeting on Electromagnetic Noise, Interference and Compatibility, Paris, France, 21-25 October 1974, AGARD, 7 rue Anelle, 92200 Nevilly Sur Seine, France, November 1974 (available from NTIS, Springfield, VA as AD-A-018980).

G. H. Hagn, "Definitions of Electromagnetic Noise and Interference," 1977 IEEE International Symposium on Electromagnetic Compatibility, IEEE Catalog No. 77CH 1231-OEMC, pp. 122-127, IEEE, New York, August 1977.

G. H. Hagn, "Units for Electromagnetic Noise Measurements," IEEE Trans. EMC (to be published).

W. R. Lauber and J. M. Bertrand, "Preliminary Urban VHF/UHF Radio Noise Intensity Measurements in Ottawa, Canada," Electromagnetic Compatibility 1977, Proceedings of 2nd

Symposium on EMC, Montreux, Switzerland, June 28-30, pp. 357-362, IEEE Catalog No. 77CH 1224-5EMC, 1977.

A. D. Spaulding and R. T. Disney, "Man-Made Radio Noise, Part 1: Estimates for Business, Residential, and Rural Areas," Report OT 74-38, Office of Telecommunications, U.S. Department of Commerce, Boulder, Colorado, (available from the U.S. Government Printing Office, Washington, D.C. 20402).

A. D. Spaulding, "Man-Made Noise: The Problem and Recommended Steps Toward Solution," Report OT 76-85, Office of Telecommunications, U.S. Department of Commerce, Boulder, Colorado, 1976, (available from the U.S. Government Printing Office, Washington, D.C. 20402).

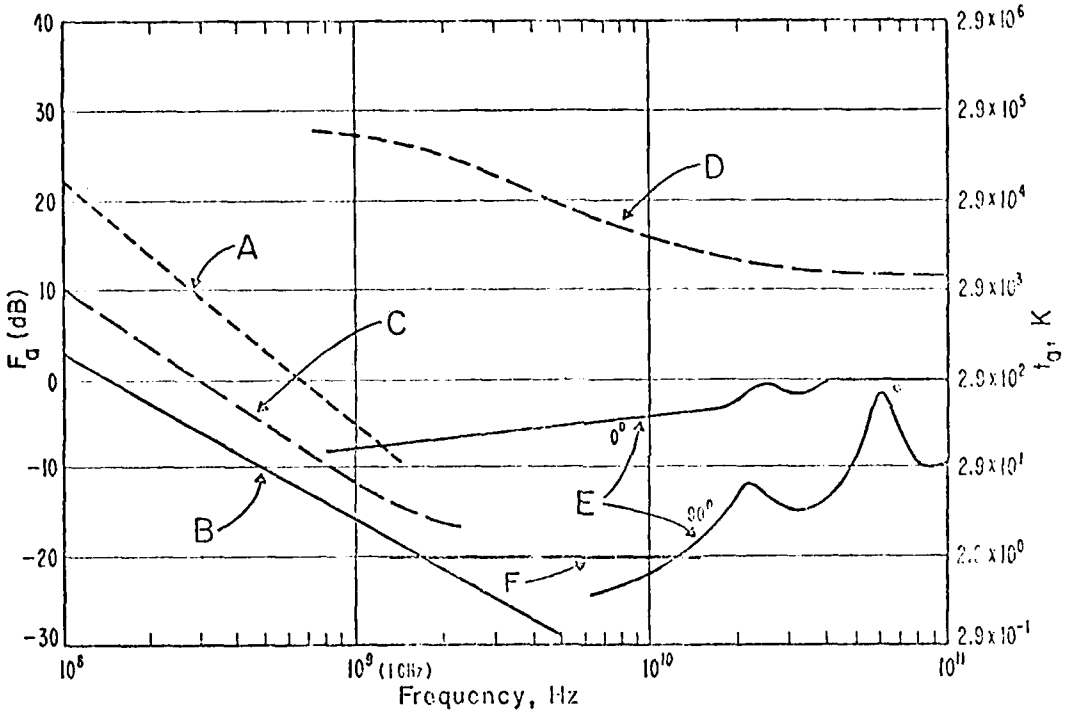


Figure 5. F_a versus frequency (10^8 to 10^{11} Hz).

A: Estimated median business-area man-made noise
B: Galactic noise
C: Galactic noise (toward galactic center with infinitely narrow beamwidth)
D: Quiet sun (½ degree beamwidth directed at sun)
E: Sky noise due to oxygen and water vapor (very narrow beam antenna); upper curve, 0° elevation angle; lower curve, 90° elevation angle
F: Black Body (cosmic background), 2.7 K

HF Noise in Space

C. M. Rush, AFGL*, D. Nelson, Aerospace Corp.,
A. L. Snyder, AFGL; V. Patterson and T. Tascione, AFGUC;
E. Ziemba, Analytical Sciences and Engineering Corp.

ABSTRACT

Radio noise of ground-based origin can be observed by a satellite-borne receiver orbiting above the F2 region. The noise that is observed is a function of receiver frequency, satellite position and local time at the sub-satellite point. A recently launched Defense Meteorological Satellite is equipped with a swept-frequency HF noise receiver and currently is providing data on a routine basis all along the satellite orbit. The receiver provides measurements of radio noise of terrestrial origin every 100 kHz in the frequency range 1.2 to 13.9 MHz. In this paper, we point out some of the prominent features of the noise environment at the height of the satellite as well as a brief discussion as to how the data can be used to infer the value of the F2 region critical frequency at the sub-satellite point.

INTRODUCTION

Radio noise measurements have been made at numerous ground-based locations for many years. These measurements were undertaken partly because noise can be a limiting factor in the performance of electromagnetic propagation systems. Ground-based measurements obviously are confined to regions determined by political and geographical considerations. With the advent of satellites, measurements of radio noise could be made within the ionosphere at all locations along the satellite orbit.

Most of the measurements of radio noise at satellite heights tend to deal with noise at frequencies other than HF. Notable exceptions are the studies undertaken by Horner and Bent (1969) utilizing observations recorded on-board the UK3 (Ariel 3) satellite and Herman et al. (1973) using data obtained by the Radio Astronomy Explorer (RAE-1) satellite. Further studies by Herman et al. (1975) attempted to show the degree to which the CCIR makes of ground-based atmospheric noise (CCIR, 1963) can be related to RAE-1 observations.

Recently Rush and Buchau (1977) pointed out the feasibility of determining the critical frequency of the F2 region (foF2) using satellite-borne noise measurements. Because the ionosphere behaves as a high-pass filter with respect to radio energy propagated through it from the ground, the lowest frequency that a satellite orbiting above the F2 region peak electron density will see, will ideally correspond to foF2. This would be the case if the ionosphere were perfectly homogeneous in latitude and longitude and noise sources of all frequencies were distributed uniformly over the surface of the Earth. Despite the fact that the ionosphere is not homogeneous nor are noise sources distributed uniformly over the Earth's surface, Rush and Buchau (1977), using data obtained by the ISIS2 satellite, were able to show that in regions of the globe where ground-based HF noise is high and the ionospheric structure does not display severe horizontal gradients, satellite measurements of noise of terrestrial origin can be used to determine foF2 generally to within one megahertz. Rush and Ziemba (1978) have expanded the study of Rush and Buchau (1977) by undertaking a detailed comparison between the value of foF2 observed by the ISIS2 satellite, the value of foF2 inferred from ISIS2 noise measurements, and the predicted values of foF2 for the appropriate time and location. Their results show a clear diurnal dependence in the usefulness of noise measurements in determining foF2. The night-time values of foF2 inferred from noise measurements tend to be within 0.33 MHz of the observed foF2. This compares to predicted values of foF2 being within 0.95 MHz of the observed values. During the daytime hours, however, the value of foF2 deduced from noise measurements tends to be on the order of 0.90 MHz away from the observed foF2. Predictions of daytime values of foF2 from the same data display similar errors, being on the order of 1.0 MHz when compared to actual observations.

Efforts are currently underway to determine the degree to which noise measurements obtained by the HF receiver operating on-board a recently launched Defense Meteorological Satellite (DMSP) compares to values of foF2 observed at a few

* Currently at the Office of Telecommunications, Institute for Telecommunication Sciences, Boulder, CO

ground-based ionosonde stations. Preliminary results obtained from this study are the subject of this paper along with some discussion on the general behavior of the HF noise intensity at the height of the DMSP orbit.

DMSP NOISE RECEIVER CHARACTERISTICS

The HF noise receiver currently operating on-board the DMSP satellite provides values of the logarithm of field strength in 100 kHz steps between the frequencies of 1.2 through 13.9 MHz. Characteristics of the receiver include:

Noise Level:	2μV
Dynamic Range:	60 db
Gain Reduction:	≈ 20 db
Effective Antenna Length:	≃ 1 m

The receiver continuously sweeps through the 128 frequency channels in 32 seconds. Thus, a value of the radio noise in a 100 kHz bin is obtained once every 32 seconds. Because the satellite contains a recording system, data are obtained throughout the entire orbit. This is in contrast to ISIS2 satellite data which were obtained only when the satellite was in the line-of-sight of a ground-based readout station.

HF NOISE DATA AS A FUNCTION OF FREQUENCY

Figure 1 gives an example of the HF noise signal strength as a function of frequency. The signal strength has been deduced by using the equation

$$E = 10^{\left[\frac{N - 232.5}{52.5}\right]}$$

where

E is the terminal voltage and
N is the number of counts in decimal.

It is readily apparent that at the highest end of the frequency scale, the data are more erratic than that at the lower end. This is characteristic of the behavior of radio noise in the topside ionosphere--at lower frequencies (below foF2) the receiver "sees" only background noise, whereas at the higher frequencies, noise of terrestrial origin reaches the satellite. The cause of the spikes in the signal strength at the very low frequency end is not certain at this time and instrumental problems cannot be totally ruled out. Figure 1 is a typical example of the behavior of the signal strength as a function of frequency as observed by the DMSP noise receiver.

Analysis of the DMSP noise data has indicated that there is a coupling of adjacent frequency channels (2.0 ⟷ 2.1, 2.2 ⟷ 2.3, etc.) but surrounding paired channels were uncoupled. It is believed that this coupling is due to a 200 kHz frequency resonance passing through the HF receiver that is centered between the channels with sufficient bandspread to overlap and couple paired channels. The source of the resonance is thought to be outside of the receiver itself. Although this coupling does not cause the signal strength in a given channel to be significantly enhanced, it does nonetheless render it difficult to determine the frequency at which noise of ground-based origin is first detected by the receiver. In order to determine this frequency, it was decided to subtract the counts in one of the coupled channels from those in the other channel. The result of subtracting all the coupled channels is shown in Figure 2 for the same data used in Figure 1. The form of Figure 2 is more in concert with expectation based on the results of the ISIS2 analysis of Rush and Buchau (1977). However, by subtracting the coupled channels, the sensitivity of the instrument is reduced, the effective channel separation becomes 200 kHz as opposed to 100 kHz and the subtracted values do not reflect absolute counts (or signal strength). This, however, does not affect the usefulness of the instrument for determining a measure of foF2 at the sub-satellite point.

foF2 INFERRED FROM DMSP NOISE DATA

The DMSP satellite that is equipped with the HF noise receiver is basically a morning-evening constant local time orbit. This means that twice a day, once in the morning and once in the evening, the satellite passes above the vicinity of a number of ground-based ionosonde stations. When the satellite was within the vicinity of a ground-based station, a comparison could be made between the value of foF2 determined by the increase in noise intensity at the satellite and the observed value of foF2 determined by the ground-based ionosonde. Although the satellite does not necessarily pass over the exact location of the ionosonde and at exactly the same time as the ionosonde is operating, the comparison is illustrative of the usefulness of determining foF2 under realistic conditions.

Thus far we have been able to isolate 22 cases where the satellite was close enough in time and space to a ground-based observation of foF2 to warrant a comparison between satellite and ground data. The results, obtained using data of the form given in Figure 2, are summarized as follows:

fin - foF2	Cases	%
0 < 0.2	8	36
<0.3 < 0.5	4	18
<0.6 < 1.0	6	27
>1.0	4	18

where fin is defined as the value of foF2 deduced from the increased noise intensity at the satellite height. A little more than one half of the cases indicate that fin is within

0.5 MHz of foF2, and a little more than 30 percent (18) of the cases show fin to be within 1.0 MHz of foF2. These results are, in general, consistent with those given by Rush and Buchau (1977). Although the analysis is still in the preliminary stage, it appears that the HF noise receiver will be useful under certain conditions in deducing sub-satellite values of foF2.

TOTAL NOISE INTENSITY AT SATELLITE HEIGHTS

As was mentioned above, the satellite is equipped with a recording system thereby enabling observations to be obtained over the entire orbit of the satellite. Also, since the satellite is basically in a sun-synchronous morning-evening orbit, by analyzing data from successive morning or evening passes, it is possible to obtain global maps of the noise intensity at the satellite altitude. For both the morning and evening passes, the satellite returns to the same general orbital position (apart from changes associated with orbit precession) every 24 hours. Thus, it is possible to obtain world maps of noise intensity every 24 hours. These maps can be used to assess the morphological behavior of the noise at satellite altitudes as a function of frequency since the receiver on-board the DMSP satellite sweeps the frequency range 1.2 to 13.9 MHz in 100 kHz steps every 32 seconds.

Current studies of the morphological behavior of the noise intensity are being undertaken using data of the type shown in Figure 1 for morning or evening passes observed over five (5) and fifteen (15) days rather than just one day. This is being done in order to obtain maps that are more representative of the average noise behavior and perhaps more in concert with the maps given in CCIR Report 322 (CCIR, 1963). The results of these studies thus far indicate a behavior that is in general in agreement with expectation. The noise intensity is higher over land areas than over the oceans, and the intensity of the higher frequencies is greater than that of the lower frequencies.

CONCLUSION

The HF noise receiver currently operating on-board a Defense Meteorological Satellite is providing data that are being used to determine the sub-satellite value of the F2 region critical frequency along the satellite orbit. Current efforts are underway in order to assess the degree to which the value of the critical frequency thus inferred, accurately corresponds to foF2 observed by ground-based ionosonde techniques. It is anticipated that above regions of the globe where the ionosphere does not display severe horizontal gradients and where HF noise sources on the ground are rather densely distributed, the receiver will provide reasonable estimates of foF2.

Data from the noise receiver are providing the basis for maps of total noise intensity at the height of the DMSP orbit (roughly 500 nautical miles). The maps can be determined as a function of frequency from greater than 1.0 to almost 14.0 MHz. Although the receiver is quite broadbanded (100 kHz), the data are useful in providing, for the first time, an ability to routinely map the noise above the F2 region peak as a continuous function of frequency over the entire globe.

REFERENCES

CCIR, 1963, International Radio Consultative Committee, Rpt 322, International Telecommunications Union, Geneva.

Herman, J. R., J. A. Caruso, and R. G. Stone, 1973, Planet. Space Sci., 21, 443.

Herman, J. R., R. G. Stone, and J. A. Caruso, 1975, J. Geophys. Res., 80, 665.

Horner, F., and R. B. Bent, 1969, Proc. Roy. Soc., A311, 527.

Rush, C. M., and J. Buchau, 1977, J. Atmos. Terrest. Phys., 39, 277.

Rush, C. M., and E. Ziemba, 1978, To appear.

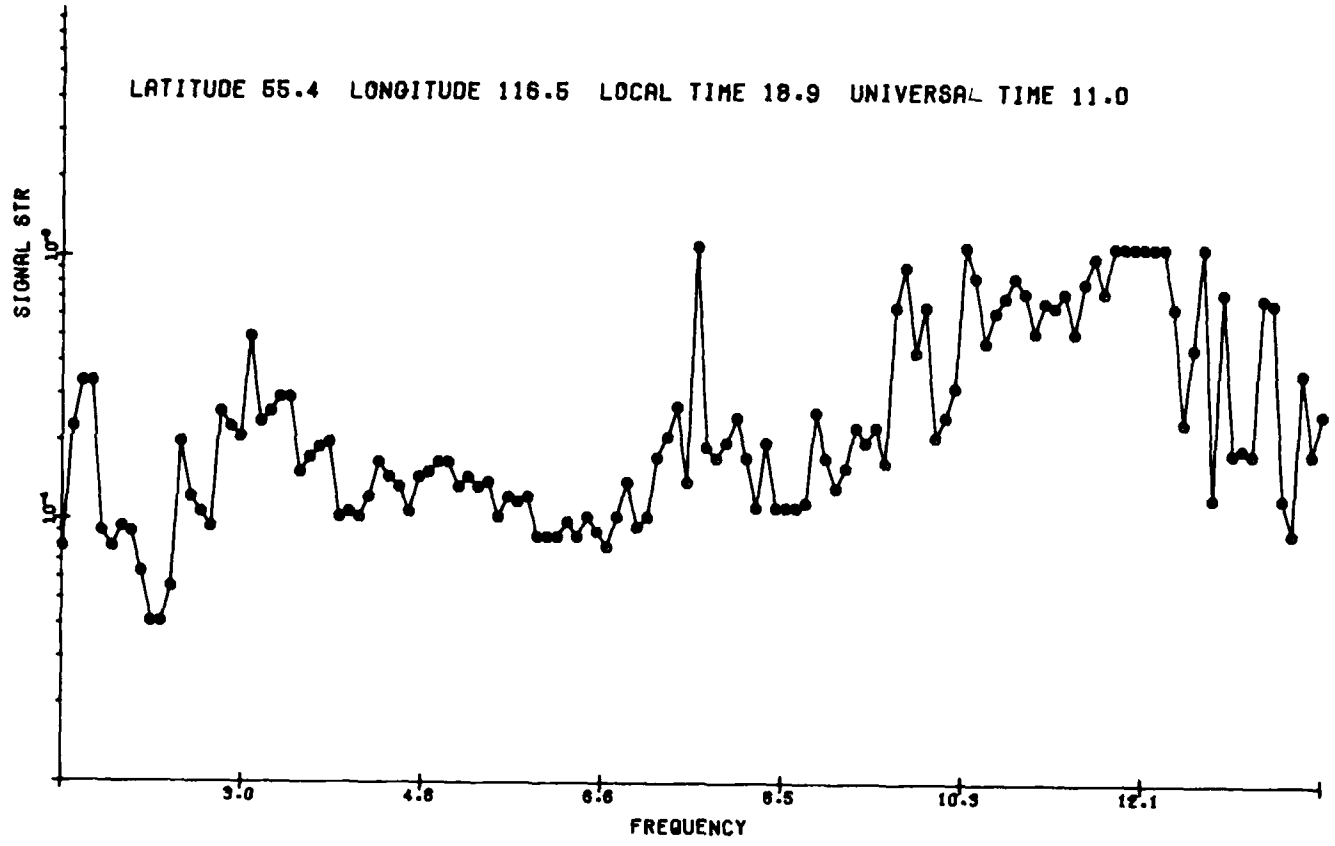


Figure 1. HF Signal Strength as a Function of Frequency.

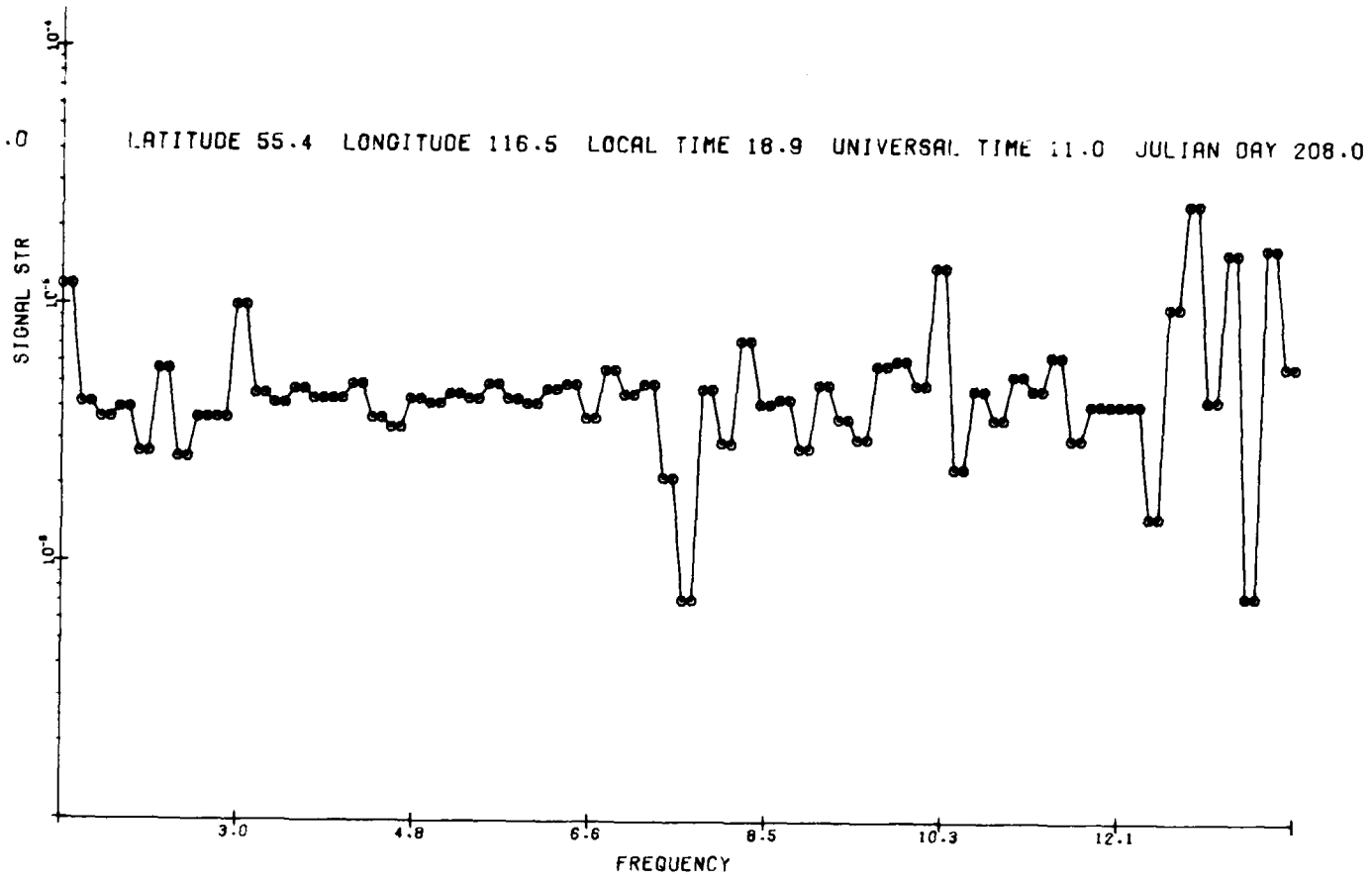


Figure 2. HF Signal Strength as a Function of Frequency Obtained after Uncoupling of Paired Channels.

ELF PROPAGATION MEASUREMENTS USING NATURAL SOURCES OF RADIATION

M. Ishaq and D. Llanwyn Jones
Wheatstone Physics Laboratory, King's College
London WC2R 2LS, England

ABSTRACT

Measurement of the frequency spectra of the radiation received from lightning flashes at global distances in the ELF band of the electromagnetic spectrum have been made in England and in Japan. The ratio of the vertical electric field to the azimuthal magnetic field (the wave impedance) has been determined as a function of wave frequency. Analysis of the wave impedance data enables the determination of the location of the source of radiation on the Earth's surface and the determination of the propagational attenuation in the Earth-ionosphere waveguide, the latter being an important parameter for the design of ELF communication systems. The attenuation is found to vary as a function of the great-circle path taken by the propagated signal, the attenuation for paths passing through high latitudes being considerably greater than that for paths extending only to middle latitudes. A theoretical explanation of the experimental data presented is suggested.

1. INTRODUCTION

The study reported here is an extension of previous ELF electromagnetic wave propagation investigations undertaken at King's College, London within the Radiophysics Group. Here we report on an experimental project designed to enable the determination of the propagation constant of the Earth-ionosphere duct at frequencies in the band 5 Hz to 60 Hz.

The background to the present work has been reported by Jones and Kemp (1970). In summary propagation within the Earth-ionosphere duct is investigated by recording on magnetic tape the vertical electric and the north-south and east-west magnetic field components emitted by lightning discharges at global distances from the observing field station. In this investigation we have employed the Chilbolton Observatory of the S.R.C. Appleton Laboratory situated near Andover, Hampshire, England. Data were recorded in the period September 1975 to April 1976. The recorded data are searched for the large amplitude transient ELF signals examined in the earlier work and the Fourier frequency spectra of these are obtained by digitising the signal waveforms and performing the Fourier analysis using a CDC 6600 computer.

The propagational attenuation is determined by analysis of the Fourier spectrum of the wave impedance of the received signals. Since the source of each signal is a current pulse it is possible to obtain propagational information as a continuous function of frequency the characteristics of the source itself being eliminated from the analysis by the wave impedance technique. The main difficulty in implementation of the technique is the fact that the attenuation at these frequencies is extremely low (0.1 to 1 db/Mm) so signals are received simultaneously from the whole globe. It is necessary to select for analysis only the largest signals which, for a short period of time, swamp the rest of the global activity. The characteristics of the sources of these largest signals are such that propagational information can be obtained only for frequencies below about 40 to 60 Hz.

2. SOURCES AND THE LOCATION OF SOURCES

The sources of the received signals were shown earlier (Jones and Kemp, 1971; Kemp, 1971) to be unusually intense lightning flashes which occur mainly in the tropics. This result has been confirmed by more recent work (Ishaq, 1977).

The range of the sources from the observing station has been determined by two methods both of which require the determination of the Fourier spectra of the (complex) wave impedance Z of the signals in the frequency band 5 Hz to some 60 Hz. Here the wave impedance is defined as the ratio of the vertical electric field to the horizontal magnetic field. In the first method the frequencies at which the modulus of the wave impedance $|Z(\omega)|$ takes maximum and minimum values enables the determination of source range using the technique proposed by Kemp and Jones

(1971). The second, related, method involves a similar analysis but using the argument of the wave impedance $-\arg Z(\omega)$ (Ishaq and Jones, 1977). We claim to be able to determine the source range with an accuracy of $\pm$ 200 km.

The azimuth (or bearing) of the received signal has been determined by the usual "crossed loop" goniometric direction finding technique which has here been applied in the frequency domain using the frequency spectra of the orthogonal horizontal magnetic components. This results in the apparent bearing of the source being found as a function of frequency. However, there seems to be no systematic variation of bearing with frequency so the mean bearing in the band 5 Hz to 60 Hz was taken to be the true bearing of the source. By this means we believe the bearing can be determined to within $\pm 5^{\circ}$.

In a recent paper (Bezrodny et al, 1977) it is suggested that these ELF signals do not propagate around the globe via great circle paths but follow paths which vary markedly with frequency. While more experimental work is clearly required here, the results of the present study do not support these theoretical predictions.

Additionally, a series of simultaneous observations have been made at our field station and at the Tottori Sand Dune laboratory in Japan in association with K. Sao's group. These observations also seem to show that the ELF signals studied follow great circle paths at least to within the accuracy of the source location technique.

3. DETERMINATION OF THE PROPAGATIONAL ATTENUATION

The attenuation of the circum globar propagation path, a (db/Mm), has been obtained as a function of frequency for a selection of 200 ELF signals, chosen on the basis of their waveforms as probably being excited by a single source lightning discharge.

$a(\omega)$ can be determined by analysis of the Fourier wave impedance spectrum using methods which have been detailed elsewhere [Kemp and Jones (1971); Ishaq and Jones (1977)].

The wave impedance, viewed as a function of frequency, is of the form of a "damped oscillation" at any selected range. The modulus and argument exhibit similar behaviour but are in "quadrature". The attenuation constant can be determined by drawing the envelopes of the maxima and minima of either the modulus or the argument of Z for each individual signal as explained in the two references cited above. This process enables a first estimate of a versus frequency to be made. The estimate can be checked and refined if necessary by computing the wave impedance spectrum using the value of $a(\omega)$ just obtained and the full mode theory of wave propagation in the Earth-ionosphere duct and comparing the theoretical and empirical data. The initial estimate of $a(\omega)$ is normally sufficiently accurate.

The propagational attenuation was thus determined for the two hundred signals which had been selected for analysis. At any one frequency considerable variability was found with attenuation values fluctuating over a 2 : 1 range at the lowest frequencies with a smaller variability at higher frequencies. This is perhaps not too surprising when it is recalled that data were obtained over an 8 month period and that signals are received from various azimuths. Greater variability is found in the c.w. data presented by Bannister (1974) at higher frequencies (45 and 75 Hz). Bannister found the short path signal strength to fluctuate mainly at night. Our signals being circum globar always propagate over mixed day-night paths.

Variations of the Attenuation with Azimuth

One of the principal reasons for undertaking the present experimental study was to determine the azimuthal variations of the attenuation constant.

An initial grouping of the data showed that a large number of the selected signals were received from near azimuth 180° corresponding to the tropical Africa storm center. Sixty signals from the data set were selected to cover as wide a range of azimuth values as possible. Unfortunately no data were received from geomagnetic west or from due (magnetic) north. This is a result of there being no persistent thunderstorm centers in the north Atlantic-south Pacific oceans or in the arctic-south Atlantic regions respectively.

The available data were plotted as functions of arrival azimuth (0° to 360°) reckoned from magnetic north with frequency as parameter. Data were sufficient to study the azimuthal variations at 5 Hz intervals from 5 to 45 Hz. It was evident from these plots that the attenuation of signals arriving from the south (azimuth 180°) was notably higher than that for signals coming from the east (azimuth 90°).

Since the data must be periodic in azimuth a linear least squares Fourier series regression analysis was made of the data as follows. At each frequency the functional dependence of a on azimuth Φ may be expressed as

$$a(\Phi) = a_o + \sum_n a_n \sin(\pi n \Phi/180) + \sum_n b_n \cos(\pi n \Phi/180)$$

where Φ is expressed in degrees and the a_n and b_n are the Fourier coefficients to be determined by the least squares principle. This determination was carried out with (a) $n = 1$ only and (b) with $n = 1$ and 2. An F-distribution test applied to each of these two analyses revealed that the regression analysis (b) was, as expected, of much greater significance. In this analysis the paucity of data near $\Phi = 270^{\circ}$ was clearly unfortunate though unavoidable.

Frequency (Hz)	5	10	15	20	25	30	35	40	45
α ($\Phi = 180^\circ$) db/Mm	0.23	.32	.37	.44	.50	.56	.61	.66	.72
α ($\Phi = 90^\circ$) db/Mm	0.15	.22	.28	.34	.41	.46	.52	.59	.65

Table (1). Variation of Attenuation with Frequency for Two Azimuths

The regression analysis confirmed the conclusion of visual analysis of the raw data, the phase of the Fourier regression curve being such that α is a minimum at $\Phi = 90^\circ$ and a maximum at $\Phi = 180^\circ$.

The values of these minima and maxima are shown in Table (1). From the results in Table (1) it can be seen that the fractional difference in the attenuation values for $\Phi = 180^\circ$ and $\Phi = 90^\circ$ decreases as frequency increases. The standard deviation of the figures given in the table also decreases from about $\pm$ 20% at 5 Hz to $\pm$ 10% at 45 Hz. There is thus considerable variability in the propagation constant estimates even after removal of the azimuth dependence.

The results of the present study confirm and extend those of earlier investigations (Jones 1967) where it was found that the average attenuation required to explain the Schumann resonances could be expressed by the following power law

$$\alpha = 0.063\, f^{0.64} \text{ db/Mm}$$

where f is the wave frequency expressed in Hz. Attenuation values calculated using this equation lie between the estimates given in Table (1).

The results also confirm the qualitative conclusions of Etcheto et al (1966) who deduced that the attenuation was considerably greater for paths crossing the polar regions than for east-west paths on the basis of a study of ELF noise intensities.

4. THEORETICAL ANALYSIS

Preliminary calculations have been made in an attempt to explain the data presented in section (3). The calculations involve the determination of the attenuation of the zero order waveguide mode of the Earth-ionosphere duct. No attempt will be made here to develop the relevant theory since this is well documented elsewhere (Wait 1970; Galejs 1972). Initial calculations have been made for two representative ionospheric electron density profiles - those of Bain and Harrison (1972) for day and Thomas and Harrison (1970) for night. These profiles represent mid latitude conditions and have been shown to predict low frequency propagation data over the VLF and LF frequency bands (Bain, 1974). Calculations have been made for these two profiles both with and without the addition of the heavy ion density profile of Pierce and Cole used in earlier calculations (see Jones 1967). The heavy ion contribution is important only at low altitudes (below 60 km). The computations were made using a "full wave" 2 x 2 matrix formulation which enables the effect of the Earth's magnetic field to be fully included. The results of these initial calculations are summarised below.

Two points which can be made immediately are (a) the heavy ion contribution is of great importance for without it the computed attenuation rates are far too low and (b) the azimuthal dependence of the attenuation is not produced by the azimuthal variation of the modal roots at any selected latitude. The latter effect is very small at ELF. For the profiles studied the difference in the computed attenuation rate for north-south and east-west paths is completely negligible at 5 Hz and amounts to only a few percent at 100 Hz e.g. at 40° latitude the largest variation found was for the Bain and Harrison profile for which the computed attenuation rate at 100 Hz for east-west and west-east propagation azimuths differed by only $\pm$ 5% from that for north-south propagation.

For both profiles studied, in the frequency band below 100 Hz the attenuation rate for north-south propagation at the equator is practically the same as that computed with the Earth's magnetic field neglected. The attenuation is greater for other latitudes increasing over the equatorial value by some 20% for the night profile, the change for the day profile being much smaller. The attenuation rate for the day profile overlies the experimental data for $\Phi = 90^\circ$ in Table (1) for mid and polar latitudes. Night time attenuation values are comparable to those for $\Phi = 180^\circ$ in Table (1) being greater by 20% at 5 Hz and smaller by a similar amount at 45 Hz.

The signals we analyse propagate along global paths, which encompass a range of latitudes varying from 0° to 52° for $\Phi = 90^\circ$ to 0° to 90° for $\Phi = 180^\circ$. We can derive only an effective attenuation rate representing an average of that found over the whole path, half of which must be a day path and half a night path.

What is required is a knowledge of the variation of the ionospheric profiles with latitude and season and data here is very meagre. However, the heavy ion region which has a dominant influence on the attenuation rates must show a strong latitude dependence since the ionising agency is cosmic rays and the flux of these is a maximum at high latitudes ($>50^\circ$) and is very small at the equator. The effect of this is to cause the calculated attenuation rate to be much lower than the measured values near the equator.

5. CONCLUSION

It is concluded that the lack of information concerning the ionospheric electron and heavy ion profiles with their latitude and seasonal dependence does not permit a comprehensive theoretical analysis of our data. We have been able to make only an indicative study. In the band 5 to 45 Hz the attenuation is larger by night and is greatly increased by the inclusion of the heavy ions at low altitude. The large attenuation at night is a result of the penetration of the waves through the ionosphere. This penetration can be anticipated to be facilitated by the fact that one of the polar regions will be in perpetual darkness for several months around the solstice (when the observations were made). In summary the conclusion is that lower effective attenuation rates are observed for signals arriving from the east than for those arriving from the south because of the effects of two factors - the variation of the cosmic ray heavy ion region with latitude (indicated by making calculations both with and without this region) and because of leakage through the winter polar ionosphere which, as yet, we have not been able to quantify.

REFERENCES

Bain, W.C. in ELF-VLF Radio Wave Propagation. D. Reidel (Boston), 1974, 151-164.

Bain, W.C. and Harrison, M.D., Proc. IEE (London), 1972, 119, 790.

Bannister, P.R., IEEE Trans. Comm., 1974, 22, 468-473.

Bezrodny, V.G., Nickolaenko, A.P. and Sinitsin, V.G., J. Atmos. Terr. Phys., 1977, 39, 661-688.

Etcheto, J., Gendrin, R. and Karczewski, J.F. Ann. Geophys., 1966, 22, 646-648.

Galejs, J., Terrestrial Propagation of Long Electromagnetic Waves, Pergamon Press (NY), 1972.

Ishaq, M., Univ. London Ph.D. Thesis, July 1977.

Ishaq, M. and Jones, D.Ll., Electron. Letters, 1977, 13, 254-255.

Jones, D.Ll., J. Atmos. Terr. Phys., 1967, 29, 1037-1044.

Jones, D.Ll. and Kemp, D.T., J. Atmos. Terr. Phys. 1970, 32, 557-566.

Kemp, D.T., J. Atmos. Terr. Phys., 1971, 33, 919-927.

Kemp, D.T. and Jones, D.Ll., J. Atmos. Terr. Phys. 1971, 33, 567-572.

Thomas, L. and Harrison, M.D., J. Atmos. Terr. Phys. 1970, 32, 1.

Wait, J.R., Electromagnetic Waves in Stratified Media. Pergamon Press (NY), 1970.

EFFECTS OF PRECIPITATING ENERGETIC PARTICLES ON AN ELF COMMUNICATION LINK

J. B. Reagan, W. L. Imhof, and E. E. Gaines
Lockheed Palo Alto Research Laboratory

T. R. Larsen
Norwegian Defence Research Establishment

J. R. Davis
Naval Research Laboratory

W. R. Moler
Naval Ocean Systems Center

Abstract. Coordinated experiments in 1975 and 1976 have been conducted under Office of Naval Research sponsorship to ascertain the effects of energetic particle precipitation from the earth's radiation belts on a communication link operating at extremely-low-frequency (ELF). Data on the intensity, spectra, and pitch-angle distribution of energetic electrons (> 160 keV) precipitating into the atmosphere were obtained with the Lockheed payload on the polar-orbiting satellite 1972-076B at the times that the U.S. Navy experimental ELF transmitter at the Wisconsin Test Facility was operating to several receiving sites around the world. Based on the satellite data, altitude profiles of ion production rates have been calculated and electron and ion density profiles in the D- and E-region ionosphere have been determined using known effective electron loss rates. The resulting electron and ion density profiles have been inputted to an ELF wave-guide-mode computer program developed at the Naval Ocean Systems Center to calculate the perturbations on the signal strength at the receiving sites. The great-circle propagations to the receiving sites have been segmented in the calculations to include the differences in ground conductivity and ionospheric conditions along the path. The calculated signal strengths have been compared with the measured signal strengths under quiescent conditions and during moderate-to-intense precipitation events. A positive correlation between electron precipitation and signal strength anomalies has been identified. The calculations indicate that the signal strength can be enhanced or attenuated by several decibels depending upon where the disturbance occurs along the path.

Introduction. Over the last 4 to 5 years extremely low frequency (ELF) signal strength recordings at frequencies in the 45 and 75 Hz bands have been made by the MIT Lincoln Laboratory, the Naval Research Laboratory (NRL) and the Naval Underwater Systems Center (NUSC) using transmissions from the U.S. Navy Wisconsin Test Facility (WTF). ELF wave data have been collected over a number of propagation paths including mid-latitude and polar paths. The transmissions from the WTF to a receiving site in Connecticut are known to experience significant signal strength reductions of up to $\approx$ 3 dB on approximately 60 to 80 nights per week. Transmissions to other stations in Maryland, Greenland, Norway and Italy display similar anomalies. It was recognized (Davis, 1974; Davis and Meyers, 1975; Larsen, 1974) that at least some of the anomalies might be due to enhanced ionospheric ionization caused by the precipitation of energetic electrons and protons into the earth's atmosphere.

For the first time in 1976 a successful correlation was made between direct particle precipitation and anomalous nighttime ELF signal strengths using the WTF transmitter (Imhof et al., 1976). In this study, energetic particle data were obtained with Lockheed payloads on the low-altitude polar-orbiting satellites 1971-089A and 1972-076B. During all 11 coordinated cases studied, significant fluxes of energetic electrons having substantially different intensities and spectral shapes were observed to be precipitating from the radiation belts into the atmosphere within the propagation path. Resulting ion-production rates in the upper D-region were calculated to be often comparable to those produced by intense solar-particle-events. Corresponding enhanced electron density profiles were at times in excess of 10^5 electrons/cm^3 at altitudes above $\sim$ 80 km. A positive qualitative correlation was therefore established, but a detailed quantitative correlation between the severity of the anomalies and particle characteristics such as intensity and spectra was not firmly established because of the limited data base used.

In a recent effort (Imhof et al. 1977), satellite data taken during special coordinated exercises in 1975 and 1976 were analyzed and compared with measured ELF transmission between the Wisconsin Test Facility and receiving stations in Maryland Greenland, Norway and Italy. Significant fluxes of electrons were measured to be precipitating into the atmosphere in the propagation path on many of these coordinated cases and corresponding signal strength anomalies were measured at the receiver sites.

Based on the measured fluxes, energy spectra, and pitch-angle distributions of the particles, ion-pair production profiles were calculated and electron density profiles subsequently obtained with the application of effective electron loss rates. Ion density profiles were derived based on the electron density profiles and ambient boundary conditions. These data were incorporated into the ELF propagation computer program to calculate the predicted signal strength behaviour for comparison with the measured signal strengths. The purpose of this comparison was to assess our ability to model such anomalous behaviour, recognizing the capabilities and limitations of the propagation program, our ability to define the large-scale geometry of the disturbance region and to derive the details of the ionospheric conditions along the path.

The waveguide computer program calculations have indicated that energetic particle precipitation events may cause a signal enhancement or degradation, depending upon where the disturbance is occurring along the path. The consistent occurrence during electron precipitation events of signal strength anomalies in either direction is supported by the measured signal strengths. The data, procedures, assumptions, and results obtained in these coordinated exercises will be highlighted in this paper. For more details the reader is referred to the report by Imhof, et al. (1977). The exercise has also identified areas for improvement in future work including better definition of the disturbance region and modifications to the computer propagation program.

Coordinated data. An example of a coordinated measurement is shown schematically in Figure 1.

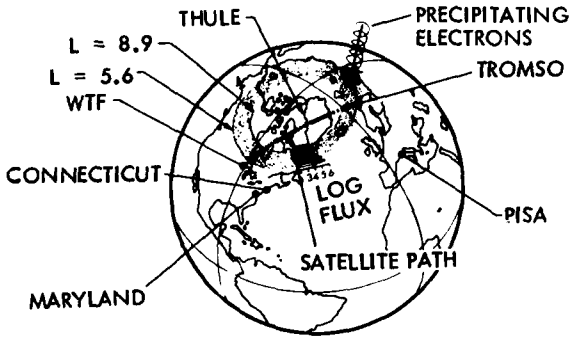


Figure 1. Schematic illustration of the coordinated experiment involving particle measurements on the polar-orbiting satellite and ELF propagation measurements from the WTF transmitter to the Tromso receiver. The electron preciptiation region with respect to the propagation path and the various other receiver sites is also shown

The propagation path from the transmitter at the Wisconsin Test Facility (WTF) to a receiver at Tromso, Norway, is illustrated. At the time of interest the satellite passed through the outer radiation belt/auroral zone region, shown as the shaded area, where intense fluxes of energetic electrons were observed to be precipitating into the atmosphere. It should be noted that the transmitted signal along the great circle path from WTR to the Thule, Greenland and Tromso, Norway receiving sites must pass through the precipitation oval once and twice, respectively. The path from WTR to the Pisa, Italy station can remain within the precipitation oval for considerable distances. The path from WTR to the Connecticut and Maryland stations is lower in latitude than the typical precipitation oval but at the times of large geomagnetic storms Imhof et al. (1976) have shown that precipitation does occur within the path, at least near local midnight, and the signal strengths are affected. The fluxes of electrons above 160 keV are shown schematically in Figure 1. The energetic particle data were obtained by the Space Sciences Laboratory at LPARL in two different ONR-sponsored satellite programs having the designations 1971-089A and 1972-076B. Each of these satellites has the advantage of being at low altitudes (750–800 km) and polar orbiting. Most of the coordinated data used in the analyses reported here were taken with the satellite 1972-076B which is spinning and the instrumentation can therefore provide good pitch angle distribution measurements. This satellite is still operational at the time of this paper.

It is important to measure the particle fluxes simultaneously with conduction of the ELF transmissions since the fluxes and energy spectra of precipitating electrons are known often to be very dynamic in time and space. Since variations with longitude also occur, data from multiple satellite crossings were used whenever available. In both satellite experiments high data coverage was achieved on a world-wide basis through use of tape recorders. On each of these satellites the LPARL payload contains an array of instruments measuring electrons, protons, and alpha particles over a broad range of energies. Of particular concern here are the data taken with the energetic electron spectrometers measuring the intensities and energy spectra of electrons from ~ 130 keV to 2.5 MeV with 256-channel analysis. Two such identical electron spectrometers are located on the oriented satellite 1971-089A at look angles of 90° and 20° with respect to the local zenith. On the spinning satellite, 1972-076B, a single spectrometer provides detailed pitch-angle distributions on the same time scale as the satellite spin period, 5 seconds. The Lockheed group has obtained data on precipitating electrons from one of these two low-altitude satellites continuously since 1971 (Imhof et al., 1973; 1974a, 1974b.

In March and April 1976 a special coordinated exercise was performed between the LPARL satellite payload and the EFL receiving stations at Maryland, Greenland, Norway, and Italy. For this coordination the payload was operated at all practical times when the satellite passed through regions near the transmission path and the transmitter and receivers were in operation. The coordination with the ELF receivers was performed by the Naval Research Laboratory.

Precipitating electron data from the satellite passes during the March–April 1976 coordinations were surveyed, plots were made of the precipitating

fluxes, and the energy spectra from the more interesting passes were obtained and subsequently processed through the LMSC energy deposition program, AURORA (Walt, et al. 1968). Maximum precipitating electron fluxes > 160 keV measured on each satellite pass at local midnight are plotted in Figure 2 as a function of time superposed on the ELF field strengths measured at the various receiving stations. The ELF transmission data were obtained at a frequency of 75 Hz predominantly during the pre- and post-midnight local time period. For reference, the time of local midnight at each station is marked with an arrow and the calculated ambient nighttime field strengths for the respective paths are shown as straight lines. To provide an indication of the geomagnetic activity, plots of the storm-time ring current index, D_{st}, are also included.

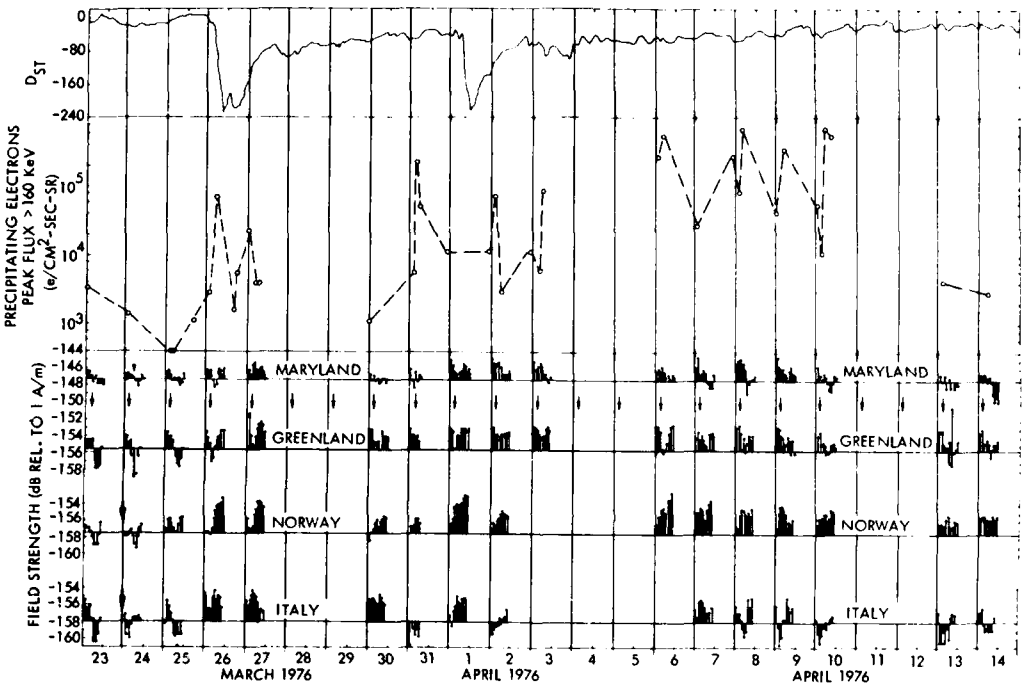


Figure 2. Plot of the peak precipitating electron flux > 160 keV as measured from the satellite and the ELF signal strengths as received at the various stations during the coordination period in March – April 1976. For comparison the D_{st} geomagnetic index is also plotted to indicate the general occurrence of magnetic storms. The calculated reference signal strength at each station is shown by a straight line. Local midnight is shown by arrows.

Two major geomagnetic storms occurred during the coordination period, on 26 March and 1 April. The precipitating electron fluxes during the main portion of these storms and also between 6 and 10 April were rather intense. Departures from the reference signal strengths, both increases and decreases, were observed at the various receiving stations. The reference was established by averaging the nighttime signal strengths at each station over the period 23 – 25 March just prior to the geomagnetic storm. One obvious fine scale feature was observed near local midnight at each station. During this time the signal strength consistently decreased, at many times to levels up to 3 dB below the pre- and post-midnight values. Such a behaviour is consistent with the more intense precipitation of electrons that occurs during storm times near local magnetic midnight.

A general slow peaking in the nighttime signal strength over this 23-day period is also seen, with each night punctuated with the signal decrease near local midnight. This slow peaking, which is evident at all the stations, begins near 23 March, peaks near 2 April, and returns to normal near 13 April. A peak signal strength increase of 4 dB over the reference

occurred at the Norway site on 1 April. It should be noted that of all the receiving stations, Tromso is the only one located continuously within the precipitation oval. This location may result in an enhancement of the excitation factor. The Maryland site exhibited the smallest signal strength reduction. Whether this slow peaking phenomena can be associated exclusively with the electron precipitation or perhaps with changes in the atmosphere/ionosphere constituents induced by the precipitation is not known at this time.

Satellite data. For several coordinated satellite passes during the period March–April 1976, the measured fluxes of precipitating electrons > 160 keV are plotted in Figure 3 as a function of invariant latitude.

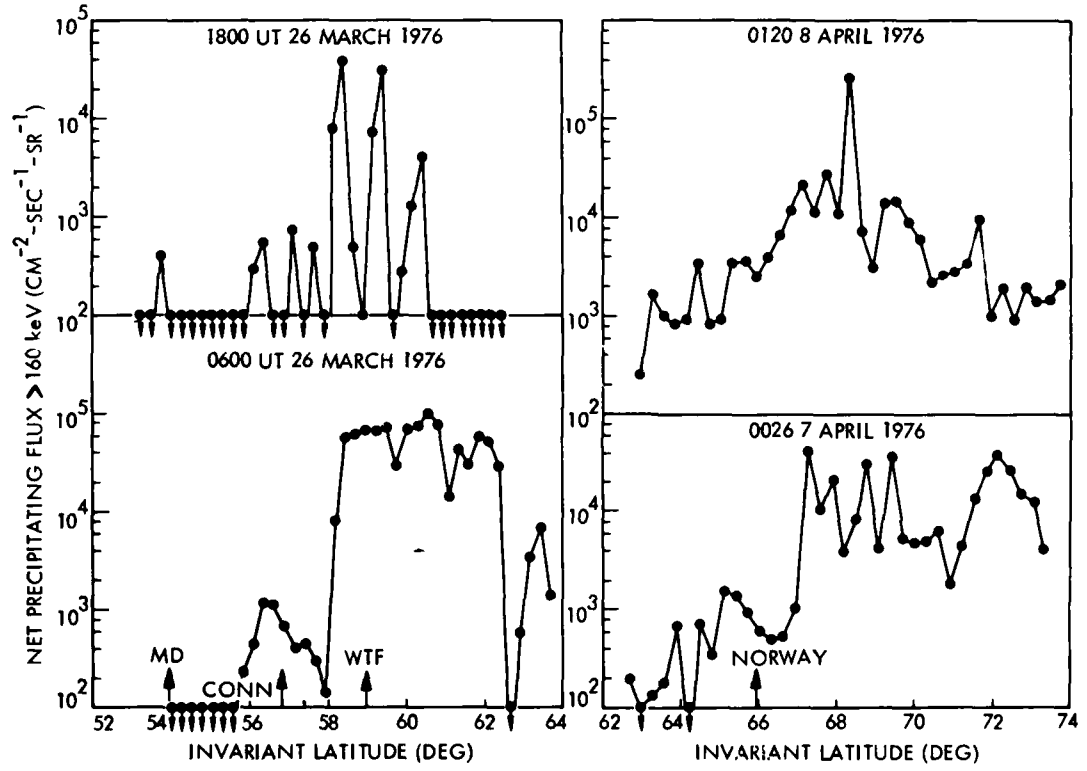


Figure 3. Measured fluxes of precipitating electrons > 160 keV plotted as a function of invariant latitude for selected passes during the coordinated exercise in March–April 1976

The latter parameter is a geomagnetic coordinate appropriate for study of energetic particle precipitation. For reference the invariant latitude of the WTF transmitter is at 59° and the receiving stations at Connecticut, Maryland, Greenland, and Norway are at 57°, 54°, 86°, and 66°, respectively. Since the waveguide program provides for segmenting the propagation path at different regions but does not allow for variation of the guide parameters perpendicular to the propagation path some type of latitude averaging must be invoked, and therefore several empirical approaches were taken. These included a linear averaging of the energy deposition profiles over the complete magnetic latitude interval between the transmitter and the receiving station as well as other procedures such as selecting the highest fluxes within an interval. For many of the analyses presented here, maximum fluxes over a latitude interval of at least a few tenths of a degree were used.

Examples of the more intense energy spectra of precipitating electrons measured during the March–April 1976 coordinations are shown in Figure 4. These measurements were all performed near local midnight, and in each case all or virtually all of the propagation path of interest was in darkness. Overall, the precipitating electron intensities are quite high but somewhat lower than many recorded during the 31 October–1 November 1972 magnetic storm event when high fluxes of precipitating electrons were observed and when pronounced ELF transmission anomalies were recorded.

For each of the electron energy spectra processed the ion-electron production rates in the atmosphere were calculated as a function of altitude, using the computer program AURORA (Walt, et al., 1968). The method consists of solving the time-independent

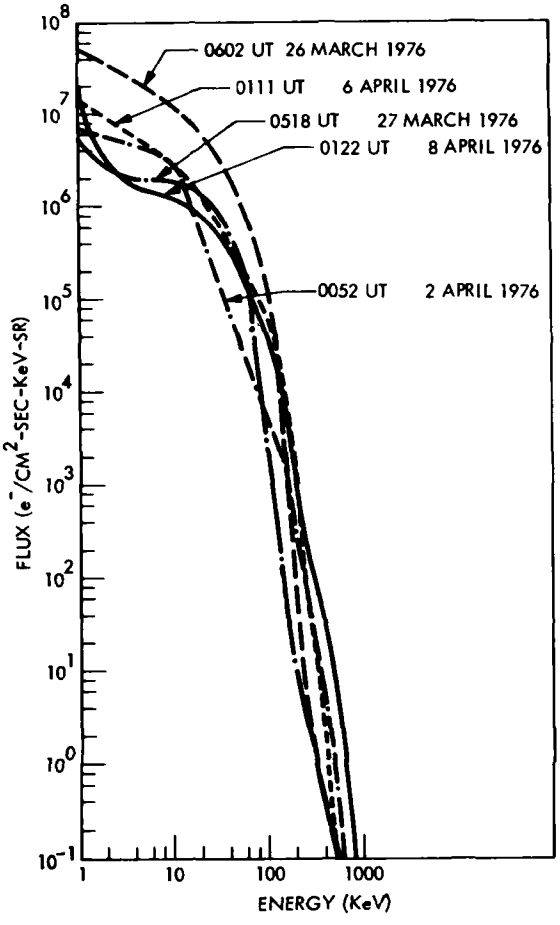


Figure 4. Examples of the energy spectra of precipitating electrons measured during the March–April 1976 coordinations

Fokker-Planck diffusion equation of electrons by use of finite difference techniques. The mirroring effects of the converging magnetic field are included. The program accepts an input electron differential energy spectrum of arbitrary shape along with any pitch angle distribution. Energy-dependent pitch angle distributions may also be applied. In such a case the program uses the input pitch angle distributions prescribed for selected energies to obtain distributions at other energies through interpolation. In the calculations the electrons are allowed to interact with the atmosphere below 300 km. The program outputs include the electron pitch-angle distribution between 0° and 180° thus including backscattered and reflected electrons at preselected energies and altitudes as well as the total energy deposition within specified altitude regions. Examples of ion-electron production rate profiles are shown in Figure 5.

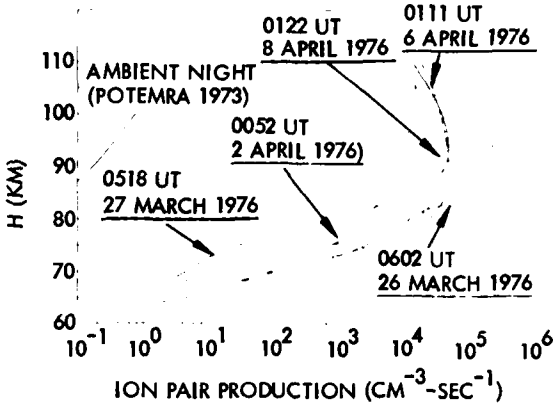


Figure 5. Ion-pair production rates obtained from the measurements of precipitating electrons during selected satellite passes in March–April 1976

The ion-pair production rates shown in Figure 5 are several orders of magnitude larger than the natural ionization rates. For example, at 70 km the cosmic ray production rates are less than 0.1 ion pair/cm^3-second and the ambient nighttime ionization at 90 km is only 0.2 ion pair/cm^3-sec (Potemra, 1973). The ion pair production rates encountered in the events of present interest are to be compared with those obtained during the very intense solar particle events in August 1972 (Reagan and Watt, 1976). During the peak of the August 1972 event the production rates over the Chatanika, Alaska radar site were less than 10^5 ion pair/cm^3-sec at all altitudes below 90 km. Of course, due to the high energy protons in the PCA event the ion pair production rates were much higher at the lower altitudes. For example, at 40 km over Chatanika at 1508 UT on 4 August 1972 the production rate was about 3×10^4 ion-pair/cm^3-sec. However, at an altitude of 90 km the production rate was only about 3×10^3 ion-pair/cm^3-sec which is to be compared with the rates in Figure 5 that are higher by one to two orders of magnitude. This difference largely reflects the higher fluxes of electrons in the tens of keV energy range during the events covered in Figure 5.

In order to obtain electron densities from the ion pair production rates it is necessary to invoke values for the effective nighttime loss rate coefficients. The effective loss rates are known to vary with the production rate at night in contrast to the daytime loss rates (Swider and Keneshea, 1972). At the times of the ELF transmission anomalies of interest here, the ion pair production rates are generally comparable to those at the times of PCA events. The loss rates obtained from PCA observations should therefore be applicable. The loss rate coefficients measured for nighttime conditions ($\chi = 98°$) during the PCA event in August 1972 (Reagan and Watt 1976) and during nighttime events in 1968 and 1969 (Ulwick,

1972) were used. At higher altitudes, data from Watt et al., 1974 were also used. Based on the published data a best-fit curve was selected.

For analysis of these data we have assumed the existence of a steady-state condition in the D-region. Such an assumption simplifies the analysis, but requires that the energy input not change substantially during typical recombination time periods which are of the order of tens of seconds. From the data acquired on a single satellite pass it is difficult to separate temporal and spatial events. The data acquired on subsequent passes (about 100 minutes later) typically indicate substantially different electron precipitation profiles but these differences may largely reflect longitude variations. From a large body of satellite data acquired by the LPARL group on the rates of precipitation of energetic electrons over a large area using the atmospheric bremsstrahlung technique (Imhof et al., 1974) it has been found that the total electron influxes typically remain rather constant during the time (~ 5 minutes) for traversal of the satellite over a given region. The conclusion drawn from the satellite bremsstrahlung measurements should be applicable to the present analyses where we are primarily concerned with the ionization produced over a rather large area. For a steady-state condition in the D-region the relation between production and loss of electrons may be written:

$$Q = \psi N^2$$

where Q = total ion pair production rate, ψ = effective electron loss rate and N = electron density. Based on this relation and the values of Q and ψ discussed above, the electron densities were obtained as a function of altitude. Examples of electron density profiles for the March–April 1976 coordinated exercise are shown in Figure 6.

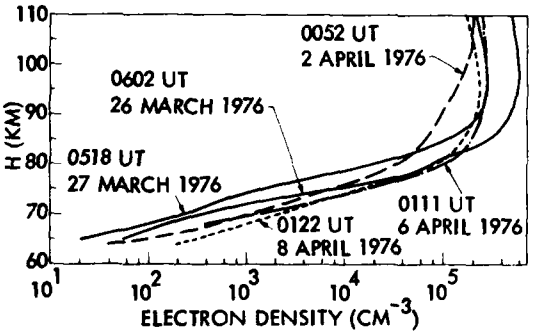


Figure 6. Examples of electron density profiles obtained from the March – April 1976 coordinated satellite passes

At altitudes above ~ 80 km the positive ion density profiles were taken to be equal to the electron densities. This altitude limit was somewhat arbitrarily selected, and may indeed be as high as 85–90 km, but the computer runs with the propagation code indicated a lack of sensitivity to the N^+ densities in this altitude region. At altitudes below ~ 40 km ambient densities for the N^+ profiles were used. In between these altitude limits a smooth curve interpolation was assumed.

The electron density profiles obtained by the above procedures were then incorporated into the ELF propagation computer program. Since at any given time the satellite measurements of the precipitating electrons were performed at only one longitude crossing it was necessary to make certain assumptions regarding the spatial extent of the enhanced ionization densities. The geometry of the transmitter, the satellite crossing, the electron precipitation oval and the receiver were illustrated schematically in Figure 1. The criteria used for calculation purposes were to assume that the maximum precipitating electron inputs measured over a 4° wide latitude band could be averaged and applied over the full longitude interval of concern. The latitude dimension is consistent with the segmentation used in the ELF computer program.

The ELF computer program for model waveguide calculations. The waveguide model computer program used in the present studies was developed at the Naval Electronics Laboratory (NELC) [now Naval Ocean Systems Center (NOSC)] and is well documented (see e.g., Sheddy et al., 1968; Pappert, 1970; Pappert and Moler, 1974). The program is based upon the formalism developed by Budden (1961) and Budden and Daniell (1965) for electromagnetic wave propagation in the terrestrial waveguide. Special emphasis has been given to its applicability in the extremely low frequency (ELF) range.

The program allows for a terrestrial waveguide with the following characteristics:

- Arbitrary electron and ion density distributions with height.
- Variable electron and ion collision frequency height profiles.
- The lower boundary of the waveguide is assumed to be a smooth homogeneous earth with arbitrary conductivity and permittivity.
- Allowance is made for the anisotropy of the ionosphere; arbitrary values for the strength and direction of the earth's magnetic field may be applied.
- Horizontal inhomogeneities (along the propagation path) may be taken into account by segmenting the path into regions having different values for the waveguide characteristics listed above.

The computer program is thus capable of treating waveguides that are reasonable model representations of the actual complex terrestrial waveguide. The program does not allow for variation of the guide parameters perpendicular to the propagation path. This simplification, acknowledged by Pappert and Moler (1974) as the major assumption in the model, should be recognized.

For propagation from WTF to Tromso, Norway with path length approximately 6 Megameters (Mm), we have computed the radiowave fields as a function of distance due to the direct wave only and rejected the contribution from the wave that propagates around the world through the antipode to Tromso. The difference in propagation length between the long and short path is approximately 34 Mm. Therefore,

only small inaccuracies are introduced by neglecting the long path component to Tromso.

The propagation characteristics of ELF waves are dependent not only upon the ionospheric conditions along the path, but also upon changes in the earth's conductivity and permittivity. The path to Norway exhibits large changes in these ground parameters with segments of sea water as well as a ≈ 1.1 Mm segment over the Greenland ice cap.

In order to take these changes into account, we have divided the path into 8 segments which we feel is a realistic, yet practicable number of segments to deal with. Appropriate values of the earth's conductivity and permittivity were used for each segment.

Several runs of the ELF waveguide mode computer program were made to study the effects on the field strength of electron density profiles measured during the March–April 1976 coordinated exercise. The results of these runs for a representative electron spectrum (0602 UT on 26 March 1976) are listed in Table 1 for transmission to each of the receiving stations. One can see that signal enhancements are predicted by the computer program for the receivers at each of the stations. The measured signal strengths averaged over the period 0500-0700 UT on 26 March 1976 also show increases over ambient by the values shown in Table 1. For the present purposes we consider it sufficient to have demonstrated a general consistency between the measurements and calculations as a baseline for evaluating the effects of energetic particle precipitation. Additional refinements in both the measurement technique and in the computer propagation code are required before a detailed quantitative comparison can be justified.

Table 1. CALCULATED FIELD STRENGTHS FOR ELECTRON PRECIPITATION EVENT 0602 HR UT ON 26 MARCH 1976

Receiving Station	Segment for Electron Precipitation (km)	Calculated Field Strengths (at 75 Hz) (in dB wrt 1 A/M)			Measured Change in Field Strength (dB wrt 1 A/M)
		Ambient	During Event	Change Due to Event	
Tromso, Norway	0 – 800	-155.6	-154.5	+1.1	+2.0
Thule, Greenland	0 – 200	-151.1	-148.7	+2.4	+0.5
Pisa, Italy	0 – 200 2400 – 6200	-158.5	-157.0	+1.5	+3.2
Connecticut, USA	0 – 300	-145.2	-144.3	+0.9	–
Maryland, USA	0 – 100	-144.6	-143.7	+0.9	0.0

The signal strengths may be either attenuated or enhanced depending upon the geometry and details of the ion pair production profiles.

- Sensitivity studies have been made to assess the dependence of the ELF signal strengths on such parameters as the electron and ion density profiles and their distribution along the propagation path. The following effects were discovered:
  - The signal strengths tend to decrease with increasing electron density at altitudes of ~ 60 km or lower.
  - The effect of a given ionization profile depends strongly on its location along the propagation path.
  - The ELF signal strengths are most sensitive to positive ion density profiles at altitudes of ~ 45 km or lower. Additional ion concentration and mobility studies and measurements studies and measurements are required to understand the sensitivities.
- The geometry for the effect of electron precipitation is very complex and as a result the following recommendations are made for future investigations:
  - New techniques for mapping electron precipitation profiles simultaneously over a broad spatial region should be used in a coordinated measurement program. Optical and x-ray imaging are the most viable methods of achieving these goals.
  - Existing ELF waveguide-mode computer programs should be modified to include treatment of variations in the electron and ion density profiles along a direction perpendicular to the propagation path.

Summary. The following major conclusions have followed from this study effort:

- From both the satellite and ELF station measurements it has been found that direct particle precipitation into the atmosphere can cause ELF transmission anomalies.

References

Budden, K. G., "Radiowaves in the Ionosphere," Cambridge University Press, 1961.

Budden, K. G., and G. J. Daniell, "Rays in Magneto-ionic Theory," J. Atmos. Terr. Phys., Vol. 27, No. 3, 395–415, March 1965.

Davis, J. R., "ELF Propagation Irregularities on Northern and Mid-Latitude Paths," in ELF-VLF Radio Wave Propagation, J. Holtet, Editor, D. Reidel Publ. Company, Dordrecht-Holland, pp. 263–277, 1974.

Davis, J. R., and W. O. Meyers, NRL Report 7923, Naval Research Laboratory, *Washington, D. C.*, 1975.

Imhof, W. L., E. E. Gaines, and J. B. Reagan, "Dynamic Variations in Intensity and Energy Spectra of Electrons in the Inner Radiation Belt," J. Geophys. Res., 78, 4568, 1973.

Imhof, W. L., E. E. Gaines, and J. B. Reagan, "Evidence for the Resonance Precipitation of Energetic Electrons From the Slot Region of the Radiation Belts," J. Geophys. Res., 79, 3141, 1974a.

Imhof, W. L., E. E. Gaines, and J. B. Reagan, "L-Dependent Peaks in the Energy Spectra of Electrons Precipitating From the Inner Belt," in Magnetospheric Physics, B. M. McCormac, Editor, D. Reidel Publ. Co., Dordrecht-Holland, pp. 129–133, 1974b.

Imhof, W. L., G. H. Nakano, R. G. Johnson, and J. B. Reagan, "Satellite Observations of Bremsstrahlung From Widespread Energetic Electron Precipitation Events," J. Geophys. Res., 79, 565, 1974c.

Imhof, W. L., T. R. Larsen, J. B. Reagan, and E. G. Gaines, "Analysis of Satellite Data on Precipitating Particles in Coordination With ELF Propagation Anomalies," LMSC Report D502063, 30 April 1976; LMSC Report D560323, 30 June 1977.

Larsen, T. R., "Preliminary Discussion of ELF/VLF Propagation Data," in ELF-VLF Radio Wave Propagation, J. Holtet, Editor, D. Reidel Publ. Co., Dordrecht-Holland, pp. 263–277, 1974.

Pappert, R. A., "Effects of Elevation and Ground Conductivity on Horizontal Dipole Excitation of the Earth-Ionosphere Waveguide," Radio Science, 5, 579–590, March 1970.

Pappert, R. A., and W. F. Moler, "Propagation Theory and Calculations at Lower (ELF) Extremely Low Frequencies," IEEE Trans. Comm., Vol. COM-22, No. 4, 438–451, April 1974.

Potemra, T. A., in Physics and Chemistry of Upper Atmospheres, B. M. McCormac, Editor, D. Reidel Publ. Co., Dordrecht-Holland, p. 67, 1973.

Reagan, J. B., and T. M. Watt, "Simultaneous Satellite and Radar Studies of the D-Region Ionosphere During the Intense Solar Particle Events of August 1972," J. Geophys. Res., 81, 4579, 1976.

Sheddy, C. H., "A General Analytic Solution for Reflection From a Sharply Bounded Aniosotropic Ionosphere," Radio Science, 3, 792–795, August 1968.

Swider, W., and T. J. Keneshea, "Diurnal Variations in the D-Region During PCA Events," AFCRL-72-0474 Special Reports, No. 144, Proceedings of COSPAR Symposium on Solar Particle Event of November 1969, 1972.

Ulwick, J. C., AFCRL-72-0474 Special Reports, No. 144, Proceedings of COSPAR Symposium on Solar Particle Event of November 1969, 1972.

Walt, M., W. M. McDonald, and W. E. Francis, "Penetration of Auroral Electrons Into the Atmosphere," in Physics of the Magnetosphere, R. Carovillano and J. F. McClay, Editors, Reinhold, New York, New York, p. 534, 1968.

Watt, T. M., L. L. Newkirk, and E. G. Shelley, "Joint Radar-Satellite Determination of the Effective Recombination Coefficient in the Auroral E-Region," J. Geophys. Res., 79, 4725, 1974.

Acknowledgments

Many individuals contributed to this successful program. Special acknowledgments are extended to Mr. R. G. Joiner of the Office of Naval Research and to Dr. T. Quinn, while at the Office of Naval Research, for their important cooperation, support and direction. Thanks go to Dr. R. A. Pappert of the NOSC for his consultation and cooperation in performing the ELF propagation calculations. The ELF signal measurements were supported by the Naval Electronics Systems Command. We acknowledge the cooperation of the Norwegian Defence Research Establishment in granting Dr. T. Larsen a leave-of-absence to engage in this ELF study activity while in residence at LPARL. The Lockheed portion of the present study was supported by the Office of Naval Research under Contract N00014-75-C-0954.

A LOW-FREQUENCY EARTH-DETACHED COMMUNICATION SYSTEM USING BALLOON-HOISTED TERMINALS

L.C. Humphrey and C.R. Roberts
General Electric Company, Electronics Laboratory
Syracuse, New York

ABSTRACT

The earth-detached propagation mode offers a means of overcoming the loss normally associated with the propagation of VLF-LF waves in the earth ionosphere waveguide when the earth boundary of the guide shows high loss, e.g., across the polar icecap. In addition, the free-space spreading loss is decreased for the surface-guided waves (whispering gallery effect) and the use of high-altitude terminals minimizes the earth's effect on antenna patterns in coupling into the mode.

One of the main parameters of interest for assessing the capabilities of the earth-detached modes is the reflection coefficient of the ionosphere as a function of incidence angle for low grazing angle at the ionosphere. This paper summarizes the results of measurements of the reflection coefficient for low grazing angles in an experiment using high-altitude balloon terminals. These measurements indicate that the earth-detached mode should be a useful communications means for medium to long ranges.

A summary of theoretical calculations and some of the unique characteristics of balloon-hoisted antennas is presented. Suggestions for overcoming problems of mode interference normally associated with VLF-LF long-distance propagation are included.

INTRODUCTION

The concept of the ionospheric waveguide or earth-detached duct at F-layer altitudes and HF (or higher) frequencies have been extensively investigated and experimentally verified[1-5]. The advantages of the F-layer, earth-detached modes include: the decrease in free-space spreading loss due to trapping, the decrease in absorption in the D-layer, and the elimination of reflection loss is at the earth's surface. The main disadvantage is the problem of coupling into the mode, which is either unreliable in the case of using fortuitous ionospheric "tilts", or high loss in the case of scattering from irregularities.

The mechanism considered in this paper is the D-layer earth-detached or "whispering gallery" mode, in which VLF and LF waves are ducted on the underside of the ionospheric D-layer. This mode reduces spreading losses and eliminates ground reflection losses that pertain to the F-layer; also there is a possibility of very rapid recovery in the event of a natural or unnatural disturbance. This mode also has the advantage that it can be launched from balloon-attainable altitudes. The primary disadvantages involved here are the absorption losses that may occur in the D-region, and the possibility of nulls that may occur due to mode interference. Both of these considerations will be examined further.

RAY DESCRIPTION OF THE WHISPERING GALLERY

At very low frequencies a waveguide characterization of the "whispering gallery" mode proves most useful. However, at frequencies above 50 kHz, an unwieldy number of modes is required, so that a multihop ray characterization becomes more efficient, and has the advantage of easier visualization. Figure 1 delineates the ray concept of the D-layer whispering gallery mode.

The figure shows an elevated duct of width, w_d, with three sets of rays. The first, indicated by a solid line, is the ray transmitted horizontally, which also represents the shortest chordal hop length and, therefore, the greatest number of chordal hops for a given transmission distance. The other two ray sets, shown

by dotted lines, bracket the energy in a duct of width, w_d. The transmitter itself is shown a distance, h_{ti}, below the reflecting D-layer, and the ionospheric height is shown to be h_i. As shown in the figure, $h_{ti} < w_d < h_i$.

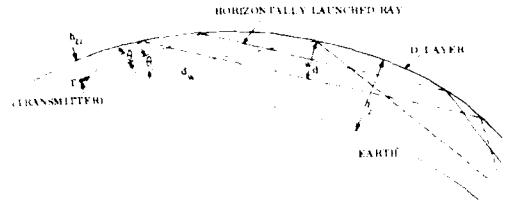


Figure 1. Ray Picture of Whispering Gallery.

A number of simplifying assumptions are made in estimating the mode loss. The spreading loss is computed assuming that the energy trapped in a duct of width, w_d, originates in the angle 2θ. The ionospheric absorption loss is obtained by multiplying the D-layer reflection coefficient by the number of hops required for the horizontally transmitted ray to reach the receiver. (The interference between multiple rays reaching the receiver is discussed later).

To relate the ducted spreading loss to the free-space loss, the area of an annulus at a distance, d_w, from the transmitter to the full duct width is used as a basis for comparison. At a total distance, D, the ratio of the area for free-space spreading to this basis is $(d/d_w)^2$. For waves trapped in the duct, the ratio is $a/d_w \sin D/a$, where a is the earth radius. The ducting gain is given by the ratio of these areas:

$$g_d = \frac{S_d}{S_{f.s.}} = \frac{\left(\frac{D}{d_w}\right)^2}{\frac{a}{d_w}\sin\frac{D}{a}} = \frac{D^2}{a d_w \sin\frac{D}{a}} \qquad (1)$$

From the geometry shown in Figure 2 we see that

$$d_w = \sqrt{a w_d} \ . \qquad (2)$$

Substitution in Eq. 1 then yields

$$g_d = \frac{D^2}{a^{3/2} w_d^{1/2} \sin\left(\frac{D}{a}\right)} \qquad (3)$$

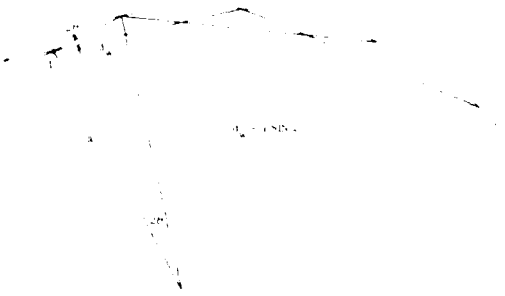


Figure 2. Ray Geometry.

From Eq. 3 it can be seen that the smaller the duct width the higher the gain. However, from the geometry of Figure 1, the minimum value of w_d for which the geometry is valid is $w_d \leq 2 h_{it}$. The maximum expected guidance is then,

$$g_d = \frac{D^3}{a^{3/2} (2h_{it})^{1/2} \sin\left(\frac{D}{a}\right)} \qquad (4)$$

The number of hops for a given ray set is obtained by dividing the total distance, D, by the chord length, $2d_w$.

$$N = \frac{D}{2d_w} \qquad (5)$$

For the special case of the horizontal ray, as shown in Figure 3,

$$dw = d_t = a \cos\psi \ , \qquad (6)$$

where ψ is the incidence angle as the ionosphere. Substituting in Eq. 5,

$$N_h = \frac{D}{2a \cos\psi} \qquad (7)$$

(The fact that the fraction may not be an integer is inconsequential for large values of D since a neighboring ray at a slightly different angle may be selected). The ionospheric absorption loss in dB is, then, the product of the number of hops times the loss in dB per hop,

$$L_T = N_h \cdot 20 \log |R| \ , \qquad (8)$$

where R is the reflection coefficient.

Figure 3. Horizontal Ray Geometry.

For near grazing angles ($\psi \approx 90^{o}$) both experimental and theoretical analysis[6-8] show the loss per hop to be given by approximately

$$20 \log |R| \approx 0.57 f \ell \cos \psi, \qquad (9)$$

where f is the frequency in kHz, and ℓ is the ionospheric scale neight in km. Substituting Eqs. (9) and (7) into (8) yields

$$L_T = 0.285 f\,(\text{kHz})\, \ell\,(\text{km}) \cdot \frac{D\,(\text{km})}{6400}\ \text{dB}. \qquad (10)$$

MODE INTERFERENCE EFFECTS

The mode characteristics described by Eqs. (10) and (4) do not include interference effects. Computer programs using mode expansions show 20 dB nulls due to mode interference effects. The same effect can be expected for interference between rays especially at large distances, as shown in Figure 4.



Figure 4. Ray Interference.

Since the reflection coefficient at near grazing angles approaches - 1, and since adjacent rays have a hop number that differs by unity, rays will interfere destructively to form severe nulls. However, such nulls are localized in altitude so that a long balloon-supported antenna will provide a space diversity effect to mitigate the effects of nulls. A series of receiving antennas, spaced along such a long vertical support and adaptively selected, should result in much less severe interference effects.

MEASUREMENT OF REFLECTION COEFFICIENT

Equation 10 shows that the average mode loss is a sensitive function of the D-layer scale height, ℓ. Reference 2 gives values of ℓ ranging from 0.5 to 3 km. These are based on experimental measurements from surface-located stations. These ground-based measurements are limited to ψ values of less than about 80^{o} and, for this reason, tend to miss sharp gradient layers that might reflect more grazing rays. To determine reflection coefficients at higher grazing angles, an experimental program was done using high-altitude balloon supported terminals.

Experimental Measurements

An experimental measurement program was performed with two balloons launched between Holloman AFB and Chico, California, a distance of 1600 km. The program was conducted under the sponsorship of the Air Force Cambridge Research Labs and utilized a 50 watt, dual-frequency pulsed transmitter in the Chico balloon and a receiving system in the Holloman-launched balloon. The Chico balloon was first launched and reached an altitude of about 118,000 feet before the Holloman balloon was released. This allowed a complete profile of measurements to be made as the Holloman balloon rose to about the same height. Measurements were made over about a 30-hour period. Details of the program and the results are summarized in References 9 and 10.

Figure 5 shows the average signal strength of the ducted mode at the higher of the two frequencies, 438 kHz. It is to be compared to predicted values of signal strength with and without focusing effects. These predictions were provided by the Radio Physics Branch of AFCRL using the profiles of Ferrano for nighttime and Mechtly for daytime[10,11,12].

Table 1 indicates the parameters used for these calculations.

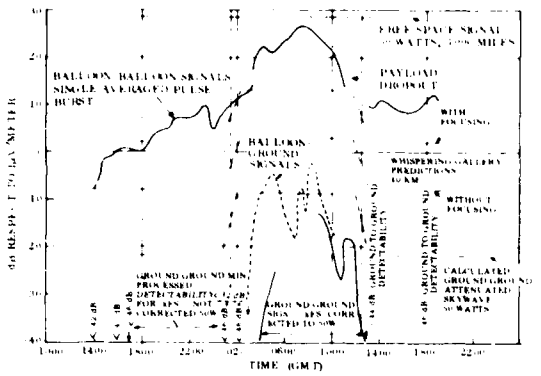


Figure 5. D-Layer Whispering Gallery Experiment - 438 kHz; Flight No. 1.

TABLE 1. MODULUS OF REFLECTION COEFFICIENT FOR WHISPERING GALLERY EXPERIMENT

Ionosphere	Profile Used	Incidence Angle	Condition	dB Loss 500 kHz	dB Loss 250 kHz
Day (60 km)	Merkis[11] (1959)	82°	Ground-Ground	48	33
		85°	40 km-40 km	38	20.5
Night (85 km)	Ferraro[12] (1959)	80°	Ground-Ground	37	31.5
		83°	40 km-40 km	32.5	26.5
Disturbed (65 km) (PCA)		82°	Ground-Ground	40.5	33
		86°	40 km-40 km	21	10.5
		89°	42.5 km-42.5 km	12	6

The free-space loss is shown over the given path for comparison purposes. Figure 6 shows the signal strength and predictions of the lower of the two frequencies, 219 kHz. The data shows good correlation with the predictions at 219 kHz and a lower loss than expected at 438 kHz. Figure 7 combines this data in the form of reflection coefficients as a function of time. The very high reflection coefficient shown with the 438 kHz signal during the daytime periods indicates, perhaps, the formation of a lower C-layer as suggested by Tsedilina and others.[5] The measurements, while preliminary, indicate a very low penetration into the D-layer resulting in very low losses. These losses during one period were reduced almost to the free-space value. The data also indicate that short-term fading effects are very small.

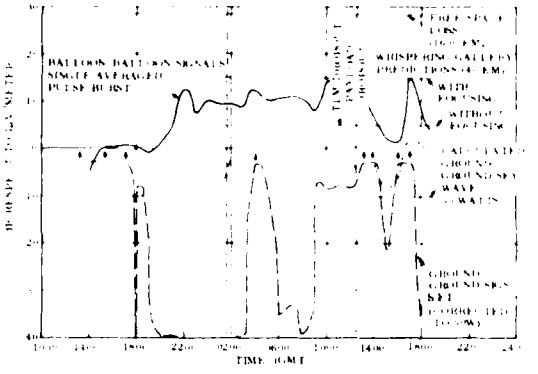


Figure 6. D-Layer Whispering Gallery Experiment - 219.5 kHz, Flight No. 1.

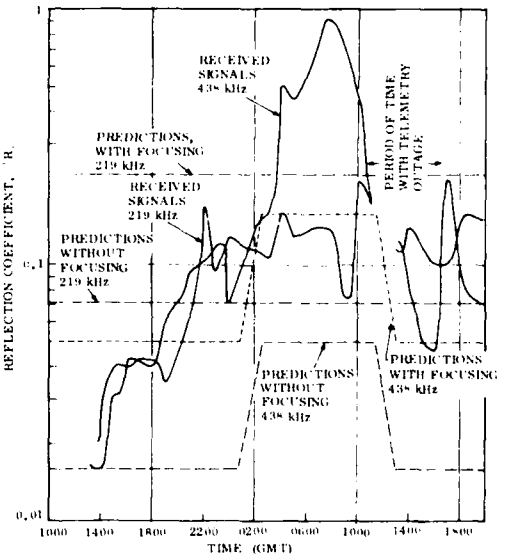


Figure 7. Reflection Coefficient as a Function of Time of Day (GMT).

We hope that more detailed measurements will be made to ascertain practical uses for certain long-range communication circuits.

REFERENCES

1. Wait, "Electromagnetics Waves in Stratified Media", Maenillan 1962.

2. Watt, VLF Radio Engineering", Pergamon Press 1967 pg. 281.

3. Tushentsova, I.A., Fishchuk, D.I., and Tsedilina, Ye.Ye., "Global Distribution of Short-Wave Absorption in the Ionosphere", Izvestiya VUZ, Radiofizika, Vol. 18, No. 9, pp. 1279-1287, September, 1975.

4. Gurevich, A.V. and Tsedilina, Ye.Ye., "Contribution to the Theory of Ultralong Propagation of Short Radio Waves", Geomagnetism and Aeronomy, Vol. 13, No. 2, pp. 242, February, 1973.

5. Tsedilina, Ye.Ye., "Investigation of the Global Properties of Ionospheric Waveguides.I", Geomagnetism and Aeronomy, Vol. 14, No. 6, pp. 851, 1974.

6. Watt, VLF Radio Engineering", Pergamon Press 1967 pg. 216.

7. R.E. Jones, Private Comm. Aug. 24, 1973.

8. J.S. Belrose, "Low and Very Low Frequency Radio Wave Propagation", in Radio Wave Propagation AGARD Lecture Series XXIX July 1968.

9. D-Layer Whispering Gallery Propagation Experiment, C.R. Roberts, General Electric Co. Technical Information Series R73ELS-17, May 1973.

10. Long Range Survivable MF Radio Communication Study Using High Altitude Whispering Gallery Modes, John Videberg and Gary Sales, August 28, 1973, AFCRL-TR-73-0552.

11. Mechtly, Rio, Skaperdas, Smith (1969) Latitude Variations of the Lower Ionosphere, Radio Sci, 4 (No. 6):517.

12. Ferraro, A.J. (1959) Experimental and Theoretical Investigations of the Low Frequency Polarization at 60, 75, and 150 Kc, Penn. State Scientific Report No. 121.

USE OF OMEGA SIGNALS FOR DETERMINING GROUP-VELOCITY CHARACTERISTICS OF THE EARTH-IONOSPHERE WAVEGUIDE

R. Grover Brown

Electrical Engineering Department and
Engineering Research Institute
Iowa State University
Ames, Iowa 50011

ABSTRACT

There is currently little experimental evidence relative to the stability of group velocity of terrestial waves in the VLF range. It is suggested that monitoring the particular Omega signals that are closely spaced in frequency could add considerably to our knowledge in this regard. Preliminary results indicate that group velocity is considerably less stable than phase velocity, especially at frequencies above 12 kHz.

INTRODUCTION

VLF wave propagation in the earth-ionosphere waveguide has been studied extensively during the past 40 years [1]. However, nearly all the reported experimental work on velocity of propagation relates to phase velocity in contrast to group velocity. This is not surprising because direct measurement of group velocity involves measurement of the envelope delay of a modulated rf signal, and this is a difficult experiment to perform precisely. The net result is that the scientific community has a wealth of information about the phase-velocity characteristics of the transmission medium and a dearth of corresponding information about group velocity.

Group velocity studies should be of interest for a number of reasons. First, there is virtually no current experimental evidence about either the long or short term stability of the transmission medium relative to group velocity. It should be of scientific interest simply to know more about this phenomenon. A better understanding of the phenomenon should also be helpful in connection with Omega lane resolution techniques. Here, beat-frequency delay is the key parameter rather than phase delay. Finally, it has recently been suggested that a relatively coarse VLF navigation system, similar to Omega but employing envelope timing, might be feasible [2]. However, before this can be pursued seriously, we need to know more about the basic limitations imposed by the stability of the transmission medium.

The Omega navigation system has recently become operational, and it offers a unique opportunity for studying group-velocity characteristics of the earth-ionosphere waveguide in the 10-14 kHz band. The technical description and status of this system are adequately covered elsewhere [3], so it will suffice here simply to mention that each of the eight stations in the system transmits bursts of 10.2, 11 1/3, and 13.6 kHz signals in a prearranged time-sequenced pattern. Each of the stations has its own highly-stable cesium-beam reference, and all transmissions from a given station are derived from the same source. Thus, they are coherent. Also, the plan for the near future is for all stations to insert an additional transmission at 11.05 kHz in one of the available time slots. Furthermore, each station is allocated a fifth frequency unique to that particular station. The net result of all this is a variety of coherent transmissions in the 10.2 - 13.6 kHz band. A modest experimental effort in monitoring these signals at a few judiciously chosen locations could add considerably to our knowledge about group-velocity characteristics of the earth-ionosphere propagation medium. Before elaborating further, though, a few tutorial preliminaries about wave propagation are in order.

GUIDED WAVE PROPAGATION

It will be recalled from wave theory that the phase and group velocities are defined as

$$\text{Phase Velocity} = \frac{\omega}{\beta} \qquad (1)$$

$$\text{Group Velocity} = \frac{d\omega}{d\beta} \qquad (2)$$

where ω is angular frequency in rad/sec and β is phase shift / unit length. For a path length d, the corresponding propagation delays are

$$\text{Phase delay} = \frac{d}{\omega/\beta} = \frac{\beta d}{\omega}$$

$$= \frac{\text{Total phase delay in radians}}{(2\pi) \times (\text{Frequency in Hz})} \qquad (3)$$

$$\text{Group delay} = \frac{d}{\left(\frac{d\omega}{d\beta}\right)} \approx \frac{\Delta(\beta d)}{\Delta\omega}$$

$$= \frac{\text{Difference in phase delays}}{2\pi \times (\text{Frequency difference in Hz})} \qquad (4)$$

The approximation indicated in Eq. (4) is the usual one of differential calculus. The two velocities have the following interpretation for amplitude modulated waves. The carrier of the wave, or fine structure, travels at the phase velocity, and the modulation, or envelope, travels at the group velocity rate. The phase and group velocities are identical if the β vs. ω curve for the propagation mode is linear through the origin. This is the case for a plane wave in free space. However, the two velocities are not the same for waveguide modes, and one must be careful to distinguish between the two in this case.

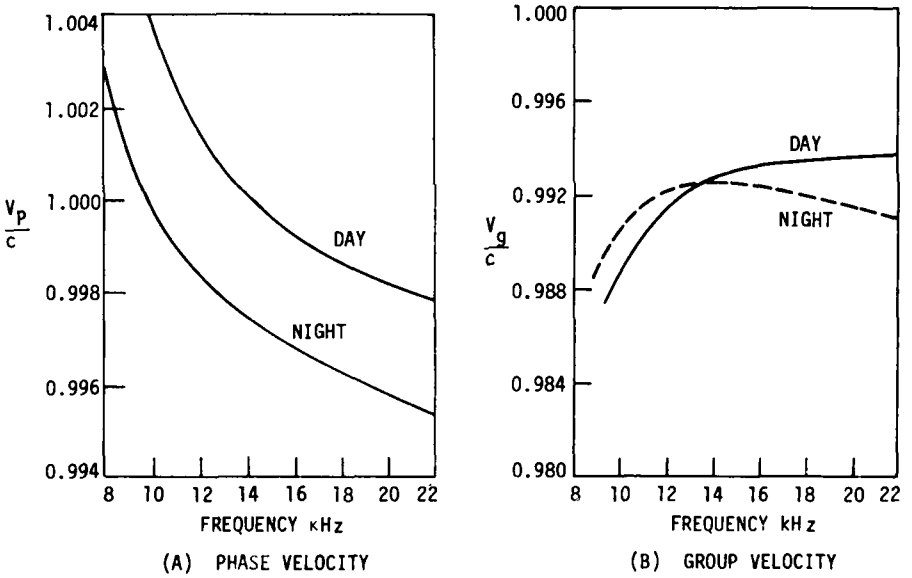


Fig. 1. Nominal phase and group velocities. (From Watt [1].)

The mechanism for terrestial VLF wave propagation is a waveguide mode with the earth and ionosphere forming the boundaries of the guide. Thus, the phase and group velocities are not the same. They have been calculated theoretically as a function of frequency for nominal day and night conditions, and curves given by Watt [1] are reproduced in Figs. 1(a) and 1(b). The group velocity curves are especially interesting because of the crossover of the day and night curves in the 12.5 - 13.0 kHz range. This suggests that if a long-range navigation system operating in this frequency range were to use envelope timing, in contrast to phase timing, there should be no diurnal shift. There are many other problems associated with such a hypothetical system, but such a system does present some interesting possibilities. One such problem is the stability of the propagation medium relative to group velocity. For example, at night when higher-order modes propagate more readily, will the group delay behave as well as phase delay? Some preliminary results indicate the answer to this is "no." However, before proceeding to limited experimental results, a few comments about experimental methods are in order.

MEASUREMENT OF GROUP VELOCITY

As mentioned previously, group velocity is the derivative of the ω vs. β curve and has a physical interpretation as envelope velocity. Thus, if one wishes to measure it referenced to a particular carrier frequency, he must either: 1) observe the time delay of the envelope directly over a known path length, or 2) obtain phase measurements at two closely-spaced frequencies and then approximate the derivative with the slope of the chord connecting two points on the ω vs. β curve. There are serious difficulties with either approach, and the problem is related to the general problem of empirical curve fitting. Using noisy data, it is generally more difficult to determine the derivative of a curve than the curve itself.

The Omega system is not well-suited for precise envelope timing, primarily because of the low pulse repetition rate and relatively low sig-

nal level [2,4]. This is not meant to disparage the system, though, because the system was not designed to be used in a direct envelope-timing mode. The net result is that one must resort to the phase measurements on multiple frequencies when using Omega for group-velocity determination.

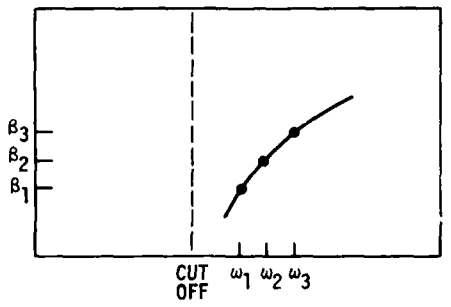


Fig. 2. Smooth-curve approximation of β vs ω curve.

We shall consider now what might be called the "smooth-curve" approach. If we assume that the β vs. ω curve is smooth, as shown Fig. 2, an obvious approach is to make simultaneous phase measurements (i.e., βd) at various frequencies, and then fit an appropriate polynomial through the measured data. The polynomial can then be differentiated and evaluated at any arbitrary frequency (within reason) to obtain group velocity (or delay) referenced to that frequency. This technique has been exploited by Brown and Van Allen [5] in an attempt to develop an algorithm for on-line compensation of the diurnal phase shift in the Omega system. Their work involved only three phase measurements at 10.2, 11 1/3, and 13.6 kHz, but the extension to four or more measurements is fairly obvious, the only question being that of how one wishes to fit the phase data.

The three-frequency case will be used to illustrate the "smooth-curve" approach. In this case, the total phase shift βd is approximated as being quadratic in ω and the coefficients chosen such that the smooth curve goes through the measured phase data. Omitting the details [5], the resulting expression for group delay is

$$T_g = \left(60\,\frac{\omega}{\omega_2} - 66\right)T_1 + \left(-100\,\frac{\omega}{\omega_2} + 105\right)T_2 + \left(40\,\frac{\omega}{\omega_2} - 38\right)T_3 \qquad (5)$$

where

ω = any arbitrary reference carrier frequency in the Omega band

ω_2 = 11 1/3 kHz

T_1, T_2, T_3 = measured phase delays on 10.2, 11 1/3, and 13.6 kHz (i.e., $T_1 = \beta_1 d/\omega_1$, $T_2 = \beta_2 d/\omega_2$, etc.)

Sample results for four reference frequencies 11.7, 11.9, 12.2, and 12.4 kHz are shown in Fig. 3. The four plots were obtained from 22 consecutive days of actual phase data for the Trinidad-N. Dakota path. These curves appear to verify that the average diurnal shift is nullified at around 12.4 kHz, which is what one would expect from the theoretical curves of Fig. 1(b). There is still a noticeable shift during the morning and evening transition periods though. Perhaps the most striking feature of the results in Fig. 3, however, is the dramatic increase in the night time "noisiness" of the group delay within just the small span of 11.7 to 12.4 kHz. A similar phenomena does not appear in the individual phase-delay plots for 10.2, 11 1/3, and 13.6 kHz (not shown); at least, if it is present it is not at all conspicuous. This would seem to indicate that the group-velocity stability at night degrades rapidly with an increase in frequency in the 11.5 to 12.5 kHz range. This might be explained by an increase in the presence of higher order modes which are known to propagate better at higher frequencies. However, this is pure speculation at this point, and we shall leave it at that for now.

One should bear in mind in the preceding smooth-curve analysis that it was assumed that a parabolic fit to the three data points represents the true state of affairs between as well as at those points. But is this really true? Suppose the wiggly solid curve shown in Fig. 4 is the actual situation rather than the smooth dotted curve. Then, our smooth-curve analysis would be meaningless relative to group velocity. The only way to ascertain the truth of the matter is to either observe the envelope delay directly or observe phase shifts at two closely spaced frequencies. We have discarded the envelope approach as not being feasible in the Omega case, so we will look next at the nearby-frequency possibilities.

CLOSELY SPACED PHASE MEASUREMENTS

It would appear that the only way to study group velocity with a high degree of confidence is to measure phase delays at two closely spaced frequencies. As will be seen presently, this is not without its problems. First, though, let us look at various possible combinations of frequencies with present and future Omega transmissions.

We note first that there is already a wealth of Omega phase data for each of the three primary navigation frequencies, 10.2, 11 1/3, and 13.6 kHz. However, these frequencies have the exact ratio 9:10:12, so it can be seen that they can hardly be considered to be relatively closely spaced, the differences being in the 10-20% range. Thus, we will ignore these data, because they would yield too coarse an approximation to the derivative of the β vs. ω curve. Next, we note that there was a period of time prior to October 1977 when both

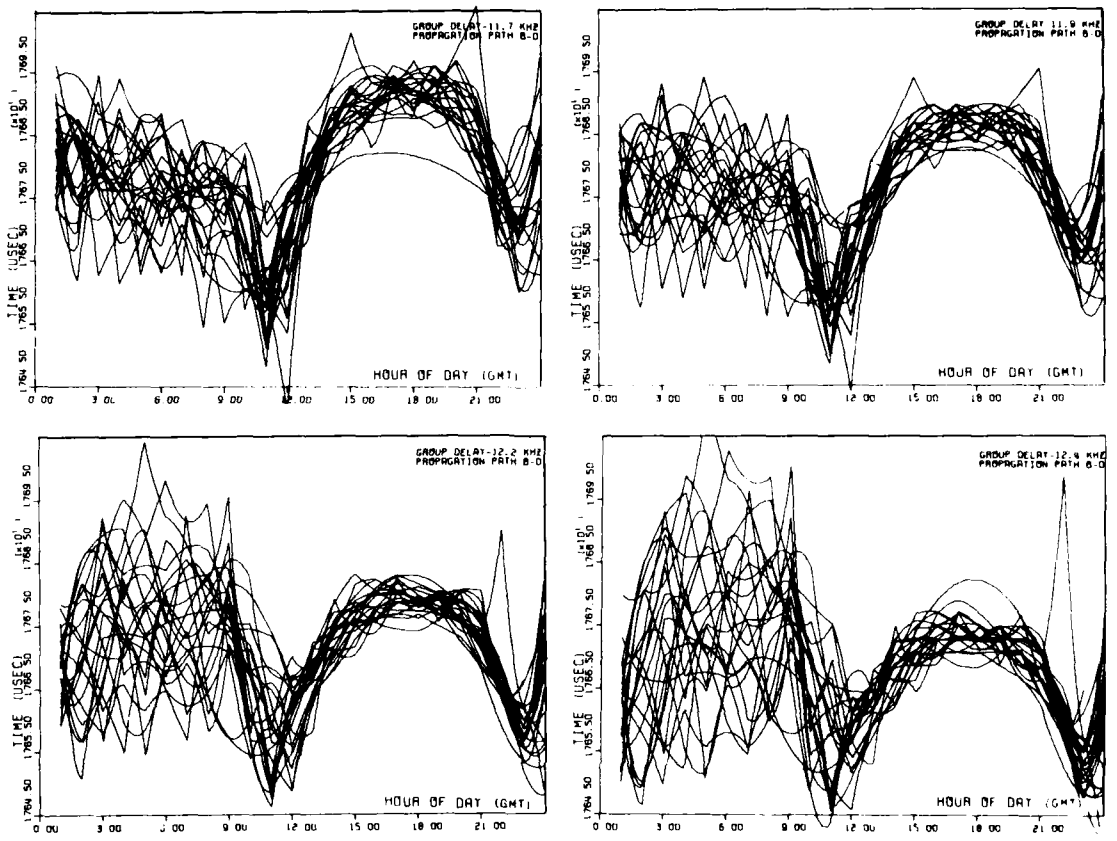


Fig. 3. Group delay versus hour of day (GMT) for various values of reference frequency. Propagation path: Trinidad-North Dakota (B-D). 10-31 March 1975. (From Brown and Van Allen [5].)

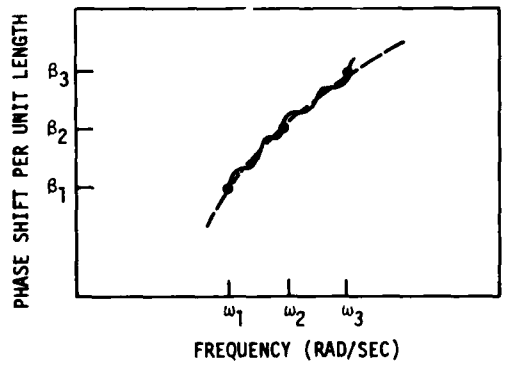


Fig. 4. Two possible β vs ω curves. Question: Which is correct, smooth or wiggly one?

the Hawaii and North Dakota stations transmitted a pair of unique frequencies that were separated by 250 Hz. The frequency difference in this case is in the 2-3% range, and phase differences of these signals should yield meaningful approximations for group delay referenced to the midfrequency. However, both stations changed over to the new format on 1 October 1977, so the 250 Hz difference transmissions no longer exist. (If any "old" 250 Hz data is gathering dust in the archives, it certainly would be interesting to analyze it.) So, looking to the future, the following possibilities exist.

1) The eventual plan is for all stations to add a fourth transmission at 11.05 kHz [7]. Thus, all stations will have a relatively closely-spaced pair at 11.05 and 11 1/3 kHz, their difference being 283 1/3 Hz. This is in the 2-3% range, so this pair should provide excellent information about group delay at about 11.2 kHz.

2) In perusing the various proposed unique frequencies for all eight stations, only those near 13.6 kHz appear to offer promise in group velocity determination. The four "uniques" of interest

are Argentina at 13.15, N. Dakota at 13.1, Japan at 13.05, and S. Pacific at 13.0 kHz. When these are differenced with 13.6 kHz, the separations range from 450 to 600 kHz. These are in the 3-5% range. So, even though the separation is somewhat larger than desired, the phase measurements should still yield meaningful results.

In passing, it might also be mentioned that similar experimental studies might be possible in the VLF communication band using those stations that remain on FSK rather than MSK. The frequency difference between "mark" and "space" is only 50 Hz, which is excellent from the derivative-approximation viewpoint. It will be seen presently, though, that this presents very severe accuracy requirements in the phase measurements. Some experimental work in the VLF communication band was recently reported by Gibbs [6] in connection with lane resolution, but no assessment of accuracy was given in the paper.

The basic accuracy problem should be apparent. We are trying to approximate the derivative as a ratio of phase difference to frequency difference, so any error in phase measurement also gets divided by frequency difference. Thus, the smaller the frequency separation, the larger the final error in the calculated group delay. Pursuing this further, consider a typical general purpose VLF tracking receiver (e.g., Tracor 599K) which compares the phase of the received signal with that of a stable reference provided locally. The output phase difference is usually presented in microseconds plus an unknown constant. Thus, if one imagines two such receivers tracking two coherent signals from the same station, the measured group delay is approximated as follows:

$$T_g \approx \frac{\theta_2 d - \beta_1 d}{\omega_2 - \omega_1} \approx \frac{\phi_2 - \phi_1}{\omega_2 - \omega_1} \qquad (6)$$

But

$$\phi(\text{in radians}) = 2\pi \times \frac{\text{total phase time delay}}{\text{Period}} \qquad (7)$$

Therefore

$$\phi_1 = 2\pi T_1 f_1 \qquad (8)$$

$$\phi_2 = 2\pi T_2 f_2 \qquad (9)$$

Presumably, our two receivers provide a measure of T_1 and T_2 (within a constant). Therefore, the group delay is given by

$$T_g \approx \frac{t_2 f_2 - t_1 f_1}{f_2 - f_1} + (\text{constant}) \qquad (10)$$

or

$$T_g \approx t_2 \frac{f_2}{\Delta f} - t_1 \frac{f_1}{\Delta f} + (\text{constant}) \qquad (11)$$

where t_1 and t_2 are the "time" readings obtained from the two receivers and are the same as T_1 and T_2 within constants. The value of the constants must be resolved by other means; or, if one is only interested in the variations of group delay, the constant in Eq. (10) and (11) is immaterial.

Note that timing errors in t_1 and t_2 are amplified by the factors $f_1/\Delta f$ and $f_2/\Delta f$ respectively. For example, if we observe the 11.05 and 11 1/3 kHz Omega signals, the amplification factors are 39 and 40, respectively. Thus, a 0.1 μ sec timing error on either frequency would produce about 4 μ sec error in the calculated group delay--a sizable error indeed. Imagine the error amplification for a 50 Hz separation and a carrier at 20 kHz!

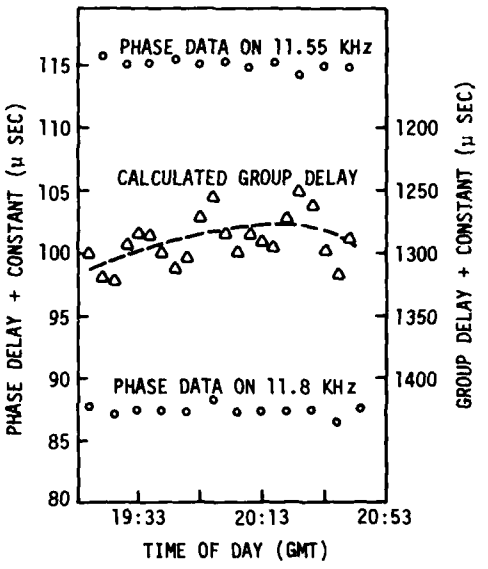


Fig. 5. Sample group delay data. Hawaii-Ames, Ia. Path in daytime, May 13, 1977.

To illustrate the technique further, a sample of phase data obtained at Ames, Iowa, from the Hawaii Omega station on 13 May 1977 is shown in Fig. 5. Hawaii was still transmitting both "uniques" at 11.55 and 11.8 kHz at that time, so Δf is 250 Hz. The two-hour span of time shown in the figure corresponds to a full daylight propagation path, so the phase data on both frequencies was relatively stable. Note, though, the calculated group delays indicated by the x's are quite noisy due to the amplification factors just mentioned. Unfortunately, these data had to be obtained with a single tracking receiver being switched back and forth between f_1 and f_2. Thus, the receiver time constant had to be small, giving rise to somewhat noisier phase data than necessary. Even so, noise amplification is inherent in the differentiation process, and one should expect it.

CONCLUSION

There is relatively little experimental evidence relative to the stability of group velocity in the VLF range. The group velocity (or delay) is the key parameter in any navigation application where the envelope or beat-frequency timing information is to be interpreted as distance. Thus, it would behove the scientific community to know more about this parameter. It appears that with only a modest effort the Omega signals planned for the near future can be used to advantage in gaining more experimental knowledge about group velocity in the 10 - 14 kHz range.

ACKNOWLEDGMENT

The author wishes to acknowledge the Iowa State University Engineering Research Institute for financial support of this project.

REFERENCES

1. A. D. Watt, VLF Radio Engineering, Pergamon Press, 1967.

2. R. G. Brown, "VLF Pulse Timing: Limitations and Potential as a Companion to Omega," Proc. of the Institute of Navigation Aerospace Meeting, Denver, Colorado, April 13-14, 1977.

3. S. H. Laurila, Electronic Surveying and Navigation, John Wiley and Sons, 1976.

4. E. R. Swanson and J. D. Adrian, "Omega Envelope Capability for Lane Resolution and Timing," Naval Electronics Laboratory Center Report TR 1901, November 20, 1973.

5. R. G. Brown and R. L. Van Allen, "Three Frequency Difference Omega," Proc. of the Institute of Navigation Aerospace Meeting, Warminster, Pennsylvania, April 27-28, 1976.

6. G. J. Gibbs, "An Automatic Omega Navigation System with VLF (USN Comm) Augmentation," Proc. of the First Annual Meeting of the International Omega Association, Washington, D.C., July 27-29, 1976.

7. E. R. Swanson et al., "Omega Format Optimization," Naval Electronics Laboratory Report TR 1966, October 8, 1975.

IONOSPHERIC PROPAGATION EFFECTS ON HF BACKSCATTER RADAR MEASUREMENTS

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

Introduction

Line-of-sight radars are capable of determining the geographic coordinates of a target by the measurement of the time delay, i.e., radar range, and azimuth and elevation angles of the reflected signal.

However, in the case of an HF backscatter radar, the target location is deduced from the signal time delay and bearing angle. This requires that an assumption be made with regard to the height of reflection of the propagated wave. Since the (virtual) height of ionospheric reflection of the radar signal is, in general, not readily known, an error in the estimation of the ground distance can be expected to exist.

The measurement of the location in space of an over-the-horizon target utilizing HF backscatter radar techniques can be in error due to the presence of ionospheric propagation anomalies such as an ionospheric tilt, i.e., gradient of electron density, and a traveling ionospheric disturbance (TID), i.e., large-scale electron density perturbation in the F-region.

An ionospheric tilt can impose an error in the angular position and ground range of an HF radar target. A TID can also introduce an angular deviation in addition to a variation in the Doppler frequency shift and a perturbation in the ionospheric reflection height of HF signals. The latter also results in a ground range measurement inaccuracy.

In this paper, estimates are made of the HF radar-positional errors that can be encountered due to ionospheric tilts and imprecise knowledge of the ionospheric height of reflection.

The characteristics of mid-latitude ionospheric tilts deduced from the spatial distribution of electron density profiles recorded during the 1959 high solar activity period are presented.

Techniques for reducing the adverse effects of the propagation medium on target-coordinate registration for an HF backscatter radar are also discussed.

Theoretical Considerations

Ionospheric Tilts

The presence of electron density gradients along an HF transmission path can introduce deviations both in azimuth and ground range.

In the derivation of the analytical expressions for the angular deviation from the great circle path and for the ground range errors imposed by an ionization tilt, it is assumed that the ray undergoes a plane mirror-type reflection in the ionosphere.

The ionospheric tilt geometry employed in this analysis is illustrated in Figure 1.

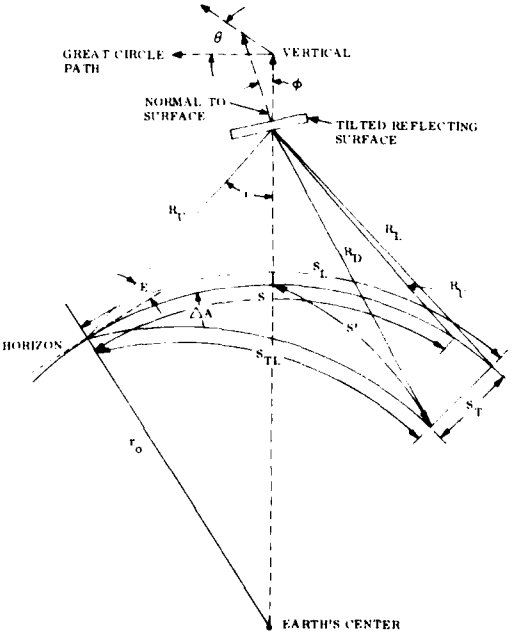


Figure 1. Ionospheric Tilt Geometry

For a ray undergoing reflection in an untilted ionosphere, the ground range, S, to the earth's surface reflection point would be attained for propagation at an elevation angle, E, and a reflection height, h. When the ray is reflected from an ionospheric region containing a tilt, it is deviated in azimuth by an amount, ΔA, and, in addition, the ground distance traversed becomes S_{TL}.

Utilizing the law of sines in spherical geometry, the azimuthal deviation of the ray can be written as

$$\Delta A = \sin^{-1}\left[\frac{\sin\left(\frac{S_T}{r_o}\right)}{\sin\left(\frac{S_{TL}}{r_o}\right)}\right] \qquad (1)$$

where r_o is the radius of the earth, and S_T is the transverse component of S_{TL}.

From the law of cosines, S_{TL} can be readily determined from

$$S_{TL} = r_o \cos^{-1}\left[\cos\left(\frac{S_L}{r_o}\right)\cos\left(\frac{S_T}{r_o}\right)\right] \qquad (2)$$

where S_L is the longitudinal component of S_{TL} (along the great circle path).

The longitudinal and transverse views of the ionospheric tilt geometry are shown in Figures 2 and 3, respectively.

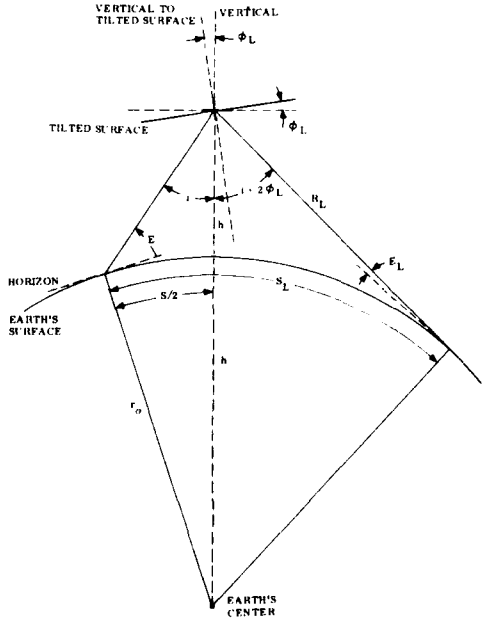


Figure 2. Longitudinal View of Ionospheric Tilt Geometry

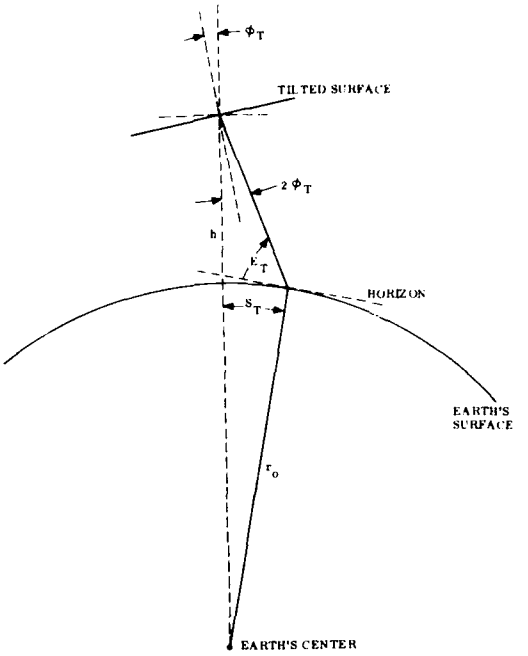


Figure 3. Transverse View of Ionospheric Tilt Geometry

It can be shown that S_L can be expressed in terms of

$$S_L = \frac{S}{2} + r_o\left[\frac{\pi}{2} - (i + 2\phi_L + E_L)\right] \qquad (3)$$

where E_L is the elevation angle, in the longitudinal plane, of the ray incident on the earth, i is the angle of incidence of the ray with respect to the vertical and ϕ_L is the longitudinal component of the tilt angle, ϕ. It is noted that the term, $i + 2\phi_L$, is the angle of reflection in the longitudinal plane.

According to Figure 2, the elevation angle, E_L, can be determined from

$$E_L = \cos^{-1}\left[\frac{r_o + h}{r_o}\sin(i + 2\phi_L)\right] \qquad (4)$$

The incidence angle with respect to the vertical is simply

$$i = \sin^{-1}\left[\frac{r_o}{R_u}\sin\left(\frac{S}{2r_o}\right)\right] \qquad (5)$$

where R_u is the slant range from the radar to the reflection point.

It can be shown from the law of cosines that R_u is obtainable from

$$R_u = \left[r_o^2 + (r_o + h)^2 - 2r_o(r_o + h) \cos\left(\frac{S}{2r_o}\right) \right]^{\frac{1}{2}} \quad (6)$$

From Figure 3, it is seen that the transverse ground distance component, S_T, is given by

$$S_T = r_o \left[\frac{\pi}{2} - (2\phi_T + E_T) \right] \quad (7)$$

where ϕ_T is the transverse component of the tilt angle and E_T is the transverse elevation angle of the reflected ray incident on the ground. From the law of sines, E_T is simply

$$E_T = \cos^{-1}\left[\frac{r_o + h}{r_o} \sin 2\phi_T \right] \quad (8)$$

The tilt angle components, ϕ_L and ϕ_T, are depicted in Figure 4 which is a top view of the tilt geometry. The parameter, θ, which is also indicated, is the angle in the horizontal plane through which the tilt angle is rotated away from the great circle plane.

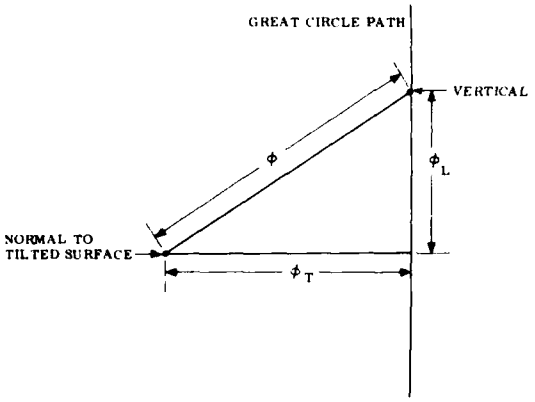


Figure 4. Top View of Ionospheric Tilt Geometry Illustrating the Angles θ and ϕ

As a first approximation, it is assumed that the tilted reflecting surface is spherical. Thus, it follows from the law of sines that

$$\phi_L = \sin^{-1}[\sin \phi \cos \theta] \quad (9)$$

$$\phi_T = \sin^{-1}[\sin \phi \sin \theta] \quad (10)$$

In determining the error in the estimation of the ground range due to the ray being reflected from a tilted ionospheric surface, it is necessary to consider the time delay between the transmitted and received signal which is a measure of the radar-target slant range. As shown in Figure 1, the true slant range is composed of

$$R = R_u + R_D \quad (11)$$

where R_D is the slant-ray path distance from the ionospheric reflection point to the ground reflection point.

The ground range error, ΔS_{TL}, is obtained from

$$\Delta S_{TL} = S_A - S_{TL} \quad (12)$$

where S_A, the apparent ground scatter distance, is given by

$$S_A = 2r_o \cos^{-1}\left[\frac{r_o^2 + (r_o + h)^2 - [(R_D + R_u)/2]^2}{2r_o(r_o + h)} \right] \quad (13)$$

From simple geometric considerations, it can be shown that

$$R_D = \left[r_o^2 + (r_o + h)^2 - 2r_o(r_o + h) \cos\left(\frac{S'}{2r_o}\right) \right]^{\frac{1}{2}} \quad (14)$$

where S' is the surface distance from the subionospheric reflection point to the ground reflection point.

By applying the law of cosines of spherical trigonometry, S' can be determined from the relationship

$$S' = r_o \cos^{-1}\left[\cos\left(\frac{S_L - (S/2)}{r_o}\right) \cos\left(\frac{S_T}{r_o}\right) \right] \quad (15)$$

Reflection Height Inaccuracy

Inaccurate knowledge of the height of ionospheric reflection of an HF propagated wave results in an error in the prediction of the geographic coordinates of a radar-target. In the derivation of the relationship expressing target ground range error in terms of the reflection height variation, it is assumed that the target is located on the earth's surface.

The range to the reflection point, R_1, shown in Figure 5, is given by

$$R_1 = \left[r_o^2 + (r_o + h)^2 - 2r_o(r_o + h) \cos\left(\frac{S}{2r_o}\right) \right]^{\frac{1}{2}} \quad (16)$$

where R_1 is one-half the radar range, R, to the target, h is the true reflection height and S is the true ground distance.

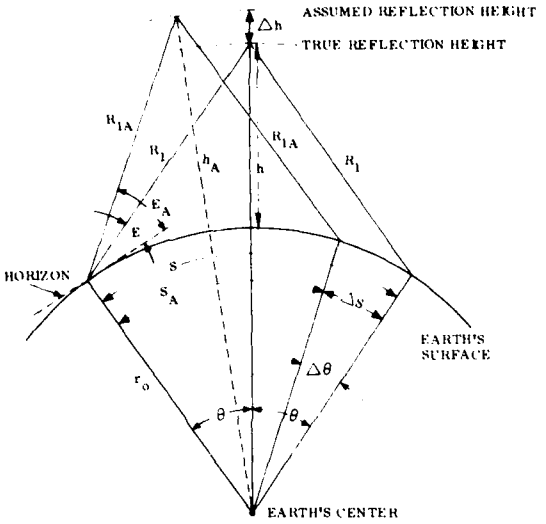


Figure 5. Geometry for Reflection Height Inaccuracy

The ground distance error, ΔS, is simply

$$\Delta S = S_A - S \tag{17}$$

where S_A is the apparent ground distance defined by

$$S_A = 2r_o \cos^{-1}\left\{\frac{1}{2r_o(r_o + h_A)}\left[r_o^2 + (r_o + h_A)^2 - R_{1A}^2\right]\right\} \tag{18}$$

and where the parameter h_A is the apparent reflection height and $R_{1A} = R_1$.

The apparent elevation angle, E_A, is given by the relationship

$$E_A = \cos^{-1}\left[\frac{r_o + h_A}{R_{1A}} \sin\left(\frac{S_A}{2r_o}\right)\right] \tag{19}$$

The expression for the true reflection angle, E, is similar to Equation (19) except that h_A and S_A are replaced by h and S.

Discussion

Errors Due to Ionospheric Tilts

The orientation of an ionospheric tilt can be described in terms of the angles ϕ which is the angle between the vertical to the earth's surface and the normal to the tilted surface and θ which is the angle in the horizontal plane through which the tilt angle is rotated away from the great circle plane.

The azimuthal deviations for ionospheric tilt angles oriented at 45° and 135° horizontal rotation angles at a reflection height of 250, and 350 km are shown in Figure 6. It is seen that, for a given tilt angle, the azimuthal deviation decreases with increasing ground distance and increases with increasing height of reflection. The lack of data points at the long distances is attributed to the ionospheric reflected rays overshooting the earth's surface.

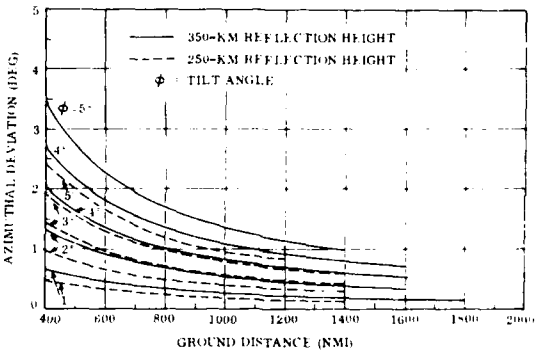


Figure 6. Azimuthal Deviation for Ionospheric Tilt Angle Oriented at 45° and 135° Horizontal Rotation Angle at a Reflection Height of 250 km and 350 km

It is of interest to note that, for the condition $\theta = 0°$ and 180°, i.e., the ionospheric tilt is located in the great circle plane, the ray path does not undergo an angular deviation.

An examination of Figures 6 and 7 reveals that maximum azimuthal deviations occurs at $\theta = 90°$ and that, for a given horizontal rotation angle, reflection height and ground distance, the magnitude of the azimuthal bending is approximately directly proportional to the ionospheric tilt angle. According to Figure 7, the azimuthal deviation, for h = 350 km, $\theta = 90°$ and ground range of 400 nmi is 0.95°, 1.90°, 2.85°, 3.82° and 4.79° for $\phi = 1°$, 2°, 3°, 4° and 5°, respectively.

An interesting feature of the azimuthal deviation data is the minor increase in the lateral displacement of the ray paths from the great circle path with increasing ground range. As shown in Table 1, for $\phi = 5°$, $\theta = 90°$ and h = 350 km, the lateral ground displacement only increases from 33.4 nmi to 35.3 nmi for a ground range change of 400 to 2000 nmi, respectively. In this range interval, the azimuthal deviation decreased from 4.79° to 1.01°.

The ground range errors introduced by ionospheric tilts are depicted in Figures 8 through 10 for the various combinations of horizontal rotation angles and reflection heights.

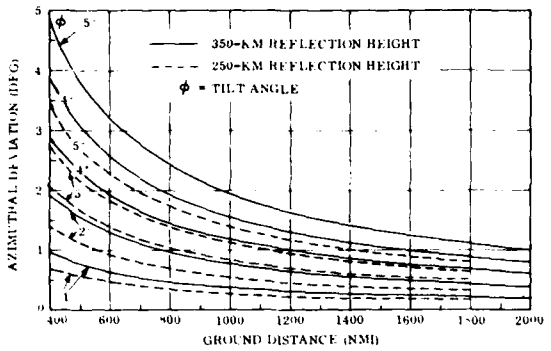


Figure 7. Azimuthal Deviation for Ionospheric Tilt Angle Oriented at 90° Horizontal Rotation Angle at a Reflection Height of 250 km and 350 km

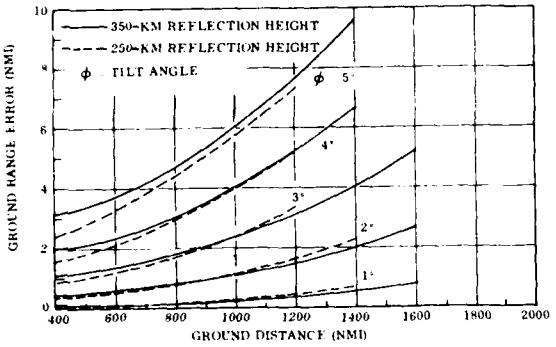


Figure 8. Ground Range Error for an Ionospheric Tilt Angle Oriented at 0° and 180° Horizontal Rotation Angle at a Reflection Height of 250 km and 350 km

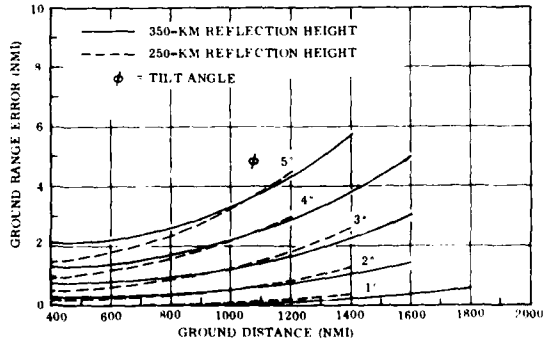


Figure 9. Ground Range Error for Ionospheric Tilt Angle Oriented at 45° and 135° Horizontal Rotation Angle at a Reflection Height of 250 km and 350 km

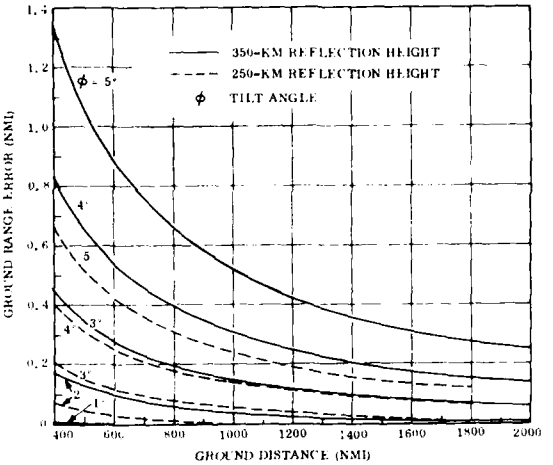


Figure 10. Ground Range Error for Ionospheric Tilt Angle Oriented at 90° Horizontal Rotation Angle at a Reflection Height of 250 km and 350 km

TABLE 1

LATERAL DEVIATION OF RAY PATH OFF THE CIRCLE PLANE FOR AN IONOSPHERIC TILT ANGLE OF 5°

Reflection Height (km)	Horizontal Rotation Angle (deg)	Target Ground Distance (nmi)	Lateral Deviation (nmi)
250	45, 135	400	16.8
		800	16.9
		1400	--
		2000	--
	90	400	23.9
		800	24.0
		1400	24.4
		2000	--
350	45, 135	400	23.5
		800	23.6
		1400	24.2
		2000	--
	90	400	33.4
		800	33.7
		1400	34.2
		2000	35.3

It is noted that, for a given set of conditions, i.e., tilt angle, reflection height and ground distance, the ground range error is a maximum at $\theta = 0°$ and 180° and a minimum at $\theta = 90°$.

It is seen that, for $\theta = 0°$ and 180° (Figure 8) and $\theta = 45°$ and 135° (Figure 9), the range error is a minimum at 400 nmi and increases with increasing ground distance. At $\theta = 90°$ (Figure 10), however,

the reverse takes place; that is, the range error decreases with ground distance.

The horizontal orientation angle at which the functional relationship is reversed is clearly evident from Figure 11 which is a plot of ground range error as a function of the angle θ for a tilt angle of 5° and reflection height of 350 km. As θ increases from 45° to 90°, the location of minimum ground range error is displaced towards increasing ground distance. For example, at θ = 55°, minimum ground range error occurs at about a ground range of 550 nmi while, at θ = 70°, the minimum is shifted to approximately 800 nmi.

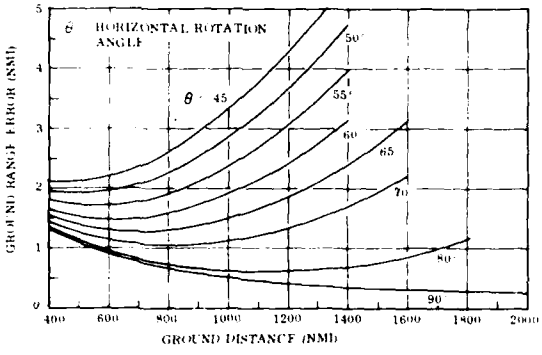


Figure 11. Ground Range Error as a Function of *Horizontal Rotation* Angle for an Ionospheric Tilt Angle of 5° at a Reflection Height of 350 km

Utilizing the data illustrating the electron density variations in the mean quiet ionosphere between latitudes 15°N and 50°N along the 75°W meridian (Wright and Fine, 1960; Wright, Wescott and Brown, 1961, 1962, 1963), an analysis was conducted to determine the magnitude of the tilts that are prevalent in the ionosphere. Sample plots of the latitudinal variation of ionospheric tilts at constant plasma frequencies are presented in Figures 12 and 13. The contours of constant plasma frequency in the altitude versus latitude plane from which the tilt angles were deduced are also presented together with h_{max}, the mean altitude of the F-layer maximum electron density.

Examples of the diurnal variation of the tilt angle are given in Figures 14 and 15. It was found that, in general, east-west tilts are for the most part less than 0.5° except at sunrise and sunset where the tilts can vary from 2° to 8° at all latitudes and all seasons. North-south tilts have a most probable value of about 2°, although between 20°N and 40°N latitude the tilts show wide variability with angles as large as 8° to 10°. During the midday hours, the tilt angles are less than 4°.

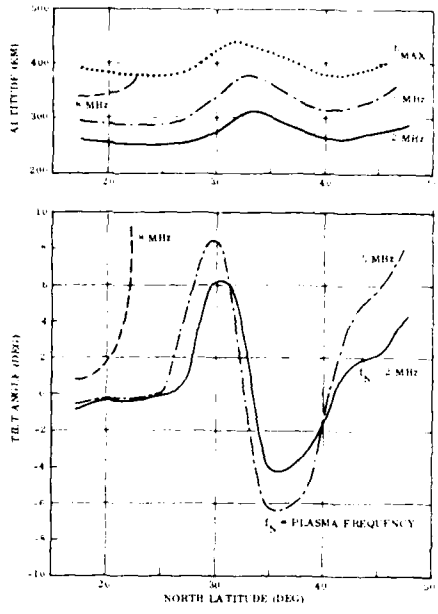


Figure 12. Latitudinal Variation of Altitude and Tilt Angles at Constant Plasma Frequencies Along the 75° West Meridian, 0000 Hours Local Time, September 1959

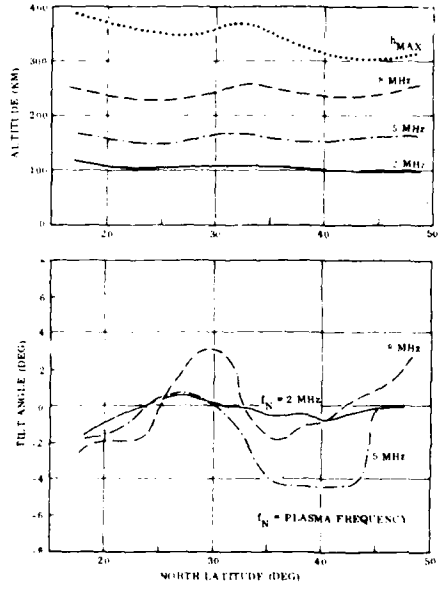


Figure 13. Latitudinal Variation of Altitude and Tilt Angles at Constant Plasma Frequencies Along the 75° West Meridian, 1200 Hours Local Time, September 1959

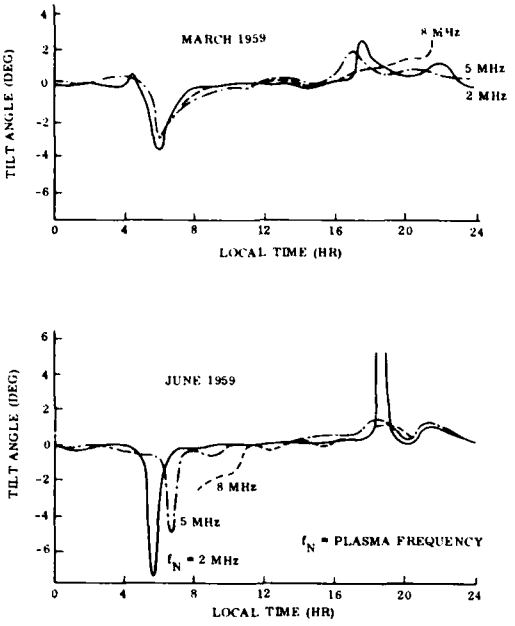


Figure 14. Diurnal Variation of Tilt Angles at Constant Plasma Frequencies at 30° North Latitude, March and June 1959

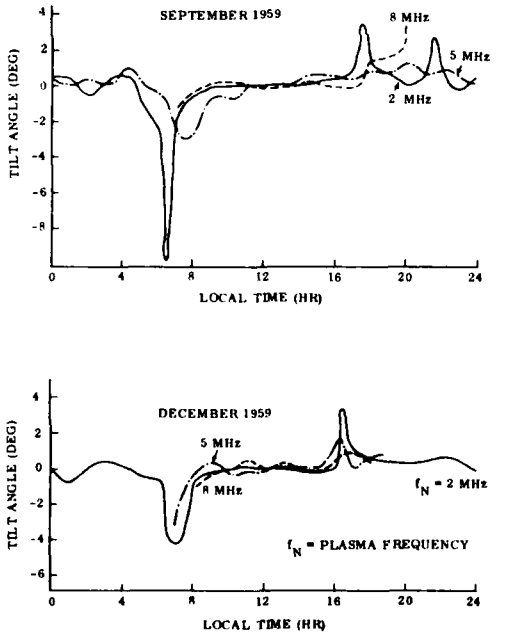


Figure 15. Diurnal Variation of Tilt Angles at Constant Plasma Frequencies at 30° North Latitude, September and December 1959

Thus, according to Figures 6 and 7, the azimuthal deviation during a major portion of the day could be less than 0.5° due to east-west tilts. At sunrise and sunset, the deviation could range between 1° and 5° at a ground distance of 400 nmi and decrease to about 0.5° to 1.5° at 1400 nmi ground distance. The azimuthal deviation for north-south tilts could vary between 1° and 2° at 400 nmi while, at 1400 nmi, it could be about 0.5°.

It follows from Figures 8 and 9 that east-west tilts could impart a ground range error on the order of 0.05 nmi and 0.25 nmi at ground distances of 400 nmi and 1400 nmi, respectively. At sunrise and sunset, however, the errors could increase to about 3.0 nmi and 9.5 nmi, respectively. For a north-south (or a south-north) path, the ground range error due to tilts could have a most probable value of 0.5 nmi and 2.0 nmi for ground ranges of 400 nmi and 1400 nmi, respectively. When the ionospheric reflection point is between 20° and 40°N latitude, the errors could be on the order of 2.0 nmi and 6.5 nmi, respectively.

Errors Due to Reflection Height Inaccuracy

A perturbation in the reflection height of HF signals could result from a traveling ionospheric disturbance moving across the transmission path. If there is an inaccuracy in the estimation of the height of ionospheric reflection, then an error is injected in the measurement of the ground range.

The ground range errors for various errors in the true reflection height of 250 and 350 km are presented in Figure 16.

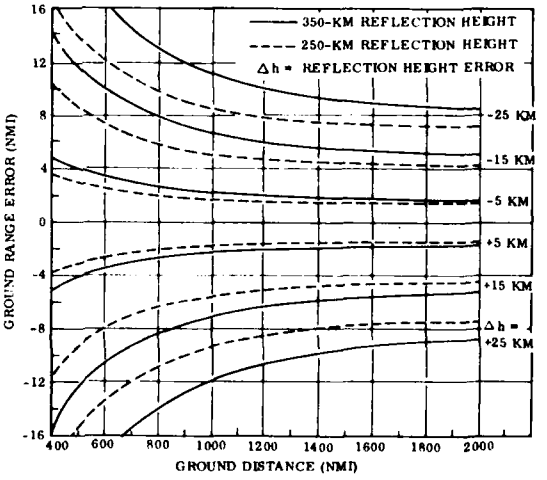


Figure 16. Ground Range Error for an Error in the Estimation of the Reflection Height

It is evident that an overestimation in the reflection height imposes an underestimation in the ground range. For a given reflection height, the error decreases with ground distance and increases

with reflection height error. For a given distance and reflection height error, the ground range error increases with increasing height of reflection.

Assuming an error in the reflection height of +15 km, it follows from Figure 16 that, for a distance of 1000 nmi, the error in the prediction of the ground range evaluates to -5.6 and -7.0 nmi for a reflection height of 250 and 350 km, respectively.

Techniques for Minimizing Errors

HF ground based transponders located within the antenna beam coverage of an HF backscatter radar can be employed to reduce the adverse effects of ionospheric tilts and travelling ionospheric disturbances, i.e., ionospheric reflection height variations, on HF radar measurements.

The presence of ionospheric tilts along the propagation path and the azimuthal deviations resulting from the tilts could be readily determined from the angle-of-arrival measurements of the transponder's radiation.

Since the geographic location of the transponders would be known, the virtual height of reflection in the ionosphere of the propagated signals could be deduced from the time delay between the transmitted and received signals.

This scheme necessitates that the mode of propagation between the radar and transponder be known, i.e., identification of the propagation paths such as single or multiple reflections or hops. HF sweep frequency oblique incidence soundings should be able to provide the data for mode identification.

The presence of an ionospheric tilt can also be inferred from oblique incidence soundings made at one or several select frequencies over the azimuthal coverage region. In the absence of tilts, the time delay to the leading edge of the ground backscatter should be approximately the same over the azimuth spread. The existence of a tilt in the path would be indicated by a variation in the time delay with bearing angle.

Conclusions

HF radar transmissions reflected from ionospheric regions containing ionization tilts could experience both an azimuthal bending and an alteration in the path length. The latter, in turn, would introduce an error in the measurement of ground distance traversed by HF signals.

The magnitude of the azimuthal deviation and the ground range error is a function of the orientation of the tilt with respect to the HF transmitted signal, the height of ionospheric reflection and the ground distance.

For a tilt oriented in the plane of the great circle path, the ground range error is a maximum. For this condition, angular deviation does not take place.

When the tilted surface is orthogonal to the great circle plane, the ground range error is a minimum while the azimuthal bending is a maximum.

The azimuthal deviation increases with increasing height of reflection and decreases with increasing ground distance. The ground range error, on the other hand, increases with increasing reflection height and ground distance. However, for the case in which the great circle plane is perpendicular to the ionospheric tilt, the ground range error decreases with ground distance.

Imprecise data on the height of ionospheric reflection of HF propagated signals induce a ground range error which increases with increasing reflection height and reflection height error and decreases with ground distance. The error brought about by this source is larger than that caused by ionospheric tilts.

The detrimental effects of the propagation medium can be minimized by utilizing ground based transponders and oblique incidence sweep frequency sounders.

Acknowledgments

Portions of the work reported in this paper were sponsored by the Air Force Electronic Systems Division under the CONUS OTH-B Prototype Radar System (PRS) Program, Contract No. F19628-75-C-0128.

References

Wright, J.W., and L.A. Fine, "Mean Electron Density Variations of the Quiet Ionosphere, March 1959", NBS Technical Note No. 40-1, February 1960.

Wright, J.W. L.R. Wescott and D.J. Brown, "Mean Electron Density Variations of the Quiet Ionosphere, June 1959", NBS Technical Note No. 40-4, May 1961.

Wright, J.W., L.R. Wescott and D.J. Brown, "Mean Electron Density Variations of the Quiet Ionosphere, September 1959", NBS Technical Note No. 40-7, April 1962.

Wright, J.W., L.R. Wescott and D.J. Brown, "Mean Electron Density Variations of the Quiet Ionosphere, December 1959", NBS Technical Note No. 40-10, March 1963.

A MODEL FOR PULSE TRANSMISSION OVER THE FADING IONOSPHERIC REFLECTION CHANNEL

Alfonso Malaga
Robert E. McIntosh
Electrical and Computer Engineering Department
University of Massachusetts
Amherst, Massachusetts

ABSTRACT

This paper is concerned with the modeling of the HF ionospheric reflection communications channel and its application to study the distortion suffered by pulses used in digital data transmission. The propagation medium is characterized by a time-variant transfer function. Results which illustrate the dependence of the transfer function on the spatial and temporal variations of the tropospheric refractive index and the electron density in the ionosphere, transmitter-receiver separation, time of the day and operating frequency are given. In addition, the frequency and time-selective fading characteristics of the HF link caused by multiple scatter phenomena are related to the stastical spatial characteristics of the refractive index of the propagation medium. This is done by obtaining expressions which relate the delay and Doppler power spectra of the HF link to the wave number spectrum of the refractive index fluctuations of the transmission medium.

The distortion of RF pulses is shown to be related to the dispersion factor (deterministic effect) and the multipath or delay spread (due to spatial variations in the refractive index) of the channel. Enhanced signal reception is shown to be possible by transmitting frequency-swept chirp pulses which compress at the receiver by as much as 40 dB.

ROCKET OBSERVATIONS OF IONOSPHERIC E-REGION IRREGULARITIES: IMPLICATIONS FOR HF MODELLING

E. P. Szuszczewicz, J. C. Holmes, and D. N. Walker

E. O. Hulburt Center for Space Research
Naval Research Laboratory
Washington, D. C. 20375

I. INTRODUCTION

There has been considerable effort to establish global models of ionospheric profiles to assess the utility of communications channels operating over the electromagnetic frequency spectrum from ELF through SHF. To a large measure, these models are intended to represent mean ionospheric conditions; but this necessitates statistical smoothing which can result in ionospheric segments with little resemblance to reality. It is important to recognize this restriction and to work toward improved ionospheric profile specifications.

The ionospheric domain with the largest data base is the F-region which is accessible to systematic global investigation by satellites. The E-region cannot be studied directly by satellites but must be probed with rockets and ground based radar and HF sounding systems. This situation has resulted in geographically localized studies of E-region phenomena and, in our opinion, has not contributed to understanding of horizontal variations and the associated impact on signal channel modelling. It is our objective to draw attention to this point and show some of the differences between ionospheric models and actual observations. We show some of our recent ionospheric rocket results and compare them with models used for prediction of HF ducting.

II. CONSIDERATIONS OF E-REGION GEOPHYSICS

The sources of ionospheric E-region irregularities can be identified with a number of important geophysical forces including neutral winds, electric fields and precipitating particles. The degree to which these forces contribute to the generation of irregularities depends not only on magnetic latitude but also on local topography. A number of special publications (e.g., Refs. 1-3) and review articles (e.g., Refs. 4-5) contain detailed treatments on morphology and geophysics as well as critical commentary on theoretical models. The emphasis here is the extent to which differences in irregular E-region profiles can be explained through geophysical processes.

Chemistry. E-region ion composition measurements have been made with rocket-borne mass spectrometers at equatorial, mid- and high-latitude sites (e.g., Refs. 6-9). A quiet (no sharp gradients of electron density) E-region is generally dominated by the molecular ions NO^+ and O_2^+ with production and loss processes that have been known for some time. Although the quiet E-region may be the best understood ionospheric domain (e.g., Ref. 5), quiet conditions do not always prevail (especially at night), and the result is highly irregular structure generically described as sporadic E.

Meteoric ions and winds. The presence of metallic ions (Mg^+, Fe^+, Si^+, Na^+, Al^+, K^+, being of major importance) at E-region altitudes is apparently a global feature which is maintained to varying degrees by ablating meteoroids[6]. Because these atomics have recombination coefficients 10^6 times smaller than the molecular species [10-13], they are subjected to long-lived transport processes and will converge by wind shear into intense layers[14-16]. This mechanism is reasonably well understood and can successfully account for much of the observed data (e.g., Refs. 1-3). The degree to

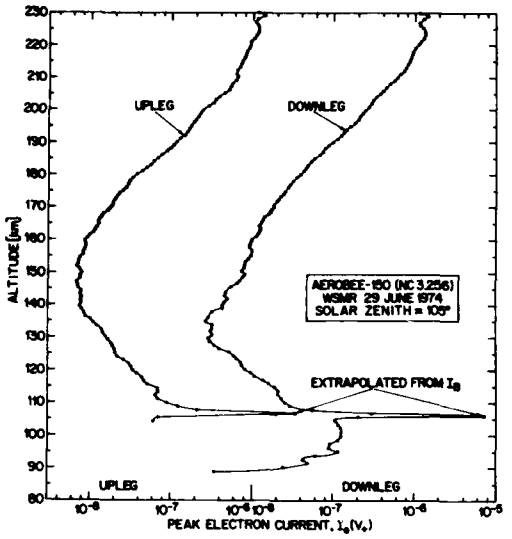


FIGURE 1

Mid-latitude profiles of relative electron density determined by plasma probe collection of electron saturation currents, $I_e(V^+)$. Conversion to absolute electron density proceeds approximately as $N_e(cm^{-3}) = 10^{11}\ I_e(V^+)$. Note the shifted current scales for the up- and downleg portions of the flight.

which convergence can take place is indicated in Fig. 1 which shows ionospheric data collected aboard an Aerobee 150 rocket launched from a mid-latitude site at White Sands Missile Range, N.M. on 29 June 1974 at 2031 MST. The payload reached an altitude of 229.5 km, and up- and downleg portions of the flight passed through ketic sporadic-E layers centered . 106.7 and 106.1 km, respectively. The flight profiles of relative electron density are measured by peak electron currents collected by an on-board plasma probe. The up- and downleg E_s thicknesses at the 0.1 N_e^{max} point were 1.16 and .73 km with corresponding peak electron densities of $3(10^5)$ and $7(10^5)/cm^3$, respectively. (The details of the measurement procedure and the roles of chemistry and winds are presented elsewhere[17].) Temperate zone E-layers are not always as simply structured as the profile in Figure 1. At times layers of ionization are observed to form out of the bottom side of the F-region and descend to lower altitudes[18]. These layers often form continuously with evidence that they are the result of ionization convergence driven by neutral winds[18-20]. The uncertainty in metallic ion distributions, coupled with the unpredictability of wind shears necessarily limit the predictive capabilities of theoretical models. In fact, the models often require parametric retrofitting to achieve agreement with experimentally determined ionization profiles.

Latitudinal variations and the equatorial electrojet. Since it is the Lorentz force $q(V_{wind} \times B)$ generated by the east-west component of horizontal wind shears which brings about ion convergence[21], it might be expected that the occurence and intensity of wind-driven sporadic E should be correlated with the magnitude of the horizontal component of the geomagnetic field. There is however a small zone centered at the magnetic equator (about 2-6° wide) where layer convergence is not expected to take place because the electrons are constrained to move along the geomagnetic field[21]. The result is a vertical polarization field which opposes the effects of strong windshear convergence. The absence of mid-latitude type E_s at the equator however does not reduce the E-region profile to conditions amenable to predictive modelling...for the polarization fields must be taken into account as well as the equatorial electrojet (e.g., Refs. 22 and 23). To focus on some of the problems at the equator, we reference Fig. 2 which presents ionospheric profiles collected by a rocket-borne probe launched from the Chilca Range (Peru; 0.5° dip) on 23 May 1975 at 2000 hr (LT). On the up-and downleg portion of the trajectory, double E-region layers were observed with thicknesses ranging from a few kilometers to 20 kilometers. (A similar observation was made during a post-midnight launch on the same night.) The most popular explanation for the formation of such layers is the wind-shear mechanism which in its present form does not operate successfully near the dip equator. However, the existence of such layers is well established,[24-26] and the interaction of associated ionization gradients with the electrojet can establish an unstable plasma state which, in turn, can generate smaller scale structure[27-30]

Before discussing instability mechanisms that might be active in the electrojet it is important to note the differences in the up- and

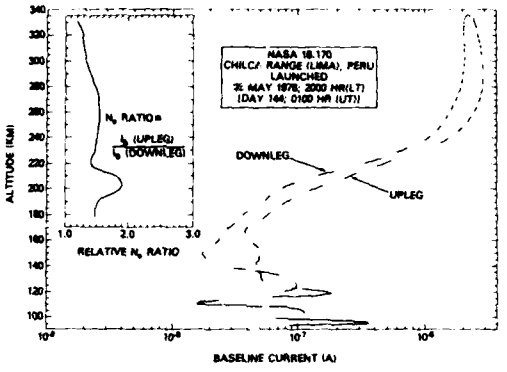


FIGURE 2

Relative electron density profiles near the dip equator as measured by baseline current I_B collected by a pulsed plasma probe[26]. $N_e(cm^{-3}) \approx 1.3(10^{12})I_B$.

downleg E-region layers which were separated by a horizontal east-west extent of 170 km. The profiles reveal significant horizontal gradients that are believed attributable to major variations in the local E-region winds[27] ...winds which are possibly unique to the topography of the Andes-Pacific Ocean interface.

Small scale structure in the equatorial electrojet. The presence of smaller scale ($1\ m \lesssim \lambda \lesssim 1\ km$) irregularities at electrojet altitudes has been known for some time (e.g., Ref. 28), with a number of mechanisms having been proposed to explain the observations (e.g., Refs. 29 and 30). Most recent attention has been given to two classes of irregularities. The first (type 1) is believed to be caused by the two-stream instability while the second (type 2) is generally associated with the gradient drift instability. The generation of gradient-drift irregularities requires a vertical background gradient in the ionospheric electron density (e.g., Fig. 2) and specific relationships between the gradient and the direction of the electrojet current[29-31]. With the appropriate conditions[24,29] sub-structure can be generated with scale sizes much smaller than a kilometer. Evidence for this type of behavior is shown in Fig. 3, a time profile of relative electron density collected during a recent equatorial investigation from Roi-Namur in the Kwajalein Atoll[32] (4 N mag). Because the data shown in the Figure is averaged over 1.5 sec time elements, only medium scale structure is evident on the larger scale gradients. However, finer resolution data to be published elsewhere shows irregularity scale sizes down to 1 meter.

High Latitude Phenomena. E-region irregularities occurring at high latitude sites are, to a large extent, found to be correlated with charged particle precipitation patterns. The convergence of the geomagnetic field in addition to large fluxes of energetic field aligned particles gives rise to charged particle interaction with neutral and ionized species at E-region altitudes. (For a review of auroral dynamics, ionospheric effects, and variation in incident energy, see Ref. 33.)

Because the patterns of precipitation are at different times spatially confined (horizontal range 5 km) or relatively widespread, the resulting ionization is similarly variable. Auroral ionization can be highly field-aligned, with dimensions parallel and perpendicular to the geomagnetic field having characteristic ratios of 5:1[34]. At other times, the auroral ionization can be sufficiently widespread to resemble a horizontally stratified layer. Maximum E-region densities can be greater than $10^6\ cm^{-3}$, but more typically peak densities are in the $10^5\ cm^{-3}$ range[34]. Characteristically the auroral E region has one major layer, ranging in thickness from approximately 10 to 30 km., a parameter controlled by the precipitating energy distribution (e.g., Ref. 34).

An example of auroral E-region profiles is shown in Figure 4. As in previous illustrations the data was collected by rocket-borne plasma probe, in this case launched from Poker Flat, Alaska 21 September 1976 at 2^h52^m (AST). From the data of Figure 4 two distinctly different E-layer density profiles separated by a horizontal distance of approximately 180km are apparent. Whereas the upleg E-layer is well-formed showing a 25 to 30 km depth at the .5 N_e^{max} points, the downleg profile is evidence of a weak diffuse pattern Preliminary results from on-board energetic particle detectors of Dr. R. A. Goldberg showed normalized counting rates reading maximum levels near 120 km altitude during the upleg portion of the flight. The energetic particle detectors covered the integrated energy ranges

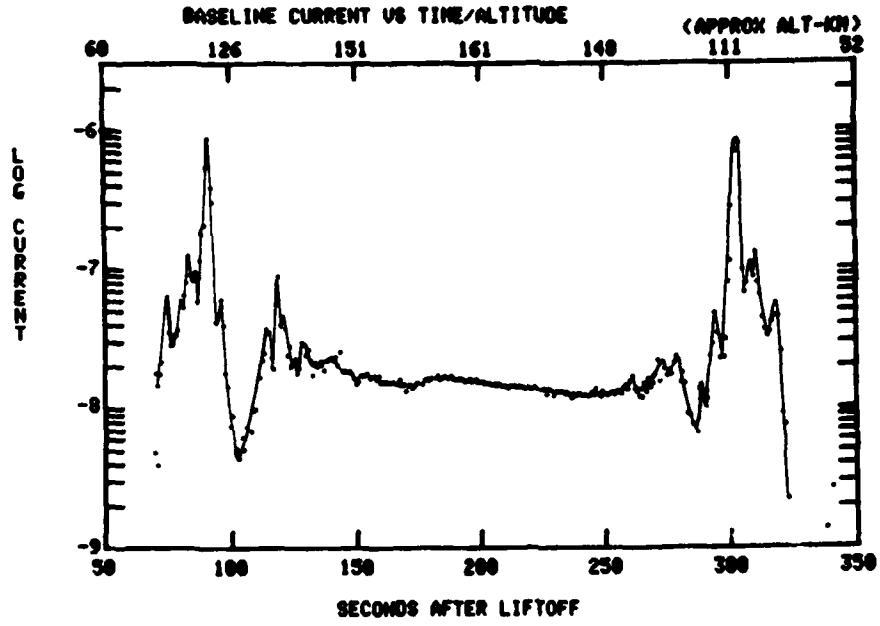


FIGURE 3

Relative electron density as a function of time after launch. 10^{-6} A→$N_e \approx 5(10^4)cm^{-3}$. Data collected over Marshall Islands in the equatorial Pacific (23 August 1977, 0 Hr 3 Min).

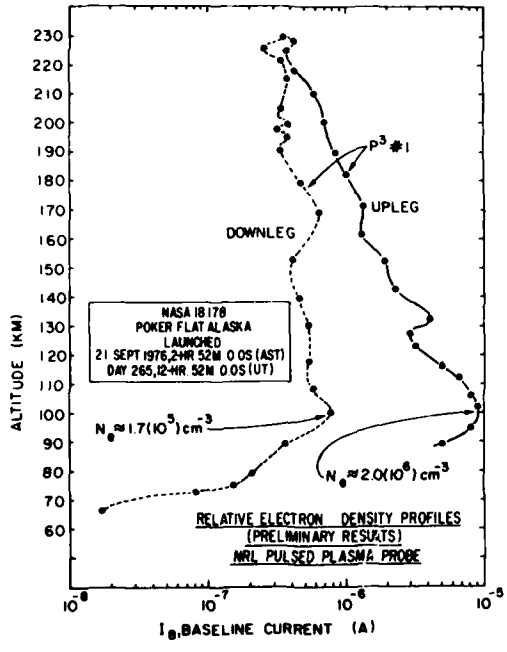


FIGURE 4

Relative electron density profiles in the auroral zone.

15-25 keV, 25-40 keV, and 40-100 keV with the largest integral flux contributions having occurred in the 15-25 keV range. These particle measurements support the basic correlation between precipation patterns and ionospheric plasma desities. Detailed correlation studies between fluxes, desities and energy deposition rates is currently being conducted.

Initial estimates of relative electron density at approximately 120 km are indicated in the figure for both upleg and downleg portions. These estimates were obtained from the plasma probe I-V characteristic and should be taken as correct only within a factor of 2. Additional details on this flight and a similar investigation conducted on 23 September 1976 at 1^h37^m (AST) are available in Ref. 35.

III. E-REGIONS MODELS AND HF DUCTING

Over the years considerable effort has gone into rocket, satellite, and radar studies of ionospheric irregularities. Irregularities cause fading of radio waves, scintillation of stellar radio sources, and fluctuations in transionospheric communication systems. In the previous section it was shown that E-region

irregularities were highly dependent upon geographic location and the prevailing geophysical conditions. In some cases the origin of the irregularities is still uncertain, although several mechanisms have been postulated. The overview and data sample were by no means intended to represent an exhaustive review of E-region phenomena, but rather to note some problems in understanding the E-region and the pitfalls that might attend the development of global models of the lower ionosphere.

This section focusses on the ionospheric models employed in the computational study of Toman and Miller[36]. They conducted theoretical analyses of HF propogation in numerically specified ionospheres. Their work introduced horizontal ionization gradients to provide favorable conditions for wave injection from the ground. It also considered ionospheric profiles with and without upper E-region ionization valleys. Of major consequence in determining ducting probability were horizontal ionization gradients and vertical electron density profiles. The authors used smoothed ionospheric profiles, noting that models representing mean ionospheric conditions are inherently restrictive. This restriction can be lessened by continued iterations between theory and experiment. We suggest that this paper is just one such iteration.

Figures 5a, and 5b are ionospheric profiles extracted from Figs. 2, and 5, respectively, of Ref. 36. The types of "no valley" profiles (A' and D' in 5a) selected by Toman and Miller [36] have no counterparts in any of the data presented and discussed in the previous section. We suggest that a profile like Fig. 5a D' most closely approximates daytime mid-latitude profiles when solar radiation "fills-in" the ionization valley and wind-shear convergence is less important (e.g., Narcisi[6] and Goldberg and Blumle[37]).

For purposes of ducting efficiency, Toman and Miller[36] found that no rays (6-21 Mhz, 0-11° launch angle elevation) were trapped in cases involving A'-type profiles. That D' profiles were found more effective in their study showed the controlling influence of peak E-region densities. While low densities in the E-region (type A') appear to hinder long-range HF ducting, it is possible that the densities can be too high resulting in the return of rays to the ground by single reflectance[38].

The model with an ionization valley in the upper E-region (Fig. 5b) is more representative of the experimental profiles presented in the previous section. However, except for the high-latitude profile, the layers were always observed to be much more irregular than the model. The

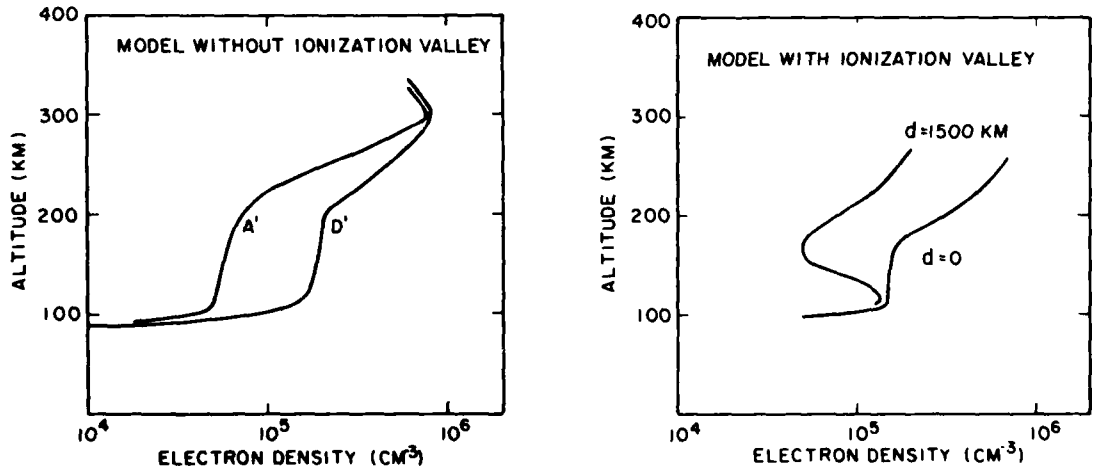


FIGURE 5

Numerically specified ionospheric profiles utilized in Ref. 36.

mid-latitude data showed an intense narrow peak superimposed on a broader, less-intense layer. The equatorial data showed multiple, high-structured E-layers. The data clearly points to some restrictions on conclusions drawn from the numerical model.

In the propagation analyses using profiles with ionization valleys, the authors[36] found that no rays were injected into the duct when the injection region horizontal gradient was zero (i.e., the case of a perfectly stratified ionosphere). They found that the number of ducted rays was greatest for large negative horizontal gradients in the injection region, and for deep valleys between the E- and F-layers. Typically, the study assumed $dN_e/dx = -6(10^{-2})elcm^{-3}/m$ at 230 km. Measured in terms of a normal horizontal range for the rocket data presented here (200 km), the suggested variation would be $N_e = 1.2(10^4)$ el cm^{-3}, a value which the data cannot argue for or against because of complications of large-scale, localized structure.

However, because horizontal density gradients are required only to achieve injection, local irregularities may well serve this purpose. Toman and Miller ducting models were smoothed and thus devoid of small scale irregularities within the duct (as opposed to within the region of injection). It remains for future study to determine the effects of irregularities for scattering successfully-injected rays out of the duct. We would expect that smooth long-ranging valley profiles would find their best experimental counterpart at mid-latitudes, at night, and over smooth earth topology.

IV. SUMMARY COMMENT

The data presented herein was collected in the nighttime ionosphere, and in every case showed the existence of an upper E-region valley. The lower E-region was generally irregular and its morphology more complicated than existing models allow.

An attempt to establish a global ionospheric model is necessarily as difficult as predicting the behavior of the variable geophysical mechanisms outlined in this paper.

Improving prediction capability will come only through improved models which will have to utilize E-region measurements having an expanded global distribution.

ACKNOWLEDGEMENTS

The authors thank the National Aeronautics and Space Administration and the Defense Nuclear Agency for their support of the research programs whence came the data for this paper. The assistance of R. Narcisi and R. Goldberg in providing additional data and in sharing and discussing their thoughts on E-region morphology is gratefully acknowledged.

REFERENCES

1) Matsushita, S., and E.K. Smith (Co-Chairmen) Third Seminar on the cause and structure of temperate latitude sporadic E, (Logan, Utah; Sept. 1971), in special issue Radio Science 7, 345-432, 1972.

2) Matsushita, S., and E.K. Smith (Guest Editors), Conference on recent advances in the physics and chemistry of the E region, (Boulder, Colorado; Aug. 1974), in special issue Radio Science 10, 229-418, 1975.

3) Matsushita, S., and E.K. Smith, (Chairmen) Proceedings of Second Seminar on the Cause and Structure of Temperate Latitude Sporadic-E, (Conference Proceedings, 19-22 June, 1968, Vail, Colorado).

4) Whitehead, J.D., "Production and prediction of sporadic-E", Rev. Geophys and Space Sci. 8, 65 (1970).

5) Strobel, D.F., "Physics and chemistry of the E-region: A review", Radio Sci. 9, 159 (1974).

6) Narcisi, R.S., "Mass spectrometer measurements in the ionosphere", in Physics and Chemistry of Upper Atmospheres, B.M. McCormac (ed.), D. Reidel Publishing Co. (Dordrecht-Holland, 1973) p 171.

7) Johnson, C.Y., "Ion and neutral composition of the ionosphere", in Annals of the IQSY, Vol. 5, Solar-Terrestrial Physics: Terrestrial Aspects, (MIT Press, 1969), p 197.

8) Goldberg, R.A., A.C. Aikin, B.V. Krishna Murthy, "Ion composition and drift observations in the nighttime equatorial ionosphere", J. Geophys Res. 79, 2473 (1974).

9) Aikin, A.C., R.A. Goldberg, and A. Azcarraga, "Ion composition during the formation of a mid-latitude E_s layer", in Space Research XIV,

(Akademie-Verlag, Berlin 1974), p. 283.

10) Oppenheimer, M., and L. Brace, "Recombination rate co-efficient of NO^+ from thermosphere daytime photochemistry", Trans. Am. Geophys. U. 57, 297, 1976.

11) Oppenheimer, M., A. Dalgarno, K. Kirby-Docken, and G.A. Victor, "Ion chemistry of the daytime thermosphere", J. Geophys. Res. (1976, in press).

12) Torr, D.G., M.R. Torr, J.C.G. Walker, L.H. Brace, H.C. Brinton, W.B. Hanson, J.H. Hoffman, A.O. Nier, and M. Oppenheimer, "Recombination of NO^+ in the ionosphere", Geophy. Res. Let. 3, 209, 1976.

13) Biondi, M.A., "Charged particle recombination processes", in Reaction Rate Handbook, DNA 1948H, P. 16-1, edited by M.H. Bortner and T. Baurer, DASIAC, GE TEMPO, Santa Barbara, Ca., 1972.

14) Axford, W.I., "Wind Shear Theory of the Formation of Temperate Zone Sporadic-E Layers", Spac. Res. 7, 126, 1967.

15) Macleod, M.A., "Sporadic-E Theory 1. Collision-Geomagnetic Equilibrium", J. Atmos. Sci. 23, 96, 1966.

16) Whitehead, J.C., "Survey of Sporadic-E processes", Space Res. 7, 90, 1967.

17) Szuszczewicz, E.P. and J.C. Holmes, "Observations of electron temperature gradients in mid-latitude E_S layers", J. Geophys. Res. 82, 5073 (1977).

18) Shen, J.S., W.E. Swartz, and D.T. Farley, "Ionization layers in the nighttime E-region valley above Arecibo:, J. Geophys. Res. 81, 5517, (1976).

19) Smith, L.G., "A sequence of rocket observations of nighttime sporadic E", J. Atmos. Terr. Phys. 32, 1247 (1970).

20) Constantinides, E., and J.F. Bedinger, "Observed redistribution of E-region ionization by neutral winds", J. Atmos Terr. Phys. 33, 461 (1971).

21) Axford, W.I., and D.M. Cunnold, "The wind-shear theory of temperate zone sporadic E", Radio Sci. 1, 191 (1966).

22) Matsushita, S., "Interrelations of sporadic E and ionospheric currents", in Ionospheric Sporadic E, E.K. Smith and S. Matsushita (eds.), (The Macmillian Co.,New York, 1962) p.344.

23) Richmond, A.C., S. Matsushita, and J.D. Tarpley, "On the production mechanism of electric currents and fields in the ionosphere", J. Geophys. Res. 81, 547 (1976).

24) Prakash, S., S.P. Gupta, and B.H. Subbaraya, "A study of the irregularities in the nighttime equatorial E-region using a Langmuir probe and a plasma noise probe", Planet. Space Sci. 18, 1307 (1970).

25) Morse, F.A., B.C. Edgar, H.C. Koons, C.J. Rice, R.F. Woodman, J.M. Pomalaza, and N.R. Teixeira, "Equion, an equatorial ionospheric irregularity experiment", J. Geophys. Res. 82, 578 (1977).

26) Holmes, J.C. and E.P. Szuszczewicz, "A versatile plasma probe", Rev. Sci. Instr. 46, 592 (1975).

27) Balsley, B.B., B.G. Fejer, and D.T. Farley, "Radar measurements of neutral winds and temperatures in the equatorial E-region", J. Geophys. Res. 81, 1457 (1976).

28) Balsley, B.B. and D.T. Farley, "Radar studies of the equatorial electrojet at three frequencies", J. Geophys. Res. 78, 227 (1973).

29) Farley, D.T., and B.B. Balsley, "Instabilities in the equatorial electrojet", J. Geophys. Res. 78, 227 (1973).

30) McDonald, B.E., T.P. Coffey, S. Ossakow, and R.N. Sudan, "Preliminary report on numerical simulation of type 2 irregularities in the equatorial electrojet", J. Geophys. Res. 79, 2551 (1974).

31) Fejer, B.G., D.T. Farley, B.B. Balsley, and R.F. Woodman, "Radar studies of anomalous velocity reversals in the equatorial ionosphere", J. Geophys. Res. 81, 4621 (1976).

32) Walker, D.N., E.P. Szuszczewicz, and J.C. Holmes, "Real-time display of ionospheric electron density profiles from a rocket-borne plasma probe: The payload, acquisition system, and the real-time results", NRL Rpt. 3649 (Nov. 1977).

33) Romick, G.J., "Review of auroral morphology, auroral dynamics, and modern observational techniques", Radio Sci. 9, 323 (1974).

34) Baron, M.J., "Electron densities within aurorae and other auroral E-region characteristics", Radio Sci. 9, 341, (1974).

35) Szuszczewicz, E.P., and J. C. Holmes, "An auroral ionospheric investigation: A preliminary report on electron density profiles", NRL Rpt. 3425 (Nov. 1975).

36) Toman, K., and D.C. Miller, "Computational study of long range high-frequency ionospheric ducting", Radio Sci. 12, 467 (1977).

37) Goldberg, R.A., and L.J. Blumle, "Positive ion composition from a rocket-borne mass spectrometer", J. Geophys. Res. 75, 133 (1970).

ANGLE OF ARRIVAL USING CONTINUOUS WAVE TRANSMISSIONS

R.N.F. BAULCH and F.C. BUTCHER
Division of Theoretical and Space Physics,
La Trobe University,
Bundoora, Victoria, 3083, Australia.

1. Introduction

The effect of the ionosphere on the angle of arrival of ionospherically reflected radio waves is conveniently monitored by measuring the phase difference between signals received on two orthogonal pairs of antennae which form a right angled isosceles triangle. If the antenna at the apex of the triangle is antenna 2 and if Δ_{12} and Δ_{32} are the measured phase differences between antenna 1 and antenna 2 and between antenna 3 and antenna 2 respectively (1 being north of 2 and 3 being west of 2) then

$$\Delta_{12} = \frac{2\pi d}{\lambda} \cos(\theta) \sin(\phi) \qquad (1)$$

$$\Delta_{32} = \frac{-2\pi d}{\lambda} \sin(\theta) \sin(\phi) \qquad (2)$$

where θ is the azimuth angle (bearing), ϕ is the zenith angle, d the separation between the antennae and λ is the wavelength of the received radio wave. However, in measuring these phase differences, two distinct classes of change occur:

(a) rapid fluctuations which are assumed to be due to the ionosphere behaving as a rough reflector (Bramley, 1951; Pfister, 1972)

(b) slow varying changes with periods of a minute or more which are associated with changes in θ and ϕ of the radio wave. For this case the angle of arrival is given by

$$\phi = \sin^{-1} \left[\frac{\lambda}{2\pi d} (\Delta_{12}^2 + \Delta_{32}^2)^{\frac{1}{2}}\right] \qquad (3)$$

$$\theta = \tan^{-1} \left(\frac{-\Delta_{32}}{\Delta_{12}}\right) \qquad (4)$$

Changes in angle of arrival may be attributed to changes in electron density gradients (tilts) in the ionosphere (Bramley and Ross, 1951; Bramley, 1953) and these may be associated with the passage of internal gravity waves (Reynolds and Morgan, 1975). Tilts may be associated with other phenomena, for example a solar zenith angle dependence in the F region has been observed (Morgan, 1972, 1974) this being more predominant near sunrise and sunset (Ross and Bramley, 1949). In all these observations data was averaged over several seconds in order to remove rapid fluctuations so that then results could be interpreted by assuming a single ray reflected from the ionosphere. Tilts have also been associated with Sporadic E (Es); both non-blanketing and blanketing and Spread F (Clarke, 1972). Since it is these tilts that effect the bearing or angle of arrival measurements, it would seem appropriate to consider possible errors in these measurements.

In this paper we discuss the problems of measuring the angle of arrival using a continuous wave transmission and, in particular, wave interference is considered when more than one ray is received. Such interference may be caused by another transmitter which has the same nominal frequency, ground wave or multiple hop propagation and it is shown that in such circumstances the error in the angle of arrival may be reduced by integrating over a suitable period. The method described may also be applicable to pulse transmissions when two or more echoes are received with time separations less than the pulse width.

Measurements of the angle of arrival were made on a CW 2.5 MHz signal transmitted by Telecom Australia from Lyndhurst, Victoria (38° 33' S, 145° 16' E (geographic)) at the La Trobe University Field Station at Beveridge (37° 28' S, 144° 56' E (geographic)) a ground distance of 71.4 km. From Beveridge, Lyndhurst has an azimuth of 145.8° (geomagnetic). The transmission was only transmitted from 09.00 to 18.00 (L.T.) and therefore was mainly an F region reflection.

2. Interference Effects

When one or more ray is received at an antenna, then, because the signals differ in phase and amplitude, they can be added

vectorially. The phase difference is due to both the difference in the angle of arrival of the rays and the difference in the Doppler shift in frequency of the radio waves.

Consider two rays of amplitude a and b whose angles of arrival are θ_a, ϕ_a and θ_b, ϕ_b respectively, and which have a phase difference α at antenna 2. The resultant amplitude and phase at each antenna may be found by considering the Argand diagrams in Figure 1 for the phase differences between antenna 1 and antenna 2 (antenna 1 is to the north of antenna 2).

From Figure 1(a) the resultant at antenna 2 is:

$$R_2 \exp(i\Delta_2) = a + b\cos(\alpha) + ib\sin(\alpha) \quad (5)$$

or

$$\tan\Delta_2 = \frac{\sin(\alpha)}{n + \cos(\alpha)}$$

and

$$R_2 = \frac{a}{n}\left(n^2 + 1 + 2n\cos(\alpha)\right)^{\frac{1}{2}}$$

where $n = a/b$. Similarly for antenna 1 Figure 1(b) gives

$$\tan\Delta_1 = \frac{n\sin(\Delta_{12a}) + \sin(\Delta_{12b})}{n\cos(\Delta_{12a}) + \sin(\Delta_{12b})} \quad (6)$$

and

$$R_1 = \frac{a}{n}\left(n^2 + 1 + 2n\cos(\Delta_{12a} - \Delta_{12b})\right)^{\frac{1}{2}}$$

where

$$\Delta_{12a} = \frac{2\pi d}{\lambda}\cos(\theta_a)\sin(\phi_a)$$

$$\Delta_{12b} = \frac{2\pi d}{\lambda}\cos(\theta_b)\sin(\phi_b) + \alpha$$

Similar expressions may be obtained for Δ_3 where Δ_3 is the phase at antenna 3. The phase differences are then $\Delta_{12} = \Delta_1 - \Delta_2$ and $\Delta_{32} = \Delta_3 - \Delta_2$ and when substituted into equations (3) and (4) give the "apparent" azimuth and zenith angles. It is found that for two interfering rays the instantaneous errors in θ and ϕ depend on the azimuth angle of the signal with the largest amplitude (i.e. the error is a function of antenna orientation). This error is found to be a minimum at

$$\theta = \frac{k\pi}{4} \qquad k = 0, 1, \ldots..7$$

and a maximum at

$$\theta = \frac{k\pi}{4} + \frac{\pi}{8} \qquad k = 0, 1, \ldots..7$$

Δ_{12} and Δ_{32} are usually integrated over a period of several seconds in order to reduce severe diversity effects. However, for two rays, errors in θ and ϕ may be still obtained since the errors are a function of α which is a function of time. Figure 2 shows the error in θ and ϕ for first and second hop ($\phi_a = 21.6^o$, $\phi_b = 10.6^o$ and $\theta_a = \theta_b = 145^o$) as a function of amplitude ratio for different integration periods. For n greater than 2, the errors are almost zero for an integrating period of 2π and it is found that for a given n, the errors vary in a damped oscillatory manner with the integrating period. This is shown in Figure 3(a) with n equal to 4. For values of n greater than four it is found that the errors in θ and ϕ are less than one degree if the integrating period is greater than $1/\Delta f$ (where Δf is the difference in the Doppler shift frequency between the two rays). Obviously the larger n the shorter the integration time needed for the same error in θ and ϕ. The integration for a period greater than $1/\Delta f$ removes the dependency of errors in θ and ϕ on the orientation of the antennae. It also removes the errors due to higher variations in the azimuth angle of the second hop (Bramley (1954) found that the azimuth variation of the second hop is likely to be an order of magnitude greater than the first hop). This is shown in Figures 3(b) and 3(c) where the error in the measured θ and ϕ are shown for second hops with a ϕ of 10.6^o and a θ of 85^o and 205^o.

For n between 1 and 2 the errors in θ and ϕ obtained from the integrated values of the phase may be quite large as seen in Figure 2. In this case averaging over a period greater than $1/\Delta f$ does not reduce the errors to an acceptable level. In most practicle cases, however, where the transmitter and receiver are separated by a distance of between 50 and 500 kms such a small amplitude ratio for interference between the first hop and ground wave or first hop and second hop from the E region should not be expected. The exception is when both non-deviative and deviative absorptions are very low and such a situation may exist when there are strong Es reflections at night.

Hence in order to obtain the true angle of arrival Δ_{12} and Δ_{32} must be integrated over a period longer than $1/\Delta f$. Typically for the E region Δf is 0.1 Hz so that the integration must be over a period greater than 10 seconds.

3. Experimental Technique

The signals from the three receiving antennae were switched sequentially through a single receiver in order to measure the phase difference between the antennae. A phase lock loop (PLL) oscillator was used as a phase reference and the signal from antenna 2 (which is at the apex of the triangle) controlled the frequency and phase of this oscillator. The PLL oscillator output was phase detected against the signals from the other antennae and these phases had subtracted from them the phase difference between the input and output of the PLL oscillator. A block diagram of the system is shown in Figure 4. Such a system has the advantage that no errors in phase occur due to any constant Doppler shift between the

received signals and our local standard oscillator. Only changes in the Doppler shift that occurred during a switching cycle caused errors and, because each antenna was sampled twelve times a second (this was the fastest switching rate allowed by the receiver), a change of 0.05 Hz per second gave an error of less than half a degree in phase (which gives less than half a degree in both θ and ϕ). Such a sustained high rate of change in the frequency is most unusual.

4. Experimental Results

Experimental results were taken from late June to late October, 1976 and in general it was found that azimuth (bearing) and zenith angles were approximately what was expected from a normal horizontally stratified medium. An example is the result obtained on the 6th of July and shown in Figure 5. It may be seen that the record may be split into two distinct periods - from 15.20 to 16.30 and from 16.30 to 18.00 (L.T.). Ionograms taken at the receiving site by a digitized ionosonde indicated that at 15.30 (L.T.) the equivalent vertical frequency of the 2.5 MHz (2.35 MHz) was being reflected from the normal E region at 98.2 kms. At 16.30 (L.T.) the E region had decayed and reflection was from an Es layer high on the E region cusp at 140 kms. By 18.00 (L.T.) the E region had completely decayed and the Es layer had descended to a height of 100 kms where it had a critical frequency of 5.1 MHz. Using the theorem of Breit and Tuve (1926),

$$\tan(\phi) = \frac{D}{2h'} \qquad (7)$$

where h' is the vertical group height and D is the distance between the transmitter and receiver. The zenith angle at 15.30 (20^o) and at 16.30 (14.5^o) give a group height of 99 kms and 139 kms respectively which are very close to the group heights measured by the digitized ionosonde. For the Es layer at 18.00 (L.T.) however, there is an apparent error of 25 kms in h' using equation (7). There are two possible explanations for this:

(1) A mixture of first and second hop was being received from a flat Es layer Assuming some Doppler shift in frequency did occur then from Figure 2 an amplitude ratio of 1.5 would be needed for such an error in ϕ. Greater variations in θ would then be explained by the amplitude ratio changing; for example, from Figure 2 a change in amplitude ratio from 1.5 to 1.6 gives a change of 10^o in θ. However, from Figure 2 much larger variations would be expected to be obtained in ϕ for such an amplitude ratio change. It is also difficult to explain how such an amplitude ratio could be sustained at this time over such a period.

(2) A more feasible explanation was that the Es layer had a real tilt along the path. Using the equation for tilts along the path

$$\gamma_L = \frac{1}{2}\left(\phi_d - \frac{\pi}{2} + \tan^{-1}\left(\frac{1}{2\tan\phi_e - \tan\phi_d}\right)\right)$$

where $\tan(\phi_d) = \tan(\phi)\cos(\theta - \theta_e)$, γ_L is the tilt along the path, θ and ϕ the measured azimuth and zenith angles, and θ_e and ϕ_e are the expected angles of arrival for a horizontally stratified medium, a 4^o tilt in γ_L is obtained for this Es layer. The slight increase in ϕ between 16.35 and 17.00 is probably due to the group height of the Es layer decreasing and as this occurs the tilt forms in the Es. However, greater variations in θ are only partially explained by a $\cot(\phi)$ dependence of $\theta - \theta_e$ for the same tilt across the path (γ_c) since it may be shown that

$$\sin(\theta - \theta_e) = \tan(\gamma_c)\cot(\phi)$$

For small tilts across the path, changing ϕ from 20^o to 15^o would be expected to give an increase of 1.4 in $\theta - \theta_e$. Between 17.05 and 17.40 (L.T.) the variations in $\theta - \theta_e$ increased by about this amount, but larger variations were observed around 16.45 and 17.50 (L.T.).

It was shown that for CW radio waves integrating the phase difference between the antennae reduced the errors in the angle of arrival due to multiple hop propagation modes; for interference between first and second hop, the maximum error for integration over a period greater than $1/\Delta f$ being one degree for an amplitude ratio of four and half a degree for an amplitude ratio of ten. This is confirmed by the results obtained both in the normal F region and in the Es layer, the latter having a tilt comparable to those obtained by Clarke (1972) who observed tilts of up to 4^o in blanketing Es.

Acknowledgements

The assistance of Dr. P.R. Hammer in the construction of the equipment was appreciated.

References

Bramley E.N. 1951 *Proc. I.E.E.*, 98, 10.

Bramley E.N. and Ross W. 1951 *Proc. R. Soc.* [A], 207, 251.

Bramley E.N. 1953 *Proc. R. Soc.* [A], 220, 39.

Bramley E.N. 1954 *Proc. I.E.E.*, 102B, 544.

Breit G. and Tuve M.A. 1926 *Phys. Rev.*, 28, 554.

Clarke R.H. 1972 *Ph.D. Thesis*, Department of Physics, Queensland University.

Morgan A.D. 1972 *AGARD Conf. Proc.* No. 115, Paper No. 28.

Morgan A.D.	1974	*J.A.T.P.*, 36, 1675.
Pfister W.	1972	*J.A.T.P.*, 34, 999.
Reynolds J.S.B. and Morgan A.D.	1975	*J.A.T.P.*, 37, 545.
Ross W. and Bramley E.N.	1949	*Nature*, 159, 132.

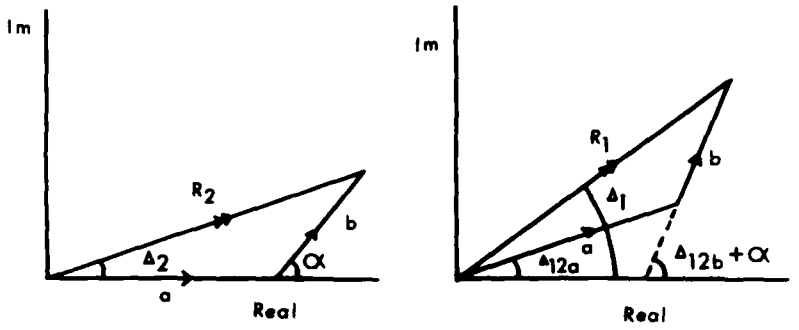


Figure 1 : Argand diagrams for the phase at antennae 1 and 2.

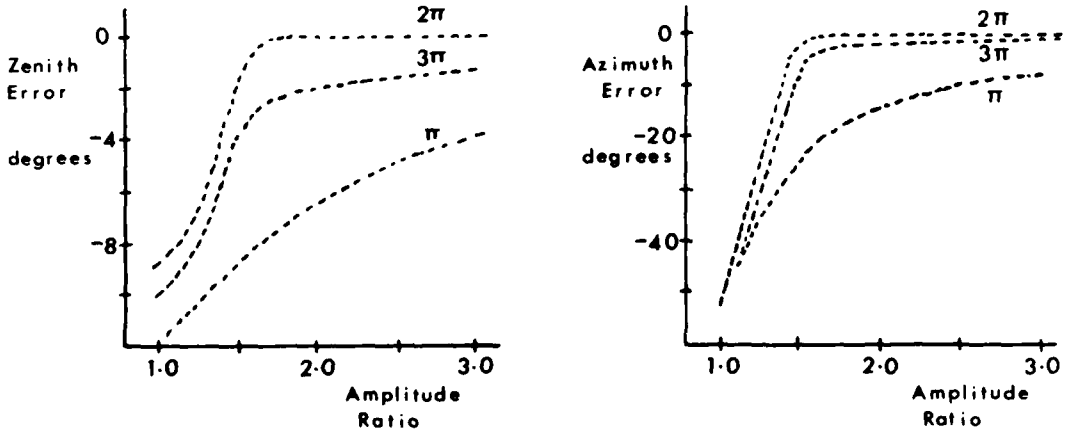


Figure 2 : Angle of arrival error as a function of n for a given integration period - 1st and 2nd hop interference.

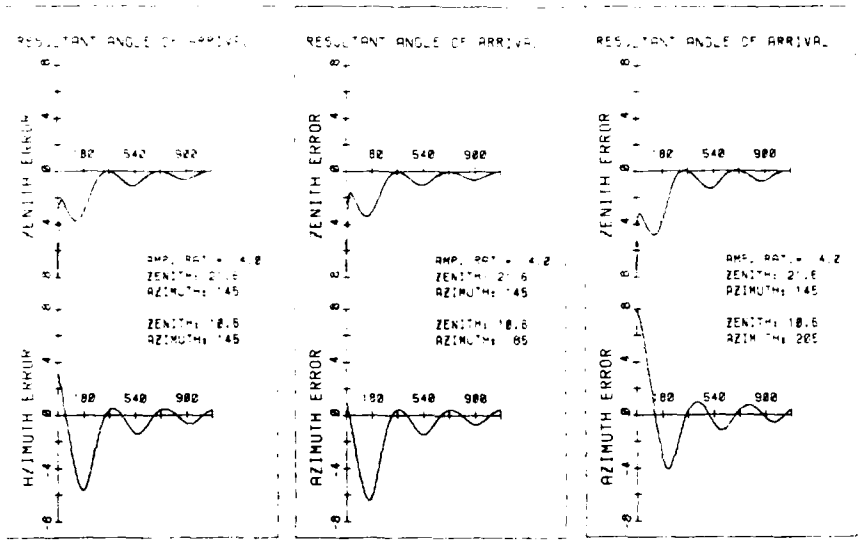


Figure 3 : Angle of arrival error as a function of integration period.

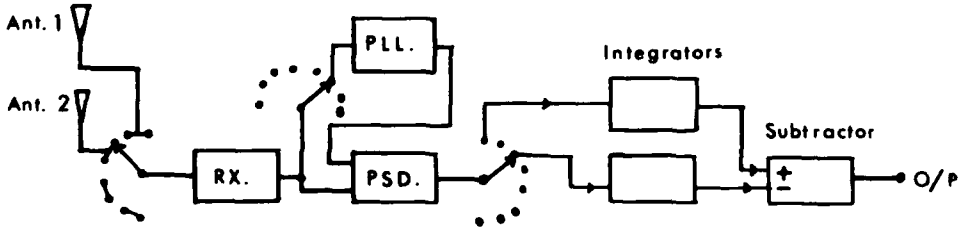


Figure 4 : Block diagram of the system for measuring the phase difference between a pair of antennae.

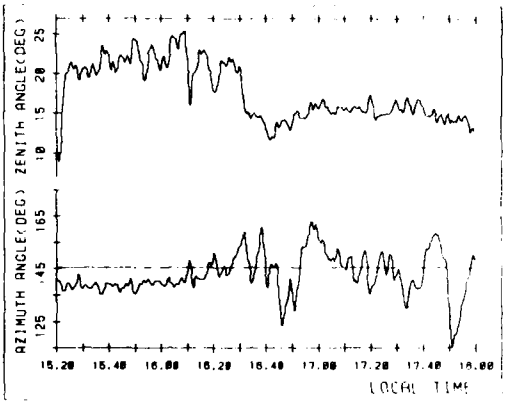


Figure 5 : Angle of arrival on 6th of July, 1976.

HF SKY WAVE DIRECTION FINDING USING A FIXED CROSSED SPACED LOOP ANTENNA ARRAY

T. C. Green, J. E. Hipp, and W. G. Guion
Southwest Research Institute
San Antonio, Texas

INTRODUCTION

Direction finding on HF radio signals which have propagated through the ionosphere (sky wave signals) has classically been accomplished using medium and wide aperture antenna arrays such as Adcocks, interferometers, Wullenwebers (CDAA), and Beverage antennas. There remains, however, a continuing requirement for a direction finder which combines both small physical size and a passive configuration to provide high accuracy, automatic DF operation agains sky wave multipolarized signals in the HF range.

In the HF range, loop antennas are typically used for small aperture direction finding applications because of their intrinsic receiving efficiency compared to other antennas with similar dimensions. However, the conventional crossed simple loop or rotating simple loop direction finder is essentially useless against sky wave signals due to pattern changes with incident signal polarization.

A polarization independent direction finding antenna can be obtained using individual loop elements connected as a coaxial spaced loop antenna. HF sky wave direction finding with a coaxial spaced loop has typically been accomplished by rotating the antenna and observing the analog display of the polarization independent nulls. With appropriate sense functions, the rotating spaced loop direction finder provides an unambiguous accurate analog display of the incident sky wave signal azimuth bearing. For modern DF applications, however, a fixed passive DF antenna is desirable which can be remotely controlled and operated by computer integrated systems.

The fixed crossed spaced loop antenna developed and patented* at SwRI has been primarily used as a computer controlled ground wave (vertically polarized signals) direction finder in applications to date. HF propagation, however, is predominately sky wave transmissions for signals originating from sources at relatively large distances. Sky waves present a much more complex problem to direction finding antennas due to the changing, unpredictable polarization and elevation angle. They may be vertically, horizontally, circularly, or elliptically polarized and are seldom stable more than a few seconds.

Recent work performed at SwRI under an internal research project consisted of both a theoretical and an experimental program to develop the fixed crossed spaced loop antenna as an appropriate sky wave direction finder. The work described herein includes (1) the theoretical development of sky wave DF and sense algorithms for the fixed crossed spaced loop antenna, (2) instrumentation of a crossed spaced loop direction finder for validating the DF and sense algorithm, (3) development of computer programs for both theory verification and data acquisition/processing, and (4) data measurement, analysis, and post-processing.

THEORY

A coaxial spaced loop antenna can be formed by combining two simple loop antennas into a simple two-element interferometer array as shown in Figure 1a. The generalized response of a single coaxial spaced loop is shown below for vertical and horizontal polarization:

$$E_T = -E_V \sin\theta \sin 2\phi + E_H e^{j\Phi_H} \sin 2\theta \sin^2\phi \qquad (1)$$

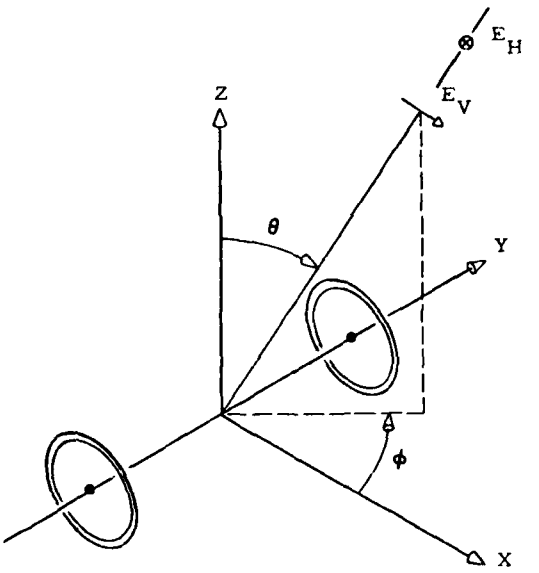
where

E_V = Amplitude of vertical component

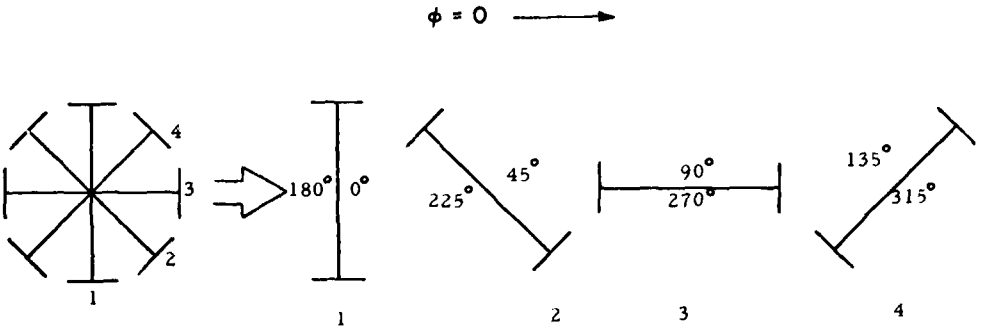
E_H = Amplitude of horizontal component

θ = Incidence angle

*Travers, D. N., Eight-Loop Antenna System and Method of Scanning Same, U.S. Patent No. 3,329,954, 4 July 1967.



a. Coaxial Spaced Loop



b. Eight-Loop Array of Crossed Coaxial Spaced Loops

Figure 1. Spaced Loop Configuration

ϕ = Azimuth angle

ϕ_H = Relative phase between vertical and horizontal component.

The spaced loop response for rotation in the azimuth plane is shown as follows:

$$E = -E_V \sin\theta \sin 2(\phi - \alpha) + E_H e^{j\phi_H} \sin 2\theta \sin^2(\phi - \alpha)$$

where α is the rotational azimuth angle.

Careful examination of the above equation shows that whenever the argument $\phi - \alpha$ is equal to 0° or 180°, that is, whenever the azimuth position of the spaced loop is aligned directly with the azimuth angle of arrival of the incident signal or aligned in the opposite direction (180°), the output voltage is 0. These are the so-called interferometer nulls of the rotating spaced loop, and are <u>independent</u> of the polarization and elevation angle of the incident signal.

Examples of the polarization and angle of elevation independence of coaxial spaced loop patterns are shown in Figure 2 where three different elevation angle conditions are compared for polarization varying from vertical to horizontal. The important feature of these patterns is that the null placement at 0° and 180° does not change for the target signal at 0° azimuth regardless of signal polarization or elevation angle. The polarization and elevation angle independence of the spaced loop interferometer nulls has been used successfully by SwRI in development of ground-based rotating spaced loop sky wave direction finders. As shown in Figure 1b, four spaced loop antennas can be arrayed in a symmetrical configuration to form an eight-loop array with each loop spaced 45° in azimuth. The development described herein is based upon the premise that the sky wave direction finding properties of the rotating spaced loop can be synthesized by sampling the outputs from individual fixed spaced loops arranged appropriately.

By expanding Equation (1) and rearranging terms, the rotating spaced loop output voltage can be re-expressed as

$$E = A_0 + A_2 \cos 2\alpha + B_2 \sin 2\alpha \qquad (2)$$

where

$$A_0 = \frac{E_H e^{j\phi_H} \sin 2\theta}{2}$$

$$A_2 = -E_V \sin\theta \sin 2\phi - \frac{E_H e^{j\phi_H} \sin 2\theta \cos 2\phi}{2}$$

$$B_2 = E_V \sin\theta \cos 2\phi - \frac{E_H e^{j\phi_H} \sin 2\theta \sin 2\phi}{2}$$

Examination of Equation 2 shows that the spaced loop pattern is band limited and the second harmonic of the rotational angle is the highest order component (for a fixed target azimuth ϕ and elevation angle θ). Therefore, from the Nyquist sampling theorem only four samples would be required to reproduce the original function. However, with the crossed spaced loop antenna arranged as shown in Figure 1b, four independent samples can be obtained over 180° of azimuth corresponding to eight samples over 360° of azimuth (twice the number required by the Nyquist theorem). From the above, therefore, it is apparent that sampling the fixed crossed spaced loop output voltages should provide sufficient information to synthesize a rotating spaced loop response and determine ϕ, the azimuth angle of arrival. Two types of technical approaches were investigated to calculate the azimuth angle of arrival from the eight spaced loop samples.

An obvious approach is to take the Fourier transform of the individual voltage samples to determine the Fourier coefficients A_0, A_2, and B_2 of the sampled wave form. Once the coefficients are determined there is a simple algebraic solution to the azimuth angle ϕ with, however, a four-way ambiguity. Another method for obtaining the Fourier coefficients uses a direct algebraic solution of Equation 2 to solve for A_0, A_2, and B_2 in terms of the antenna output voltages (phasors).

The inherent four-way ambiguity in the spaced loop response pattern (pattern null positions) is a result of the two-way ambiquities in both the simple loop nulls and the interferometer nulls which form the spaced loop pattern. By adding simple loop voltages (phasors) to the analysis, the simple loop nulls can be readily identified and eliminated. Furthermore, by comparing the spaced loop voltages and simple loop voltages in both amplitude and phase, the correct interferometer null representing ϕ, can also be identified.

Both Fourier transform and direct algebraic solution for the Fourier coefficients perform equally well for computer simulated data. Figure 3 shows examples of the computed Fourier coefficients for three polarization and elevation angle conditions from a fixed azimuth target signal (e.g., ϕ = 37.5°). The important point to note is that although the coefficients change in amplitude (and phase), the target azimuth solution for each of these sets is the same, namely 37.5°.

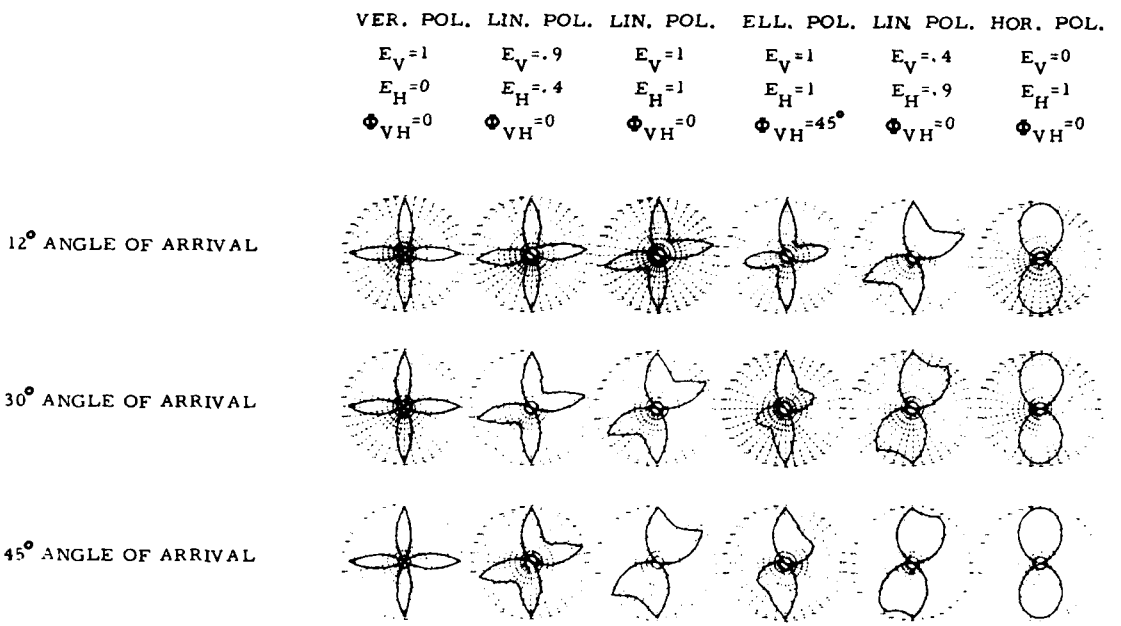


Figure 2. Coaxial Spaced Loop Patterns as a Function of Signal Polarization and Angle of Incidence

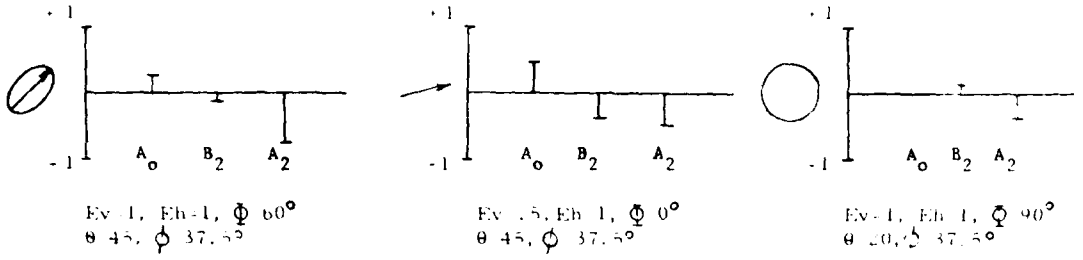


Figure 3. Examples of Fourier Coefficients Sets for Fixed Target Signal

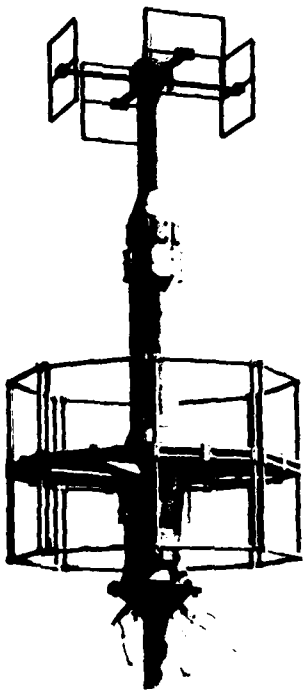


Figure 4. Coaxial Spaced Loop Antenna

FIXED SPACED LOOP INSTRUMENTATION

To implement the sky wave DF concept using a fixed spaced loop antenna, a crossed spaced loop breadboard antenna was assembled and installed at one of the SwRI HF test ranges. Figure 4 is a photograph of the breadboard DF antenna installed approximately 30 feet above ground (on top of a building). The eight-loop array, consisting of four spaced loop antennas, is visible on the lower bay of the antenna. Four simple loop antennas are shown at the upper position on the antenna main mast providing the outputs necessary for resolving ambiguities in the calculated bearing. The required simple loop output voltages can also be derived from the individual loops of the spaced loop antennas, but for construction simplicity separate simple loops were used. Each spaced loop antenna in the array has physical dimensions of 60 inches (d) x 40 inches (h) x 20 inches (w). The individual spaced loop antenna patterns are created by connecting the outputs of the diametrically opposite simple loop antennas in parallel opposition, while the two simple loop antenna outputs are formed by connecting two diametrically opposite simple loop antennas in parallel assistance.

Figure 5 is a block diagram of the complete equipment group assembled for the HF sky wave DF measurements.

Varactor diode tuning for the spaced loop antennas and preamplifiers for each individual antenna function (six total) are contained within the antenna mast. Varactor diode tuning was used for an expedient and low cost method of obtaining improved DF antenna sensitivity. Computer controlled digital tuning would be used in an operational configuration. The tuning voltages for the individual spaced loop antennas are provided at the operator's console.

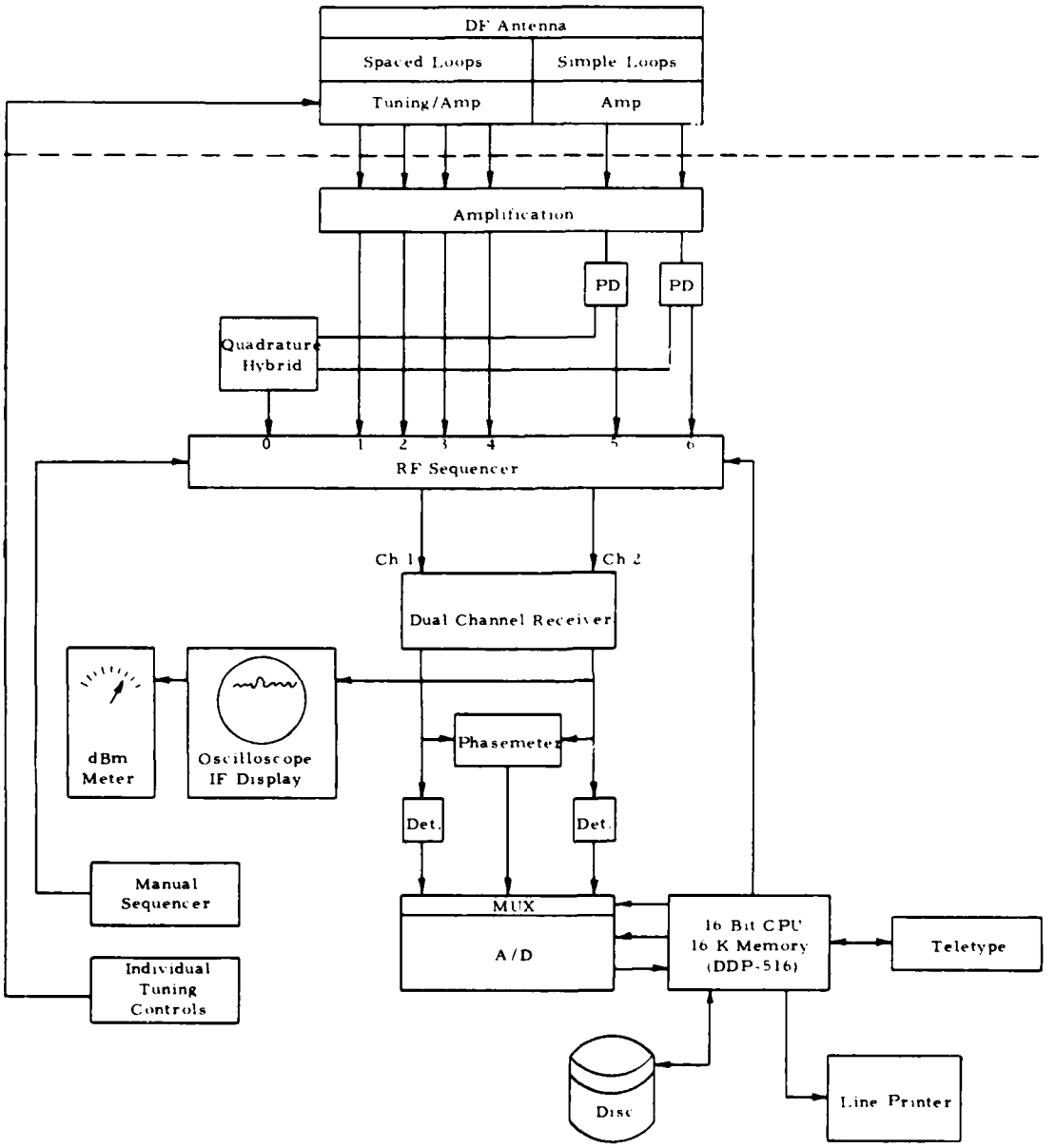


Figure 5. Block Diagram--HF Sky Wave DF Instrumentation

Six individual coaxial cables carry the outputs of the antennas inside the building to the RF sequencer and processor. A reference signal is provided by quadrature summing the amplified output of the two simple loop antennas and thus synthesizing an omni-directional antenna. The omni signal and the outputs of the six individual DF antennas are connected to seven inputs of a solid state RF sequencer. The RF sequencer can operate either under manual sequencer control or under computer sequencing control. The dual-channel receiver outputs are supplied to the phase meter and to a pair of IF detectors.

Outputs from the two detectors and the phase meter are supplied to an analog multiplexer and to an A-to-D converter. A Honeywell DDP-516 computer with 16 K word memory supplies the controlling functions to perform automatic data taking. The DDP-516 controller determines the RF sequencer state, the signal selected by the multiplexer, and the A-to-D conversion timing. The sequencer is operated at a frame speed of 60 per second allowing fixed spaced loop operation against high speed ICW and FSK signals.

The SwRI developed fixed spaced loop DF and sense algorithms were used in a DDP-516 computer to calculate the azimuth angle of arrival for each frame of data (six antenna voltage measurements). Additional program development was also necessary in order to interface the DDP-516 computer to the high speed RF sequencer and fixed spaced loop RF inputs. A Fortran callable subroutine was written in assembly language to control the RF switching, the multiplex switching, the A/D conversion, and timing of these events. In addition, this subroutine stores the results in the 516 computer memory.

SKY WAVE DF MEASUREMENTS

A series of ground wave measurements were initially accomplished to validate both the DF antenna and the basic algorithms. Next, a series of sky wave DF measurements were performed using targets of opportunity. Test objectives for the sky wave measurements included:

(1) Demonstrate fixed spaced loop capability to accurately determine azimuth angle of arrival for multipolarized sky wave signals at elevation angles from 0° to 60°.

(2) Demonstrate DF algorithm capability to perform under high speed ICW, FSK and similar modulations.

Commercial stations were typically used for the sky wave DF measurements and consisted primarily of fixed shore-to-ship broadcast stations operating in manual Morse (ICW) at speeds of nominally 18 to 30 words per minute. WWV in Fort Collins, Colorado and other known AM stations in the HF range were also used from time to time to check system performance. These stations were selected because of the capability of obtaining station identification and therefore the station coordinates. From the station coordinates the true great circle bearing from the SwRI laboratories to the station and the range were computed.

For each station, 500 acceptable frames of data were obtained. Acceptance criteria for each frame of data was that all antenna amplitudes must be within a predetermined linear range of the receiver and detector circuits. The computer would then automatically terminate the data collection process and produce a histogram of the 500 bearings distributed over 0° to 360° in 5° resolution cells. The operator then performed post-processing on the bearing histogram display to obtain the average bearing to the target. Figure 6 shows an example of the 0° to 360° histograms obtained at the termination of the data collection process as well as the post-processing analysis performed by the DF operator. The operator observes the dominant bearing peak on the 0° to 360° histogram and selects a reprocessing window (on the TTY terminal) for additional analysis. The average bearing is then recomputed for only the data within the selected azimuth window excluding all other data. In this manner an accurate refined bearing typically with bearing errors and standard deviations less than 5° can be obtained on sky wave signals independent of target range, elevation angle, and polarization.

For the tests described herein, ninety-four 500 frame average bearings on sky wave signals were obtained on signals over the 5 to 16 MHz range and at target ranges from 835 to 12,555 km. The average sky wave bearing accuracy demonstrated by the fixed crossed spaced loop antenna can be computed from this data collection. Figure 7 shows a histogram of the bearing error distribution for the averaged bearings. The mean error for this data group is 0.98° with a standard deviation of 4.31°. The range and azimuth distribution of the target signals (with respect to the SwRI laboratories in San Antonio, Texas) are also shown in Figure 7.

In conclusion, the fixed spaced loop antenna using the sky wave DF and sense algorithms described herein provides a versatile and accurate HF sky wave direction finder. The antenna is compatible with shipboard and ground-based environments where antenna size and weight must be constrained. The inherent use of high speed computer processing for the sky wave DF algorithm provides additional capabilities for using the fixed spaced loop sky wave direction finder in both civilian and military intercept and DF requirements as well as radio propagation research and development.

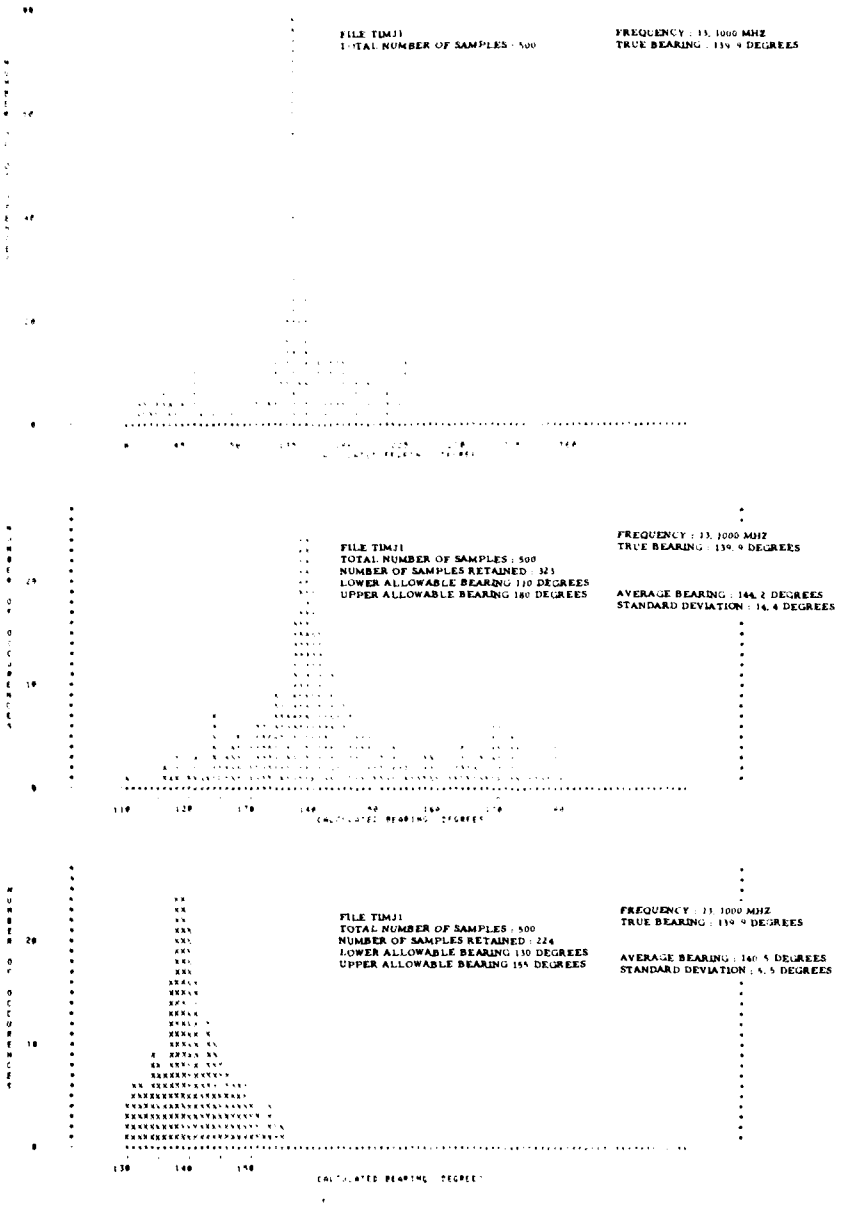
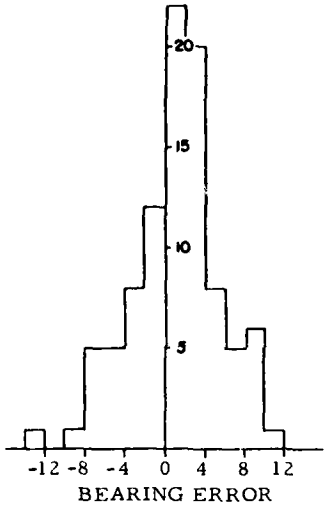
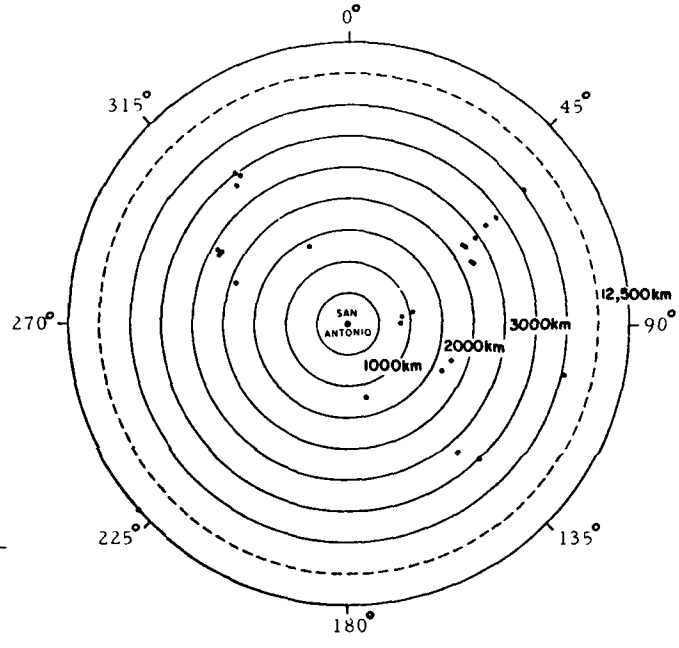


Figure 6. Example DF Processing for Average Bearing



a. SUMMARY HISTOGRAM OF AVERAGE BEARING ERRORS FOR 94 FILES OF 500 FRAME DATA.



b. RANGE / AZIMUTH DISTRIBUTION OF TARGETS.

Figure 7. Fixed Spaced Loop Sky Wave DF Measurements

THE EFFECT OF SIGNAL MODULATION IN THE APPLICATION OF A WAVE-FRONT LINEARITY TEST IN HF DIRECTION-FINDING

D. W. Rice, G. O. Venier and G. Atkinson
Department of Communications
Communications Research Centre
Ottawa, Ontario. K2H 8S2

Abstract Wave interference arising from multiple propagation modes constitutes a significant source of error to HF direction-finding systems, particularly for apertures of a few wavelengths or less. Various workers have attempted to circumvent this difficulty by taking bearing snap-shots at times when propagation was believed, on the basis of a wave-front linearity test, to be nearly unimodal. These short periods of quasi-unimodal propagation have been postulated to result from the independent amplitude fading of the various modes present. It is suggested here that a factor contributing to the observation of apparently unimodal propagation on short time scales is the presence of amplitude modulation on the transmitted signal. Provided the reciprocal of the modulation bandwidth is similar to or less than the mode transit-time differences, the modulation will tend to shorten the time interval between events in which a signal arrives via one mode with large amplitude, coincident with low amplitudes of the other arrivals.

Measurements have been made over a one-hop path using a 6 element phase-measuring interferometer of 663 metres aperture. Both unmodulated CW and SSB-modulated test signals were transmitted. For the CW signal, it is found that wave-front nonlinearities change on a time scale of a few seconds, whereas for the modulated signal, wave-front nonlinearities change on a much shorter time scale which is determined mainly by the modulation.

INTRODUCTION

Wave interference arising from multiple propagation modes constitutes a significant source of error in HF direction-finding, particularly for those systems with apertures of a few wavelengths or less. One suggested way to circumvent this difficulty is to take bearing measurements only at times when propagation appears to be nearly unimodal (i.e. dominated by one ionospheric mode). Probably the earliest attempt at this was the 'Dot-Lock' technique (1), in which bearings were taken on the leading edge of an amplitude-keyed signal. This same technique was later extended to other types of modulation (2). It has also been noted that it may be possible to take advantage of mode amplitude fading or scintillation, which provides periods of a basically plane-wave condition at the receiving array (2), (3).

Ionospheric channel measurements (4) indicate that the ionosphere is well represented for communications purposes as a summation of statistically-independent Rayleigh-fading components, one for each resolvable propagation mode. On this basis, there is a significant probability that the amplitude of one mode will be large, coincident with low amplitudes of the other modes. If bearings are taken at these instants of quasi-unimodal propagation, then the bearings will be relatively free from the effects of wave interference.

The measurements reported here support these notions. It is found that, for a CW signal, wave-front nonlinearities have a time-domain behaviour on a time scale consistent with measured mode fading rates. It is also found that, for an amplitude-modulated signal propagated over a time-spread, multi-path ionospheric channel, the time scale of changes in wave-front nonlinearity is commensurate with that of the modulation.

THE TWO-MODE CASE

Multi-mode interference can be very complex in its details; here we consider only the two-mode case which illustrates most of the properties of interest. Two-mode interference is most easily visualized as the weaker mode

producing corrugations in the otherwise plane equi-phase surface of the stronger mode (5). Of course, the mode which dominates may change, but the picture remains the same, except that the average phase slope about which the corrugations are measured changes from that corresponding to one mode to that of the other.

It can readily be shown that the deviation ϕ from the phase of the stronger wave is given by (6)

$$\tan \phi = \frac{r \sin \theta}{1+r \cos \theta} \qquad (1)$$

where r is the ratio of mode amplitudes ($r \leq 1$), and θ is given by

$$\theta = kx(\cos \psi_1 - \cos \psi_2) - \omega_d t \qquad (2)$$

Here k is the wave number (2π/wavelength), x is the distance along the array, ψ_1 and ψ_2 are the angles between the wave normals of the two signals and the axis of the array, ω_d is the radian frequency difference between the two signals, and t is the time. ω_d is not zero in general because of ionospherically-induced Doppler shift.

A measure of wave-front nonlinearity is the mean-squared phase deviation

$$\sigma_\phi^2 = \overline{[\phi(r)]^2} = \frac{1}{2\pi} \int_0^{2\pi} \phi^2(r)\, d\theta \qquad (3)$$

where ϕ is given by Equation (1). The integration over 2π in θ is equivalent to spatial integration over one period of the interference pattern. A numerical evaluation of Equation (3), for various values of the mode amplitude ratio r, yields the curve shown in Figure 1.

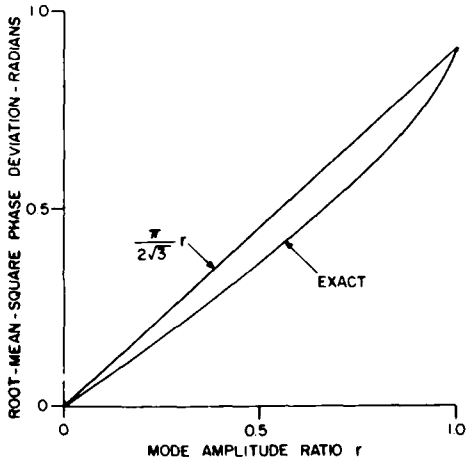


Figure 1. Root-mean-square phase deviation across an antenna aperture as a function of mode amplitude ratio.

As a good approximation, it is seen from Figure 1 that

$$\sigma_\phi \simeq \frac{\pi}{2\sqrt{3}}\, r, \quad 0 \leq r \leq 1 \qquad (4)$$

that is, the rms phase deviation σ_ϕ is a linear function of the mode amplitude ratio r. Thus the rms phase deviation σ_ϕ should exhibit the same statistics as the mode amplitude ratio r.

If a CW signal is transmitted, the time-domain behaviour of the rms phase deviation σ_ϕ should then correspond to that of the ionospheric channel. On the other hand, if there are rapid amplitude variations because of signal modulation, then r may be thought of as the ratio of the magnitudes of the modulation envelopes as transmitted via the two modes, with the weaker mode signal a time-shifted and amplitude-scaled version of the other. In this case, provided the time delay is at least of the order of the reciprocal of the modulation bandwidth, the time-domain behaviour of the ratio r, and therefore of the rms phase deviation σ_ϕ, depends primarily on the modulation, with a slowly-varying ionospheric component superimposed.

If there are more than two modes present, the details become more complicated. However, the addition of more modes is not likely to significantly alter the time scale of changes in wave-front nonlinearity.

EXPERIMENTAL ARRANGEMENTS

The measurements reported here were made over a 911-km path between Sept Iles, Quebec, and Ottawa, Ontario. Signals were received at Ottawa on a 6-element linear-interferometer array aligned in a direction perpendicular to the propagation path. (The array and associated instrumentation were part of the HF Direction-Finding Research Facility at the Communications Research Centre (7)). Inter-element spacings of 15.2, 30.5, 68.6, 167.6 and 381.0 metres were used, for a total aperture of 662.9 metres. Each of the six elements in the array were connected to separate, phase-matched receivers with quadrature product-detector outputs. The receiver outputs were scanned and digitized, with an interval between scans of 610 microseconds.

The transmitted signal consisted of a 5-second burst of CW, followed immediately by a 5-second burst of suppressed-carrier single-sideband modulation in which the modulating waveform was a multi-level pseudo-random sequence which repeated every 64 milliseconds. (SSB was used rather than AM in order to maximize the utilization of the rather restrictive 800 Hz bandwidth of the receiver system). The important statistical parameter of this signal was that its amplitude had an essentially impulsive autocorrelation function. The transmission of this combination of signals was repeated once every two minutes, except that the schedule was interrupted every 20 minutes to make oblique

ionograms over the path.

The frequency of operation was between 7.5 and 8.0 MHz, and was chosen on the basis of the ionogram information to maximize the range of mode delays present over the path. The measurements reported here were made on one day between 0900 and 2100 Eastern Standard Time. During this time, the time delay spread amongst the E, F1, and F2 modes was in the range 0.8 to 1.5 milliseconds.

After measurement of the received signal phase at the six element locations, it was necessary to resolve the 2π ambiguities inherent in the readings from the widely-spaced elements by extrapolating outward and refining the phase slope one element at a time, starting with the unambiguous phase slope obtained from the two most closely-spaced elements. Once the phases were resolved in this way, a least-squares-best-fit straight line and the rms deviation of the measurements from this straight line were computed. This latter computation corresponds to one determination of the σ_ϕ of Equation (3), except that this experimental σ_ϕ measurement is from a rather limited number of spatial points, and the total aperture may not be sufficient to include a complete period of the interference pattern. For these reasons, the experimentally-measured σ_ϕ will be somewhat biased towards low values. On the other hand, noise and other measurement errors tend to add a fixed constant to σ_ϕ^2. The minimum σ_ϕ observed was about 3 degrees; this is believed to be the contribution from instrumental errors.

EXPERIMENTAL RESULTS

The variation of the measured rms phase deviation σ_ϕ, over a 5-second period, is shown in Figure 2 for a CW signal. The rms deviation varies with a time scale of the order of one second. This is commensurate with the rate to be expected from mode amplitude fading. The normalized auto-covariance function (correlation coefficient) for the data of Figure 2 is shown in Figure 3. The covariance function has the general appearance of a damped cosinusoid, with the first zero crossing at 1.7 seconds. This means that low values of rms deviation, corresponding to the dominance of one mode, should persist, once present, for a few hundred milliseconds or more. Conversely, if one is observing a high rms phase deviation, it may be necessary to wait a similar length of time for a quasi-single-moded condition to reappear.

It should be noted that in cases in which the period of the interference pattern is larger than the array aperture, another source of significant time variation in the observed rms phase deviation is the drift of the interference pattern across the array. Drift of the interference pattern is a manifestation of differential Doppler shift between modes. While the issue is not crucial to the conclusions reached here, other observations (6) do support the view that, for mid-latitude circuits and apertures of the size used here, mode amplitude fading is the dominant cause of phase front decorrelation, although undoubtedly drifts due to differential Doppler shifts also contribute.

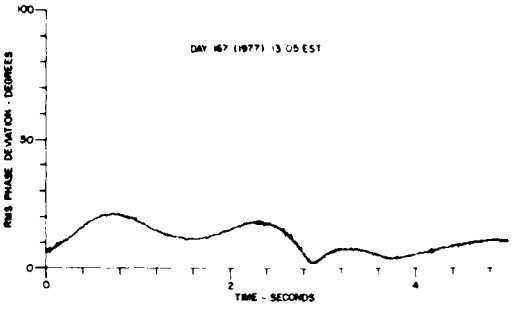


Figure 2. Measured rms phase deviation as a function of time for a CW transmitted signal. Time of observation 13:05 EST, 16 June 1977.

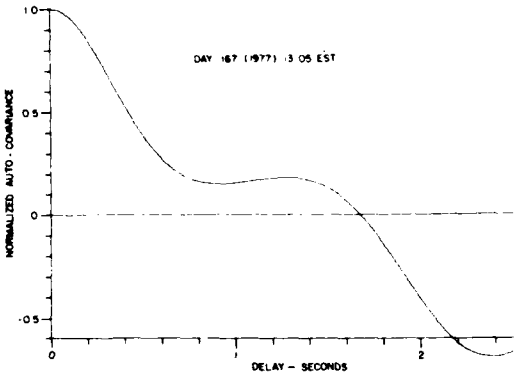


Figure 3. Normalized auto-covariance function for the data of Figure 2.

In sharp contrast to the result of Figure 2, the time variation of the rms deviation σ_ϕ for the modulated portion of the signal is shown in Figure 4. Note in particular that low values of rms deviation recur much more frequently, but they do not persist. The normalized auto-covariance function for the data of Figure 4 is shown in Figure 5. The correlation is small

almost everywhere, except for impulses recurring at multiples of the basic period of the pseudo-random sequence used as the modulating signal. The magnitude of the impulses dies away at longer times, corresponding to the decorrelation produced by the ionospheric channel, as shown in Figure 3.

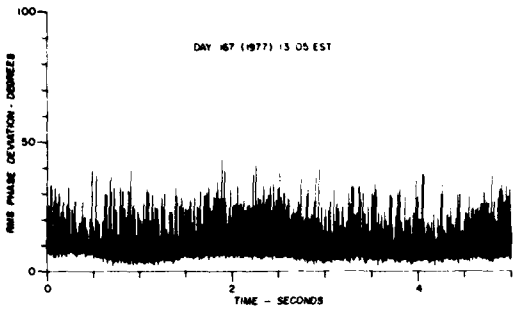


Figure 4. Measured rms phase deviation as a function of time, for a modulated signal. Time of observation 13:05 EST, 16 June 1977.

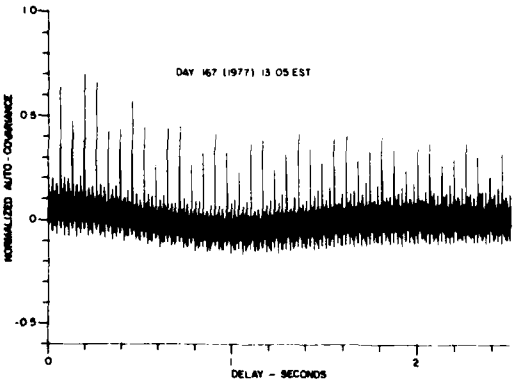


Figure 5. Normalized auto-covariance function for the data of Figure 4.

As a summary of the entire 12 hours of measurements, the normalized auto-covariance function obtained by averaging the corresponding power spectra for all of the CW data is shown in Figure 6, while Figure 7 shows the corresponding auto-covariance function for the data from the modulated signal. Again, the marked difference in time scale between the results for the CW and modulated-signal cases is clearly illustrated.

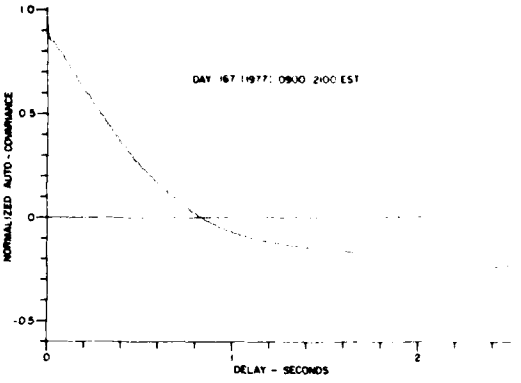


Figure 6. Normalized auto-covariance function for all CW data obtained over the 12-hour period from 0900 - 2100 on 16 June 1977.

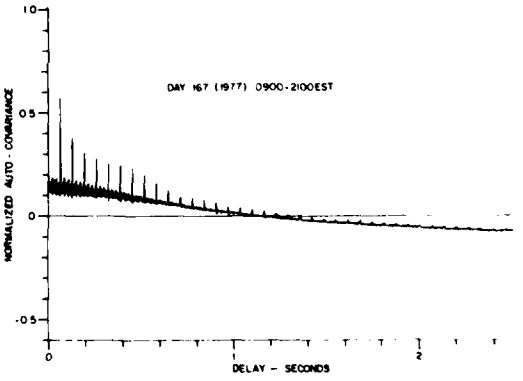


Figure 7. Normalized auto-covariance function for all modulated-signal data obtained over the 12-hour period from 0900 - 2100 EST on 16 June 1977.

DISCUSSION

From the results presented here it is clear that signal modulation can have a profound effect on the time-domain statistics of phase-front observations. In fact, modulation may impart a benefit to direction-finding by shortening the waiting time between phase-linear events. This benefit may be expected to be

largest for signals which have significant amplitude modulation, with a bandwidth of the order of or greater than the reciprocal of the time delay between ionospheric modes. It could best be exploited by a system which takes bearing measurements on a similarly short time scale.

ACKNOWLEDGEMENT

This work was supported by the Department of National Defence of Canada.

REFERENCES

1. Anonymous. Radio Direction Finder. Wireless Eng. 13, 443 (1936).

2. Treharne, R.F. Vertical Triangulation Using Skywaves. Proc. IREE Aust., p. 419, (Nov. 1967).

3. Sherrill, W.M. A Survey of Interferometry for Ionospheric Propagation Research. Radio Science, Vol. 6 #5, 549, (May 1971).

4. Watterson, C.C., J.R. Juroshek, and W.D. Bensema. Experimental Confirmation of an HF Channel Model. IEEE Trans. Commun. Tech. COM-18, 792 (Dec. 1970).

5. Hayden, E.C. Propagation Studies Using Direction-Finding Techniques. J. Res. NBS-D, Radio Propagation, Vol. 65D, 197 (May-June 1961).

6. Rook, B.J. Study of the Behaviour and Stability of Phase Fronts on Short Time Scales, CRC Report, to be published.

7. Rice, D.W. and E.L. Winacott. A Sampling Array for HF Direction-Finding Research. CRC Report #1310, (Oct. 1977).

MEASURED SEA PATH SKY WAVE/GROUND WAVE SIGNAL POWER RATIOS OVER THE 2-10 MHZ RANGE

J. E. Hipp and T. C. Green
Southwest Research Institute
San Antonio, Texas

INTRODUCTION

In June of 1973, Southwest Research Institute recorded a series of 230 oblique ionospheric soundings over the Gulf of Mexico as part of a Naval Electronic Systems Command funded program. The result of these measurements was an extensive set of sea path ionograms recorded under a wide variety of measurement parameters, including signal frequency (from 2 to 10 MHz), Greenwich Mean Time (essentially around the clock for a period of five days), range between the transmitter and receiver (50-1000 km), and several combinations of transmit and receive antenna polarizations. These recorded ionograms allowed the relative amplitudes of sky wave and ground wave propagation modes to be analyzed as functions of the measurement parameters. The objective of measuring the relative ground wave and sky wave energies was to qualitatively relate the presence of ground wave to measurement parameters and to quantitatively relate the sky wave/ground wave energy ratio to the measurement parameters. A detailed description of the analysis summarized in this discussion is presented in the final report* for the NAVELEX program.

DATA MEASUREMENT

The soundings were performed with an FM CW ionospheric sounding receiver located at Port Isabel about 100 meters from the shore, and a similar FM CW ionospheric sounding transmitter located on an SwRI boat traveling in the Gulf of Mexico. Figure 1 is a photograph of the Southwest Research Institute boat showing the vertical and horizontal transmitting antennas. Selectable vertical and horizontal antennas were employed at both the receive site and on the transmitting boat.

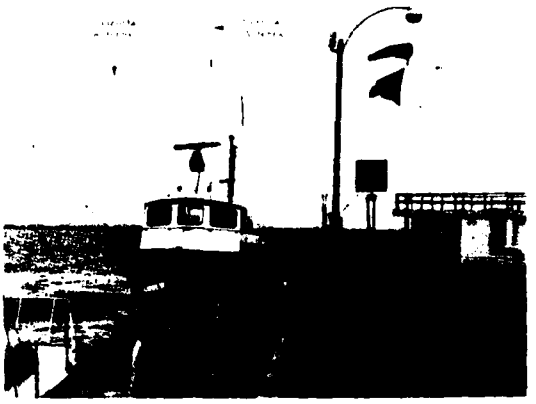


Figure 1. Transmit Antennas

Figure 2 is a partial map of the Gulf of Mexico indicating the traversal path of the transmitter equipped boat. The total path was traveled over a time period of five days and allowed a wide distribution of propagation path length and time of day as shown in Figure 3. Once each hour, soundings were recorded for each of the four possible combinations of the transmitter and receiver polarization with sounder frequency varying continuously from 2 MHz to 10 MHz. The range of the ship was determined by maintaining an on-board Loran-A navigation system. At the receive site, each received ionogram was recorded and annotated on analog magnetic tape for later processing.

DATA ANALYSIS

After completing the sea path ionogram measurements, the individual recorded ionograms and relevant measurement parameters were digitized. Each digitized ionogram was

*Hipp, J. E. and T. C. Green, "Measured Sea Path Sky Wave/Ground Wave Signal Power Ratios over the 2-10 MHz Range," Southwest Research Institute, Task Summary Report XXXIII for U. S. Navy Contract N00039-72-C-1275, 23 January 1976.

then processed using spectrum analysis to separate ground wave and sky wave modes and to determine the relative energies of the ground wave mode and the sky wave modes as functions of the sounder frequency. All sky wave modes were summed to provide a total sky wave energy value for comparison against the ground wave mode. The results for each ionogram were plotted and stored on a computer disc file for further processing. Figure 4 is a sample resulting plot of sky wave/ground wave energy ratio as a function of sounder frequency. For comparison, a photograph of the corresponding ionogram is also shown, indicating that the ground wave exists from about 2 MHz to about 8 MHz, and that sky wave modes exist from about 4 MHz through 9.5 MHz. The computerized plot of sky wave energy relative to ground wave energy verifies the frequency ranges of sky wave and ground wave existence. However, in contrast to the ionogram photograph, the plot quantitatively presents intensity of sky wave mode energy relative to ground wave energy. A total of about 230 ionograms were recorded, processed, and stored in this manner.

DATA SUMMARY

To statistically summarize the data, it was necessary to categorize the data into groups of similar range and time of day. The intervals chosen for time of day groups were determined by examining photographs of the ionograms and choosing as intervals for typical daytime and nighttime those intervals during which the sky wave modes were stable. The time intervals associated with sunset and sunrise were identified by the extent of rapidly changing propagation mode occurrences. The resulting time intervals are shown in Figure 5 superimposed on the original time/range distribution shown in Figure 3. With the time intervals thus established, the resulting distribution can be subdivided into a number of obvious range intervals for statistical analysis. The processed ionograms were combined in this manner to provide statistical analyses of sky wave/ground wave propagation as functions of frequency and combinations of transmit/receive polarization.

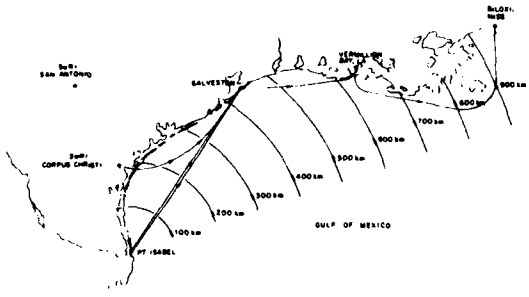


Figure 2. Route of Transmitter Equipped Boat

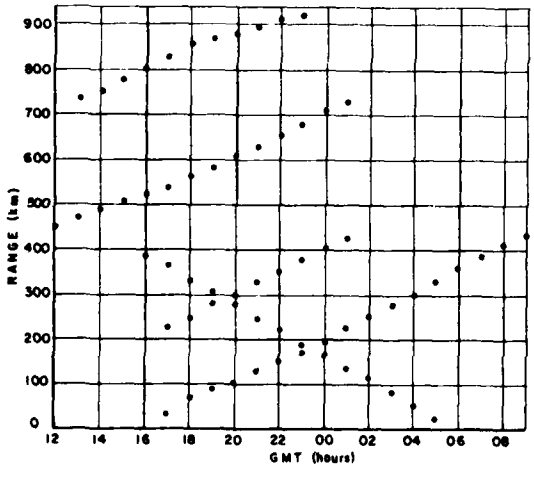


Figure 3. Range/Time Distribution of Ionosphere Soundings

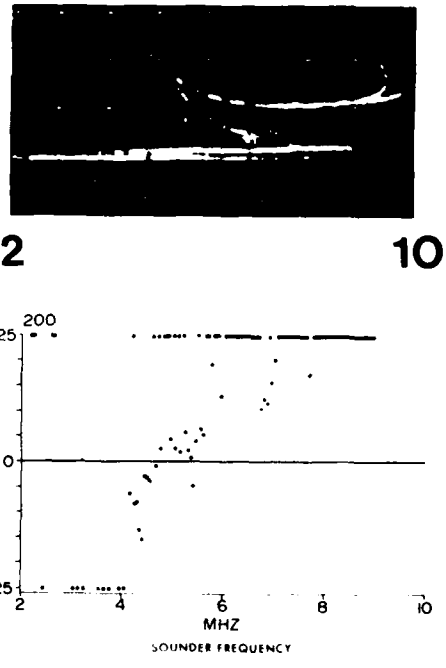


Figure 4. Ionogram Photograph and the Corresponding Plot of Energy Ratio vs Sounder Frequency

A nighttime data group and daytime data group with similar range parameters are presented here as examples of the statistical summaries performed. Figure 6 is a set of histograms summarizing the analysis of the 200-400 km daytime data group and contains four sets of histograms, one set for each combination of transmit/receive polarization. It should be emphasized that the polarization descriptions apply only to the respective antennas, and are not indicative of the polarization of the signal propagating through the ionosphere. Each histogram is displayed as a function of sounding frequency in 200 kHz intervals from 2 MHz to 10 MHz. In addition, each histogram is normalized to vary between 0 and 100% of the total number of ionograms that make up the data base for each histogram. Each set of histograms contains four histograms describing occurrences of:

(1) Ground wave only (GW energy/ SW energy $\geq$ 10 dB)

(2) Sky wave only (SW energy/ GW energy $\geq$ 10 dB)

(3) Mixed sky wave and ground wave (i.e., sky wave energy within 10 dB of the ground wave energy)

(4) No detectable signal

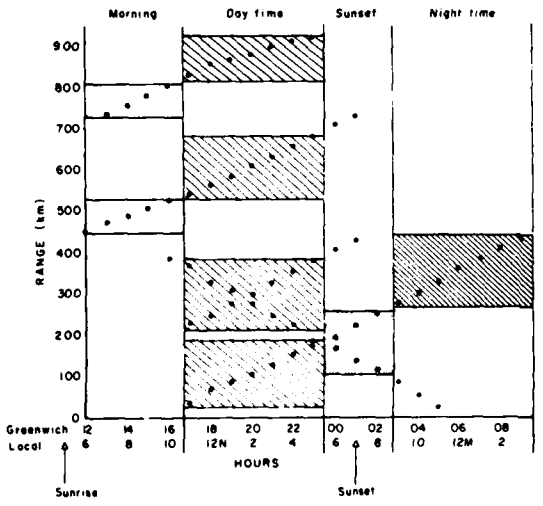


Figure 5. Ionogram Distribution for Time/Range Classification

Daytime Data Set

Daytime propagation is characterized by a high degree of ionospheric ionization, producing the familiar E and F modes as well as high level D region absorption. The daytime ionosphere routinely supports sky wave propagation for paths up to 1000 km and frequencies up to and above 10 MHz. However, the presence of high level D region absorption produces severe sky wave attenuation for frequencies below about 4.5 MHz.

As seen in Figure 6, the ground wave signal is predominant from 2 to 10 MHz for vertical transmit/vertical receive. The analysis results of the terminal-to-terminal transmission characteristics for the two mixed receive/transmit polarizations are roughly equivalent. Any observable differences between the two mixed polarization conditions can be attributed to mechanical configuration differences in the horizontal transmit antenna and the horizontal receive antenna. For the mixed antenna polarization case, the ground wave signal is predominant from 2 to 4.5 MHz and again above 9 MHz, with sky wave and mixed signals dominating for the intervening frequencies. In the horizontal transmit/horizontal receive data group, the received signal is primarily ground wave from 2 to 4 MHz and sky wave from 4 to 10 MHz.

Nighttime Data Set

In contrast to the daytime ionosphere, the nighttime ionosphere is characterized by a slowly decaying ionization and a corresponding decrease in the critical frequency of the F layer. The E layer, which is completely dependent upon the sun's rays for replenishment, is not present during the nighttime period. Examination of individual nighttime ionogram photographs taken during this period reveals that the critical frequency varied between 5 MHz and 7 MHz, depending on the time of night. Due to the absence of D region absorption, the ionosphere will support sky wave propagation down to near 2 MHz, depending on the antenna polarizations and the terminal-to-terminal path range.

Figure 7 is a set of nighttime summary histograms that is analogous to the daytime set of histograms. As shown in Figure 7, the ground wave signal is predominant from 2 to 9 MHz for vertical transmit/vertical receive combinations. Ground wave signal is also predominant from 2 to 4 MHz and from 6 to 9 MHz for the mixed antenna polarizations. During mixed antenna polarizations, sky wave and mixed sky wave/ground wave signals are prevalent from 4 to 6 MHz. For the horizontal transmit/horizontal receive combination, the ground wave signal is predominant from 2 to 3 MHz and the sky wave signal is predominant from about 4 MHz through 6 MHz. During the

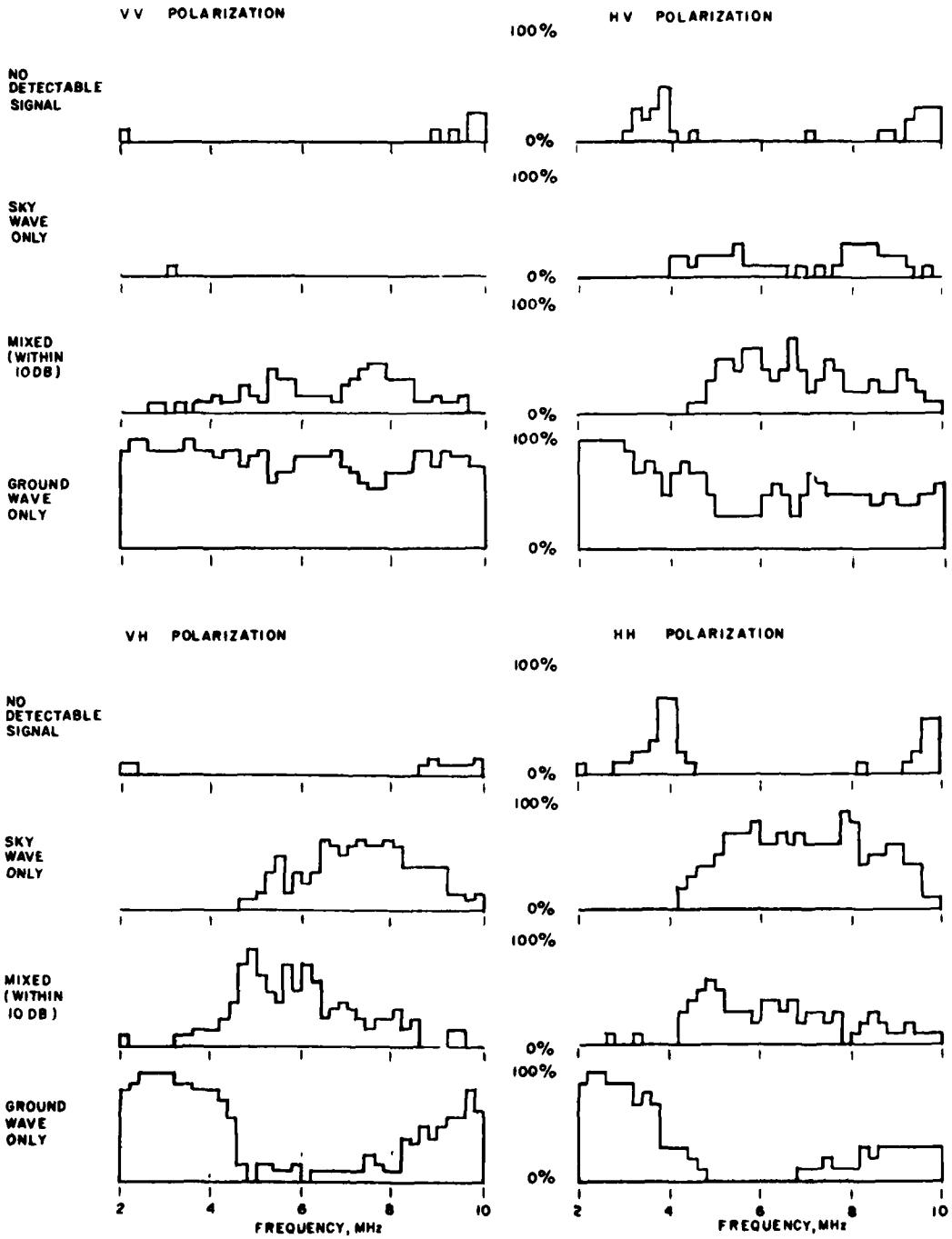


Figure 6. Summary Histograms for 200-400 KM Daytime Data Group

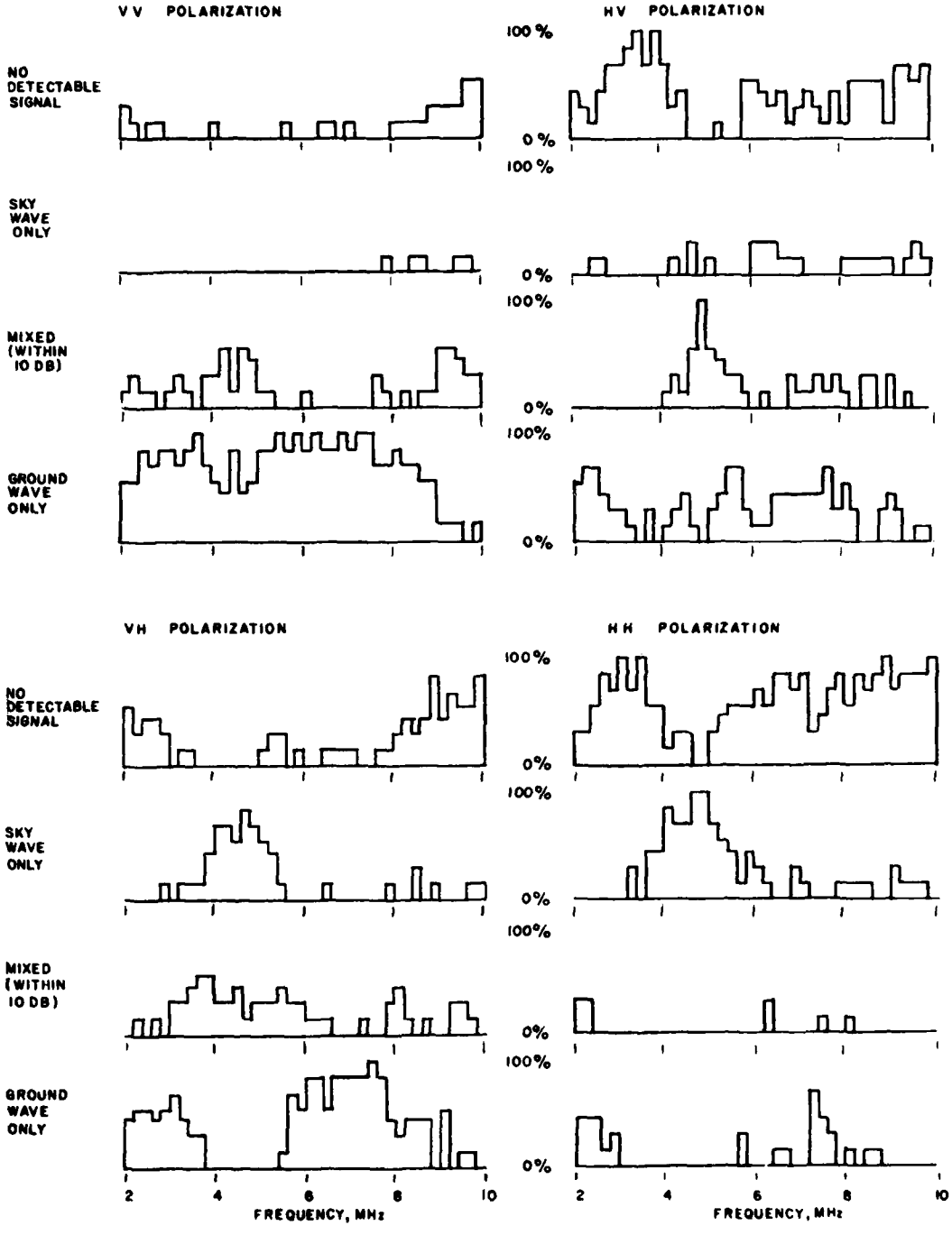


Figure 7. Summary Histograms for 250-450 KM Nighttime Data Group

nighttime interval, a much higher incidence of no detectable signal was registered than was observed during the daytime interval. The increase of no detectable signal is attributed largely to the increase in radio communications traffic during the nighttime period and to the absence of D layer absorption for distant unrelated transmissions.

CONCLUSIONS

During the course of the NAVELEX program, all of the time/range data groups shown in Figure 5 were statistically analyzed by the procedure described above. The conclusions resulting from the entire set of analyses are given below.

1. Ground wave propagation is shown to be generally predominant (greater than 10 dB above the sky wave signal) for the majority of the measured data set in the 2-4 MHz frequency range and for terminal-to-terminal path distance less than 400 km. The ground wave predomination for the above frequency and range constraints is shown to be generally true throughout the 24-hour diurnal period.

2. At ranges greater than 400 km, insufficient data were obtained to provide conclusions over the complete 24-hour diurnal period. However, the ionograms obtained in these longer ranges clearly show the pronounced effects of the sun on the ionosphere and the resulting pronounced effects on HF sky wave propagation. For example, ground wave propagation predominance does not occur for frequencies greater than approximately 4 MHz for daytime conditions due to E layer sky wave propagation. On the other hand, for nighttime conditions, ground wave propagation will be predominant at frequencies above the F layer critical frequency (f_c); however, significant radiated power must be applied due to the high surface wave attenuation in the 6-10 MHz range.

3. In the 4-10 MHz range mixed sky wave/ground wave propagation as well as sky wave only propagation can be anticipated to occur over a large percentage of the 24-hour diurnal period.

4. Ground wave and sky wave signal intensities are greatly affected by the transmitting and receiving antenna polarizations with the ground wave signal enhanced by vertical-to-vertical polarization combinations and the sky wave signal significantly enhanced for horizontal-to-horizontal polarization combinations. For the vertical-to-vertical polarization combination, for example, the ground wave signal is shown to be predominant over the complete 2-10 MHz range for terminal-to-terminal distances of less than 400 km. On the other hand, horizontal-to-horizontal antenna polarization combinations show ground wave propagation predominance only in the 2-4 MHz range for the same distance ranges. The cross polarization conditions for the transmit and receive antennas (i.e., VH and HV) show similar propagation results generally between the vertical-to-vertical and horizontal-to-horizontal extremes.

5. In the 0-200 km range, ground wave propagation is shown to be predominant over roughly the 2-5 MHz range independent of the antenna polarization combination and the time of day.

6. The sunrise and sunset periods often provide good support for 2-4 MHz sky wave propagation, resulting in ground wave contamination even at ranges less than 400 km.

7. The computed sky wave/ground wave power ratios from the measured ionograms agree generally with theoretical predictions (Freres,* et al). The disagreement, when it occurred, always indicated a higher relative level of sky wave intensity. This disagreement could arise from two sources:

   (a) The sunspot number of 80 used in the theoretical model would result in higher relative sky wave intensity.

   (b) The actual gain patterns of the vertical receive and transmit antennas used during the series of measurements may have been significantly different from the theoretical model, thus providing less relative gain for the sky wave signal components.

*Freres, C. H., D. B. Sailors, and T. L. Taylor, "Radio Propagation Over the Sea in the 2-to-10-MHz Range," NELC/Report 1526, Naval Electronics Laboratory Center for Command Control and Communications, San Diego, CA 92152, 21 November 1967.

DIGITAL IONOSONDE STUDIES AT FORT MONMOUTH, NEW JERSEY

F.J. GORMAN, JR., AND H. SOICHER
COMMUNICATIONS/AUTOMATIC DATA PROCESSING LABORATORY
US ARMY ELECTRONICS COMMAND
FORT MONMOUTH, NJ

ABSTRACT

Recent developments in the area of digital ionosondes and computer technology have enabled systems to be developed in which virtual heights and echo amplitudes obtained may be translated directly into true height profiles, sky maps and frequency management charts.

Real time ionogram evaluation will enable users to make highly accurate judgements as to the proper selection of frequencies for short and long range HF communications. Libraries of previous ionograms make ionospheric predictions considerably more accurate. The spectral analysis of data taken utilizing an array of receiving antennas allows the user to spatially locate traveling ionospheric disturbances. This system gives the specific location of the echo's reflecting surface and its motion with respect to the user's position.

INTRODUCTION

Because of the complexity and variability of the ionosphere, extensive measurements and the evaluation of many geophysical parameters are necessary for the understanding of dynamic effects. Much data analysis and classification must precede attempts to explain trends and variations. Gathering the multidimensional information requires sophisticated, reliable measuring instruments and systems. The Digisonde (Digital Ionospheric Sounding System consisting of an on-line computer, drift attachment and auxiliary equipment) determines overhead ionospheric structure in real-time. It also has the capability of determining relative amplitude and phase information of its emitted signals.

The ionosonde has a single quartz crystal oscillator as a source for all generated frequencies. Phase coding of the selected transmitted pulse will unambiguously select ionospherically reflected echos. Variable height resolution, coherent detection, large dynamic range are programmable either locally or remotely by a computer.

Data processing within the ionosonde permits the display of transit time with an accuracy of 1.5 km [Bibl, K., et al., 1971] and the echo amplitudes can be measured and displayed to an accuracy of 1 dB. These amplitudes can be used for absorption measurements, and the ionosonde can also operate at a fixed frequency for special studies such as an interferometric system at which incidence angles can be measured as a function of time.

An integral part of the ionosonde system is a chassis which processes conventional ionograms. Raw ionograms are fed into what is known as the Automatic Ionogram Compression and filtering unit. Each transmitted frequency is evaluated and a unique reflecting layer height determined. The echo's amplitude at that height is also displayed. This listing of frequencies, heights and amplitudes is then fed directly into the computer. It is also displayed on a printer in a conventional ionogram format as well as structured in such a way that it can later be presented for frequency management studies.

The computer referenced here, is considered to be a general purpose mini-computer and an ideal choice for on-line real time analysis. The program used in the computer utilizes the lamination technique devised by Jackson [1971]. Each layer is divided into a series of laminations, each having its own set of parameters. The rate of change of slope is considered constant throughout each lamination. The profile is a parabola in either plasma frequency or a natural logarithm of plasma frequency.

Frequency management charts are displayed by playing back long term records of compressed and reformatted ionograms that the Automatic Ionogram and Compression and filtering unit has stored on magnetic tape. The E-layer and F-layer are evaluated independently and displayed side by side on printer. In this type of display, echo amplitudes and frequencies are plotted as a function of time. This record yields a convenient display for frequency

selection in the case of a user requesting assistance for a choice of optimum transmitting frequency for his communication equipment. Thus, the least attenuated signal can be determined. The ionosphere is probed continuously at short intervals and the optimum communication frequency is determined. The effects of gravity waves and other phenomena on this optimum frequency is determined.

The installation of a drift attachment introduces a new technique under which four closely spaced antennas can be positioned such that inhomogeneities and motions in the ionosphere can be measured. The main features of this development are the recording of phase as well as the signal amplitude and the introduction of complex spectral analysis. It is now possible to describe the various components of a signal reflected by the ionosphere in terms of their angle of arrival and Doppler Shift of individual components. In contrast to this, an ionogram pictures the average signal delay of the various components as a function of frequency. It can be easily visualized that a combination of the ionosonde with the updated drift technique has significant advantages for research as well as for practical purposes. This type of program will yield the true height of each reflecting layer as well as its specific location within the ionosphere. The picture will also be dynamic in the sense that it will be updated periodically. Thus, motions of reflecting layer will be tracked as it passes over the ionosonde sight.

The drift attachment scans four antennas and includes an on-line multichannel spectrum analyzer. The complex spectrogram data of the four antennas reflection points for each spectral line window within which sufficient coherent energy has been reflected.

DATA

The following figures graphically display examples of data collected at Fort Monmouth. These figures are typical of the data recorded over the last year of continuous operation at our field site.

An example of an ionogram as it is recorded by the ionosonde and then processed by the Automatic Ionogram Comparison and filtering unit (AIC) is shown in Figure 1. This ionogram was taken at 1430 EST on 23 Oct 77. The echo amplitudes are displayed by numbers indicative a 1 dB difference in signal strength. Every 10 frequency steps a preface line indicates time of day, day of year, frequency coverage and receiver settings. In this particular case, the ionogram covers a range of frequencies from 2 to 8 MHz. In the raw ionogram, there are echos at various altitudes. The most significant feature of the cleaned ionograms is that it assigns one unique altitude for each transmitted frequency.

The cleaned ionogram is then fed into the computer. The computer translates the virtual height information into the true height of each reflecting layer. Figure 2 is a copy of the computer printout of an ionogram. There is a listing of virtual heights and frequencies at the beginning of the printout. The computer then tablulates the frequencies, electron densities and the true height of each reflecting layer.

Figures 3 and 4 are printouts of compressed ionograms covering two separate time periods. Figure 3 — March 22 through March 28, 1977, and Figure 4 — September 21 through September 26, 77. The E-layer and F-layers are displayed separately. The echo amplitudes and frequencies are displayed as a function of time. The diurnal behavior of the ionosphere is clearly projected. Critical frequencies in the F-layer varies from 2 to 8 MHz. Each vertical line is a specially formatted ionogram, and a new ionogram is initiated every 15 minutes during this period.

By viewing these charts it is clear which frequencies would give optimum communication performance. Each echo is displayed with its amplitude indicated by number whose signal strength differs from an adjacent by 1 dB. It is apparent that frequencies that yield the strongest echo amplitude would be optimum for communication channels. Predictions are possible by studying libraries of previous ionograms.

FUTURE EXPERIMENTS

The Drift System presently operates on an off-line basis. Spectral evaluation of echos from the antenna array will be fed directly into the on-line mini-computer. At the present time, sky maps are generated off-line on a large computer. There is an active program underway which will enable this evaluation to take place in real time.

A program for real-time ray tracing is also underway. Real time calculations of transmission paths through the ionosphere for various elevation angles will be incorporated into the ionosonde facility.

REFERENCES

Bibl, K., et al., "Digital Ionosondes for Monitoring the Ionosphere," AFCRL-71-0507, 1971.

Jackson, John E., "The P'(f) to N(h) Inversion Problem in Ionospheric Soundings," Goddard Space Flight Center," Greenbelt, Maryland, 1971.

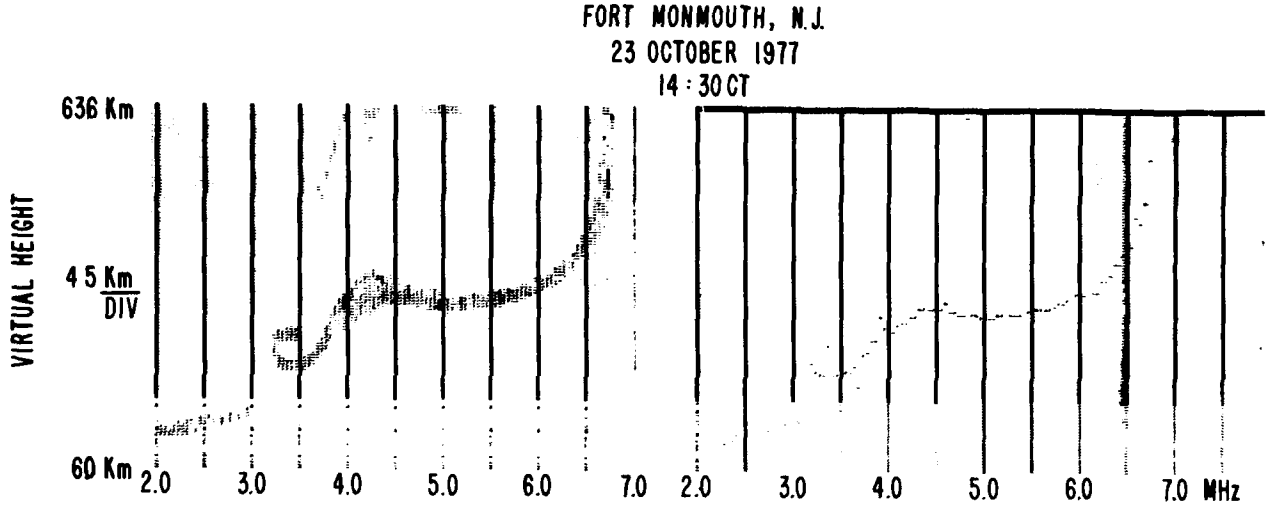


Fig. 1

Raw Ionogram (left) and processed ionogram (right), 23 Oct 77, 1430 EST.

ANALYSIS OF IONOGRAMS AT FT. MONMOUTH

INVERSE CUBE, PARABOLIC IN LOG (N)

INPUT DATA

YEAR 1977 MONTH 10 DAY 19 LOCAL STANDARD TIME 1240 DAY NO. 293

DEPTH OF E-VALLEY = 0.20 MHZ

O TRACE

F	HP	F	HP	F	HP	F	HP	F	HP
0.284	80.0	1.500	102.0	1.750	110.0	2.000	112.0	2.500	115.0
3.000	125.0	3.260	172.0	3.320	280.0	3.400	248.0	3.500	235.0
3.700	232.0	3.900	238.0	4.100	240.0	4.300	248.0	4.500	250.0
5.000	255.0	5.500	255.0	6.000	257.0	6.500	262.0	7.000	265.0
8.000	280.0	9.000	305.0	10.000	335.0	11.000	385.0	11.700	430.0

DIP AND FHS AT 80. KM COMPUTED FROM FIELD G.
DIP= 68.9839 FHS= 1.4968 DLAT= 40.2500 DLONG= -74.0600

F	ELCC	HEIGHT
0.2840	1000.1345	80.0000
1.50000	27900.0000	94.9975
1.75000	37975.0000	97.6356
2.00000	49600.0000	99.8549
2.50000	77500.0000	103.3947
3.00000	111600.0000	107.9155
3.26000	131782.2344	115.7691
3.32000	136677.7656	134.6002
3.40000	143344.0000	138.3937
3.50000	151900.0000	142.2444
3.70000	169756.0000	149.1712
3.90000	188604.0156	155.4240
4.10000	208444.0000	161.2370
4.30000	229276.0156	166.7542
4.50000	251100.0000	171.9912
5.00000	310000.0000	183.0253
5.50000	375100.0000	191.6675
6.00000	446400.0000	198.5058
6.50000	523900.0000	204.5660
7.00000	607600.0000	210.0488
8.00000	793600.0000	220.5818
9.00000	1004400.0000	232.5402
10.00000	1240000.0000	246.6302
11.00000	1500400.0000	264.5140
11.70000	1697436.0000	280.3761

CALCULATION OF HMAX AND NMAX ASSUMING PARABOLIC PEAK OF THE FORM
N = NMAX(1-(H-HMAX)(H-HMAX))/CONSTANT

HMAX = 395.7 KM AND NMAX = 0.2370+07 EL/CC.
THE RATIO OF THE LAST TWO SLOPES IS 0.879

ALT.	DENSITY	ALT.	DENSITY	ALT.	DENSITY	ALT.	DENSITY
290.0	0.181E+07	300.0	0.191E+07	310.0	0.200E+07	320.0	0.208E+07
330.0	0.215E+07	340.0	0.222E+07	350.0	0.227E+07	360.0	0.231E+07
370.0	0.234E+07	380.0	0.236E+07				

STOP

Fig. 2

Computer Print-out of Computed N(h) Profile, 19 Oct 77, 1240 EST.

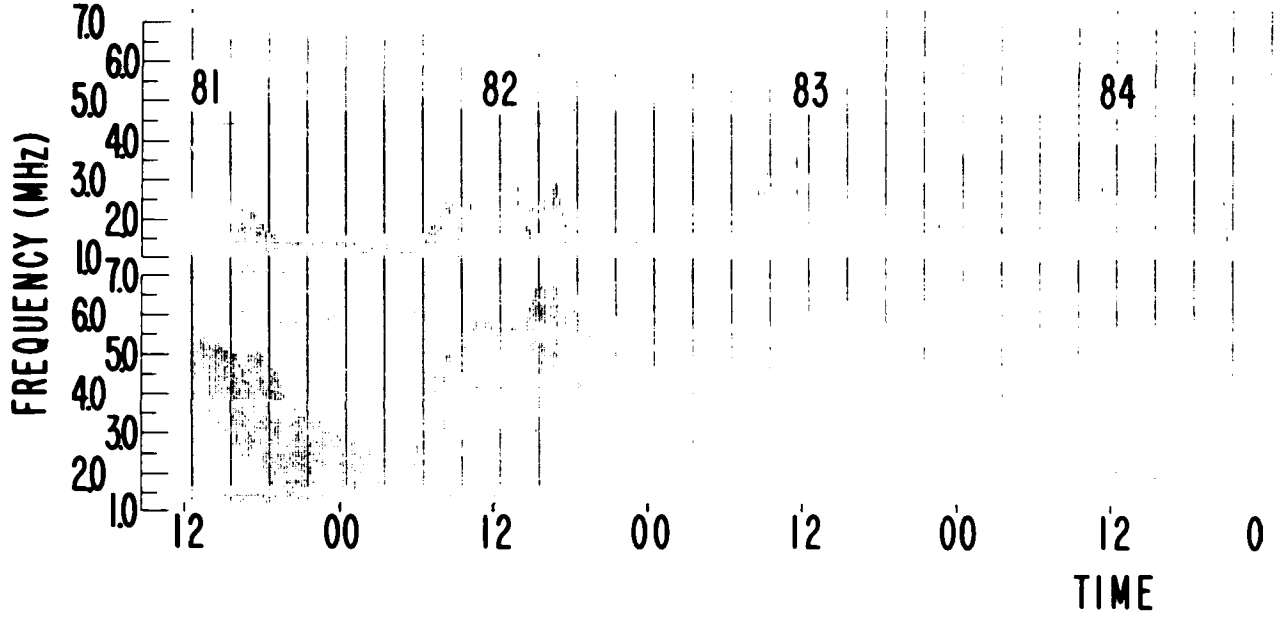


Fig. 3

Diurnal variation of E- (top) and F- (bottom)— layer usable frequencies, March 22-28, 1978. Abscissa indicates time at 15-minute intervals, ordinate indicates frequency. Numerical characters indicate relative signal amplitudes at the various frequencies.

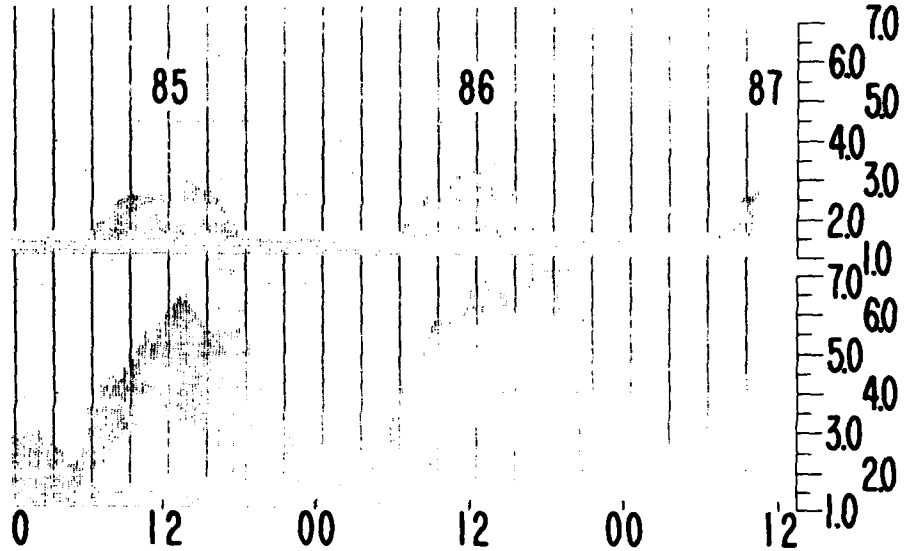


Fig. 3 (Cont'd)

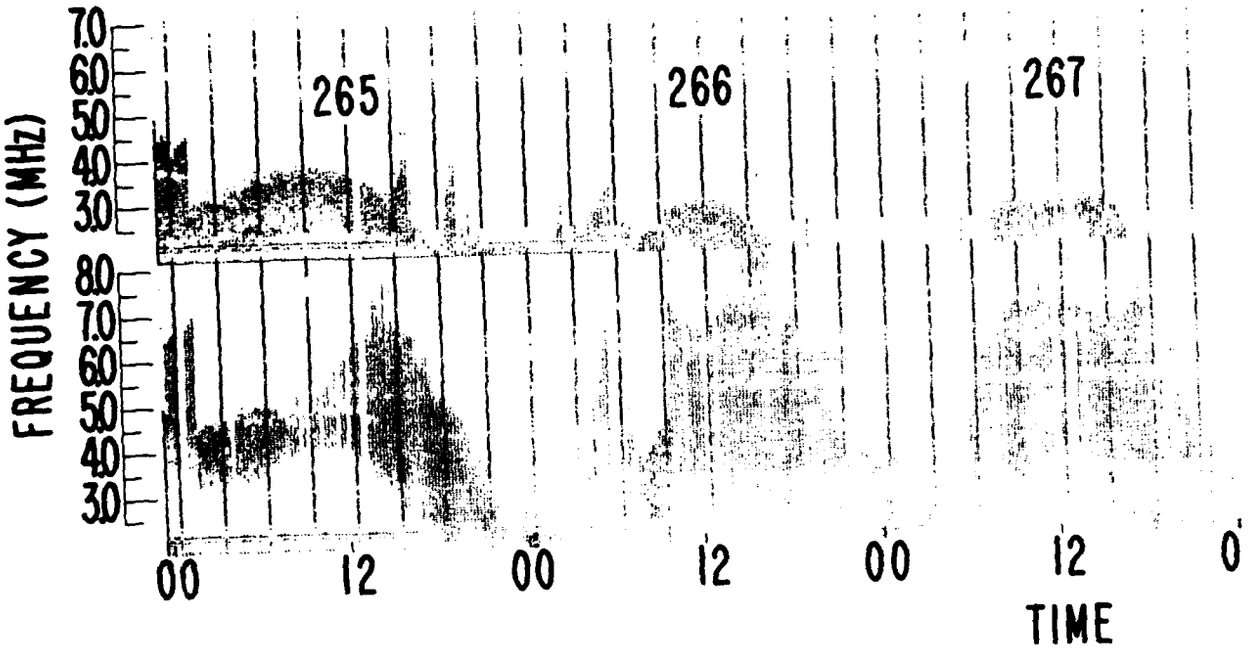


Fig. 4

Same as Fig. 3, but for time period 21-26 September 1977

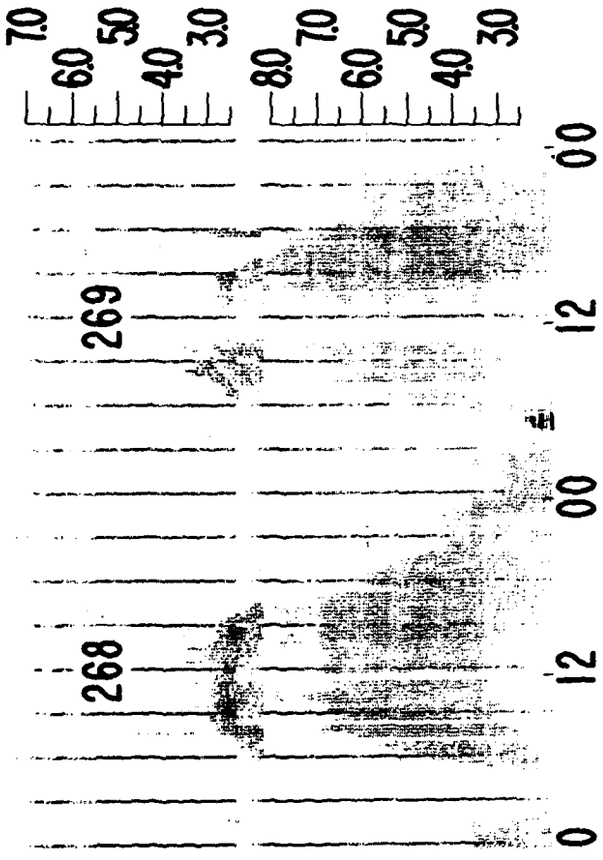


Fig. 4 (Cont'd)

STATE-OF-THE-ART FOR THE ASSESSMENT AND CONTROL OF ENVIRONMENTAL CHARGING OF SPACECRAFT

Alan Rosen
TRW Defense and Space Systems Group
Redondo Beach, California

INTRODUCTION

During recent years, a large number of synchronous satellite programs have reported anomalous behavior, degradation, and occasional "burnout" or loss of spacecraft subsystems. It is now believed that a large fraction of the anomalies are due to electrostatic charging of the various spacecraft surfaces to kilovolt potentials and their subsequent discharge. The arc discharge and associated electromagnetic pulses then interact with various portions of the spacecraft electronic system to generate "logic trips" or "burnout" of sensitive components.

Significant progress has been made in recent years in understanding the plasma environment and plasma-spacecraft interaction that charges up spacecraft to the multikilovolt levels.[1,2] This understanding is now being translated into spacecraft design guidelines, aimed at reducing or eliminating the susceptibility of spacecraft to environmentally induced arcing malfunctions. The purpose of this paper is to discuss the state-of-the-art in the area of spacecraft charging and the degree to which this knowledge can be translated into spacecraft design guidelines.

ASSESSMENT OF THE DANGER

One of the most important questions facing spacecraft designers is whether charging represents a significant hazard. Differential voltages in the multikilovolt range, generally do no harm to a spacecraft. Most often the resulting vacuum arcs do not cause damage. Occasionally, the arc may cause an interruption of service or damage to a circuit; depending on the magnitude of the arc, the coupling coefficient between the arc and the circuit and the susceptibility of that particular circuit. The arc characteristics depend on environmental factors as well as spacecraft configurational factors. Table 1 lists the environmental and spacecraft configurational factors that are important in assessing the effects of charging on a spacecraft system. Worst case arcing, in most cases, represents a significant hazard to a spacecraft, and preventative actions on the part of the spacecraft designer are necessary. Table 2 shows some typical values of induced voltages associated with radiated fields and arcs striking cables directly. In general, arcs striking cables directly represent a more hazardous phenomena than induction by radiated fields.

Table 1. Environmental Factors and Spacecraft Configurational Factors Important in Assessing the Hazard due to Spacecraft Charging

ENVIRONMENTAL FACTORS	
• ELECTRON AND ION FLUXES AND ENERGY DISTRIBUTIONS	• POSITIONAL DEPENDENCE (LATITUDE, LONGITUDE)
• ANGULAR DEPENDENCE (OMNIDIRECTIONAL, FIELD ALIGNED)	• TIME DEPENDENCE (SHORT TERM, LOCAL TIME, SEASONAL, SOLAR CYCLE

CONFIGURATIONAL FACTORS	
• SURFACE CONFIGURATION - EXPOSED DIELECTRIC AREAS - EXPOSED METALLIC AREAS - GROUNDING OF METALLIC COMPONENTS - APERTURES/CAVITIES	• THERMAL EFFECTS - ACCEPTIBILITY OF NEW MATERIALS THERMAL SHORTING DUE TO GROUND STRAPS - ARC DISCHARGE BURNOFF OF VACUUM DEPOSITED ALUMINUM - CONTAMINATION OF THERMAL CONTROL SURFACES BY ARC DISCHARGE BY-PRODUCTS
• ELECTROMAGNETIC COMPATIBILITY - LOCATIONS OF POSSIBLE ARC DISCHARGES - COUPLING INTO HARNESS CABLES - EMI COUNTERMEASURES - EMC VERIFICATION TESTING	• OPTICAL EFFECTS OPTICAL SIGNALS FROM ARC DISCHARGES CONTAMINATION OF OPTICAL SURFACES BY ARC DISCHARGE BY PRODUCTS BURNOFF OF OPTICAL COATINGS BY ARC DISCHARGES
• SYSTEM DESIGN - MISSION REQUIREMENTS (VOLTAGE CONTROL) - WEIGHT/COST TRADEOFFS (SHIELDING) - LIFETIME CONSIDERATIONS - STABILIZED/SPINNER	• MECHANICAL EFFECTS COULOMB FORCES DUE TO ELECTROSTATIC CHARGE BUILDUP AND DISCHARGES - MAGNETIC FORCES DUE TO DISCHARGE CURRENTS - "ION ENGINE" FORCES DUE TO ARC DISCHARGES BURNOFF OF ADHESIVES CHARRING OF INSULATORS

It is extremely risky to design and build spacecraft without undertaking a thorough assessment of the spacecraft charging hazard, and the design elements that are most likely to reduce that hazard.

DESIGN FACTORS

The design of spacecraft to operate correctly in a charging environment requires, as a minimum, adherence to a set of design guidelines similar to those illustrated in Table 3.

These design guidelines should become firm requirements whenever they impact the survival of the spacecraft. For example, the proper ground-

Table 2. Arc Discharge Parameters and Typical Voltages Induced into Spacecraft Cables

ARC DISCHARGE PARAMETERS

	METAL-TO-METAL ARCS	DIELECTRIC-TO-METAL ARCS
ASSUMED ARCING AREA	2000 cm^2	20,000 cm^2
CAPACITANCE (50 pf/cm^2	.1 μf	1 μf
STORED ENERGY AT 10 KV	5 Joules	50 Joules
STORED CHARGE AT 10 KV	10^{-3} Coulomb	10^{-2} Coulomb
DURATION	1 μs	10 μs
PEAK CURRENT	1000 Amperes	1000 Amperes

CALCULATION OF INDUCED VOLTAGE BY ARCS STRIKING CABLES DIRECTLY

PEAK CURRENT: 1000 Amperes

ATTENUATION: 100 dB (10^5:1 ratio of currents)

IMPEDANCE: 5 kΩ Typical

INDUCED VOLTAGE: $V = IR = 1000\ 10^{-5} \times 5000 = 5$ Volts Induced into Cable

CALCULATION OF INDUCED VOLTAGE BY RADIATED FIELDS

ELECTRIC DIPOLE MOMENT: $p = qd = 10^{-3} \cdot 10^{-4} = 10^{-7}$ coulomb-meters

ELECTRIC FIELD: $E = p/(4\pi\epsilon_0 r^3)$, assume $r = 1$ m

$= 10^{-7}/(1.1 \cdot 10^{-10} \cdot 1) = 900$ volts/m

INDUCED B FIELD: Assume plane wave in space

$H = E/Z_0$ where $Z_0 = 377$ ohms

$B = \mu_0 H = \mu_0 E/Z_0 = \frac{4\pi \cdot 10^{-7} \cdot 900}{377} = 3 \cdot 10^{-6}$ weber/m^2

INDUCED VOLTAGE: Assume voltage pickup area, A, of 1 meter length by 1 cm height and a risetime, τ, of 10^{-7} seconds

$V = \frac{AB}{\tau} = \frac{10^{-2} \cdot 3 \cdot 10^{-6}}{10^{-7}} = 0.3$ volts

Table 3. General Design Requirements and Criteria

GROUNDING

GROUND ALL BOXES TO PLATFORM.

GROUND CABLE SHIELDING AS FREQUENTLY AS POSSIBLE.

PROVIDE GOOD GROUNDS TO STRUCTURE FOR ALL METALLIZED LAYERS IN THERMAL BLANKETS.

GROUND ALL ISOLATED OR INSULATED METAL STRUCTURES, E.G., THE ALUMINUM HONEYCOMB IN THE SOLAR CELL PANELS.

SHIELDING

PROVIDE ADEQUATE SHIELDING FOR EXPECTED ELECTRIC FIELD LEVELS AND SPECTRA.

PROVIDE ADEQUATE SHIELDING OF CABLING AND CONNECTORS.

USE TWISTED PAIR WIRING AND COMMON-MODE REJECTION TECHNIQUES WHERE NECESSARY.

CIRCUIT DESIGN

EACH INTERBOX WIRE SHOULD BE GROUNDED AT EACH BOX FOR FREQUENCIES HIGHER THAN THE INTENDED PURPOSE FOR THAT WIRE.

FILTERING: CIRCUITS SHOULD BE DESIGNED TO MINIMIZE REQUIRED BAND WIDTHS OR MAXIMIZE REQUIRED RISE TIMES ON INTERBOX WIRING.

CHARGE BALANCE

REDUCE VOLTAGE STRESS LEVELS AT SPECIFIC LOCATIONS AS DETERMINED BY CIRCUIT SUSCEPTIBILITY BY SELECTING THE PROPER SURFACE MATERIAL AND RATIO OF CONDUCTOR TO INSULATOR.

APERTURES AND SLITS

CLOSE OFF ALL APERTURES AND SLITS TO REDUCE VOLTAGE STRESS LEVELS AT SPECIFIC LOCATIONS.

GENERAL

PERFORM VERIFICATION TESTING TO ASSURE THE INTEGRITY OF GROUNDS, SHIELDS, CIRCUIT DESIGN, AND CHARGE BALANCE

INCLUDE HIGH-INTENSITY, HIGH-FREQUENCY (ARC DISCHARGE) SOURCES IN SPACECRAFT EMI ANALYSES. ONCE THIS IS DONE PROPERLY, STANDARD EMI PROBLEM-SOLVING TECHNIQUES MAY BE BROUGHT TO BEAR ON EACH PROBLEM AREA.

MINIMIZE EXPOSED INSULATED SURFACE AREAS TO REDUCE THE OCCURRENCE OF DIELECTRIC-TO-METAL ARCS, E.G., USE GROUNDED CONDUCTIVE COATING ON SOLAR CELLS AND EXPOSED MYLAR THERMAL BLANKET SURFACES

ing of metallic surfaces reduces the amplitude of the maximum arc that can occur on the spacecraft. The grounding procedure should be carefully controlled to assure the integrity of the ground in the space environment, for the lifetime of the mission. Figure 1 shows a number of TRW techniques used to ground vacuum deposited aluminum (VDA) films on thermal blankets. Questions such as the minimum size of ground strap, ratio of ground strap to area, exposed edges, etc. should be part of the design requirements imposed on the spacecraft designer.

There are five major elements to the design of spacecraft so that they survive and operate successfully in a charging environment:

1. Environmental Characteristics - generally defined by an environmental specification. The environmental specification may be used to set requirements on industrial spacecraft fabricators to assure that the spacecraft operates successfully within the specified environment.

2. Analytical Models - translate the environmental parameters and spacecraft configurational parameters into a quantitative measure of the hazard to the spacecraft.

3. Design Guidelines and Procedures - design guidelines are deduced from analytical model calculations and environmental models that predict the differential voltage stresses and the magnitude and location of environmentally induced arcs. The design guidelines are aimed at reducing the amplitude of the arcs, and adding sufficient shielding and filtering to the circuits to assure that malfunctions do not occur.

4. Verification Testing - a spacecraft test-verification program serves to test and verify that the design guidelines have been effectively applied and that the spacecraft is indeed immune to environmentally induced arcs.

5. Spacecraft Charging Housekeeping Monitors - are designed to yield engineering data directly applicable to the design and operation of the spacecraft. Since in most cases, it is impossible to eliminate arcing on synchronous spacecraft, the monitors perform routine housekeeping functions to verify and localize the arcing mechanisms, measure the coupling of radio frequency interference (RFI) into spacecraft harnesses, and increase the efficiency and reduce the cost of the analysis of operational problems.

The following sections present a discussion of each of the five major design factors that assure successful spacecraft operations.

ENVIRONMENTAL SPECIFICATIONS - SPACECRAFT CHARGING

The environmental specification is a detailed specification of the spacecraft plasma environment. It, more than anything else, is the key document that determines the hazard to the spacecraft and the design guidelines and procedures. There are two preliminary versions of an environmental specification for the geo-

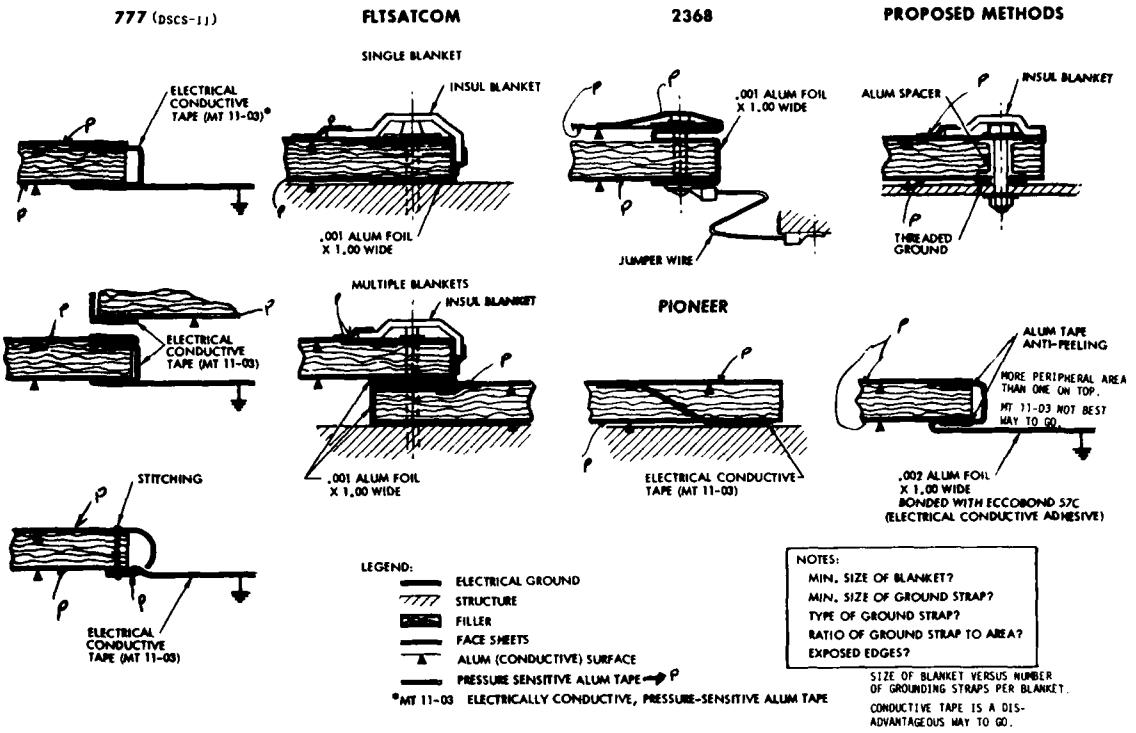


Figure 1. Insulator Grounding Techniques

synchronous plasma environment, one prepared for the Defense Nuclear Agency,[3] the other by NASA Lewis Research Center.[4] These documents are important to spacecraft sponsor agencies as well as industrial spacecraft designers and fabricators. Spacecraft sponsor and user agencies generally impose the requirements on the spacecraft fabricators—that the spacecraft shall be designed to operate successfully within the environment specified by the environmental specification. Since there are cost implications associated with this requirement, it is important that they be incorporated into the program at inception. One of the major problems associated with spacecraft charging immunization activities is that in the past, industrial fabricators have not been required to design spacecraft to withstand the plasma environment, and, therefore, have not included the cost of such undertakings; they have, therefore, been extremely reluctant to undertake an immunization program without additional funds.

Substorm Models; The Basis for Environmental Specifications

The ATS-5 and ATS-6* spacecraft provided the most complete in situ data on the geosynchronous orbit environment. Complete spectra of both negative and positive particles were generated continuously during the lifetime of the satellites. For purposes of charging analyses, it is convenient to characterize the plasma in terms of Maxwellian energy distributions, i.e., omnidirectional fluxes of ions and electrons and their equivalent temperatures. Frequently, a two component plasma consisting of cold and hot particles is a satisfactory representation; however, the possibility that a third higher energy component in the MeV range should be included has been discussed from the viewpoint that these particles have much greater penetration capabilities. Figure 2 shows the variation of equivalent Maxwellian fluxes and temperatures of electrons and protons for 2 January 1970. These fluxes may be used as a forcing function to compute the potential induced onto a spacecraft. Figure 3 shows the computed response of the DSCS II structure potential and two other surfaces for this "forcing function."[5]

Other environmental features that should be included in an environmental model are field aligned fluxes, a "warm" plasmapause bulge and possible longitudinal effects.

Intense field aligned fluxes of electrons and protons have been observed. These cause a charging pattern different from omnidirectional fluxes. The time dependence associated with these fluxes is important because different configurations on spacecraft have chargeup time constants that vary from a fraction of a second

*The ATS-6 plasma experiment has become inoperative as of May, 1977.

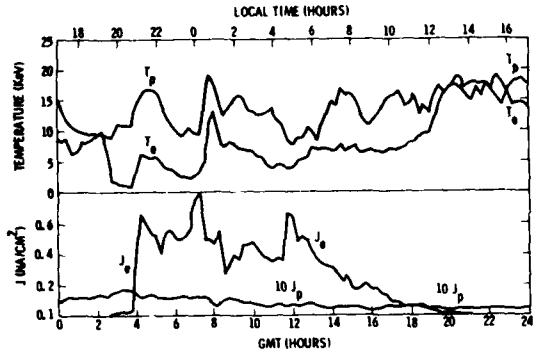


Fig. 2. ATS-5 Environmental Data for 2 January 1970 (Maxwellian Distributions)

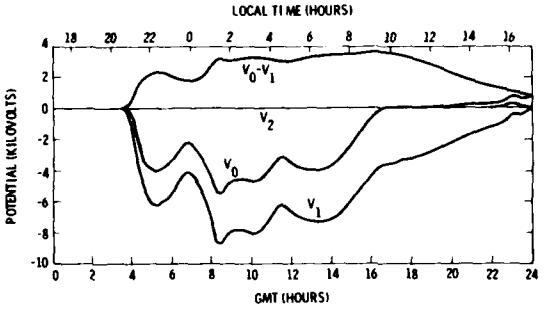


Fig. 3. Spacecraft Potentials in Response to 2 January 1970 Environment

to many minutes.

The "warm" plasmapause bulge in the evening local time sector, has been identified as a possible contributor to spacecraft charging effects.[6]

Geosynchronous orbit longitude effects have been observed in the sense that supposedly identical spacecraft do not have similar anomalies. A geomagnetic latitude difference resulting from the longitudenal displacement has been suggested as a possible explanation. The complex geometry of the earth's tilted magnetic field with respect to the solar wind during the yearly orbit period affects the shape of the magnetosphere and must be considered in specifying the substorm model.[7] The long term variations of the substorm model and the statistical description of its parameters are important in defining many of the spacecraft lifetime design requirements.

Ground based magnetometer stations also provide an important environmental data base. The Canadian magnetometer stations operated by G. Rostocker are an example of an experimental data gathering array. Figure 4 is a magnetogram supplied by Rostocker which was an essential element in establishing the correlation of the failure of the DSCS II Flight 1 spacecraft with an extremely unusual environmental condition.[8]

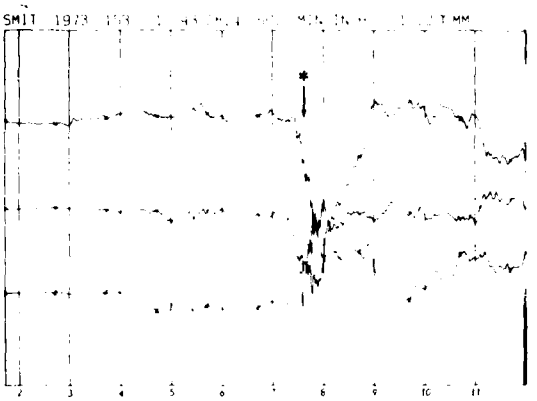


Fig. 4. All three components of the magnetic field recorded at Fort Smith, N.W.T., Canada, during the first part of the day on which the DSCS II Flight 1 spacecraft ceased operation, June 2, 1973 (Day 153)

A number of programs for performing experiments in space are being planned to provide information pertinent to the environmental specification. The Spacecraft Charging at High Altitudes (SCATHA) program is currently the most relevant space experiment program. Its objectives are to provide engineering information regarding materials responses and electromagnetic interference (EMI) data as well as environmental data relevant to future improvements of space systems.

Table 4 outlines the types of information received from in-orbit experiments. Environmental measurements and actual in-flight verification

Table 4. Experimental Programs in Space

TYPE OF INFORMATION	COMMENTS
ENVIRONMENT DEFINITION FLUXES AND ENERGY SPECTRA TEMPORAL, ANGULAR AND SPATIAL DISTRIBUTIONS SEASONAL AND SECULAR DEPENDENCES	ATS-5 AND ATS-6 ARE CURRENTLY PROVIDING DATA. DATA REDUCTION, ANALYSIS AND DISSEMINATION IS IMPERATIVE FOR MORE COMPLETELY INSTRUMENTED PAYLOADS (E.G., SCATHA). ALSO MULTIPLE SPACECRAFT DATA IS NEEDED.
ENGINEERING MEASUREMENTS MATERIALS PARAMETERS MATERIAL DEGRADATIONS EMI ENVIRONMENT, COUPLING CONTAMINATION EFFECTS ACTIVE POTENTIAL CONTROL	SCATHA WILL PROVIDE MUCH OF THIS TYPE OF DATA. ADDITIONAL DATA IS REQUIRED.
DESIGN VERIFICATION EMC EFFECTIVENESS CONTAMINATION COUNTERMEASURES (THERMAL, OPTICAL, PARTICULATE) MECHANICAL INTEGRITY THERMAL DESIGN LONG-TERM EFFECTS	ANOMALOUS SPACECRAFT BEHAVIOR IS ONLY VERIFICATION DATA. SCATHA WILL PROVIDE MUCH DATA, BUT INCORPORATION OF HOUSEKEEPING MONITORS ON EVERY GEOSYNCHRONOUS SPACECRAFT IS IMPORTANT.

of successful design are of course only performed in situ. In many cases, engineering data can be obtained in ground based tests. However, test configurations cannot simulate conditions exactly, and therefore require extrapolations that need to be verified in space. Costs and schedule considerations also enter into the tradeoffs as to the types of tests to be performed in space or on the ground. For example, long-term degradation data is needed now for design efforts currently under way. This information can be extrapolated from accelerated exposure tests, but should be verified in space. Ground based laboratory tests might not be the most cost effective approach in some cases. The costs of a multi-year vacuum chamber test could be prohibitive.

ANALYTICAL MODELS

Analytical modeling of spacecraft, together with an environmental model, provides a method of predicting the location, magnitude and polarity of differential voltage stresses and predicting the currents and voltages induced into wiring harnesses from arc discharges that result from the voltage stresses. Simplified analytical models have been instrumental in leading to design modification of the external surfaces of spacecraft.[5] Modifications such as grounding of thermal blankets and other metallic components, closure of apertures, redistribution of conductive and dielectric external surfaces, and reselection of dielectric surface materials have been successfully applied to spacecraft programs to reduce the occurrence and magnitude of environmentally induced arcs.

An additional feature of the analytical model is its capability to perform parametric studies of the effects of using different materials. The importance of the materials parameters such as photoemission, secondary emission, and resistivity on the spacecraft chargeup response are clearly indicated by model analyses.

Figure 5 shows a simplified charging model of a typical spacecraft that includes illuminated and dark dielectric exterior surfaces. As an electrical circuit, the current continuity equation requires that the summation of all incident currents be zero and thereby defines all of the potentials on the spacecraft. These potentials must, of course, be consistent with the plasma

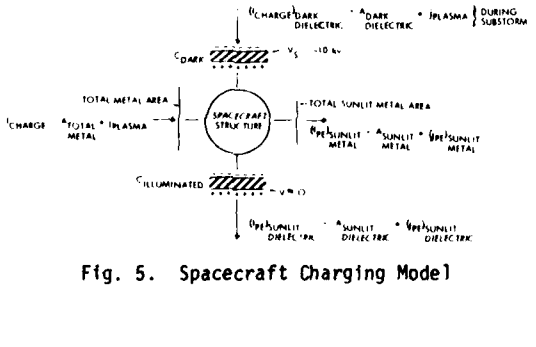


Fig. 5. Spacecraft Charging Model

environment, and this is the source of the complexity of the analytical modeling problem. A first order solution by G. T. Inouye[5] assumed that the Langmuir Mott-Smith equations for the sphere in a Maxwellian plasma apply. Since this treatment ignores emissions from the sphere itself (secondaries and photoemission), these were accounted for separately.

Self-consistent solutions of the "external" problem, the description of the plasma densities, and potentials surrounding the spacecraft must satisfy the collisionless Vlasov equations and Poisson's equation:

$$\vec{v} \cdot \nabla f + \vec{a} \cdot \nabla_v f = 0,$$

and

$$n(\vec{r}) = \iiint f(\vec{r},\vec{v})\, dv_x\, dv_y\, dv_z = \nabla^2\phi$$

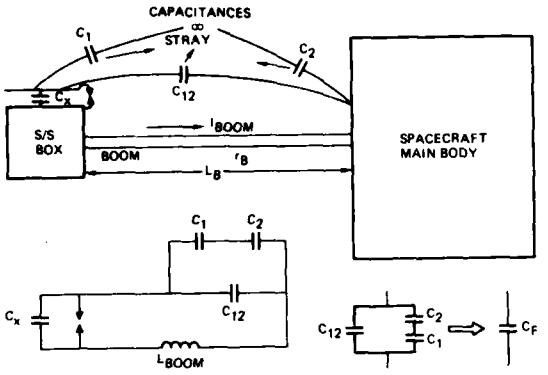
Many problems associated with the solution to these equations are currently being worked on by L. W. Parker,[9] E. C. Whipple,[10] J. G. Laframboise,[11] and by S^3 Inc. for NASA LeRC.[12]

Elegant and costly computer programs are involved in even the simplest geometry, and a cost-effective but adequately accurate approach to analytical modeling must be developed that is usable by the spacecraft designer.

Another area of analytical modeling that has received relatively less attention so far is that of prediction of currents and voltages induced into wiring harnesses from arc discharges that result from spacecraft charging. Two processes and, therefore, two additional models are involved: the physics and characterization of the arc discharge itself and the conductive or electromagnetic coupling of the discharge energy to the wire within the spacecraft. The first of these, the arc discharge analytical model, suffers from a lack of directly applicable basic research. Much research has gone into metal-to-metal arcs since this information is crucial in designing high-powered switches and circuit breakers, as well as long-life telephony type switches. The data base for the present application to dielectric-to-dielectric arcs seems to be minimal.

For the modeling of the arc discharge coupling to harness cables, there are a number of electromagnetic coupling analysis programs. A computer program developed at TRW for electromagnetic compatibility (EMC) has been modified to include arc discharge sources for TRW programs as well as for the Jet Propulsion Laboratory (JPL) Voyager spacecraft. The computer program developed for the Air Force, IEMCAP, is another possible analysis tool. Figure 6 is an example of an arc discharge model incorporated into the Voyager analytical model.[13]

Further sources of analytical models, particularly applicable to chargeup and discharge are those developed for EMP (Electromagnetic Pulse) and SGEMP (System Generated Electromagnetic



$$I_{BOOM} = I_{ARC}\left(\frac{X_{CX}}{X_{LB} + X_F + X_{CX}}\right)$$

$$= I_{ARC}\left(\frac{1/2\pi fC_X}{2\pi fL_B + 1/2\pi fC_F + 1/2\pi fC_X}\right)$$

$$= I_{ARC}\,[C_F/(C_F + C_X)]\,(4\pi^2 f^2 C_{eq} L_B + 1)^{-1}$$

$$= I_{ARC} \cdot P_1(f)$$

P(f) = LOW PASS (TYPE = 1)

G' = INBAND GAIN = $C_F/(C_F + C_X)$

$f_c = (2\pi\sqrt{L_B C_{eq}})^{-1}$

ORDER = 2

$C_{eq} = C_F C_X/(C_F + C_X)$

$V_{BOOM} = I_{BOOM} \cdot X_{LB}$

$V_B = I_{ARC} \cdot P_1(f) \cdot P_2(f)$

$P_2(f)$ = HIGH PASS (TYPE = 2)

G_1'' = INBAND GAIN = $2\pi L_B f_c$

$f_c = 10^9$ (ARBITRARILY HIGH)

ORDER = 1

L_B = INDUCTANCE OF BOOM GROUND STRAP

Fig. 6. Boom Replacement Current Model for Voyager

Pulse) purposes. These have generally been funded by the Air Force Weapons Laboratories (AFWL) and the Defnese Nuclear Agency (DNA). The usefulness of the available analytical modeling techniques depends on the cost effectiveness of the analysis in terms of direct relevance to the spacecraft design. For example, extreme accuracy to a fraction of a percent is useless for most purposes. Inaccuracies of 50 (3.5 db) to 100 percent (6 db) could be tolerable in terms of the countermeasures that might be incorporated in any case such as shielding and filtering. On the other hand, some scientific spacecraft missions would be totally invalidated if the analytical predictions are off by 0.1 percent (1 volt out of 1000 volts). Therefore the usefulness of a given analytic tool depends on the applications of the analyses as well as their accuracy and costs.

Material Characterization and Development of New Materials

The susceptibility of a spacecraft to the space plasma environment depends, in part, on the selection of spacecraft materials and their material characteristics. Table 5 lists some of the properties of materials that are important in the analytical modeling calculations. Unfortunately the parametric values are not known for many materials commonly used in spacecraft design. A number of material characterization studies have recently been initiated. These have generally been coupled with material development programs aimed at finding suitable spacecraft materials that also alleviate the problems resulting from spacecraft charging.[14-17]

Table 5. Properties of Materials Subjected to the Substorm Environment

BULK CONDUCTIVITY
BASE MATERIAL, REFERENCE CONDITION (NOT ILLUMINATED, NO CHARGED PARTICLES DEPOSITED, REDUCED LEVELS OF ELECTRIC STRESS).
BASE MATERIAL, ENHANCED BY PHOTOCONDUCTION EFFECTS.
BASE MATERIAL, ENHANCED BY BOMBARDMENT INDUCED CONDUCTIVITY EFFECTS.
BASE MATERIAL, ENHANCED BY HIGH ELECTRIC STRESS EFFECTS.
SURFACE CONDUCTIVITY
BASE MATERIAL, REFERENCE CONDITION.
BASE MATERIAL, ENHANCED BY PHOTOCONDUCTION EFFECTS.
BASE MATERIAL, ENHANCED BY BOMBARDMENT INDUCED CONDUCTIVITY EFFECTS.
BASE MATERIAL, ENHANCED BY HIGH ELECTRIC STRESS EFFECTS.
SECONDARY ELECTRON EMISSION RATIOS
BASE MATERIAL AT LOW LEVELS OF ELECTRON DEPOSITION CURRENTS.
BASE MATERIAL AT HIGH LEVELS OF ELECTRON DEPOSITION CURRENTS.
BULK POTENTIAL GRADIENT BREAKDOWN LEVEL
SURFACE POTENTIAL GRADIENT BREAKDOWN LEVEL
SOLAR ABSORPTIVITY
INFRARED EMISSIVITY

Figures 7, 8 and 9 are examples of the types of materials testing data obtained at TRW.[16,17] Figure 7 shows the results of a thermal blanket ground strap test for durability under surge currents. Figure 8 shows the nonlinear bulk resistivity of Kapton with voltage stress and also the lowering of resistivity with illumination with UV. Figure 9 shows nonlinear resistivity with voltage stress. In addition, it shows an apparently permanent hysteretic change in bulk resistivity under irradiation with an electron beam.

Table 6 provides a representative list of materials that are presently under development to avoid or to reduce spacecraft chargeup. Two different approaches to the problem have been considered. In the first, the exterior materials of the spacecraft surface are made conducting so that the deposited and created charge carriers are transported to another location on the spacecraft to prevent the occurrence of discharges. The stored electrical energy is minimized so that discharges, even if they should occur, operate with a small energy content.

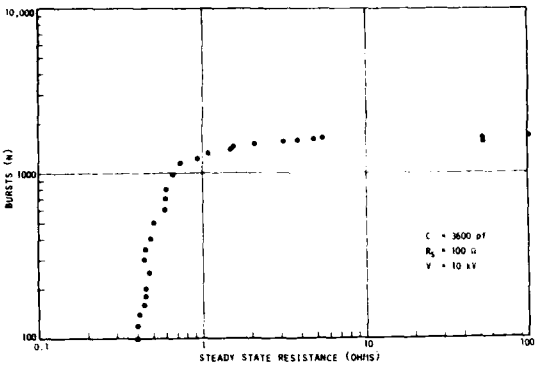


Fig. 7. Resistance of VDA/NOMEX/ground strap sample 1 as a function of the number of current bursts. Initial surge current = 100 amperes and total charge throughput = 36 microcoulombs.

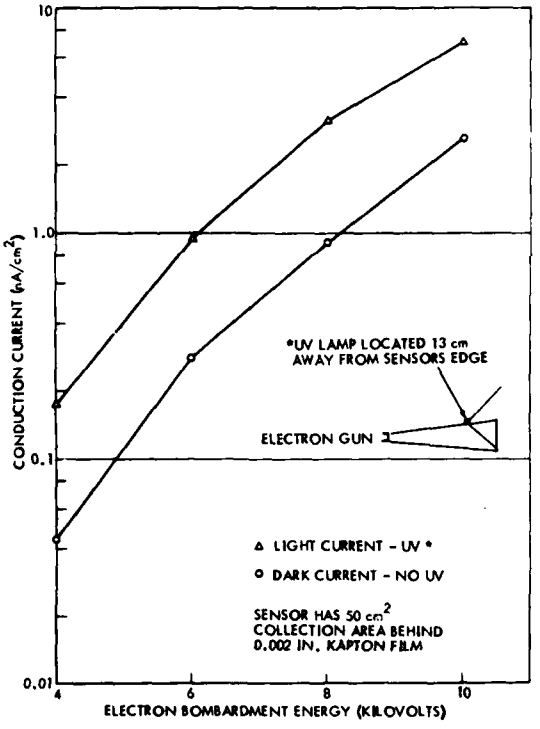


Fig. 8. Conduction current density through Kapton as a function of incident electron bombardment voltage for UV absent and UV present conditions.

In the second approach, the possibility of arc discharges is acknowledged; however, the spacecraft material mix and electrical circuitry are so configured in order to avoid harm as a result of these discharges.

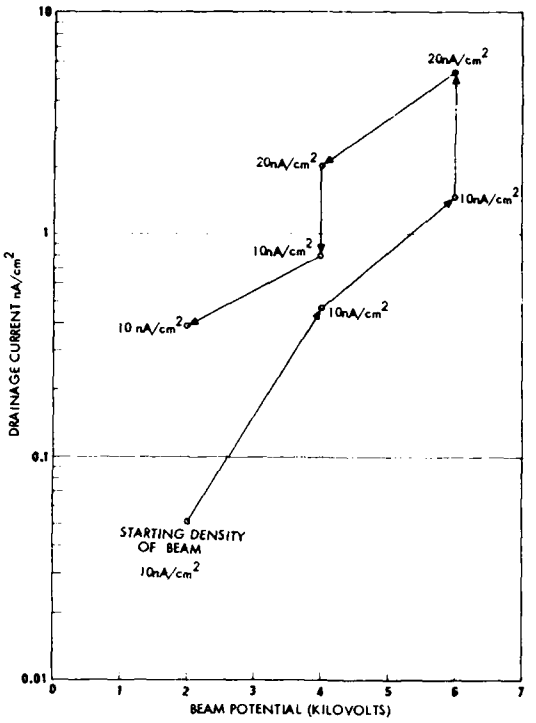


Fig. 9. Electron drainage current density as a function of electron beam acceleration voltage for deposition flux densities of 10 nA/cm^2 and 20 nA/cm^2.

Table 6. Spacecraft Surface Materials Under Development for Spacecraft Chargeup Conditions

POLYMERIC FILMS
• POLYMERIC FILMS WITH BULK ADDITIVES (INCREASED BULK CONDUCTIVITY)
• POLYMERIC FILMS WITH SURFACE ADDITIVES (INCREASED SURFACE CONDUCTIVITY)
— INDIUM TIN OXIDE (In_2O_3/SnO_2) LAYERS (SPUTTER DEPOSITION)
— INDIUM OXIDE, ALUMINUM OXIDE (In_2O_3, Al_2O_3) LAYERS (VAPOR DEPOSITION)
— COPPER GRID (PHOTOETCHED)
— SILVER GRID (SILKSCREENED)
(NOTE: SUBSTRATE POLYMER FILMS ARE KAPTON, MYLAR, FEP TEFLON)
CONDUCTING PAINTS
• CONDUCTIVE ORGANIC POLYMERS
• CONDUCTIVE INORGANIC BINDERS
• CONDUCTIVE PIGMENTS
CONDUCTING FABRICS
• SILICA FABRICS (ASTRO QUARTZ)
• OSR AND SOLAR CELL COATINGS
— CONDUCTING FILM ADDITIVES
• INDIUM TIN OXIDE (In_2O_3, SnO_2) LAYERS (SPUTTER DEPOSITION)
— GLASS FILM ADDITIVES
• LITHIUM BOROSILICATE GLASS
• LITHIUM ZINC ALUMINOGERMANOBORATE GLASS

TEST VERIFICATION

Having followed the design guidelines, a spacecraft test verification program may be undertaken to assure that the spacecraft circuits are adequately shielded and filtered, that the grounding of all conductive elements are intact, and that capacitive configurations have been minimized or eliminated. Tests for spacecraft charging effects may be performed at the material, component, subassembly, and integrated spacecraft level.

At the material level the data are used for material selection and for analytical prediction of charging effects. At the component and spacecraft levels, testing is performed to verify the design and also to verify that fabrication is correct. Test data at these levels also serve as supportive and corroborative information for analytical model predictions of subsystem and system performance in space. These test objectives are shown in Fig. 10.

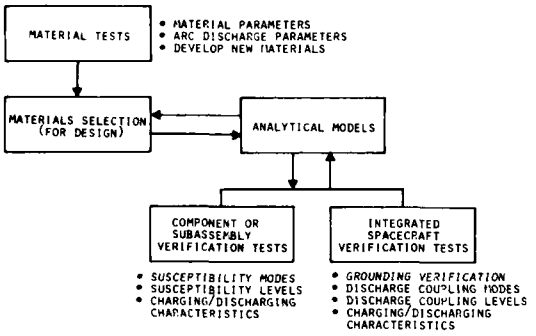


Fig. 10. Test Objectives

Because of the importance of test data, test specifications should be written to maximize the usefulness, quality and quantity or types of data obtained. The efforts of the Test and Evaluation Working Group of the NASA/USAF spacecraft charging investigation are channeled towards the development of common test specifications that are useful to the entire interested community.[18] MIL Standard 1541, Sec. 6.5.2.4.1 "Electrostatic Discharge," is the most applicable existing EMC specification dealing with arc discharge effects, but is totally inadequate. It does, however, provide a framework to incorporate the present concerns.

At the present time, as can be expected for a relatively new area, the requirements and techniques for testing spacecraft charging immunity are not clearly defined. Various organizations are producing test specifications as they deem appropriate. At TRW, tests have been performed at the material and component level on five separate spacecraft programs and at the integrated system level on two programs. Further, we were heavily involved as consultants on JPL's Voyager spacecraft immunization and verification test effort.

The conceptual design of the tests performed on the Voyager spacecraft demonstrates a crucial role of analytical modeling in reducing test costs in schedule time and manpower. The computerized EMC program modeled for the Voyager spacecraft was modified to include arc discharge sources located at various positions on the spacecraft. A preliminary analysis defined the most susceptible source locations as well as the most susceptible victim harness cable wires. Test excitation location and diagnostic points were minimized to only those locations crucial to the verification of the model predictions.

Another crucial concern in the verification testing of flight equipment is that none of the components are overstressed during the tests. The danger exists that a marginal overstress condition might not be revealed until after launch. An awareness of these concerns, the incorporation of adequate diagnostics, and the concommitant use of analytical modeling are techniques that must be implemented.

Materials testing raises a number of questions as to the nature of the tests being performed. For example, "What are the implications of the following?"

- Testing with monoenergetic electron unidirectional beams
- Absence of ions
- Effects of vacuum walls, chamber size, vacuum level
- Size and shape of test samples
- Cleanliness of test samples
- Combined effects of plasma and UV irradiation
- Accelerated test levels.

SPACECRAFT CHARGING HOUSEKEEPING MONITORS

It is impossible to completely eliminate the occurrence of arcs on present day synchronous altitude spacecraft systems. At best, the spacecraft system may be made immune to the effects of environmentally generated arcs. Unfortunately the plasma environment is not completely understood and the long term effects of the space environment on material characteristics are known to a lesser degree than many of the material characteristics themselves.

Spacecraft charging housekeeping monitors would perform routine housekeeping functions to monitor problem areas during flight. Simple harness noise monitors may be used to verify and localize the arcing mechanism and measure the coupling of RFI into the harness. Advantages of such monitors are (1) problems on spacecraft will be identified before they become sufficiently

serious to cause component or spacecraft failures. (The low level threshold of the detector would permit correlation of smaller events with minor changes in housekeeping data.) (2) Problems that are peculiar to a given spacecraft could be identified and analyzed. (3) The data base that is generated would be relevant to the study of the effects of substorms on spacecraft and would permit a meaningful correlation with ground station observations and observations of other space systems.

CONCLUSIONS

There has been much progress during the past few years in understanding spacecraft charging phenomena and in developing design procedures and standards for successful operation of spacecraft in the environment. There are, however, serious gaps in our knowledge; gaps that have a direct effect on our ability to design and fabricate spacecraft that are impervious to environmentally generated arcs. At this time there is not available a standard, simple and direct methodology for determining (1) approximate spacecraft voltage stresses, (2) expected arc discharge characteristics, (3) expected coupling of arcs to spacecraft subsystems, and (4) a measure of the susceptibility of some standard spacecraft subsystems/components.

The Three-Dimensional Dynamical Study performed by S^3 is a very important contribution to the field.[12] It is a particularly useful tool for scientific space systems such as the SCATHA program. However, it is not useful to the designer of synchronous, operational spacecraft systems. First of all, it is too complicated; second, it is most likely more accurate than required by spacecraft designers; and thirdly, it is probably more costly to use than most projects are willing to pay. In addition, there are significant gaps in the development. Namely, the dynamics of the arc discharge and the coupling of the arc discharge into the spacecraft. The tasks that must be performed, as soon as possible, are

1. Evaluate the accuracy requirement of spacecraft charging stress calculations. Simultaneously develop a simple, standard approximate methodology for estimating voltage stresses and their locations on spacecraft.

2. Characterize the coupling of arc discharges into spacecraft circuits:

   a. Develop simplified model (e.g., electromagnetic susceptibility program such as the TRW Specification and Electromagnetic Compatibility Analysis Program [SEMCAP] computer code).

   b. Define experimental test data requirements to implement the program:

      1) Models of arc discharge source

      2) Models of arc discharge coupling

      3) Models of susceptibility of subsystems/components.

   c. Perform analyses in terms of worst case waveforms.

The available data—characterizing arc discharges—is spotty, incomplete and not reliable. A definitive study in this area is essential if we are to understand the coupling sources and the techniques to make spacecraft impervious to them. The immediate tasks that must be performed in this area are experimental in nature. They include

1. Evaluation of existing data

2. Evaluate difference between test simulation sources and actual "inflight" arcs

3. Carefully define the experiments that should be performed

4. Perform laboratory experiments using an array of standard spacecraft materials

5. Generate tables or handbooks of data.

A spacecraft design guideline specification is now under development. There are two additional areas that require attention. The first relates to methods of performing design verification tests on spacecraft to assure they are impervious to environmentally generated arcs and the second relates to the requirement for development of standard inflight housekeeping monitors to determine the susceptibility of operational spacecraft to environmentally generated arcs.

ACKNOWLEDGMENTS

The work reported here represents the various efforts of the USAF, NASA Lewis Research Center, JPL, the University of California at San Diego, and TRW. It is not possible to enumerate all those who contributed significantly to this paper. I am particularly indebted to M. H. Bunn of the SAMSO DSCS Project Office; R. Lovell and N. J. Stevens of NASA Lewis Research Center; Bill Shipley, Tom Gindorf, A. Whittlesey, and A. Beck of JPL, for their support and guidance in identifying many of the important spacecraft charging tasks. At TRW, the many discussions with G. T. Inouye, N. L. Sanders and J. M. Sellen, Jr. have been invaluable in developing a deeper understanding of the whole field—without their support this paper would not be possible.

REFERENCES

1. Proceedings of the Spacecraft Charging Technology Conference, Ed. C. P. Pike, R. R. Lovell, AFGL-TR-77-0051 and NASA TMS-73537, 24 February 1977.

2. Spacecraft Charging by Magnetospheric Plasmas, Progress in Astronautics and Aeronautics, Vol. 47, Ed. Alan Rosen, Pub. by

AIAA in cooperation with MIT Press, 1976.

3. A Preliminary Specification of the Geosynchronous Plasma Environment, DNA 3951T, September 1976.

4. Provisional Specification for Satellite Time in A Geomagnetic Substorm Environment, NASA TMX-73446, October 1976.

5. Inouye, G. T., "Spacecraft Charging Model," Paper 75-255, AIAA 13th Aerospace Sciences Meeting, Jan. 20-22, 1975, Pasadena, Ca.; also *J. of Spacecraft & Rockets, 12*, No. 10, Oct. 1975, pp. 613-620.

6. Reasoner, D. L., W. Lennartsson and C. R. Chappell, "Relationship between ATS-6 Spacecraft-Charging Occurrences and Warm Plasma Encounters," Spacecraft Charging by Magnetospheric Plasmas, Progress in Astronautics and Aeronautics, Vol. 47, Ed. Alan Rosen, Pub. by AIAA in cooperation with MIT Press, 1976.

7. Rosen, Alan, "Spacecraft Charging Problems," Physics of Solar Planetary Environments, Proceedings of the International Symposium on Solar-Terrestrial Physics, Vol. II, Edited by D. J. Williams, P. 1024-1038, Amer. Geophys. Union, 1977.

8. DeForest, S. E., R. L. McPherron and G. Rostoker, "The Possibility of an Environmental Origin for TDAL Anomalies of the DSCS II Satellite," Program 777 Anomaly Investigation, TRW Report No. 9670-REP-050-01, Contract No. FO 4701-C-0280, 22 March 1974.

9. Parker, Lee W., "Calculation of Sheath and Wake Structure About a Pillbox-Shaped Spacecraft in a Flowing Plasma," Proceedings of the Spacecraft Charging Technology Conference, Ed. C. P. Pike, R. R. Lovell, AFGL-TR-77-0051 and NASA TMX-73537, 24 February 1977.

10. Whipple, E. C., "Theory of the Spherically Symmetric Photoelectron Sheath and Comparison with the ATS-6 Observation of a Potential Barrier," Spacecraft Charging by Magnetospheric Plasmas, Progress in Astronautics and Aeronautics, Vol. 47, Ed. Alan Rosen, Pub. by AIAA in cooperation with MIT Press, 1976.

11. Laframboise, J. G. and S. M. L. Prokopenko, "Numerical Simulation of Spacecraft Charging Phenomena," Proceedings of the Spacecraft Charging Technology Conference, Ed. C. P. Pike, R. R. Lovell, AFGL-TR-77-0051 and NASA TMX-73537, 24 February 1977.

12. Katz, I., et al., "Three-Dimensional Dynamic Study of Electrostatic Charging in Material," Contract No. NAS3-20119 prepared for NASA LeRC, August 1977.

13. Inouye, G. T., A. C. Whittlesey, S. R. Ponamgi, G. B. Cooperstein, and A. K. Thomas, "Voyager Spacecraft Electrostatic Discharge Immunity Verification Tests," presented at the Symposium on the Effects of the Ionosphere on Space and Terrestrial Systems, Arlington, Va., Jan. 24-26, 1978.

14. Lenh, W. L., "A Conductive Spacecraft Materials Development Program," Proceedings of the Spacecraft Charging Technology Conference, Ed. C. P. Pike, R. R. Lovell, AFGL-TR-77-0051 and NASA TMX-73537, 24 February 1977.

15. Gilligan, J. E., R. E. Wolf, and C. Ray, "Electrically Conductive Paints for Satellites," Proceedings of the Spacecraft Charging Technology Conference, Ed. C. P. Pike, R. R. Lovell, AFGL-TR-77-0051 and NASA TMX-73537, 24 February 1977.

16. Hoffmaster, D. K., G. T. Inouye and J. M. Sellen, Jr., "Surge Current and Electron Swarm Tunnel Tests of Thermal Blanket and Ground Strap Material," Proceedings of the Spacecraft Charging Technology Conference, Ed. C. P. Pike, R. R. Lovell, AFGL-TR-77-0051 and NASA TMX-73537, 24 February 1977.

17. Sellen, J. M., Jr., "Electrical Equilibrium of Conducting and Insulating Materials in the Presence of Energetic Electrons and Ultraviolet Light," Spacecraft Charging by Magnetospheric Plasmas, Progress in Astronautics and Aeronautics, Vol. 47, Ed. Alan Rosen, publishedby AIAA in cooperation with MIT Press, 1976.

18. Lovell, R. R., N. J. Stevens, Wayne Schober, C. P. Pike, and William Lehn, "Spacecraft Charging Investigation: A Joint Research and Technology Program," Spacecraft Charging by Magnetospheric Plasmas, Progress in Astronautics and Aeronautics, Vol. 47, Ed. Alan Rosen, published by AIAA in cooperation with MIT Press, 1976.

DIFFERENTIAL CHARGING OF NONCONDUCTING SPACECRAFT

Lee W. Parker
Lee W. Parker, Inc.
252 Lexington Road, Concord, Massachusetts 01742

ABSTRACT

The differential charging of a nonconducting spacecraft is modeled numerically by following charged-particle trajectories in a self-consistent space-charge-less sheath. In the presence of a plasma flow, but independent of any photoelectric or secondary emission, a potential difference between the front and wake surfaces of the spacecraft is generated, resulting in an asymmetric sheath and in the creation of a potential barrier for electrons. The potential difference can amount to volts in the ionosphere, and kilovolts in the solar wind, that is, large compared with the potentials typically generated by photoemission alone. As in the more familiar case of photoelectric charging, the asymmetric sheath and potential barrier produced by the plasma flow can lead to erroneous interpretations of experiments measuring space electric fields and low-energy particle spectra. In an example of photoelectric emission, a sunlit area on an otherwise dark surface becomes positively charged by the emission, and is found to acquire a potential more than twice the emitted energy relative to the dark surface. This effect is associated with the physics of the "terminator" bounding the sunlit and dark areas.

1. INTRODUCTION

Differential spacecraft charging takes place when the spacecraft surface is partly or entirely insulating and the charged-particle fluxes vary from point to point over the surface. In the relatively familiar case of photoelectric emission from a sunlit insulated area (cf. Grard, 1973), due to electrons escaping from it the sunlit area tends to become positively charged relative to the surrounding dark areas. Another mechanism of differential charging, which is less familiar and appears to have been discussed little if at all in the literature, is that due to the relative motion between a nonconducting spacecraft and the external plasma (e.g., a spacecraft in the ionosphere or in the solar wind). The fluxes of ambient ions and electrons on the wake surface are not the same as on the front surface. For high velocities of relative motion compared with the mean ion thermal velocity, whether this occurs in the ionosphere (due principally to spacecraft motion) or in the solar wind (due principally to plasma motion), there is a significant differential in the ion fluxes, but a negligible differential for the electrons. Since the net current density must vanish locally at each surface point in the steady state, this plasma-flow effect leads to a larger negative equilibrium potential on the wake surface than on the front surface. If there is photoemission as well on the front surface (as in the solar wind), this differential charging is enhanced. The present paper is concerned with these effects. As shown below, differences between the front and wake surface potentials amounting to many kT/e (where T is the temperature, k is Boltzmann's constant, and e is the electron charge), together with a potential barrier for electrons, can be generated by the motion. The potential differential may be expected to be of the order of volts (tens of kT) in the ionosphere, and of the order of a kilovolt (100 kT) in the solar wind. In the ionosphere and solar wind, this differential can therefore be much larger than that generated by photoemission alone. In the magnetosphere, however, the plasma-flow effect is relatively weak.

As of the mid-1960's there was already a considerable literature on the subject of estimating spacecraft potentials, using simplified models without considering trajectories and assuming perfectly conducting spacecraft (see reviews by Brundin, 1963; Whipple, 1965; Samir and Willmore, 1966). These relatively crude models for estimating spacecraft potentials assumed either (a) very thin sheaths such that the "planar approximation" could be used, or (b) very thick sheaths (for small objects in the ionosphere) but with radial symmetry so that the simple Langmuir theory could be used. The crude estimates were not concerned with differential charging; they sufficed for treating effects where the detailed structure and asymmetry of the sheath (potential barriers, for example) were not considered important.

From the mid-1960's onward, spacecraft have increasingly sampled the magnetosphere and solar wind, where the spacecraft conditions are altered in several ways important for differential charging. First, there is an increase of sheath thickness to the order of the spacecraft dimensions, as opposed to the thin sheaths encountered in the ionosphere; this thickness is governed essentially by photoelectrons and secondary electrons from the surface, the space plasma contribution being typically relatively weak. A second circumstance is that many spacecraft surfaces are partly or entirely nonconducting (e.g., composed of glass-covered photocells or insulating thermal control blankets), which becomes important when the sheath is thick. A third circumstance is that nonuniform particle fluxes occur over the spacecraft surfaces, e.g., due to photoelectron and secondary emission and to plasma flows. The combination of the foregoing circumstances leads to differential charging of the spacecraft surfaces, which can have deleterious effects as follows.

If the differential charging is severe enough, spacecraft malfunctions can occur due to electrical discharges on the insulating surfaces (Fredricks and Scarf, 1973; Rosen, 1976). The appearance of hot magnetospheric plasmas in the kilovolt temperature range impacting on the dark surface, combined with photoelectric emission on the sunlit surface, make possible intense differential charging such that tens-of-kilovolts differentials can appear (see spacecraft charging symposia by Grard, 1973; Rosen, 1976; Pike and Lovell, 1977).

Even if the differential charging is not so strong, say no more than tens to hundreds of volts of differential, it can interfere with measurements of, say, weak ambient electric fields or low-energy particle spectra (Grard, 1973; Whipple and Parker, 1969). An interesting feature of differential charging as it affects low-energy electron measurements is that it can create electron potential barriers which can return emitted photoelectrons or secondary electrons to the surface and lead to erroneous interpretations of the data (Fahleson, 1973). This type of electron potential barrier is distinct from and should not be confused with the more familiar space-charge potential minimum of the order of a volt which can be produced by emitted-electron space charge and is not due to differential charging. The space-charge potential minimum has been studied theoretically by Soop (1972, 1973), Schröder (1973), Parker (1976b), Whipple (1976a), and Rothwell et al. (1977); it can, however, be less important than the barriers produced by differential charging effects. Discussing a well-documented experiment on the ATS-6 geosynchronous satellite, Whipple (1976b) infers that photoelectrons and secondary electrons from the spacecraft surfaces are reflected from a potential barrier which is much too large to be due to space charge but must be associated with some kind of differential charging (Whipple, 1976ab). A similar potential barrier due to differential charge is that produced artificially by an attractive electron trap mounted on a repulsive spacecraft, which can cause secondary-emission currents to be incorrectly interpreted as plasma currents (Whipple and Parker, 1969). In the present paper, a potential differential and a potential barrier are shown to be producible by a plasma flow. This can lead to difficulties of interpretation of solar wind measurements such as those of Rosenbauer (1973), and may be responsible for singularities observed by photoelectron detectors in the ionosphere (W. K. Peterson, private communication, 1977).

The procedure for theoretically predicting sheath asymmetries and potential barriers is generally complicated in that it requires particle trajectory calculations as well as a three-dimensional sheath description for a realistic treatment (Parker, 1970, 1973, 1976a, 1977, 1978; Parker and Whipple, 1967, 1970).

In this paper we present results of sample calculations of differential charging due to both plasma flow and photoemission, primarily addressing the asymmetric sheath and potential barrier produced by the plasma flow. These results may be considered preliminary, because the photoemission is considered separately rather than simultaneously. However, the differential charging is enhanced by photoelectric emission on the front surface. Space charge is neglected compared with surface charges as sources of the field; this neglect has been shown to be justified (Soop, 1973). Hence, while the predictions may not be quantitative for an actual spacecraft, they are conservative and indicate what may be expected: (a) in the ionosphere for small insulated objects, small meteoroids, or small parts of a spacecraft (e.g., a painted antenna) located within the wake region of a moving spacecraft, and (b) in the solar wind for an entire spacecraft.

In the example of photoelectric emission, a sunlit area on an otherwise dark surface becomes positively charged by emission, and is found to acquire a potential more than twice the emitted energy. This effect is associated with the physics of the "terminator" bounding the sunlit and dark areas.

In Section 2 we indicate briefly the nature of the numerical methods used, which are presented in more detail by Parker (1978).

Results are discussed in Section 3.

2. NUMERICAL APPROACH

The model chosen to represent a spacecraft is a truncated cylinder of approximately equal height and diameter, as shown in Fig. 1 where it is called (for brevity) a "pillbox." It is assumed that the spacecraft surface charge and

the electric field around the spacecraft are axially symmetric, and that the electric field in space is given by the solution of Laplace's equation on a grid of points in r-z space (including the pillbox-shaped spacecraft). This means that when the potential distribution on the grid points has been computed (cf. Parker and Whipple, 1970), the potential and electric field at any point in space may be obtained by interpolation. This allows ion and electron trajectories to be computed within the region of space spanned by the grid. The outer boundary of the grid represents "infinity." The method used for computing the field allows the use of variable grid intervals (see Parker, 1976a, 1977, 1978; Parker and Whipple, 1970).

The normal component of the ion or electron flux may be computed at any point on the pillbox surface by evaluating a triple integral in velocity space, and following trajectories backward in time to determine their origin and therefore the value of the integrand. This represents the "inside-out" method originated by Parker (1964). Details are given by Parker and Whipple (1970), and more generally by Parker (1976a, 1977, 1978). An alternative approach is the outside-in method (Parker, 1978). For the calculations to be discussed, the conditions at the spacecraft surface are represented by a discrete set of grid points and associated surface areas, as illustrated (by 12 points) in Fig. 2.

For the present purposes, the differential potential and charge distributions on the spacecraft surface may be determined by two different approaches. In one approach we may determine the potential at each local grid point by a relaxation process until the net current density is zero. This involves "cutting-and-trying", whereby the surface potentials (12 values as illustrated in Fig. 2) are first given assumed values, and later successively corrected in accord with the signs and magnitudes of the resulting set of net current densities. The surface potentials represent the "inner" boundary condition. In the other approach, a time-dependent method is used whereby the surface charges at the surface grid points are updated after each time-step, during which the net current densities are calculated. The next potential distribution is computed using the new surface charges as sources. In the computations to be described next, a grid of about 400 points was used, of which 24 were used to represent the spacecraft surface points. Further details are given by Parker (1978).

3. SAMPLE SOLUTIONS

A. Plasma Flow

In one of the problems treated here, we assume the nonconducting spacecraft to be in a flowing plasma, with the plasma flow along the axis, from the bottom (front region) toward the top (wake region) in Fig. 1. The plasma is taken to be ionized hydrogen and is assumed to have a velocity of flow four times larger than the most probable ion thermal velocity (ion "Mach number" = 4). Since the unperturbed ion flux to the wake surface is about 9 orders of magnitude smaller than the corresponding ion flux to the front surface, and since the electron fluxes are about the same to the front and wake surfaces, there will be a significant differential between the equilibrium potentials at the front and wake surfaces (see below). The effects of photoemission for the pillbox geometry are treated in the next section.

Figure 3 shows equipotential contours around the spacecraft, obtained by numerical solution, labeled by encircled numbers representing dimensionless values of the potential (in units of kT/e, where T is the plasma temperature). The errors in the solution shown are estimated to be under 10 percent. These potentials are obtained from Laplace's equation, where the surface potentials are obtained by the relaxation method discussed by Parker (1978), under the requirement of zero net current density at all surface points.

There are three regions of characteristic behavior of the potential, the "top" or "wake", the "side", and the "bottom" or "front". In the top region, the potentials are of the order of -10kT/e. This large negative value (about -17kT/e at the surface) is associated with the reduction in ion flux due to the flow. In the side region, the potentials are of the order of -3kT/e; this is essentially the order of the equilibrium potential when there is no flow ($-\ln\sqrt{m_i/m_e}\ kT/e$). In the bottom region, the potentials are of the order of -kT/e, i.e., less negative than the side, because of the enhancement of the ion flux due to the flow. (Adding photoemission here would make the front potential still less negative.) The surface points are thus not equipotential. Note that there is a saddle-point in the front region, that is, a potential barrier for electrons. The barrier height is between 10 percent and 20 percent higher than the potentials at the nearest (front) surface points. This feature is caused by the interaction between the relatively-large-magnitude wake potentials and the relatively-low-magnitude front potentials. The dashed part of the contour labeled "-3.0" near the side surface indicates that there is more complicated fine structure (variations of potential along the side surface) than is shown in the figure. The potentials along the top surface fall off to the right. The potentials along the bottom surface first fall with radius, then rise sharply as the corner is approached. This may be a "corner effect."

On the basis of these results, one would expect a relatively small body in the ionosphere, such as a thin antenna or boom painted with nonconducting paint, or a painted or insulated object in the very near wake of a

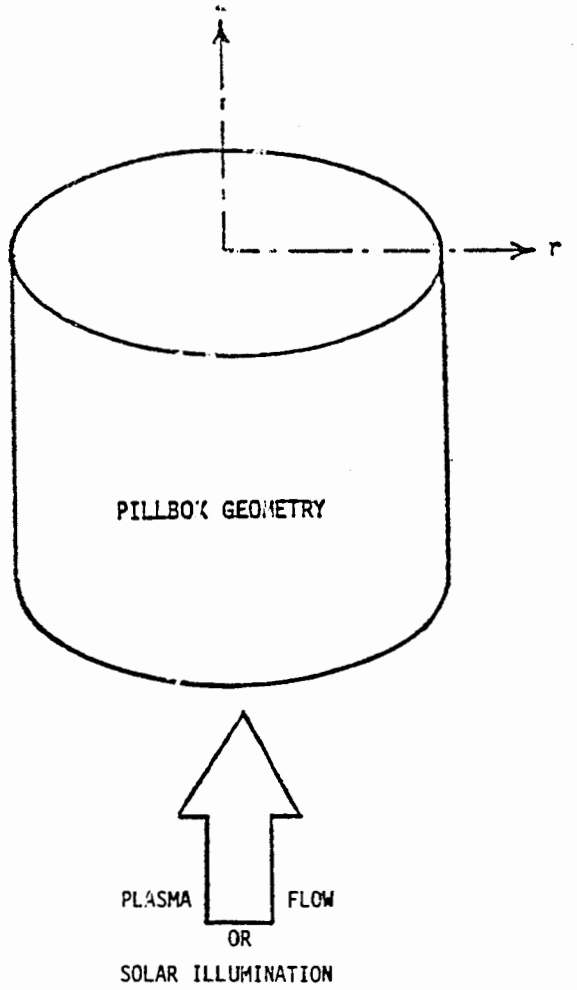


FIG. 1. SPACECRAFT AND PLASMA-FLOW OR SOLAR-ILLUMINATION GEOMETRY

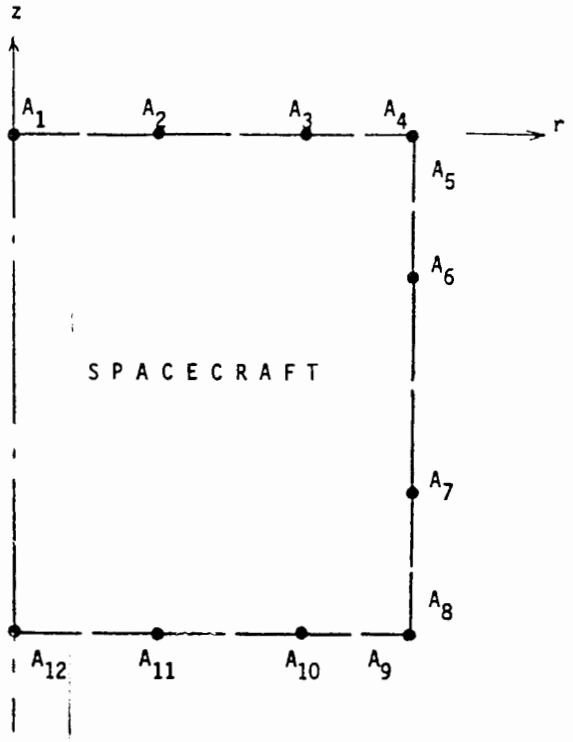


FIG. 2. SURFACE AREAS ASSOCIATED WITH SURFACE POINTS ON SPACECRAFT

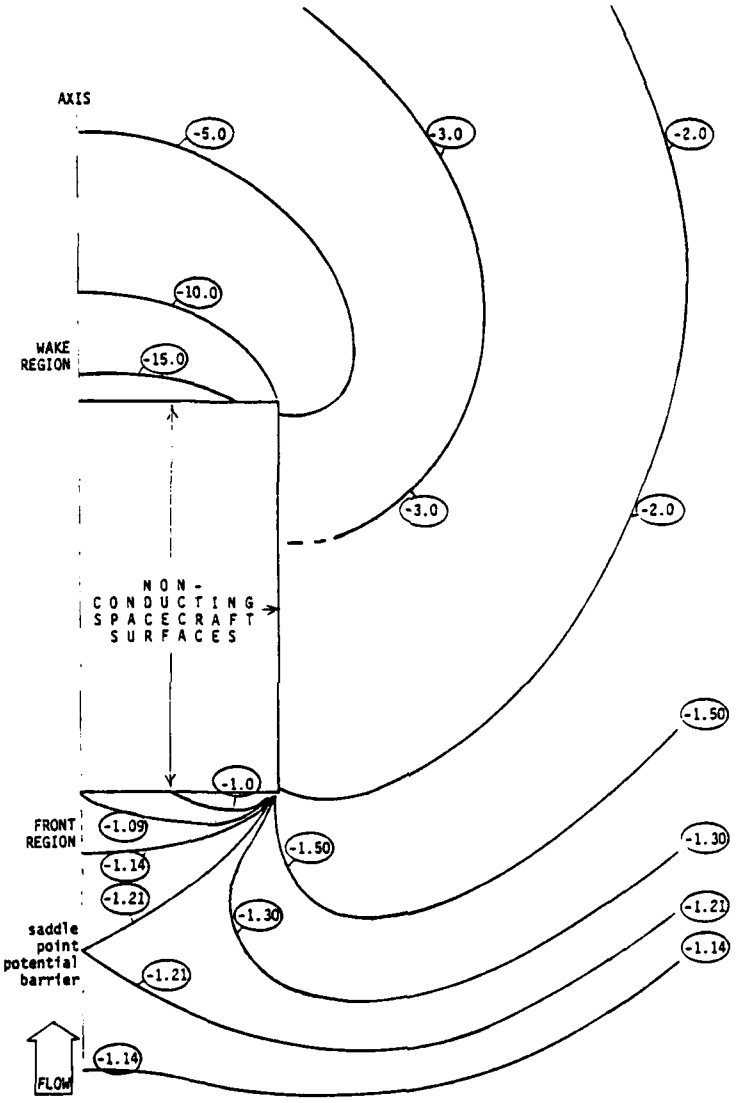


FIG. 3. DIFFERENTIAL CHARGING OF NONCONDUCTING SPACECRAFT BY PLASMA FLOW (EQUIPOTENTIAL CONTOURS IN V)

spacecraft (or the spacecraft surface itself if it is a dielectric), to become highly negatively charged, to potentials of the order of volts in the ionosphere.

In the solar wind, the above calculation could apply to an entire spacecraft since the Debye length is large. However, the ion Mach numbers are of the order of 10 rather than 4, which would lead to negative dimensionless wake potentials an order of magnitude larger than the wake potentials shown in Fig. 2. (It is shown by Parker (1978) that the potential difference in units of kT/e generated by the flow should be of the order of the square of the ion Mach number of the flow.) This means that for T=10ev in the solar wind, one may have kilovolt potential differences between the wake and front surfaces. The electric fields due to this differential charging may significantly disturb measurements of low-energy plasma electrons, for example on the HELIOS spacecraft (Rosenbauer, 1973).

B. Photoemission

In the second problem treated here, the bottom surface of the pillbox is assumed to be illuminated by the sun (along the axial direction) and to emit photoelectrons, while the sides and top of the pillbox are dark and non-emitting. (The axial direction of illumination is appropriate for maintaining axial symmetry. In the solar wind the plasma flow and solar illumination are in the same direction.) The ambient plasma contributions are not considered simultaneously with the photoemission. The photoelectrons are assumed to be emitted isotropically and monoenergetically, with 1 ev of kinetic energy. All points and their associated areas on the bottom surface are emitting except for the corner point, for instance, Nos. 10, 11 and 12 in Fig. 2. In the actual problem the "terminator" was put at $R=0.95R_0$. (The terminator is not put exactly at the corner $R=R_0$, for numerical reasons.) The time-dependent method is used, together with the outside-in method discussed by Parker (1978). As zero time is no charge on any surface. As time from zero, emitted photoelectrons face at the terminator, moving behind the bottom surface potential. As this potential value of the order of a volt energy), the photoelectrons no longer all escape to infinity, but begin to return to the spacecraft. (It is of interest that the potential is found to rise significantly above one volt - see below.)

Figure 4 shows the time-behavior of the potential of the center point on the bottom surface (for instance, No. 12 in Fig. 2). The potential in volts is plotted against dimensionless time, where the scale time t_0 in seconds may be written as $1.11\ V_0/(R_0J_0)$, where V_0 is the kinetic energy of emission in volts, R_0 is the scale dimension of the emitting area in cm, and J_0 is the emitted current density in picoamp per cm^2. Thus, for a spacecraft radius R_0=50 cm, photoemission current density J_0=1000 picoamp/cm^2, and emission energy V_0=1 volt, the scale time is $t_0=2.22 \times 10^{-5}$ sec. (The ordinate in Fig. 4 scales with the ejection energy.) The potential along the bottom surface is not uniform, but is maximum in the center and falls off with radius (see Fig. 5, which indicates an approximate drop of 30 percent over the illuminated area). The central potential rises to an approximate plateau of slightly over 2 volts in about one characteristic time t_0, and, after an interval of approximately constant behavior between t_0 and $3t_0$, continues to increase but much more slowly than the initial rate.

It is a curious fact that the potential should rise to more than 2 times the ejected energy potential-equivalent, rather than to exactly the ejected energy as may be expected purely on the basis of conservation of energy without consideration of the nature of the trajectories, surface, and potential distribution. The computer results show a transfer of electrons from the illuminated areas to the small dark area associated with the corner point; that is, emitted electrons are pulled back, but cross the local terminator at $R=0.95R_0$ and hit the dark corner area rather than return to other points on the emitting area. The charge on this corner area continues to increase negatively, while the illuminated-surface charge increases positively. No charges are deposited on the top and side dark surfaces. A similar build-up to more than the ejection energy has been observed in computer experiments performed by De and Criswell (1977) and by Pelizzari and Criswell (1977) in studies of the photoelectric charging of locally-sunlit areas in the dark lunar terminator region. A possible explanation of this "excess" charging phenomenon is given here as follows.

The effect is [illegible] of electrons from the illuminated area of a [illegible] with no compensating [illegible] from an external plasma. It also [illegible] the returning electrons "sticking" where they hit. After the initial potential buildup, despite the deposition of negative charge on the dark side of the terminator, the surface potential falls off monotonically from the central value but remains positive as one goes into the dark region. That is, the surface gradient (tangential electric field) remains finite and continuous across the terminator. (If the emitting area were a conductor, the surface gradient would be discontinuous and singular across the terminator.) Thus, there is a finite interval straddling the terminator, such that within this interval electrons can cross the terminator from a sunlit point (moving "uphill") to a dark point where they are held fast, without using up their kinetic energy. A finite rate of transfer across the terminator from the sunlit area to the dark area can thus occur as long as the surface potential gradient is finite at the position of the terminator, regardless of how high the central potential of the sunlit

area becomes. The transfer rate should approach zero as the gradient at the terminator approaches infinity. Whether this process is self-limiting, that is, whether the gradient at the terminator becomes infinite within a finite time, is presently unknown. The key point is probably that the sunlit area cannot be strictly an equipotential surface. It may be approximately so over most of its area, due to electron transport tending to maintain equipotentiality (De and Criswell, 1977; Pelizzari and Criswell, 1977), but the potential should fall off in the vicinity of the terminator.

Figure 5 shows a few equipotential contours around the spacecraft. (Only half of the spacecraft is shown, since it is symmetric about the axis.) The potential contours are taken from the solution of the foregoing problem, at the time $t=2t_0$. At this time the bottom-surface potential is approximately at its plateau value (Fig. 4) of about 2 volts, and is undergoing its smallest rate of change. The [illegible] field is a nonuniform disk of [illegible] bottom surface, with [illegible] left side of the [illegible] charge [illegible] potentials are [illegible] of the disk since [illegible] constant was assumed [illegible] to the free-space value (no polarization effects).

ACKNOWLEDGMENT

This work was supported by the Air Force Geophysics Laboratory.

REFERENCES

Brundin, C.L., Effects of charged particles on the motion of an earth satellite, AIAA J., 1, 2529, 1963.

De, B.R., and D.R. Criswell, Intense localized photoelectric charging in the lunar sunset terminator region. 1. Development of potentials and fields, J. Geophys. Res., 82, 999, 1977.

Fahleson, U., Plasma-vehicle interactions in space - some aspects of present knowledge and future development, in Photon and Particle Interactions with Surfaces in Space, edited by R.J.L. Grard, p. 563, D. Reidel, Dordrecht, Holland, 1973.

Fredricks, R.W., and F.L. Scarf, Observations of spacecraft charging effects in energetic plasma regions, in Photon and Particle Interactions with Surfaces in Space, edited by R.J.L. Grard, p. 277, D. Reidel, Dordrecht, Holland, 1973.

Grard, R.J.L., editor, Photon and Particle Interactions with Surfaces in Space, D. Reidel, Dordrecht, Holland, 1973.

Parker, L.W., Numerical methods for computing the density of a rarefied gas about a moving object, Allied Research Associates, Inc. Report AFCRL-64-193, 1964.

Parker, L.W., Theory of the external sheath structure and ion collection characteristics of a rocket-borne mass spectrometer, Mt. Auburn Research Associates, Inc. Report AFCRL-71-0105, 1970.

Parker, L.W., Computer solutions in electrostatic probe theory, Mt. Auburn Research Associates, Inc. Report AFAL-TR-72-222, 1973.

Parker, L.W., Computation of collisionless steady-state plasma flow past a charged disk, Lee W. Parker, Inc. Report NASA CR-144159, 1976a.

Parker, L.W., Theory of electron emission effects in symmetric probe and [illegible] sheaths, Lee W. Parker, [illegible] TR-76-0294, 1976b.

Parker, L.W., [illegible] structure [illegible] Proceedings [illegible] Technology Conference, C.P. Pike and R.R. [illegible], Joint Air Force-NASA Report AFGL-TR-77-0051 and NASA TMX-73537, 1977.

Parker, L.W., Potential barriers and asymmetric sheaths due to differential charging of nonconducting spacecraft, Lee W. Parker, Inc. Report AFGL-TR-78-0045, 1978.

Parker, L.W., and E.C. Whipple, Jr., Theory of a satellite electrostatic probe, Ann. Phys., 44, 126, 1967.

Parker, L.W., and E.C. Whipple, Jr., Theory of spacecraft sheath structure, potential, and velocity effects on ion measurements by traps and mass spectrometers, J. Geophys. Res., 75, 4720, 1970.

Pelizzari, M., and D.R. Criswell, Patchy photoelectric charging of the lunar terminator terrain, EOS Trans. Amer. Geophys. Union, 58, 1180, 1977.

Pike, C.P., and R.R. Lovell, editors, Proceedings of the Spacecraft Charging Technology Conference, Joint Air Force-NASA Report AFGL-TR-77-0051 and NASA TMX-73537, 1977.

Rothwell, P.L., A.G. Rubin, and G.K. Yates, in Proceedings of the Spacecraft Charging Technology Conference, edited by C.P. Pike and R.R. Lovell, p. 389, Joint Air Force-NASA Report AFGL-TR-77-0051 and NASA TMX-73537, 1977.

Rosen, A., editor, Spacecraft Charging by Magnetospheric Plasmas, Vol. 47, Progress in Astronautics and Aeronautics, AIAA-MIT Press, 1976.

Rosenbauer, H.R., Possible effects of photoelectron emission on a low energy electron experiment, in Photon and Particle Interactions with Surfaces in Space, edited by R.J.L. Grard, p. 139, D. Reidel, Dordrecht, Holland, 1973.

Samir, U., and A.P. Willmore, The equilibrium potential of a spacecraft in the ionosphere, Planet. Space Sci., 14, 1131, 1966.

Schröder, H., Spherically symmetric model of the photoelectron sheath for moderately large plasma Debye lengths, in Photon and Particle Interactions with Surfaces in Space,

edited by R.J.L. Grard, p. 51, D. Reidel, Dordrecht, Holland, 1973.
Soop, M., Report on photosheath calculations for the satellite GEOS, Planet [illegible] 20, 859, 1972.
Soop, M., Numerical [illegible] turbat[illegible] [illegible] edited [illegible] Reidel, Dor[illegible]
[illegible]. The equilibrium electric [illegible]tential of a body in the upper atmosphere and in interplanetary space, Ph.D. thesis, George Washington University, 1965. (Also NASA Tech. Note X-615-65-296, 1965).
[illegible] [illegible] sheath [illegible] with the ATS-6 [illegible] potential barrier, J. Geo[illegible], 601, 1976a.
[illegible], Jr., Observation of photoelectrons and secondary electrons reflected from a potential barrier in the vicinity of ATS-6, J. Geophys. Res., 81, 715, 1976b.
Whipple, E.C., Jr., and L.W. Parker, Effects of secondary electron emission on electron trap measurements in the magnetosphere and solar wind, J. Geophys. Res., 74, 5763, 1969.

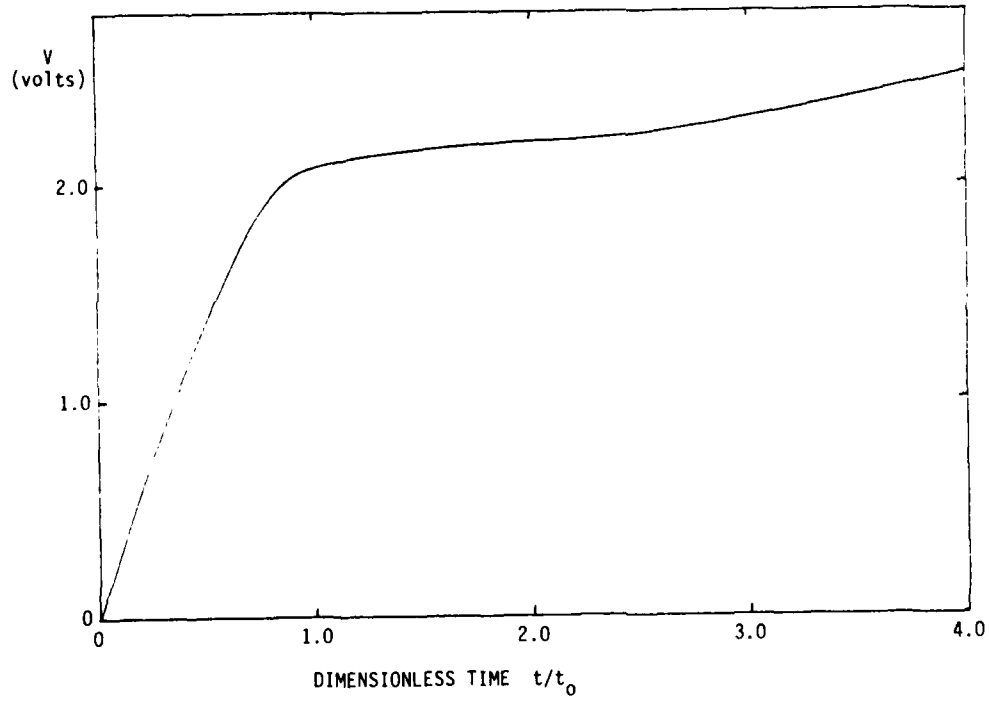


FIG. 4. POTENTIAL OF CENTER POINT VERSUS DIMENSIONLESS TIME

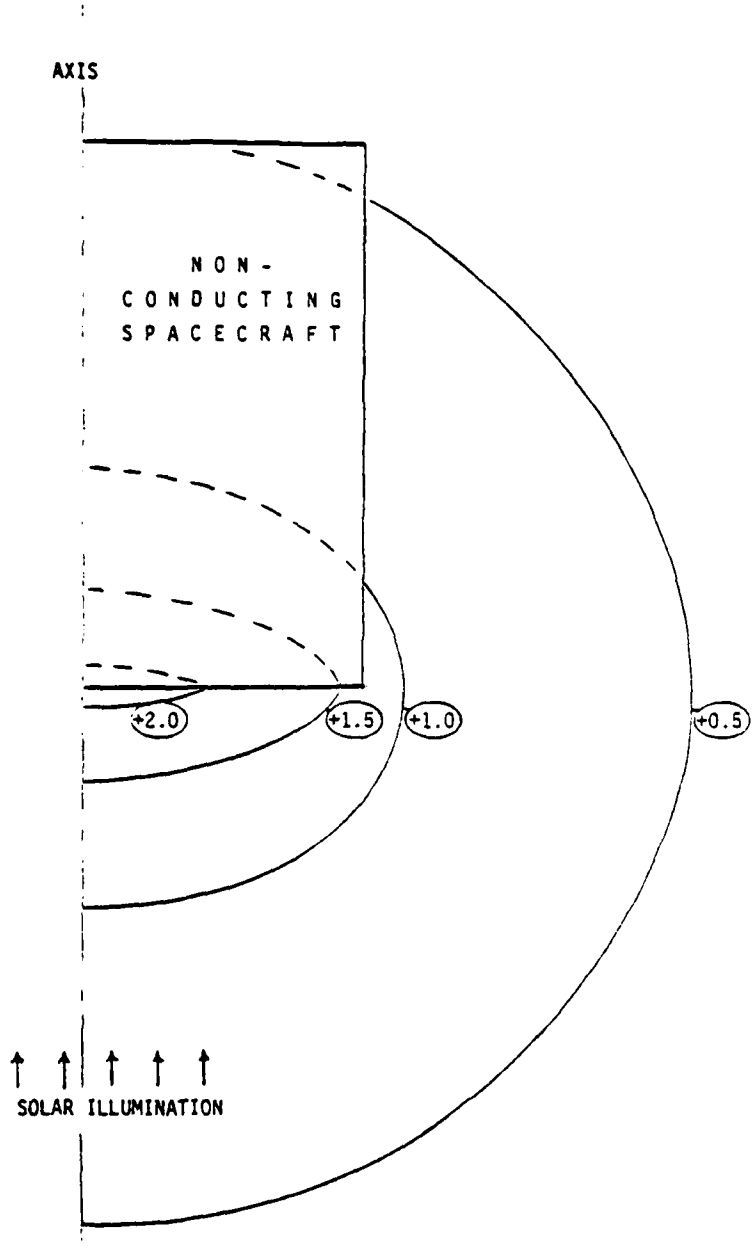


FIG. 5. DIFFERENTIAL CHARGING OF NONCONDUCTING SPACECRAFT BY PHOTOEMISSION - EQUIPOTENTIAL CONTOURS IN UNITS OF EMISSION KINETIC ENERGY - time=$2t_o$

ON THE PERTURBED ENVIRONMENT OF METALLIC BODIES IN SPACE PLASMAS

Jean - Pierre J. Lafon
Observatoire de Meudon
Department Recherches Spatiales
92190 - Meudon (France)

The exploration of the terrestrial environment (from the ionosphere to the solar wind) requires an accurate theory of the behaviour of a biased conductor in a collisionless plasma: a strongly self-consistent and non linear inhomogeneous sheath grows around the body. This sheath depends strongly on the potential of the body, which is self-consistent with the sheath structure when it is not fixed (floating potential). In any case the sheath governs the collected currents and even, if the body is an antenna, it modifies its characteristics of wave emission. Indeed, ignoring these phenomena may lead to erroneous interpretations of measurements in space.

The structure of the sheath and the collected currents have been investigated in a number of particular cases by various authors (a review is given if ref. 1) using different (physical or mathematical) approximations depending on the cases investigated. For cylindrical and spherical conductors in non magnetized plasmas the most elaborate (but also the most complicated theories are those developed by Bernstein and Rabinowitz,[2] Moskalenko,[3] and Laframboise[4]. These theories are concerned with plasmas with only two components (ions and electrons); moreover it is assumed that the surface of the conductor absorbs all particles striking it, i.e., the surface effects are neglected. However a surface effect such as electron photoemission should not be neglected in some cases; it may even be dominant. Other effects such as the reflection of some of the particles striking the conductor or the emission of some particles by the conductor may also be important. Besides there are few works concerning the case of magnetized plasmas and in this case the approximations used are rather crude[1].

The equations governing the sheath structure are a VLasov equation for the distribution function F_s of the particles of each species s together with Poisson's equation which determines the electric potential in terms of the F_s. Even in spherical or cylindrical symmetry the solution of these equations is a difficult problem because they are strongly non linear and self - consistent.

Without crude approximations they can be solved only using numerical iteration in which the main problem is always the classification of the particle orbits. New we have found a formalism in which the problem appears to be a particular case of a more general problem which consists in the analysis of some systems satisfying continuous sets of conditions in some continuous sets of states [1, 5]. We have proved a general theorem and we have derived from it a systematic method for solving this problem. The method is well fit for numerical computations and even reduces them to the minimum when they are necessary. In particular it is a powerful tool for solving the sheath problem for conductors absorbing, emitting or partially reflecting charged particles, even in some magnetized plasmas, for arbitrary theoretical or experimental models of polyenergetic or menoenergetic particle distribution functions.

Of course, strictly speaking absolute self - consistence requires complete radial (cylindrical or spherical) symmetry for exact solution. However, as a property of Poisson's equation (which smoothes the space

charge irregularities), rather strong asymmetries can be taken into account with very good accuracy (see fig. 3 to 8 and ref. 8 for instance), provided that they do not disturb too much the symmetry of the potential (which is the only thing truly necessary).

Using our new method we have investigated the behaviour of metallic bodies reflecting some of the particles striking them (rather different effects are produced depending on the polarization of the body)[6,7] and that of metallic bodies emitting photoelectrons in space plasmas[8] for which experimental data are now available[9] (space experiments GEOS and ISEE). We have also investigated the case of cylindrical bodies in some weakly magnetized plasmas [10]. Of course it is not possible to discuss the results here; however they are discussed in detail in the above mentioned publications[1,5,6,7,8,10], and the figures shown hereafter illustrate some characteristics of the sheath structure under various circumstances.

The field of applications of the theorem and the method is of course not limited to the investigation of plasma sheath around metallic bodies. In particular they can be used for investigating all plasma sheath problems which can be stated under a similar form (for instance those concerning steady state double layers...).

REFERENCES

LAFON J. - P. J. - Etude d'une classe de systems obeissant a un ensemble continu de conditions dans un ensemble continu d'etats. These de Doctorat d'Etat es Sciences. Universite de Paris VII. 1976. Rpt. Observatorire de Meudon DESPA 159 bis.

BERNSTEIN I. B. and RABINOWITZ I. N. 1959, Phys. Fluids, 2, 112-121.

MOSKALENKO A.M. - 1064, XVth Int. Astron. Cong., Warsaw, Poland, Sept. 7-12.

LAFRAMBOISE J. - Theory of Cylindrical and Spherical Langmuir Probes in a Collisionless Maxwellian Plasma at Rest. Rpt n^{o}100 of the Institute for Aerospace Studies of the University of Toronto, Canada (UTIAS 100), 1966.

LAFON J.-P.J. - 1977, Journal of Mathematical Physics, 18, 1178-1187.

LAFON J.-P.J. - 1975, Plasma Phys., 17, 741-740 and 1175.

LAFON J.-P.J. - 1975, Plasma Phys., 17, 741-756.

LAFON J.-P.J. - 1976, Radio Science, 11, 483-493.

GRARD R.J.L. - 1973, J. Geophys. Res., 78, 2885-2906.

LAFON J.-P.J., 1973, J. Of Plasma Phys. 10, 383-396.

Distribution function of the particles that leave a metallic body in a plasma :

B_e denotes the distribution function of the emitted particles (photoemission, secondary emission, thermoemission ...)

γB_d denotes the distribution function of the diffusely reflected particles (it depends on that of the particles striking the body)

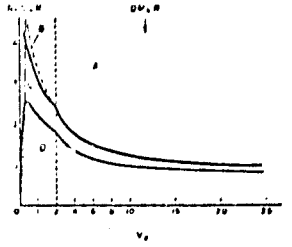
γ denotes the coefficient of specular reflection

$$\gamma[\Pi_{x,v}] + \gamma\alpha_d[\Pi_{x,v}] + \gamma\alpha_d[\Pi_{x_m,v}] + \gamma\alpha_d\gamma[\Pi_{x_m,v}] + \gamma\alpha_d\gamma[\Pi_{x_{mm},v}] + \gamma\alpha_d\gamma[\Pi_{x_{mm},v}] + \alpha_e[\Pi_{x_{mm},v}]$$

- Particles specularly reflected after being emitted by the probe
- Particles specularly reflected after being diffusely reflected
- Particles coming from the sheath edge and specularly reflected by the probe
- Particles diffusely reflected by the probe
- Particles emitted by the probe

Square brackets denote a function equal to 1, when ξ, E are the co-ordinates of a point of the region denoted between the brackets, and equal to zero otherwise.

Figure 1



Effect of partial reflection (reflection coefficient γ) of some of the particles striking a cylindrical body. $T_{ions} = T_{electrons} = T$, p, V_p, q denote respectively the temperature of the particles, the radius of the body (in Debye lengths), the potential of the body (in kT) and the charge of a particle; index a for particles such that $q_a V_p < 0$, b for those for which $q_b V_p > 0$.
The current collected is OML (Orbital Motion Limited) i.e. independent of the sheath structure in the regions indicated in the following table :

particles a	C + D	D	B + C + D
particles b	everywhere	everywhere	OML
	$\gamma_a = 0$	$\gamma_a \neq 0$	$\gamma_a = 0$
	$\gamma_b = 0$	$\gamma_b = 0$	$\gamma_b \neq 0$

Figure 2



Figure 3

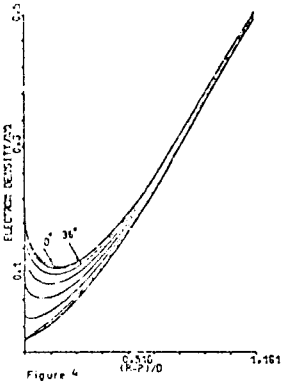


Figure 4

Figs 3 and 4 - Spherical body emitting photoelectrons in a plasma. Electron density profiles in planes through the center of the sphere and at a regularly increasing angle (18° /step) with the direction of the ligth.($T_{electrons} = T_{ions} = T$, radius of the body = Debye length)

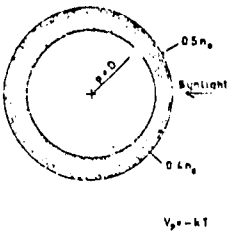


Figure 5

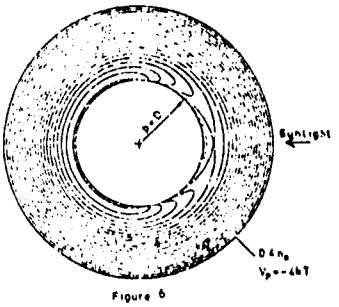


Figure 6

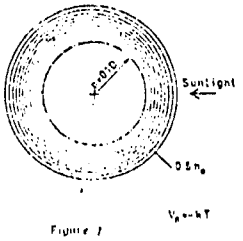


Figure 7

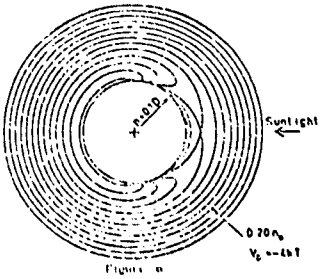
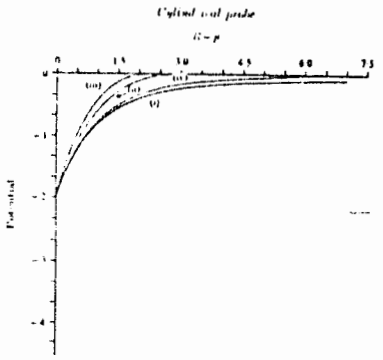


Figure 8

Figs 5 to 8 - Spherical body emitting photoelectrons in a plasma. Iso - electron density curves; $T_{ions} = T_{electrons} = T$ $T_{photoelectrons} = 5T$; Radius of the body = p (in Debye lengths)



Contraction of a cylindrical sheath by an axial magnetic field. Curves i,iv,ii,iii correspond respectively to the values 0.,0.1, 0.6,2. of the ratio (Debye length/larmor radius); (T_{ions} = $T_{electrons}$; radius of the body = Debye length)

Figure 9

VOYAGER SPACECRAFT CHARGING MODEL CALCULATIONS

N. L. Sanders and G. T. Inouye
TRW Defense & Space Systems Group
Redondo Beach, California

INTRODUCTION

The large fluxes of energetic electrons and the possibility of a relatively low cold plasma density within 10 R_J of Jupiter imply that a spacecraft flying by the planet can charge up to very high, multikilovolt, potentials. There is some evidence that the Pioneer Jupiter spacecraft could have charged up significantly, particularly in the wake of the Jovian satellite, Io. In that case, several spacecraft anomalies such as false commands occurred near Jupiter.

The present study was undertaken as part of an effort by the Jet Propulsion Laboratory (JPL) to immunize the Voyager spacecraft from deleterious effects of large voltage stresses between different components that could result from differential chargeup near Jupiter. In the part of the study which will be discussed in this paper, a spacecraft charging model for the Voyager spacecraft was generated. Both the spacecraft potential and the electrostatic stress across several different dielectrics in the Jupiter radiation belts were analyzed. Figure 1 shows the Voyager spacecraft and some of the components discussed in this paper are identified.

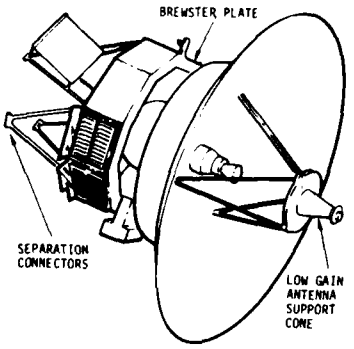


Fig. 1. Voyager Spacecraft

The test program used to verify the immunization of Voyager to charging effects at Jupiter is described in a companion paper.[1]

An important aspect of the present work was the examination of the effect of changes in several of the parameters involved in the computations. The results obtained using two different Jupiter environmental models were compared. Furthermore the effect on the results caused by varying some of the relevant material parameters over reasonable ranges of values were examined.

MODELING OF THE VOYAGER FOR SPACECRAFT CHARGING

A spacecraft charging model for Voyager was generated. The spacecraft charging model, shown in Fig. 2, is an electrical circuit representation of the spacecraft from which electrostatic stresses across individual components may be calculated. In the model various dielectric

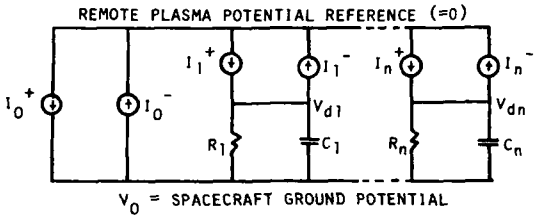


Fig. 2. Spacecraft Charging Model

surfaces are represented by capacitances, C_i, and leakage resistances, R_i. The spacecraft ground potential, V_0, and the various dielectric surface potentials, V_{di}, are obtained by solving the multinode network equations for current continuity. The charging and discharging current generators, I_i^+ and I_i^-, connected from the remote zero potential reference to spacecraft ground and to each dielectric surface, are defined by the plasma parameters, by the surface material properties, and the effective exposure areas of each surface to the plasma and to solar UV. Each dielectric area which is expected to be charged uniformly and to form an equipotential surface must be considered as a separate capacitor.

Metallic Surfaces

For Voyager, because the metallic areas are an overwhelmingly large proportion of the total surface, the spacecraft potential, V_o, is determined directly by I_o^+ and I_o^- and is influenced negligibly by the remaining currents. Thus, each dielectric surface potential, V_{di}, may be computed directly once V_o is known, rather than by solving a set of simultaneous equations. The electrostatic stress across a given dielectric is, then,

$$V_{stress} = V_d - V_o.$$

The Voyager geometrical factors for current collection were generated from examination of the Proof Test Model (PTM) and Flight 1 Voyager spacecraft and reference to many photographs and discussions with JPL personnel. Table 1 is a summary of the effective areas of metallic surfaces for plasma collection and for solar UV. The actual area of each individual component is modified by a factor to account for its "acceptance angle" or its "field-of-view" of the ambient plasma if it is less than 2π steradians. For irradiation by solar UV, the projection of each area onto the plane perpendicular to the sun line was taken to be its effective area.

Table 1. Total Effective Area of Conducting Surfaces for Exposure to UV and to Omnidirectional Plasma

	EFFECTIVE ILLUMINATED AREA (cm^2)			
	PLASMA		UV	
1. ANTENNAS				
A. PLASMA RADIO ASTRONOMY (PRA)	12,600		392	
B. HIGH GAIN ANTENNA ASSEMBLY	185,600		106,000	
TOTAL		198,200		109,400
2. SCIENCE BOOM AND EXPERIMENTS				
A. SCIENCE BOOM	9,400		1,660	
B. EXPERIMENTS	16,200		4,040	
TOTAL		25,600		5,700
3. MISSION MODULE AND ATTACHMENTS				
A. MISSION MODULE	72,500		0	
B. SUPPORT STRUTS	12,000		0	
C. SCIENCE CALIBRATION TARGET/ THERMAL RADIATOR	5,760		0	
TOTAL		90,300		0
4. RTG'S AND BOOM, MAGNETOMETER BOOM CANISTER				
A. RTG UNITS (3)	24,200		6,930	
B. RTG BOOM	10,500		[illegible]	
C. MAGNETOMETER BOOM CANISTER	[illegible]860			
D. PHOSPHOR BRONZE PORTION OF LANYARD	[illegible]			
[illegible]TAL	[illegible]			

[illegible] than it [illegible]. Note that [illegible], the A_{UV} is 50% of [illegible] the estimated values for Voyager [illegible] nearly midway between the disc and the sphere. Taking the diameter of the equivalent spacecraft, i.e., a sphere with equal A_{plasma}, its capacitance to space may be calculated as:

$$C = 2\pi\varepsilon_o D(MKS) = 186 \text{ pf.}$$

All Voyager thermal blankets were coated with a black Sheldahl conducting paint. In addition to the black Sheldahl coated thermal blankets, all painted surfaces were considered to be conducting if over a conducting surface. A thin paint layer over a conductor forms as good a conductor as Sheldahl over an insulator. The resistance of Sheldahl coating, measured to be about 10^6 ohm per square, corresponds to a resistivity of 10^3 ohm-cm if the coating thickness is assumed to be 0.4 mil. A square meter of surface at 10^{-9} amps/cm^2 would collect 10^{-5} amperes, and this current would give only a 10 volt drop across 10^6 ohm. Considering a square centimeter of 4 mil thickness of painted surface over a conductor, its bulk resistance is 10^{10} ohm ($\rho \sim 10^{12}$ ohm-cm). A plasma charging current of 10^{-9} amperes, again, would cause only a 10 volt drop across this resistance. Since the graphite epoxy material of the large high gain antenna dish has a resistivity in the order of 1 ohm-cm as compared with 10^3 ohm-cm for Sheldahl, the white paint (less than 4 mil thick) over it may be considered to be metallic for spacecraft charging purposes.

Dielectric Surfaces

For Voyager, with the elimination of solar cells by the use of the Radioisotope Thermal Generator (RTG) power sources and the extensive use of conductive Sheldahl-coated thermal blankets, only an extremely small portion of the external surface area remained under the category of dielectrics.

Table 2 is a summary of Voyager dielectric area parameters. Thirteen different items, some multiple, are listed. For each, a capacitance, to spacecraft ground, and a leakage resistance

Table 2. Summary of Dielectric Area Parameters

1 [illegible]				
[illegible]			2,440	588
[illegible]	14	$4 \cdot 10^{16}$	64	16
INSULATING COLLARS AROUND PLASMA RADIO ASTRONOMY (PRA) ANTENNAS (2, EACH)	8.6	$4.2 \cdot 10^{16}$	41	0
10. SEPARATION CONNECTORS (6; EACH)	4.05	$6.6 \cdot 10^{14}$	14.5	0
11. TEFLON RINGS ON LOW ENERGY CHARGED PARTICLES (LECP)	20.9	$1.27 \cdot 10^{16}$	200	75
12. MAGNETOMETER BOOM SECTION (90; EACH)	4.6	$1.57 \cdot 10^{17}$	205	14.9
13. MAGNETOMETER BOOM LANYARD SECTION (90; EACH)	1.87	$2.8 \cdot 10^{18}$	32	8.9

are estimated. In addition, effective areas for plasma charging and solar UV irradiation, are shown. These are corrected in the same manner as for metallic surfaces, for field of view and for solar exposure.

SPACECRAFT CHARGING ANALYSIS

Spacecraft Potential

The equation used for determining the spacecraft potential V_o, was the single node current continuity equation

$$(J_e - J_{bs} - J_{se} - J_i - J_{si})A_p - J_{uv} \cdot A_{uv} = 0.$$

The above equation defines the current sign convention such that the accumulation of negative charge is positive. The terms are defined as follows:

J_e = plasma electron current collected per unit area

J_{bs} = backscattered electron current density

J_{se} = emitted secondary current density due to impinging electrons

J_i = plasma ion current collected per unit area

J_{si} = emitted secondary electron current density due to impinging ions

A_p = effective surface area [illegible] collecting plasma [illegible]

J_{uv} = electron [illegible]

[illegible] spacecraft

[illegible] current terms are defined in [illegible] spacecraft potential and discussed [illegible] final report for the JPL study.[2] The equations used were approximations to the current density to and from the spacecraft for a plasma having a Maxwellian energy distribution. For example, the electron and ion current terms used,

$$J_e = J_{oe}\, e^{V/Te} \text{ and } J_i = J_{oi}\, e^{-V/Ti}$$

in the "repulsive" case, are correct for any shape probe in a thermal plasma regardless of the thickness of the plasma sheath. However, the terms

$$J_i = J_{oi}\left(1 - \frac{V}{T_i}\right) \text{ and } J_e = J_{oe}\left(1 + \frac{V}{T_e}\right)$$

which were used in the attractive case are thick sheath approximations for a spherical probe. Furthermore the real environment was approximated by multiple Maxwellian distributions. The validity of this approximation is discussed in the section on environment.

Isolated Dielectrics

On Voyager, the major portion of the spacecraft exterior surface is conductive and is at [illegible] spacecraft potential, V_o. Because of the [illegible] of dielectric surface areas, [illegible] assuming that V_o is known rather than [illegible] large number of simultaneous equations [illegible] current continuity equation for each [illegible] dielectric, then, is

$$(J_e - J_{bs} - J_{se} - J_i - J_{si}) \cdot [\text{illegible}] = \left(\frac{1}{R} + C\frac{d}{dt}\right) \cdot \left(V_o - V_d\right)$$

The term on the right side of the equation includes the leakage current due to the bulk resistance, R, and the displacement current due to the capacitance, C, of the surface to spacecraft ground. Both of these terms depend on the difference between the dielectric surface potential, V_d, and V_o. All of the terms on the left side of the equation depend on V_d rather than on V_o.

ENVIRONMENT

In order [illegible] the particle [illegible] sources [illegible] distributions. [illegible] characterization of [illegible] in terms of Maxwellian [illegible] required that three temperature ranges, cold (eV), medium (keV) and hot (MeV) be used.

Thus each plasma current term in the current continuity equation is composed of the sum of three components, cold, medium, and hot for which there are corresponding temperatures, i.e.,

$$J_e = J_{oe(cold)}\, e^{\frac{V}{Te(cold)}} + J_{oe(medium)}\, e^{\frac{V}{Te(medium)}} + J_{oe(hot)}\, e^{\frac{V}{Te(hot)}}$$

The plasma parameters for the cold plasma distribution were derived from NASA JPL/ARC Document No. 660-24, "Jupiter Charged Particle Environment for Jupiter Orbiter Probe 1981/1982 Mission." These parameters are primarily based on Pioneer 10 and 11 plasma probe observations.[3,4] These results have been difficult to extract from the data due to penetration of high energy particles into the instrument and also because of possible spacecraft-plasma interactions. As a result, the data leading to the NASA 660-24 cold plasma model has generated a significant amount of controversy.[5,6] For

this reason a second cold plasma model based on plasma densities derived by Axford and Mendis[7] was also included. In that case we have used the assumptions of Scarf[8] that the ion and electron temperatures are equal at 10 eV and that the plasma density is reduced in the wake of Io by three orders of magnitude. A comparison of the two models is shown in Figs. 3 and 4.

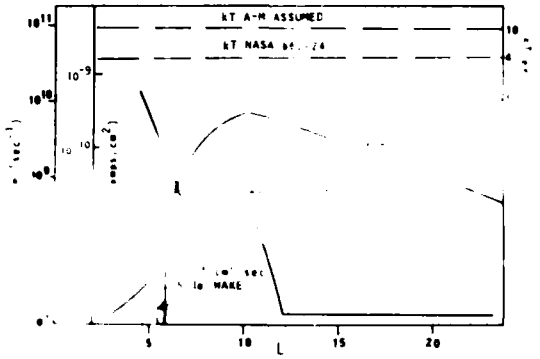


Fig. 3. Jupiter Cold Plasma Electron Parameters

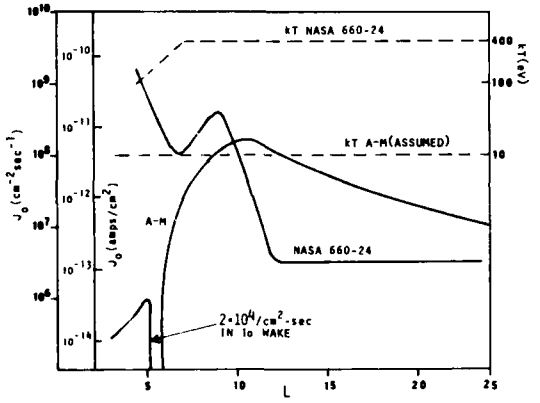


Fig. 4. Jupiter Cold Plasma Ion Parameters

For the medium and hot ranges, we have fit the Jupiter high energy particle fluxes as given in NASA 660-24 with the sum of two Maxwellian distributions. The resulting plasma parameters for the medium and high energy electrons and ions are shown in Figs. 5 and 6. The adequacy of this approximation is shown in Figs. 7 and 8. In those figures, points computed for the high energy flux using the double Maxwellian approximation are superimposed on the curves from NASA 660-24.

Since the Voyager remains within ±15° of the magnetic equator, corrections to the currents used due to latitude effects were relatively small and were neglected.

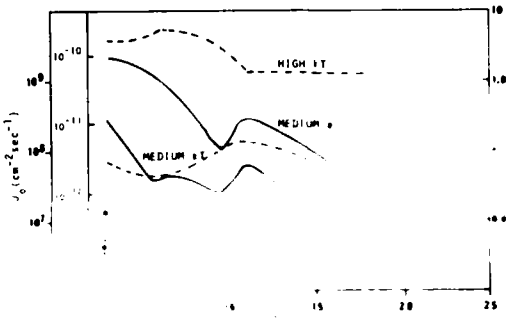


Fig. 5. Jupiter Medium and High Energy Electron Parameters

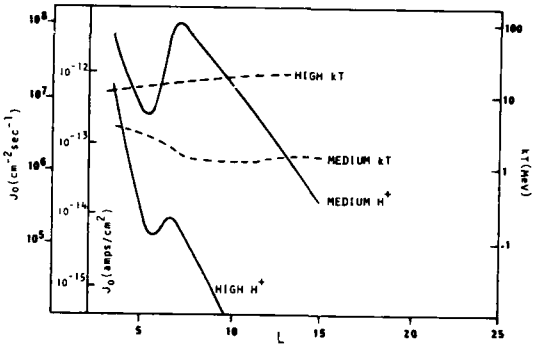


Fig. 6. Jupiter Medium and High Energy Proton Parameters

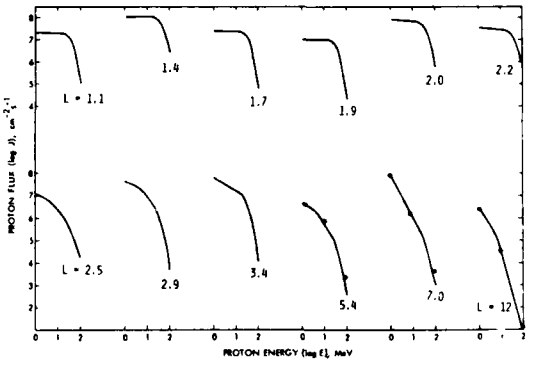


Fig. 7. Equatorial Proton Energy Spectra. Data Points show Double Maxwellian Approximation.

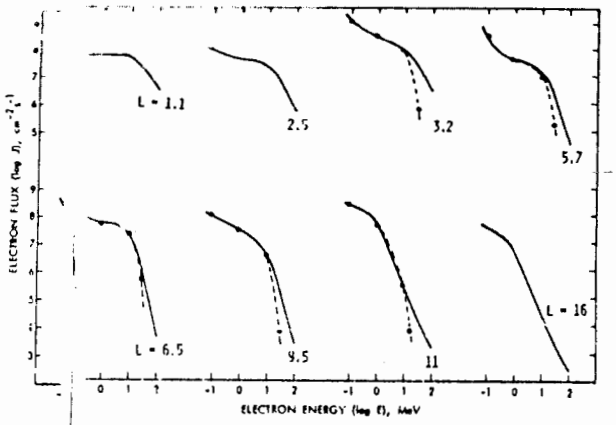


Fig. 8. Equatorial Electron Energy Spectra. Data Points show Double Maxwellian Approximation.

COMPUTATIONS

The computations were performed by means of a small Hewlett Packard 9830 calculator programmed with an iterative technique using human interaction. The details of each current expression used in the computations are given in the final study report to JPL.[2]

The spacecraft ground potential as well as the potential and stress on three different dielectrics were computed for the Voyager "near" approach to Jupiter. In these analyses, the Voyager JST trajectory[9] was used. The three dielectrics chosen were the Brewster Plate, the Separation Connector and the Low Gain Antenna Support Cone. These were selected because the first two are always in shadow and the other is always in sunlight. For the dielectric stress computations, static, time independent worst case analyses were performed. Furthermore, a parameter analysis was performed to indicate the effect of the various parameters involved in the computations and to underscore the areas of uncertainty in the results. The effects of the following parameters were examined by changing the parameter values over the ranges indicated:

1. The photoelectric current at Jupiter J_{ouv} was varied from 11.1 to 111 pa/cm^2. This corresponds to 0.3 to 3 na/cm^2 at the earth, corrected for distance by the inverse square dropoff. These are reasonable values of photoelectron currents and are consistent with the range of photoelectron currents measured in the laboratory by Grard.[10]

2. The resistivity, ρ, of the Brewster Plate was varied from 10^{16} to 10^{18} ohm-cm.

3. The secondary emission factor, S_i, for incident ions, i.e., the secondary emission current per unit of incident ion current, was varied from 0.25 to 4.

Furthermore, a comparison is made of the results obtained using the cold plasma specified in NASA 660-24 with those obtained using the Axford-Mendix (A-M) cold plasma density profiles.[7]

RESULTS

Spacecraft Ground Potential

The potential on the conducting portion of the spacecraft is shown in Fig. 9. The results

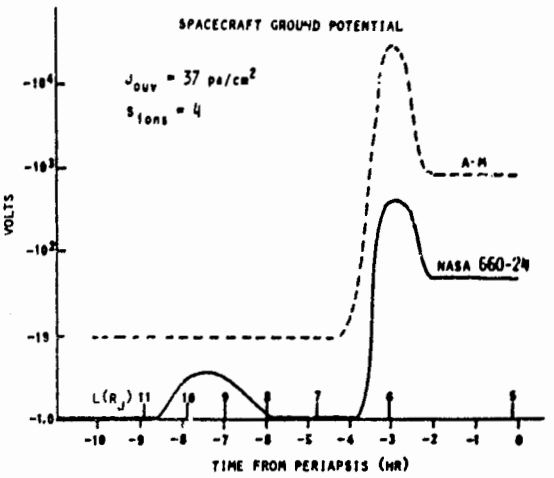


Fig. 9. Voyager Charging for JST Trajectory Jupiter Approach

shown are for J_{ouv} = 37 pa/cm^2 and S_i = 4. The conducting area of the spacecraft is shown in Table 1. The area of the surface exposed to the sun, A_{uv}, is 123,200 cm^2 and the area exposed to the plasma, A_{plasma}, is 351,700 cm^2. We note that for both NASA 660-24 and A-M cold plasma the spacecraft potential reaches a maximum negative value at the L shells corresponding to the Io flux tube. In the case of the A-M model the spacecraft potential is estimated to be 24,000 volts negative, and in the NASA 660-24 model we estimated a potential of 700 volts negative. The details of the potential vs time (or L value) curve can be understood by examining the cold plasma electron parameters given in Fig. 3. In the A-M case J for electrons is greater than J_{ouv} for all L values greater than L = 6 so that the spacecraft remains negative. This is not the case in the NASA 660-24 model. The increase in spacecraft charge at the low L values is due to the dominance of the medium energy electron current over the photoelectron current. The charging is more severe with A-M since the medium energy electron plasma dominates the cold plasma as well.

Brewster Plate Stress

The stress on the Brewster Plate, i.e., the difference in potential between the Brewster Plate and the spacecraft ground potential was computed (see Fig. 10). Since the Brewster Plate is always in shadow during Jupiter encounter A_{uv} = 0. The area exposed to the plasma

BEST AVAILABLE COPY

was determined to be A_{plasma} = 240 cm^2 and, R, the resistance of the Brewster Plate was estimated to be 9.11 x 10^{11} ohm. The effect of resistance on the stress is examined in the next section.

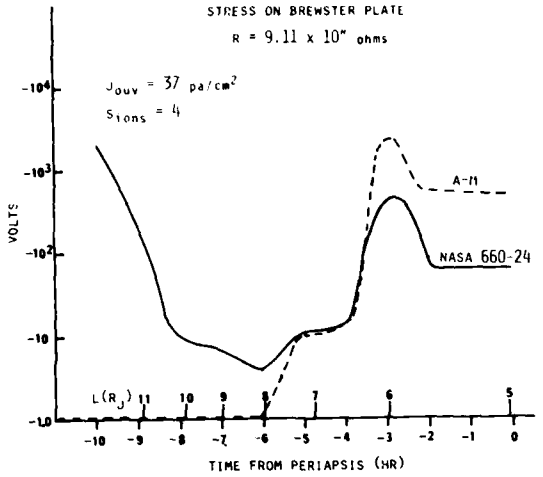


Fig. 10. Brewster Plate Charging for JST Trajectory Jupiter Approach

The NASA 660-24 model predicts the largest stress at high L values reaching 2000 volts negative at t = -10 hr, whereas in the A-M model the stress is very small at high altitudes. This is due to the fact that in the A-M model, the cold electron current is almost an order of magnitude greater than the photoelectron current at these high L values, and therefore the photoelectron cutoff resulting from the shadowing of the Brewster Plate has little effect on the potential. In the case of NASA 660-24 the presence of the photoelectron current keeps the spacecraft ground clamped and its absence in the case of the shadowed dielectric results in a large negative chargeup.

At low L values the stress computed in both models is qualitatively similar reaching a maximum of minus 2400 volts for A-M and minus 600 volts for NASA 660-24.

In the case where J_{ouv} = 37 pa/cm^2 and R = 9.11 x 10^{11} ohm the potential difference between the Brewster Plate and the spacecraft ground is never sufficiently large to produce breakdown electric fields. This would require about 5000 volts across the 2 mil thick Brewster Plate if we assume a breakdown electric field of 10^6 volts/cm.

Parameter Study of Brewster Plate Stress

The results obtained in the previous section are very sensitive to the choice of resistivity of the material as well as the choice of photoelectron current. We have examined the effects of these parameters on the Brewster Plate stress at L = 5.9 and t = -3 hr, i.e., the time of encounter with the Io flux tube. Figure 11 shows that stress can vary by as much as two orders of magnitude by varying the resistance over two orders and by almost one order of magnitude over the range of photoelectron currents studied. In Fig. 12 we note that the effect of S_i, the secondary emission factor for ions, on the stress is less significant.

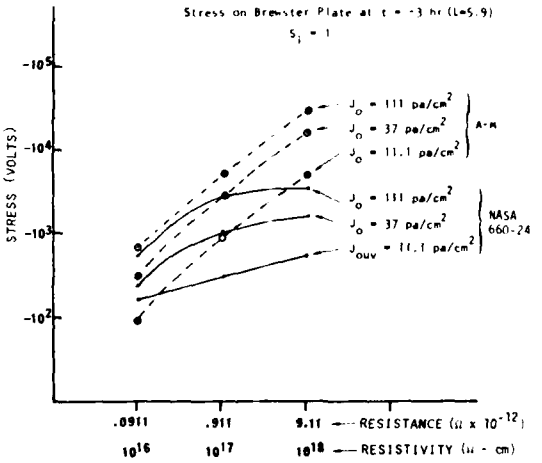


Fig. 11. Effect of Resistivity and Photoemission Current on Brewster Plate Stress

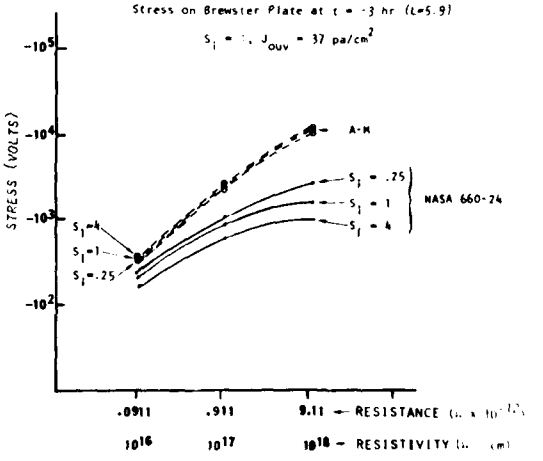


Fig. 12. Effect of Secondary Emission Factor for Ions on Brewster Plate Stress

A summary of the results of the parameter study is shown in Fig. 13. In that figure those values of Brewster Plate stress sufficiently large to cause a breakdown are shown shaded. It is clear from the matrix that the A-M model predicts that breakdowns are likely at t = -3 hr if the resistivity of the Brewster Plate is about

J_{ouv} \ S_{tons}	$\rho = 10^{16}$ ohm-cm, R = 9.11·10^10 ohm: 0.25	1	4	$\rho = 10^{17}$ ohm-cm, R = 9.11·10^11 ohm: 0.25	1	4	$\rho = 10^{18}$ ohm-cm, R = 9.11·10^12 ohm: 0.25	1	4
11.1	* 73	78	170	740	750	700	6,400	6,000	5,300
	** 180	170	54	380	300	170	720	480	210
37	* 260	260	410	2,500	2,400	2,400	17,000	16,000	14,000
	** 240	230	190	1,300	1,100	590	2,500	1,700	850
111	* 630	620	570	5,600	5,500	5,000	50,000	30,000	25,000
	** 560	520	370	3,200	2,500	1,100	6,000	3,900	1,400

*A-M Stress
**NASA 660-24 Stress
All stresses are negative voltages.

$S_{tons}J_i$ = CURRENT OF SECONDARY ELECTRONICS PRODUCED BY THE ION CURRENT J_i INCIDENT ON A NEGATIVE SURFACE.

J_{ouv} = PHOTOELECTRON CURRENT DENSITY IN pa/cm^2.

ρ = RESISTIVITY OF BREWSTER PLATE.

Plasma Parameters at t = -3hr (L=5.9)	J_e (na/cm^2)	kT_e (eV)	J_i (pa/cm^2)	kT - (eV)
NASA 660-24 Cold	40	4	0.93	300
A-M Cold	8.8×10^{-3}	10	2.1×10^{-6}	10
Medium	37	4.3×10^{5}	0.22	2.3×10^{6}
Hot	1.56	4.6×10^{5}	.05	16×10^{6}

Fig. 13. Parameter Study of Brewster Plate Stress (volts) at t = -3 Hours

10^{18} ohm-cm, or if the photoelectron current is sufficiently high and the resistivity is 10^{17} ohm-cm. The NASA 660-24 model only predicts a discharge under the most unfavorable set of circumstances, i.e., $\rho = 10^{18}$ ohm-cm, $S_i = 0.25$ and $J_{ouv} = 111$ pa/cm^2.

Since it is very unlikely that the resistivity of the Brewster Plate is greater than 10^{16} ohm-cm both models indicate that no arc discharges should be anticipated.

Low Gain Antenna Support Cone (LGASC) Stress

The low gain antenna support cone (LGASC) is located in the center of the main dish and a significant fraction of its surface is exposed to the sun. It is shown in Fig. 3 that its effective area for photoelectron current is A_{uv} = 261 cm^2 and for exposure to the plasma is A_{plasma} = 540 cm^2. It is also shown that as a result of its length, even for a resistivity as low as 10^{16} ohm-cm, the resistance of the support cone is 1.3×10^{16} ohms to the spacecraft structure. For this reason the support cone can charge up to relatively high voltages before leakage currents become significant.

The stress on the LGASC three hours before periapsis is shown in the matrix in Fig. 14 for $R = 1.3 \times 10^{16}$ ohm. We note that in this case the stress is most sensitive to the cold plasma model and the amount of photoelectron current. For J_{ouv} = 111 pa/cm^2 (the highest "reasonable" photoelectron current), both the spacecraft and the LGASC are clamped regardless of the cold plasma model used. However, for smaller photoelectron currents, the stress can become as high as +13,000 volts in the case of A-M cold plasma and as high as -1,100 volts for NASA 660-24. Although the electric field across the length of the LGASC is not sufficiently large to predict an arc discharge directly to spacecraft structure, the high potential differences predicted by the A-M model could be a source of discharges to cables which might come in close proximity to the LGASC.

$R = 1.3 \times 10^{16}$ ohm, $\rho = 10^{16}$ ohm-cm

J_{ouv} \ S_{tons}	0.25	1	4
11.1	* +10,000	+7,200	+4,400
	** -1,100	-730	-250
37	* +13,000	+12,000	+9,000
	** +430	+300	+530
111	* 0	0	0
	** +5	+4	+3

$S_{tons}J_i$ = CURRENT OF SECONDARY ELECTRONS PRODUCED BY THE ION CURRENT J_i INCIDENT ON A NEGATIVE SURFACE.

J_{ouv} = PHOTOELECTRON CURRENT DENSITY IN pa/cm^2.

ρ = RESISTIVITY OF LGASC

R = RESISTANCE OF LGASC

*A-M Stress
**NASA 660-24 stress
Plasma parameters at t = -3 hr (L=5.9)

Fig. 14. Parameter Study of LGASC Stress (volts) at t = -3 Hours

The Separation Connector, as the Brewster Plate, is not illuminated by the sun during Jupiter encounter. Its resistance, however, has been estimated to be relatively high, R = 6.6×10^{14} ohm, for $\rho = 10^{16}$ ohm-cm. Therefore, like the LGASC, the connector can attain high potential differences with respect to the spacecraft structure. The surface area to the plasma was estimated to be A_{plasma} = 14.5 cm^2.

The stress on the separation connector three hours before periapsis is shown in Fig. 15. The potentials with respect to ground are

$R = 1.3 \times 10^{16}$ ohm, $\rho = 10^{16}$ ohm-cm

J_{ouv} \ S_{tons}	0.25	1	4
11.1	* -16,000	-13,000	-8,300
	** -2,600	-1,700	-60
37	* -38,000	-34,000	-24,000
	** -4,700	-3,000	-1,200
111	* -62,000	-56,000	-43,000
	** -8,200	-5,300	-1,900

$S_{tons}J_i$ = CURRENT OF SECONDARY ELECTRONS PRODUCED BY THE ION CURRENT J_i INCIDENT ON A NEGATIVE SURFACE.

J_{ouv} = PHOTOELECTRON CURRENT DENSITY IN pa/cm^2.

ρ = RESISTIVITY OF SEPARATION CONNECTOR

R = RESISTANCE OF SEPARATION CONNECTOR

*A-M Stress
**NASA 660-24 Stress
Plasma parameters at t = -3 hr (L=5.9)

Fig. 15. Parameter Study of Separation Connector Stress (volts) at t = -3 Hours

seen to be uncombortably high for all parameter values examined in the case of the A-M model and also quite high for the NASA 660-24 case. Snap-in metallic covers have been installed to eliminate possible problems with the separation connectors.

SUMMARY

The charge on the spacecraft structure and the stress on exposed dielectric surfaces is critically dependent on a number of parameters. Among those which are known with insufficient accuracy, are the cold plasma currents, the photoelectric emission currents and the resistivities of the dielectrics.

If the NASA 660-24 cold plasma model is correct, the Voyager spacecraft ground potential can be expected to be significantly smaller than those predicted by the model deduced from A-M densities. In all the dielectric chargeup cases examined, the A-M density model led to spacecraft ground at lower altitudes, particularly in the wake of Io. On the other hand, the NASA 660-24 model led to the greater voltage stress for L > 8 Jupiter radii. It has been demonstrated that uncertainties in the resistivity of the material and the photoelectron current emitted from the material critically affected the results of differential potential stresses. The secondary emission of ions were shown to play a somewhat smaller role for the range of parameters examined.

Uncertainties in material and environmental parameters such as those discussed above makes further refinements in the charging calculations appear to be unwarranted. Some generalizations concerning the Voyager charging can, however, be made. The spacecraft structure is expected to "chargeup" in the wake of the Io flux tube to the negative potentials as high as 25,000 volts. Furthermore, the analysis indicates that negative spacecraft potentials from 50 volts (NASA 660-24) to a couple of thousands of volts (A-M) can be anticipated at all times from two hours before periapsis to periapsis.

If the breakdown electric field is assumed to be 10^6 volts/cm, the stress on the Brewster Plate is not expected to become sufficiently high to produce arcing. The stress will, in any case, be limited by leakage currents through the relatively low leakage resistance anticipated. Since the resistance of the LGASC is relatively high (1.3×10^{16} ohm) the LGASC could be expected to chargeup to fairly high voltages relative to the structure. For the A-M cold plasma the stresses could be as high as 13,000 volts. A similar situation occurs with the Separation Connectors. In this case large stresses are predicted with both cold plasma models. For A-M cold plasma the stress can reach as much as 60,000 volts and as high as 8300 volts using NASA 660-24. Appropriate countermeasures were implemented on the Voyager spacecraft to prevent deleterious effects due to the possible chargeup described above.

ACKNOWLEDGMENTS

The work reported here was performed by the authors as TRW consultants to the Jet Propulsion Laboratory, California Institute of Technology under Contract NAS7-100 sponsored by the National Aeronautics and Space Administration. Significant guidance and support was provided by the JPL Voyager Team. We are especially indebted to Dr. J. Wiggins of TRW for the modeling of Voyager surfaces for UV and plasma exposure.

REFERENCES

1. Inouye, G. T., A. C. Whittlesey, S. R. Ponamgi, B. D. Cooperstein, and A. K. Thomas, "Voyager Spacecraft Electrostatic Discharge Immunity Verification Tests," presented at the symposium on The Effect of the Ionosphere on Space and Terrestrial Systems, Arlington, Va., Jan. 24-26, 1978.

2. Sanders, N. L., G. T. Inouye and A. Rosen, "Voyager Spacecraft Charging Immunization Support to JPL," TRW Report No. 31440-6002-RU-00, May 2, 1977.

3. Frank, L. A., et al., *J. Geophys. Resh.*, *81*, 457, 1976.

4. Intriligator, D. S. and J. H. Wolfe, *Geophys. Letters*, *1*, 281, 1974.

5. Scarf, F. L., "Plasma Physics and Wave-Particle Interactions at Jupiter," Jupiter, Ed. T. Gehrels, U. of Arizona Press, 870-895, 1976.

6. Grard, R. J. L., et al., *Geophys. Letters*, *4*, No. 6, 1977.

7. Axford, W. I. and D. A. Mendis, *1974 Ann. Rev. Earth and Planet. Sci.*, *2*, 419.

8. Scarf, F. L., "Plasma Physics Phenomena in the Outer Planet Magnetospheres," Exploration of the Outer Solar Systems, Progress in Astronautics and Aeronautics, 50, 131-156, 1976.

9. JPL IOM 312/77.1-13-PAP, to A. L. Lane from P. A. Penzo, "Red Spot Location and Magnetic Dipole Parameters for MJS77 JST and JSX Trajectories," 9 Feb. 1977.

10. Grard, R. J. L., "Properties of the Satellite Photoelectron Sheath Derived from Photoemission Laboratory Measurements," *J. Geophys. Res.*, *78*, 16, June 1, 1973, p. 2885-2906.

PREDICTIONS OF HIGH-VOLTAGE DIFFERENTIAL CHARGING ON GEOSTATIONARY SPACECRAFT

J.G. Laframboise and S.M.L. Prokopenko
Physics Department, York University
Toronto, Canada M3J 1P3

ABSTRACT

Two separate calculations are described. The first is a numerical prediction of upper bounds on negative floating potentials of shaded spacecraft surfaces, using a local-current-balance formulation. Results include the following: (a) by influencing the velocity-space cutoff boundaries for incident ion fluxes, the spacecraft geometry and sheath potential profile (particularly shaded-sunlit asymmetries in the latter) have large influences on shaded-surface potentials, which may exceed -20kV in certain circumstances (b) for electrically isolated surfaces in shaded cavities, negative floating potentials may exceed those on convex surfaces (c) in some conditions, two distinct floating potentials are predicted, leading to the possibility of "bifurcation phenomena" in which adjacent isolated surfaces made of the same material may follow different charging histories. The second calculation described is a more elaborate two-dimensional simulation, presently under development.

1. INTRODUCTION

The performance of many satellites in geostationary orbit has been degraded by anomalous events which include frequent spurious spacecraft commands and in some cases permanent damage. These events invariably appear to involve electrical discharges caused by differential charging of spacecraft surfaces to large relative potentials. The latter condition in turn is known to result from the relatively large average energies (up to a few keV) of the charged particle environment at geostationary orbit altitude, particularly in disturbed magnetospheric conditions. Since photoelectron emission from sunlit surfaces tends to compensate for incident electron fluxes, thereby holding sunlit-side surface potentials close to space potential in most cases, an estimate of differential charging magnitudes can be obtained by simply calculating floating potentials of electrically isolated shaded surfaces, relative to space potential. In the present work, we have attempted to obtain upper bounds on such potentials, which in cases of interest are usually highly negative, because these bounds constitute "worst cases" for design purposes, and also because unlike more exact calculations, they can be obtained from simple current-balance calculations. Furthermore, it is sufficient to consider local current balance only, because this corresponds to an electrically isolated surface element, which is also a "worst case" for differential charging. To calculate these bounds, we have extended a calculation by Knott[1], of the floating potential of a spherically symmetric geostationary-altitude satellite in eclipse. To investigate geometrical effects, we have replaced Knott's use of the Mott-Smith and Langmuir[2] orbit-limited current expression for collection of Maxwellian ions by a unipotential sphere, by the corresponding expression for an infinite cylinder; both expressions have been shown[3] to be upper bounds for collisionless ion collection as a function of local surface potential, for three- and two-dimensional collectors, respectively, regardless of collector shape, sheath potential, or potential of other parts of the collector. This replacement causes a large decrease in ion collection and a correspondingly large increase in negative shaded-side floating potentials (Sec. 3). Another important ion-current restriction may be caused by "effective-potential barrier"[3,4,5] or "angular-momentum selection" effects, in which the presence of less-negative sunlit-side potentials produces dipole and higher moments in the sheath potential[6], causing steepening and contraction of the potential well surrounding the shaded side (Fig.1). A similar steepening effect will also occur if an isolated shaded surface element is surrounded by adjacent shaded surfaces which for any reason have less negative potentials (Fig.2). The most extreme possibility would be a potential profile which was equal to space potential almost to the spacecraft surface, then fell discontinuously to surface potential. This limit would correspond to a "planar sheath" situation in which the ion collection on any shaded convex surface would be given by just the ion random flux. This amounts to a further ion current restriction which produces even larger increases in negative shaded-side floating

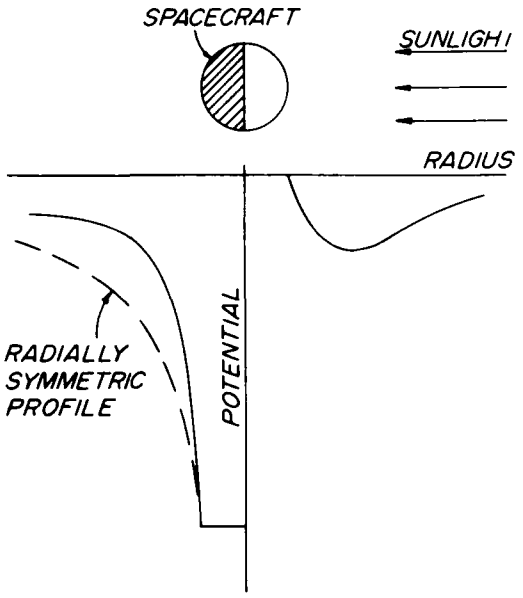


Figure 1: Steepening of shaded-side potential profile for a spacecraft with an insulated surface, after Fahleson[6].

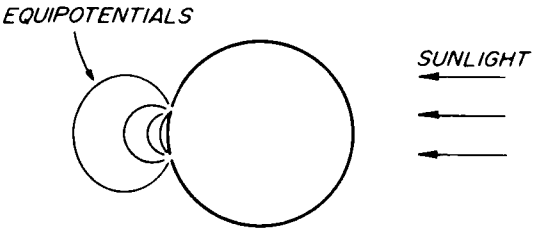


Figure 2: Conductive spacecraft with shaded isolated surface patch.

potentials (Sec. 3). This situation corresponds to a velocity-space cutoff boundary for incident ions which is "one-dimensional"; the cutoff boundaries corresponding to spherical and infinite cylindrical collectors are, respectively, "three-dimensional" and "two-dimensional"[3] (Sec.2).

We also show (Sec.3) that if shaded cavities containing isolated surfaces exist on a spacecraft, negative potentials on such surfaces may surpass even these predictions. In some cases, more than one possible floating potential results from the calculation; this implies the possibility of "bifurcation phenomena" in which adjacent isolated surfaces of the same material may arrive at different floating potentials as a result of differences in their charging histories (Sec. 3).

We have also modified Knott's calculation in another way, by including currents due to electron backscattering (Sec.2). These currents will tend to decrease net electron collection, thereby making floating potentials less negative than otherwise (Sec. 3). A process not included by either Knott or ourselves is secondary electron emission due to ion impacts; this will also tend to make floating potentials less negative.

We also describe (Sec. 4) a more elaborate two-dimensional numerical simulation which is presently under development.

2. THEORY

The ambient electron energy distributions used in the present work are a model quiet-time spectrum (Knott[1], Fig. 1) and a model disturbed spectrum (Knott[1], Fig.2b) based on measurements by Shield and Frank[7] and DeForest and McIlwain[8], respectively. Both of these distributions, and also the ambient ion distribution, are assumed isotropic. The disturbed spectrum was chosen from the three used by Knott because it has a higher average electron energy (~10keV) than the others. In using it, we have changed it as follows: in the energy ranges $0.5\ \text{keV} \le E \le 10\text{keV}$ and $10\text{keV} \le E \le 40\ \text{keV}$, we have replaced Knott's differential energy spectrum by $\sqrt{2} \times 10^{9} E^{-\frac{1}{2}}$ and $\sqrt{2} \times 10^{9}\ E^{-3/2}$ electrons/cm^2 sec steradian keV, respectively, where E is energy. These relations are simpler than those indicated by Knott, and they also bring the model spectrum into closer agreement with the data on which it is based. We therefore believe that they may have been the ones actually used by Knott, and that the corresponding parts of Fig. 2b in his paper may be incorrectly plotted. The resulting expressions for orbit-limited fluxes (particle current densities) J_e, J_i, J_{sec} and J_{scat} of ambient electrons, ambient ions, escaping secondary electrons and escaping backscattered electrons, corresponding to three-, two-, and one-dimensional cutoff boundaries, have been given by Prokopenko and Laframboise[9].

In using these expressions we have made further approximations as follows. We have introduced a linear rise from 0 to 1eV in both spectra,

and an upper cutoff at 50keV in the quiet-time spectrum, in order to avoid divergent integrations[9]. We have assumed[1] that the ambient ions are Maxwellian with a temperature T_i = 1keV, and that the random ion to electron flux ratio J_{io}/J_{eo}=0.025. For the secondary electron fractional yeld $\delta(E)$, we have used, following Knott[1], the relation of Sternglass[10]:

$$\delta(E) = 7.4\delta_{max}(E/E_{max})\exp[-2(E/E_{max})^{\frac{1}{2}}]. \quad (1)$$

We have used values of δ_{max} and E_{max} obtained from Gibbons[11], Hachenberg and Brauer[12], and Willis and Skinner[13].

The process of electron backscattering, which was not included in Knott's calculations, becomes important at incident electron kinetic energies larger than those for which secondary emission is dominant. For the backscattered electron fractional yield η, we have fitted the results of Sternglass[14] and Palluel[15] with a relation of the form:

$$\eta(E) = A - B\exp(-CE) \quad (2)$$

where the coefficients A,B, and C are functions of the atomic number Z of the surface material. We have evaluated A,B, and C for each surface material considered (Sec.3) by substituting Sternglass' and Palluel's values of η at 0, 1 and 16 keV, into Eq.(2). In all cases, η is a very slowly varying function of E. For compound surface materials, we have used an atomic number given by a weighted average of those of each constituent.

There exist more recent measurements of η[16,17] which give generally larger values than those of Sternglass or Palluel, especially for electrons having near-tangential incidence. However, we have found these results to be too fragmentary for our purposes, and we have therefore used Sternglass' results throughout. Presumably we have therefore underestimated η, and our predicted floating potentials in Sec. 3 will therefore be slightly more negative than more realistic corresponding values.

When surface potential $\varphi_s > 0$, not all secondary and backscattered electrons will escape. To calculate flux escaping, we have assumed [14,18] for ease of calculation, that both secondary and backscattered electrons are emitted with Maxwellian velocity distributions having thermal energies $E_{sec} = kT_{sec} \approx 3$ eV, and $E_{scat} = kT_{scat} \approx (0.45 + 2 \times 10^{-3}Z)\bar{E}$ eV, respectively, regardless of the form of the incident velocity distribution. Here, $\bar{E}$ is the average incident electron energy. We have further assumed that escape of emitted electrons is orbit-limited, i.e. that no barriers of effective potential[3,4,5] or negative barriers of electric potential exist on the shaded side. Fahleson[6] has pointed out that such barriers are likely to exist on the sunlit side independently of any space-charge effects, if substantial shaded-sunlit differences exist in φ_s.

The floating potential(s) of an isolated shaded surface element is(are)now given by the zero(s) of the function

$$J_{net} = J_i - J_e + J_{sec} + J_{scat}. \quad (3)$$

3. RESULTS AND DISCUSSION

Table 1 shows floating potential values obtained by numerical solution of the equation $J_{net} = 0$, where J_{net} is given by Eq. (3). The most important feature of Table 1 is the very large floating potentials which are evident in disturbed conditions in the presence of the two- and one-dimensional velocity-space cutoffs. The dramatic differences which exist among these results are evidence that spacecraft geometry and sheath potential shape are important influences in determining floating potentials. It is important to note that as floating potential becomes more negative, it also becomes more sensitive to the presence of small amounts of high-energy electrons. This means that if a spacecraft should encounter conditions that are "more disturbed" than those given by Knott's spectrum 2b, the values in Table 1 most likely to be significantly exceeded are those for the one-dimensional cutoff. This implies that for design purposes in which worst-case information is desired, it is important to do calculations with the "most disturbed" electron spectra available.

In obtaining these results, we have made no attempt to calculate the time needed to approach the steady-state conditions which they represent. In general, the most negative potentials correspond to a balance between the smallest currents, and will therefore involve the longest charging times.

Also evident in Table 1 are situations in which the current-voltage characteristic of the surface has three roots. For these to occur, it is necessary that δ_{max} be substantially greater than one, and that the incident spectrum contain a sufficient proportion of electrons in the energy range where secondary emission is a maximum. The centre root never represents a possible floating potential, because it is "unstable" in the sense that a small change in surface potential would cause a net current collection of a sign which would drive the surface potential away from this root to one on either side. A further consequence of such a situation is discussed below.

In comparison with the results of Knott[1], Table 1 includes the further addition of backscattered electron flux (Sec. 2) and therefore represents a more realistic physical situation. In most cases, the net effect of backscatter is a moderate reduction of negative floating potentials. In some cases, the reduction is large, as in the case of a gold surface exposed to the "quiet" spectrum. In several other cases, all associated with the quiet spectrum, backscattering changes a multiple-root to a single-root situation[9]. As indicated in Sec. 2, we have probably underestimated backscattered fluxes, and we have also (Sec. 1) ignored secondary electron emission caused by ion

Table 1

Floating potentials of shaded surfaces of geostationary-altitude spacecraft, using the same incident spectra as assumed by Knott[1], with secondary and backscattered electron emission included, with three-, two-, and one-dimensional velocity-space cutoffs corresponding to orbit-limited ion collection in spherical, infinite cylindrical and planar symmetries, respectively.

Material	Secondary Electron Emission Data		Back-scatter Data	Spectrum 1 "Quiet"			Spectrum 2b "Disturbed"		
	δ_{max}	E_{max} (eV)	η(5000eV)	Floating Potential(volts)			Floating Potential(volts)		
				3-dimen.	2-dimen.	1-dimen.	3-dimen.	2-dimen.	1-dimen
Gold	1.45	800	.42	-34.2	-34.3	-34.9	-3470	-6430	-15,450
Aluminum	.97	300	.16	-1410	-2140	-5390	-6770	-11,500	-21,770
Aluminum with Oxide Coating	2.60	300	.12	+4.6 -630* -750	+3.2 -490* -1560	+1.9 -420* -4900	-6610	-11,360	-21,610
Quartz	2.50	420	.12	+3.9	+2.6	+1.5 -640* -4120	-6310	-10,960	-21,130
Aquadag	.75	350	.08	-1560	-2380	-5890	-7090	-12,010	-22,350
Beryllium Copper	2.20	300	.31	+4.7	+3.2	+1.9 -560* -3430	-5740	-9920	-19,890
Beryllium Copper Activated	5.00	400	.31	+8.0	+6.0	+4.2	+4.4 -900* -3950	+3.0 -830* -7580	+1.6 -770* -17,670
Teflon	3.00	300	.09	+5.2	+3.6 -620* -1440	+2.2 -490* -4960	-6690	-11,500	-21,780
Kapton	2.10	150	.07	+3.9 -170* -1580	+2.6 -170* -2440	+1.4 -170* -6040	-7180	-12,130	-22,520
No secondary or backscattered electrons	—	—	—	-1860	-2830	-6680	-7550	-12,690	-23,130

* Unstable

impacts. Both of these effects would tend to further reduce negative potentials. However, such changes are likely to be small. The results in Table 2 should probably be regarded as consistent with observations of potentials reaching -19kV on the ATS 6 spacecraft, as reported by Whipple[19].

Figures 3-7 show current -voltage characteristics for some of the situations in Table 1. Figure 3 shows a "typical" single-root situation in which secondary and backscatter contributions do not change the general shape of the net current curve. Figure 4 shows the above-mentioned case of gold exposed to the quiet spectrum, in which the backscatter contribution changes a large predicted negative floating potential to a much smaller value. Figure 5 shows a triple-root situation. Figure 6 shows the disappearance of a triple-root situation because of backscatter. In Fig. 7, secondary electron current is sufficient by itself to prohibit a negative floating potential.

We now examine situations which may arise in the case of spacecraft which have shaded cavities containing electrically isolated interior surfaces. Figure 8 shows an idealization of such a spacecraft. We wish to show that the effects of surface concavity may cause ion collection to be reduced more than net electron collection at an interior point such as B, relative to an exterior point A; such a situation would result in floating potentials more negative than those of Table 1. To demonstrate this possibility, we first note that in the presence of an isotropic ambient plasma, incident fluxes to any surface depend only[3] on the locations, in velocity space, of the cutoff boundaries inside of which the orbits of ambient particles can connect "from infinity" to the surface. Figure 8 shows a set of the associated "cutoff orbits". We see from Fig. 8 that the included angle between cutoff orbits has been reduced in going from A to B for ions but not for electrons, for which orbits tangential to the surface are shown as reaching both A and B. Accordingly, the incident ion current contribution for the energy shown will also be reduced, but the electron contribution will not. This picture is invalid for higher-energy electrons at B, whose orbits are straighter and will have a greater tendency to connect back to the interior surfaces of the cavity. Even though such higher-energy orbits will generally have lower populations than lower-energy orbits, it is not clear whether the relative current reduction at B will be greater for ions or for electrons. However, this argument is intended to demonstrate only the possibility that the bounds in Table 1 will be exceeded. On the other hand, this possibility will be enhanced by the effects of secondary and backscattered electrons, which will tend to be recollected inside any cavity, rather than escaping into space, thus tending to increase net electron collection and driving floating potentials more negative. This effect will be strongest for backscattered electrons because their higher emission energies will cause them to have straighter orbits. To draw firm conclusions will require detailed numerical simulation. An additional feature of cavities is their generally higher outgassing pressures, which will increase any tendencies for arcing to occur. More negative floating potentials may also result if the ambient electron distribution contains beam-like constituents[20] which happen to be directed into a cavity. Especially severe arcing problems are known to have occurred between electronic components mounted inside a cavity at one end of the DSCS spacecraft.

Finally, we discuss some further implications of the multiple-root results shown in Table 1 and Figure 5. Consider a situation involving two or more adjacent but isolated spacecraft surfaces which are made of the same material, and whose external conditions change with time, as in the case of time-varying ambient distributions, or a spacecraft rotation which carries these surfaces from sunlight into shadow. Such a situation might involve the continuous evolution of a single-root into a multiple-root situation, and the possibility would then arise of a "bifurcation" phenomenon in which different surface elements followed different potential histories, with a correspondingly large potential difference arising between them. Again, detailed numerical simulations (Sec. 4) are necessary in order to find out if such phenomena can actually occur.

4. TWO-DIMENSIONAL NUMERICAL SIMULATION

A numerical simulation program is being developed[21], having the following features:

a) infinite circular cylindrical geometry with angle-dependence. Reasons for such a choice include the following: A realistic model must be at least two-dimensional because the asymmetry between sunlit and shaded surfaces is a key feature of the differential charging problem. Cylindrical geometry is the simplest two-dimensional geometry, and it is also a useful approximation to many spacecraft shapes.

b) ability to include effects of: velocity distributions of incident particles, photoelectrons, secondary electrons, backscattered electrons, and gun emissions; internal current pathways including surface conductive layers.

c) quasistatic time-dependent iteration. In this procedure, sheath potential changes during particle transit times are ignored. This leads to the following iteration scheme: An angle-dependent surface potential is chosen. Poisson's equation is then solved to provide a radius- and angle-dependent static sheath potential [see (d) below]. Particle orbits are then followed numerically in this potential, yielding surface charging rate as a function of angle [orbit-following is, however, not used to provide space-charge densities for Poisson's equation; see (d) below]. These rates are then

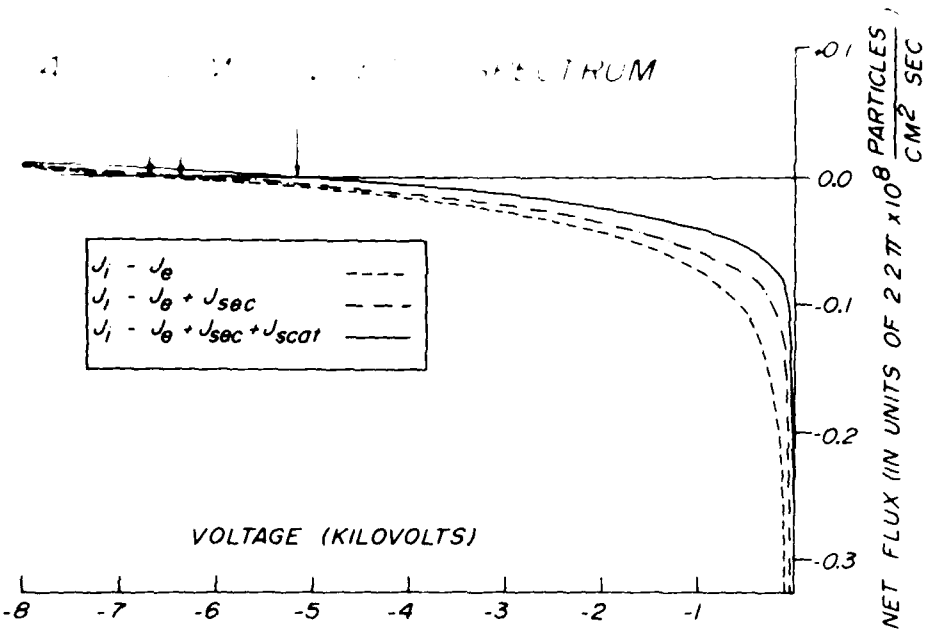


Figure 3. Current-voltage characteristic for aluminum in "quiet" conditions, with a one-dimensional velocity-space cutoff. In Figs. 3-7, the zeros of the characteristics are indicated by arrows.

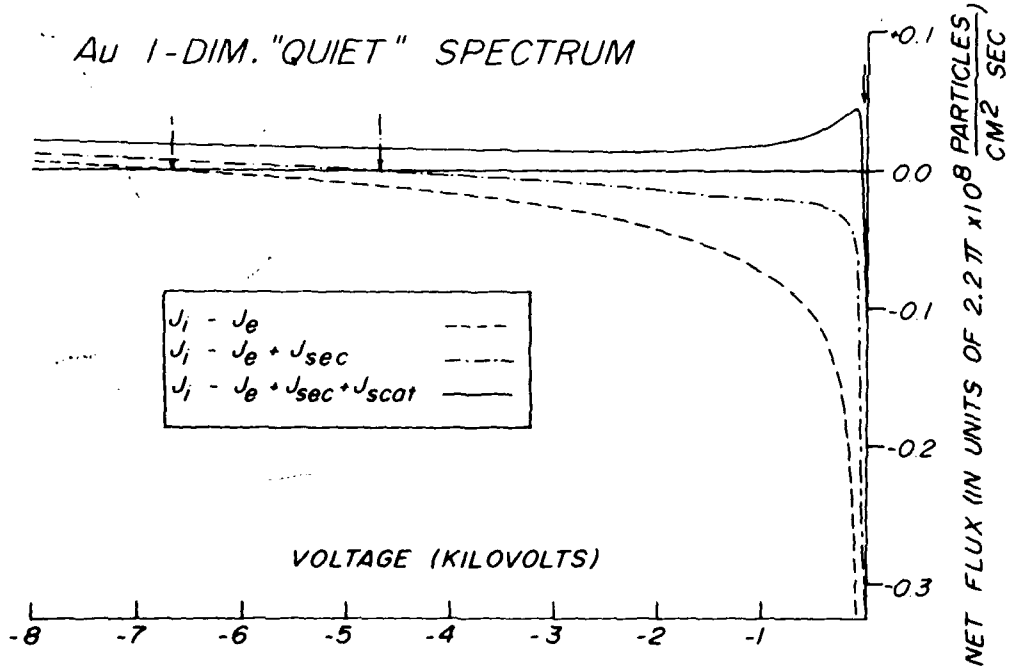


Figure 4: Current-voltage characteristic for gold in "quiet" conditions, with a one-dimensional velocity-space cutoff.

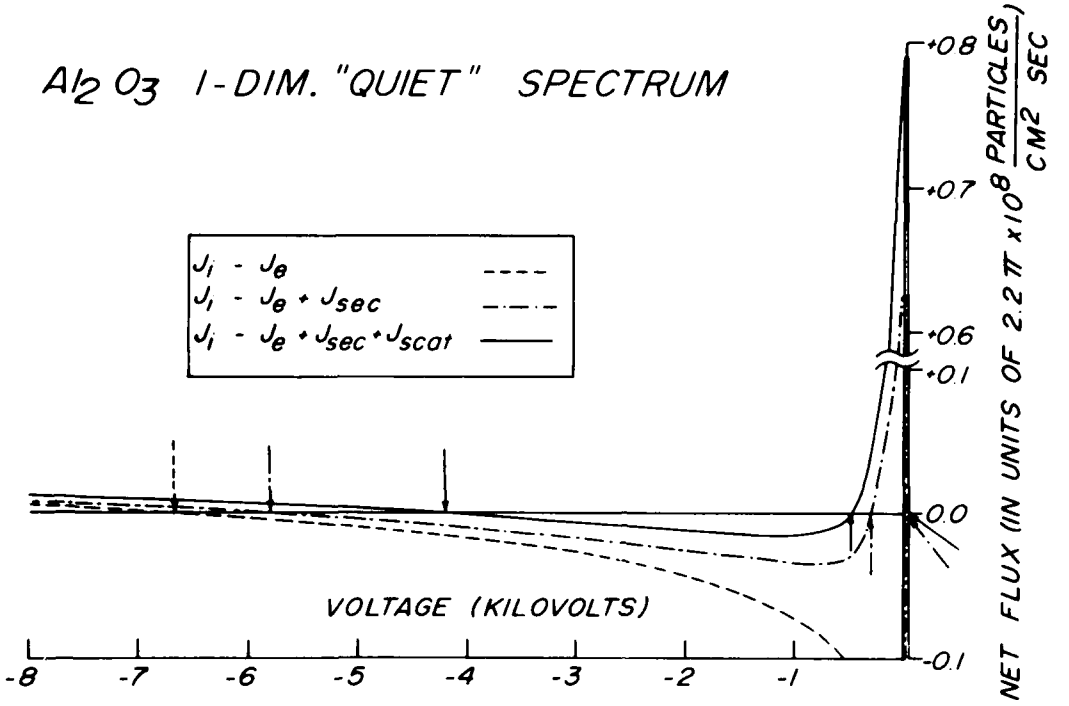


Figure 5: Current-voltage characteristic for aluminum oxide in "quiet" conditions, with a one-dimensional velocity-space cutoff.

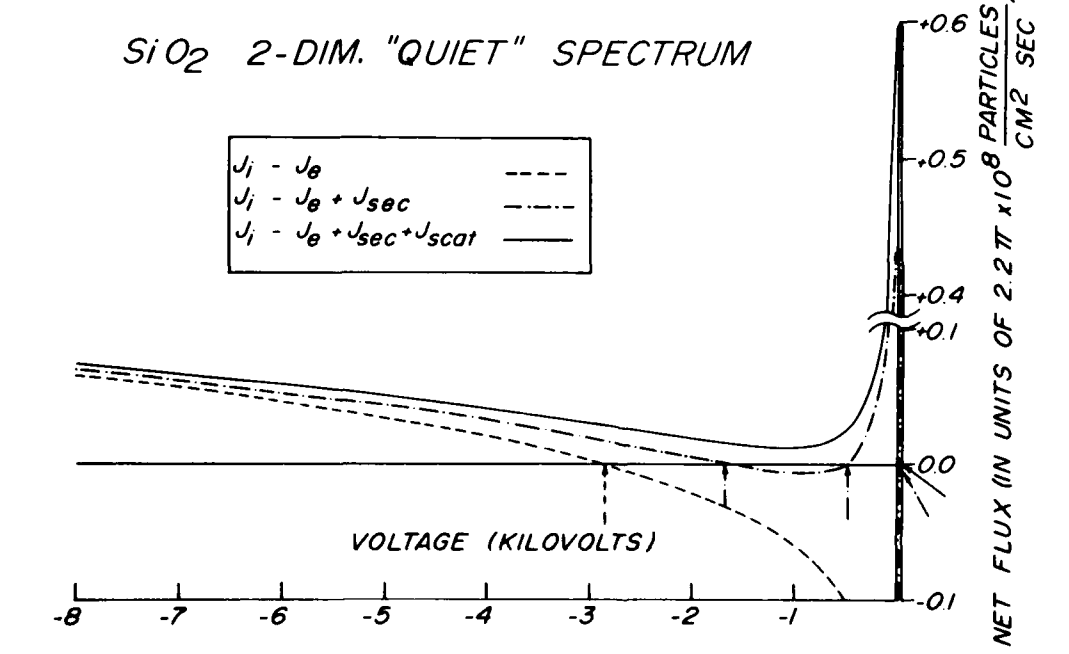


Figure 6: Current-voltage characteristic for quartz in "quiet" conditions, with a two-dimensional velocity-space cutoff.

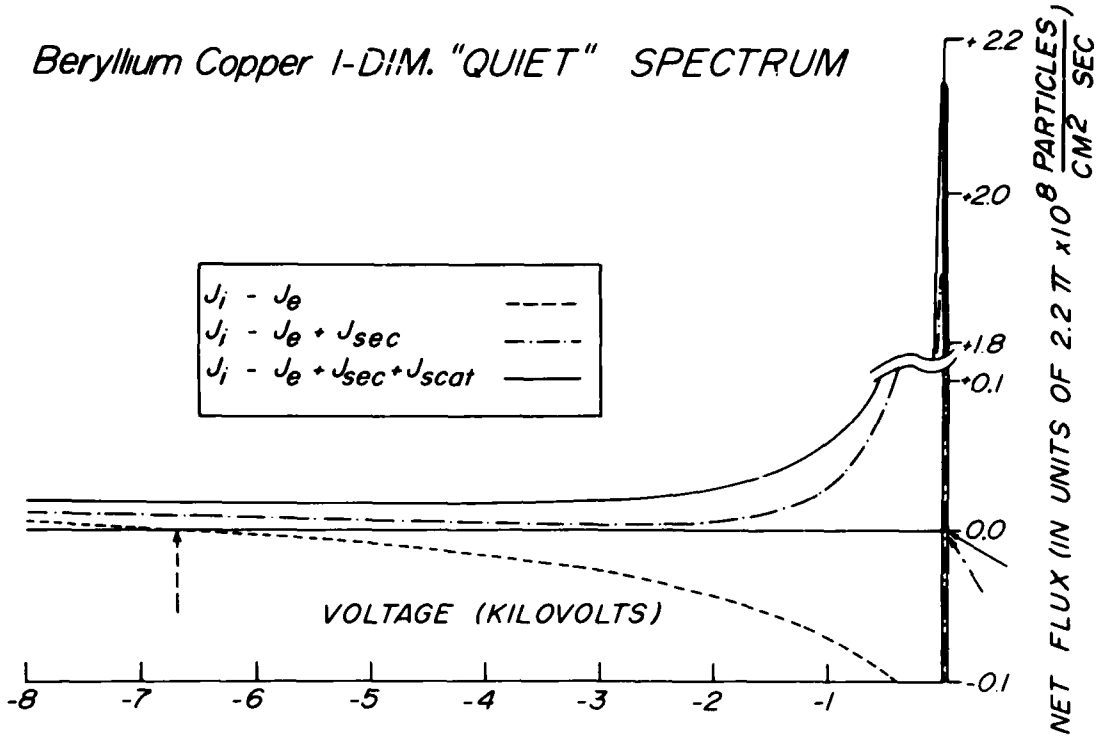


Figure 7: Current-voltage characteristic for activated beryllium-copper in "quiet" conditions, with a one-dimensional velocity-space cutoff.

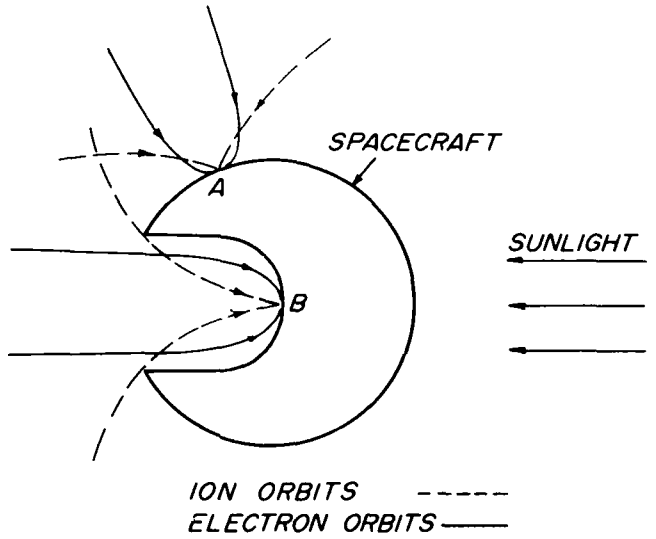


Figure 8: Spacecraft with shaded isolated cavity, showing incident ion and electron orbits with energies close to the lowest for which collection of each species is possible. The orbits shown are cutoff orbits, defined as the most nearly tangential orbits for which incident particles of a given energy are able to reach a given point on the spacecraft surface, having tangential velocity component in a given direction.

averaged over any conductive sector, and any currents transferred internally (including those through any surface layers having finite conductivity) are subtracted off. The resulting net charging rates are then used to carry forward one time-step, yielding new surface potentials. This process is then repeated until a steady-state condition results, or, in a situation in which external conditions vary with time, is repeated to follow such time-dependence.

d) use of simplified space-charge density expressions, rather than numerical orbit-following, in solving Poisson's equation for sheath potentials. This approximation is expected to yield large savings in computer time. Furthermore, space-charge effects are likely to be small enough in most cases that any reasonable approximation for space-charge density will produce very little loss of accuracy. This statement is most valid for ambient particles, since these have a typical Debye length of order 10 meters. Some examples of density approximations are described and reviewed by Laframboise and Prokopenko[21]. Such approximations must ultimately be validated by comparison with a few carefully chosen exact calculations. A relatively small amount of numerical orbit-following is still used to calculate surface currents; see (c) above.

e) use of isotropic or beam-like (monokinetic) ambient velocity distributions, or some superposition of these, which can be used to approximate any ambient distribution as closely as desired.

ACKNOWLEDGMENTS

We are indebted to H. Cohen, Jen-Shih Chang, and L.W. Parker for valuable comments. This work was supported by the U.S. Air Force Office of Scientific Research under grant number AFOSR-76-2962.

REFERENCES

1. Knott, K. (1972) The equilibrium potential of a magnetospheric satellite in an eclipse situation, Planet. Space Sci. 20:1137-1146.
2. Mott-Smith, H., Jr., and Langmuir, I. (1926) The theory of collectors in gaseous discharges, Phys. Rev. 28:727-760.
3. Laframboise, J.G., and Parker, L.W. (1973) Probe design for orbit-limited current collection, Phys. Fluids 16:629-636.
4. Bernstein, I.B., and Rabinowitz, I.N. (1959) Theory of electrostatic probes in a low-density plasma, Phys. Fluids 2:112-121.
5. Laframboise, J.G. (1966) Theory of spherical and cylindrical Langmuir probes in a collisionless, Maxwellian plasma at rest, Univ. of Toronto, Institute for Aerospace Studies, Rep.100.
6. Fahleson, U. (1973) Plasma-vehicle interactions in space: Some aspects of present knowledge and future development. In: Photon and Particle Interaction with Surfaces in Space, R.J.L. Grard, Editor, D. Reidel Pub. Co., Dordrecht, Holland, pp. 563-569.
7. Shield, M.A., and Frank, L.A. (1970) Electron observations between the inner edge of the plasma sheet and the magnetosphere, J. Geophys. Res. 75:5401-5414.
8. DeForest, S.E., and McIlwain, C.E. (1971) Plasma clouds in the magnetosphere, J. Geophys. Res. 76:3587-3611.
9. Prokopenko, S.M.L., and Laframboise, J.G. (1977) Prediction of large negative shaded-side spacecraft potentials. In: Proceedings of the Spacecraft Charging Technology Conference, C.P. Pike and R.R. Lovell, Editors, Air Force Surveys in Geophysics, No.364, Hanscom Air Force Base, Mass., Rep. AFGL-TR-77-0051/NASA TMX-73537, pp. 369-387.
10. Sternglass, E.J. (1950) Secondary electron emission and atomic shell structure, Phys. Rev. 80:925-926.
11. Gibbons, D.J. (1966) Secondary electron emission. In: Handbook of Vacuum Physics, A.H. Beck, Ed., Pergamon Press, Oxford, Vol.2, pp. 301-395.
12. Hachenberg, O., and Brauer, W. (1959) Secondary electron emission from solids, Adv. Electronics Electron Phys. 11:413-499.
13. Willis, R.F., and Skinner, D.K. (1973) Secondary electron emission yield behaviour of polymers, Solid State Comm. 13:685-688.
14. Sternglass, E.J. (1954) Backscattering of kilovolt electrons from solids, Phys. Rev. 95:345-358.
15. Palluel, P. (1947) Composante rediffusée du rayonnement électronique secondaire des métaux, Compt. Rend. 224:1492-1494.
16. Thomas, S., and Pattinson, E.F. (1970) Range of electons and contribution of backscattered electrons in secondary production in aluminum, J. Phys. D: Appl. Phys. 3:349-357.
17. Darlington, E.H., and Cosslett, V.E. (1972) Backscattering of 0.5-10 keV electrons from solid targets, J. Phys. D: Appl. Phys. 5:1969-1981.
18. Chung, M.S., and Everhart, T.E. (1974) Simple calculation of energy distribution of low-energy secondary electrons emitted from metals under electron bombardment, J. Appl. Phys. 45:707-709.
19. Whipple, E.C., Jr. (1977) Modeling of spacecraft charging. In: Proceedings of the Spacecraft Charging Technology Conference, C.P. Pike and R.R. Lovell, Editors, Air Force Surveys in Geophysics, No.364, Hanscom Air Force Base, Mass., Rep. AFGL-TR-77-0051/NASA TMX-73537, pp. 225-235.
20. DeForest, S.E. (1977) The plasma environment at geosynchronous orbit. In: Proceedings of the Spacecraft Charging Technology Conference, C.P. Pike and R.R. Lovell, Editors, Air Force Surveys in Geophysics, No.364, Hanscom Air Force Base, Mass., Rep.AFGL-TR-77-0051/NASA TMX-73537, pp.37-52.
21. Laframboise, J.G., and Prokopenko, S.M.L. (1977) Numerical simulation of spacecraft charging phenomena. In: Proceedings of the Spacecraft Charging Technology Conference, C.P. Pike and R.R. Lovell, Editors, Air Force Surveys in Geophysics, No.364, Hanscom Air Force Base, Mass., Rep. AFGL-TR-77-0051/NASA TMX-73537, pp. 309-318.

CHARGED-AREA EFFECTS ON SPACECRAFT DIELECTRIC ARC DISCHARGES

K.G. Balmain, P.C. Kremer and M. Cuchanski
Department of Electrical Engineering
University of Toronto
Toronto, Canada M5S 1A4

INTRODUCTION

A great quantity of evidence exists in support of the postulate that magnetic-substorm electrons cause arc discharges to occur on synchronous-orbit spacecraft [De Forest, 1972; Fredricks and Scarf, 1973; Rosen, 1975, 1976a, 1976b, 1976c]. Extensive laboratory simulation studies [Adamo and Nanevicz, 1976; Balmain, 1973; Balmain, Orszag and Kremer, 1976; Hoffmaster and Sellen, 1976; Stevens, Lovell and Gore, 1976] have demonstrated the existence of strong and sometimes destructive arc discharges. These range from tiny micro-discharges on dielectric specimens exposed to the focussed beam of a scanning electron microscope, to destructive lightning-like arcs up to 30cm in length traversing the surface of large, thin sheets of metal-backed dielectric. The only existing theory of the arc breakdown has been put forward by Meulenberg [1976], a theory which has become widely accepted and which describes the breakdown between a buried layer of impacted electrons and the electron-depleted surface of the material. Because this theory is one-dimensional, its applicability is primarily to the initial breakdown, while the propagation of the breakdown across the dielectric sheet very likely proceeds in the manner described by Gross [1957, 1958].

The main effect of discharge propagation apparently is to mobilize a large fraction of the available embedded-electron charge into a single arc which emits a burst of electrons with an accompanying downward pulse of current into the metal backing. Therefore, it might seem reasonable to expect that the dielectric specimen surface area would control the peak pulse current. The experiments to be described involve variations in surface area at two very different scales, one with specimens of varying area (of the order of $1cm^2$) exposed to a fixed flood beam of electrons, and the other with varying electron beam cross-sectional areas (of the order of $10^{-4}cm^2$) incident on specimens with much larger areas.

MACRO-DISCHARGES

Large-scale dielectric specimens were mounted in a scanning electron microscope (SEM) which had been modified to produce an approximately uniform flood beam up to 10cm in diameter. In the experiments to be described, the accelerating voltage was 20kV and the beam current density was 1 to 2 $\mu A/cm^2$ incident on the dielectric surface. Shown in Fig. 1 is the experimental apparatus, including a 250 MHz bandwidth oscilloscope for display of the discharge current pulse.

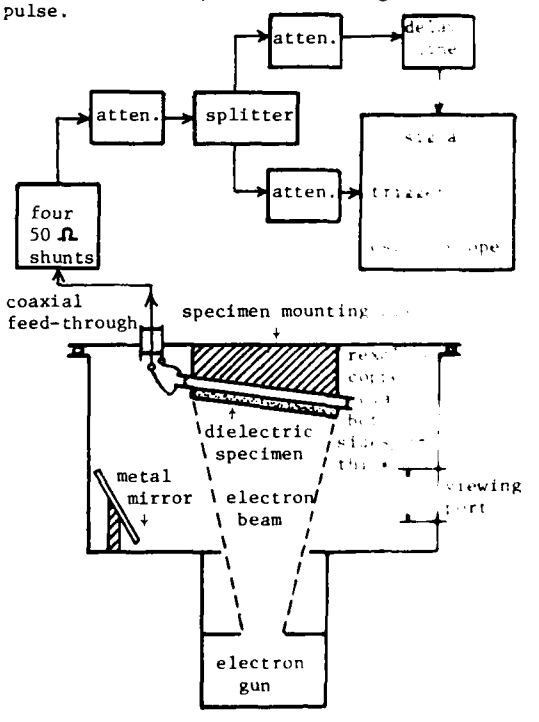


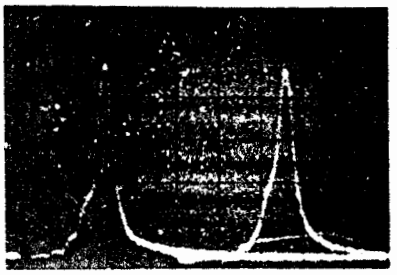
Fig. 1 The discharge current measurement system using the flood-beam modified SEM.



(a) Mylar specimen 48x26x0.12 mm, showing central illuminated region and arc going to lower edge.



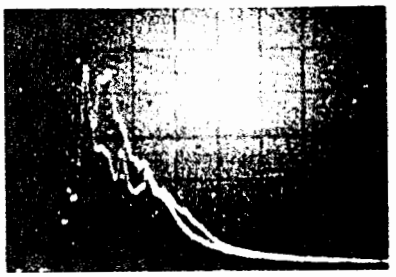
(b) Same specimen as at left, showing arc traversing illuminated region and going to upper edge.



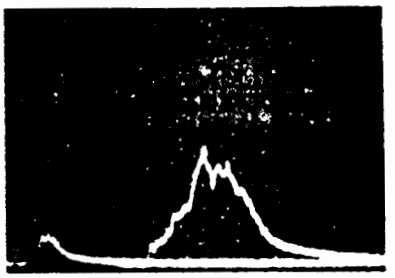
(c) 100 A peak current pulses on a 10 cm^2 Mylar specimen 0.1 mm thick. Hor. scale 50 ns/div.



(d) Quartz fabric discharges of 1.6 A and 1.3 A peak on a specimen of 2 cm^2. Hor. scale 50 ns/div.



(e) Silvered-Teflon discharges of 20A peak on a specimen of 10 cm^2. Hor. scale 50 ns/div.



(f) Aluminized-Kapton discharges on a 5 cm^2 specimen, at 100 ns/div. Large pulse: 9A peak.

Fig. 2 Examples of macro-discharges on various materials, for an incident-beam accelerating voltage of 20kV.

A t
2(a) as
central
The othe
are not
which ap
charge,
permits
lectric
central
filament
arc can
indicati
strong
emission

pical discharge arc can be seen in Fig.
"lightning stroke" extending from the
egion to the lower edge of the specimen.
rays extending outward to the edges
rcs but rather luminescent streaks
ear for a short period after each dis-
hen the reduced negative surface charge
he incident electrons to reach the die-
urface with appreciable energy. The
right spot is illumination from the
of the electron gun. In Fig. 2(b) an
e seen crossing the bright spot, thus
g that the filament illumination is not
ough to discharge the specimen by photo-

The strongest discharge currents measured to date, 100 A peak, are shown in Fig. 2(c). Also shown in Fig. 2(d), (e) and (f) are typical discharges on quartz fabric, silvered-Teflon second-surface mirror, and aluminized-Kapton thermal blanket outer layer.

One silvered-Teflon specimen (not shown) had its central region covered with an electrically conductive material. This produced the unexpected result of increased complexity in appearance of the discharge arc, together with the occurrence of multiple-peaked current pulses.

In order to study the anticipated specimen-area effects, experiments were conducted in which a specimen was sequentially halved in area,

BEST AVAILABLE COPY

at each step being exposed to the flood beam of electrons. For each step at least six discharge current pulses were recorded and their peak currents averaged. The silvered-Teflon result of Fig. 3(a) for the first run (in which the specimens were alternately square and rectangular) suggests the existence of a definite power-law relationship between peak current and area. A second run (square specimens only) was not significantly different.

Similar experiments on copper-clad Mylar revealed the possibility of thickness dependence. Fig. 3(b) shows that the thicker specimen exhibits by far the greater discharge current. The reasons for this are not clear although conceivably they could involve occasional electron beam penetration of the thin specimen.

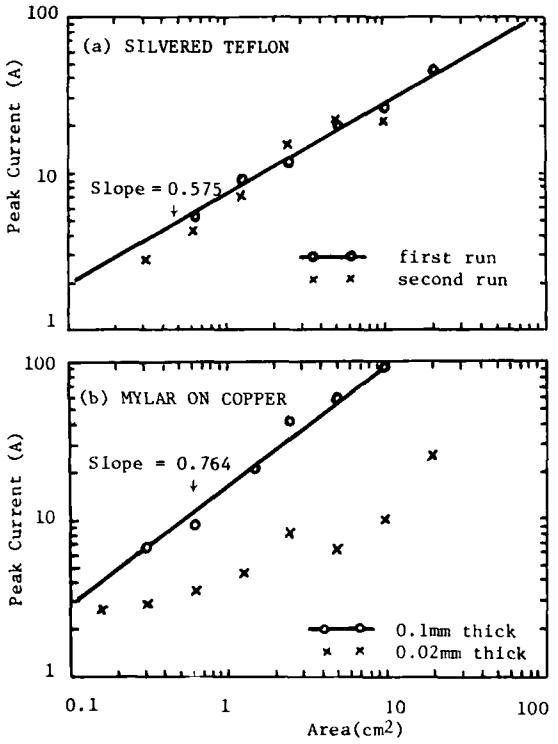


Fig. 3 The scaling of discharge peak current with specimen area for a 20kV beam.

It was observed that the discharge pulse duration is not simply related to the area. In the silvered-Teflon experiments the pulse duration (at half peak current) increased very erratically from 15ns to 100ns with increasing area. However, for the 0.1mm Mylar, the duration increased from 15ns to 25ns, again erratically. It is difficult to draw conclusions from these pulse duration observations although it can be said that the dielectric arcs bear little resemblance to capacitor discharges.

MICRO-DISCHARGES

Defocusing of a stationary (non-scanned) electron beam in an SEM was found to be an effective means of charging a very small, measurable area on a dielectric specimen. Measurement of the beam diameter was achieved by using the beam to implant a "stripe" of electrons across the specimen. A secondary-electron image of the charged stripe was made by using the photographic scan mode of the SEM and examining the specimen after a brief exposure to air in order to neutralize excess surface charge. Knowing the magnification, it was straightforward to measure the stripe width and hence the beam diameter.

The micro-discharge current pulses were measured using a sampling oscilloscope with a 4GHz bandwith, as shown in Fig. 4. A careful

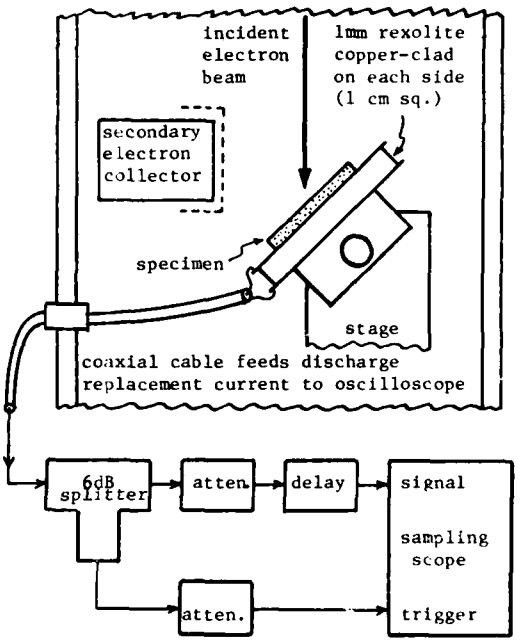


Fig. 4 Arrangement of components in specimen vacuum chamber of scanning electron microscope, and instrumentation for measuring micro-discharge current pulses.

setting of the oscilloscope trigger threshold made possible the selection of the strongest pulses, and the frequent discharge occurrence made possible the recording of relatively smooth pulse envelopes (an example of which is shown in Fig. 5) by taking one sample per pulse and stepping the sampling time through the duration of the pulse. The smoothness of the measured pulses indicates that the discharge phenomenon is highly repeatable within a small region of the dielectric surface. During each experiment the

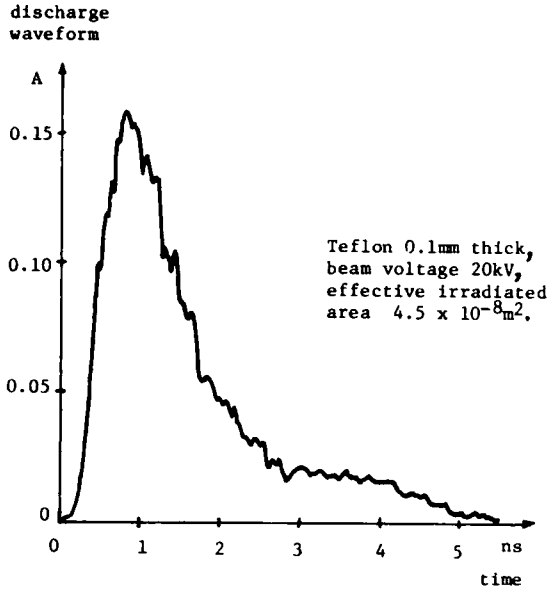


Fig. 5 A sampled current-pulse envelope.

position of the beam was changed very slowly to avoid "fatiguing" any point on the specimen.

Because of the slow beam movement described above, the effective charged area was estimated to be larger than the beam cross-section, and the area was computed as shown in Fig. 6.

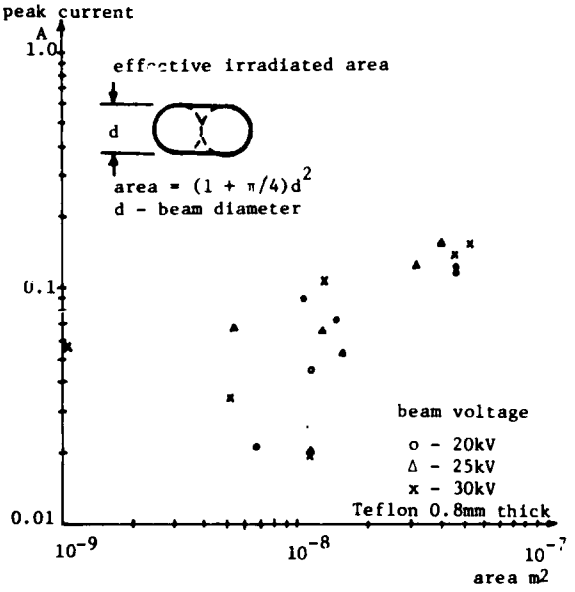


Fig. 6 Micro-discharge peak current as a function of charged area, with beam voltage as a parameter.

In the figure the experimental points show a definite tendency for peak current to increase with charged area, but the points exhibit no well-defined dependence on beam voltage. If beam voltage effects exist, they are not so pronounced as charged-area effects or the effects of using different materials such as Mylar, as indicated in Fig. 7.

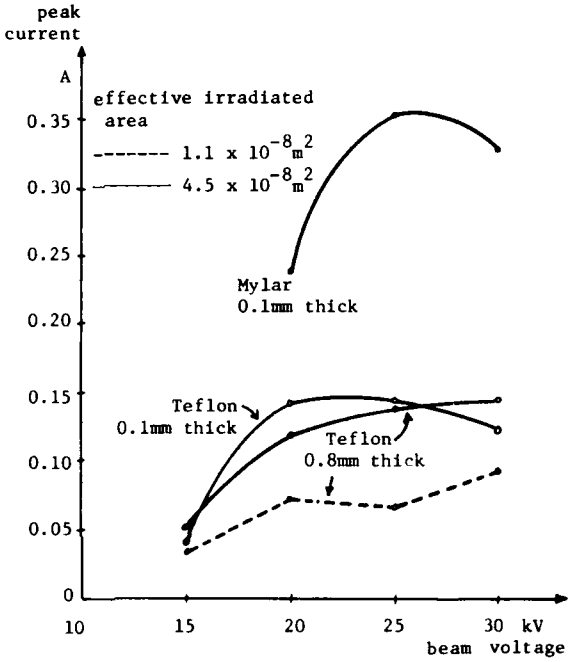


Fig. 7 Micro-discharge peak current dependence on material, specimen thickness, charged area and beam voltage.

The frequency of occurrence of discharge pulses is strongly dependent on beam voltage, as shown in Fig. 8. The low frequency of occurrence with Mylar is offset by high peak discharge currents, in terms of its potential as a source of electromagnetic interference.

The variations in pulse shape are summarized in Figs. 9 and 10, in terms of rise time and pulse width. Rise time is defined as the time interval between the 10% and 90% points, and pulse width is defined as the time interval between points at 1/3 the peak value. The rise time appears to increase with beam voltage, independent of area, while the pulse width shows a tendency to increase with area. Otherwise it is clear that pulse shape is characterized mainly by a high degree of variability.

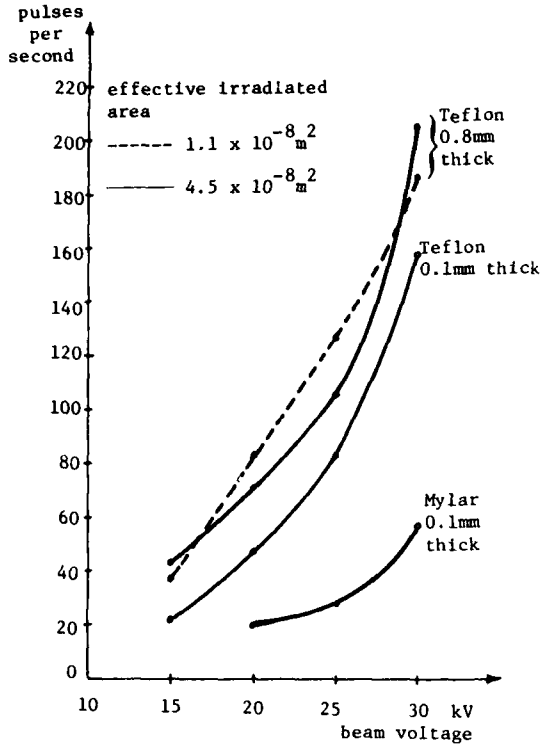


Fig. 8 Micro-discharge frequency of occurrence.

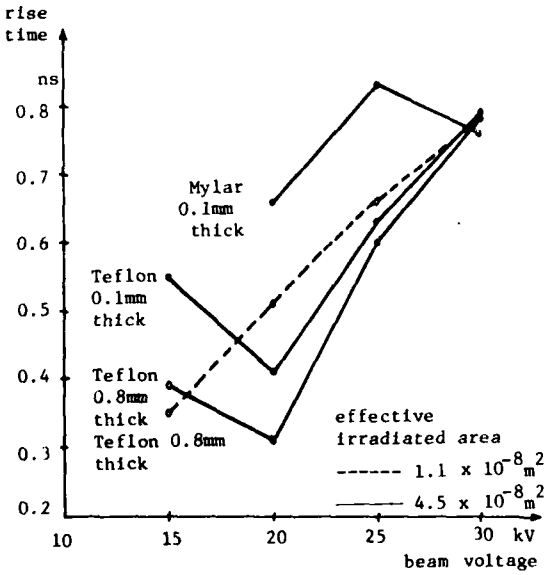


Fig. 9a Micro-discharge current pulse average rise time.

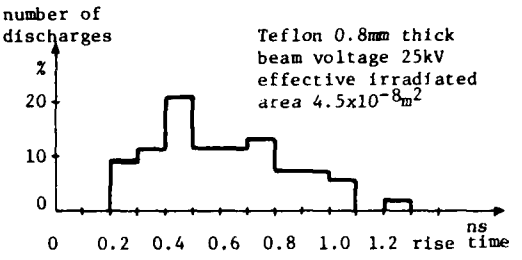


Fig. 9b Typical rise time histogram, calculated from 57 pulse envelopes.

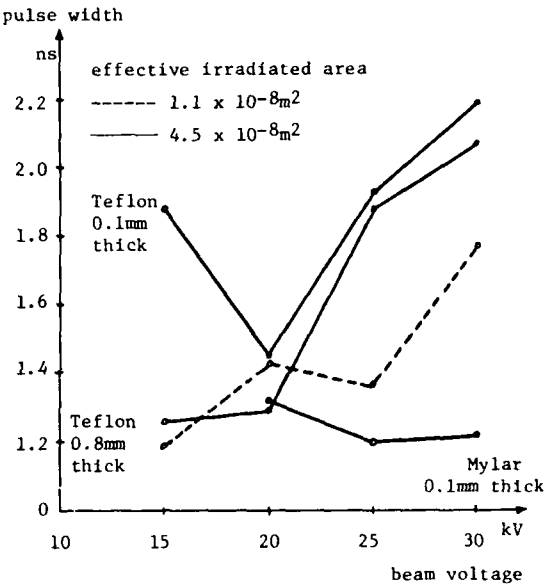


Fig. 10a Micro-discharge average pulse width.

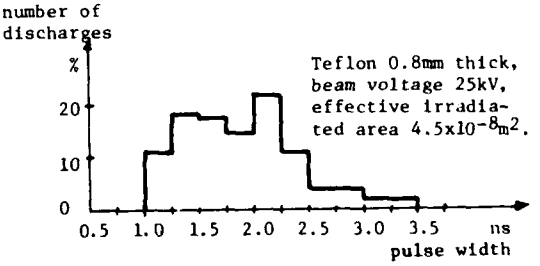


Fig. 10b Typical pulse width histogram, calculated from 57 pulse envelopes.

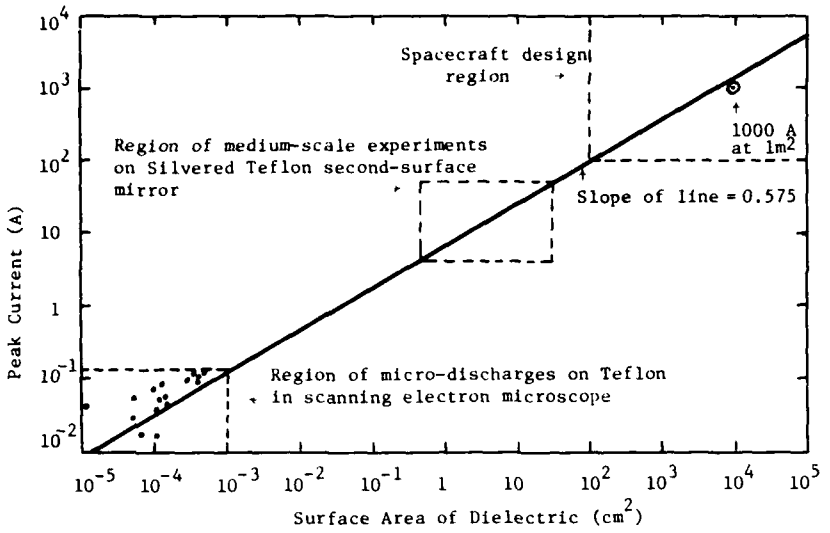


Fig. 11 Extrapolation of macro-discharge peak-current scaling law for Teflon (from Fig. 3a). The dots indicate micro-discharge data for Teflon (from Fig. 6, without distinction as to beam voltage).

CONCLUSIONS

A definite tendency for discharge peak current to increase with increasing charged area has been identified for both macro-discharges and micro-discharges. The experimental data from both regimes is plotted on the same graph in Fig. 11, in which the extrapolated macro-discharge straight-line characteristic can be seen to pass close to the micro-discharge experimental points. The fit of this line to the micro-discharge points would be very close if the effective charged area were larger by a factor of about two, a realistic possibility. The existence of this simple power-law relationship between current and area is the first concrete evidence that macro-discharges and micro-discharges are just different orders of magnitude of the same basic phenomenon.

Because there is no indication of flattening out of the current-versus-area characteristic, extrapolation to larger effective charged areas may be justified. The extrapolated characteristic of Fig. 11 passes above 1000 A at $1m^2$, suggestive of extremely strong electromagnetic interference effects for spacecraft-size metal-backed dielectric sheets.

For Mylar the existence of a simple current-versus-area scaling characteristic appears much less certain. Within the ranges of parameters tested, compared to Teflon, Mylar exhibits less frequent but stronger discharges.

In general, it can be concluded that standard and modified scanning electron microscopes can provide discharge data over very wide ranges of parameters. Therefore, larger and more expensive vacuum chambers appear to be unnecessary for much of the experimentation necessary to gain an understanding of surface discharge arcs and to develop new and better spacecraft materials.

ACKNOWLEDGMENTS

The authors acknowledge with pleasure the contributions of M. Palfreyman, J.V. Gore, W.L. Lehn and W.C. Nixon. The research was supported by the Communications Research Centre, Department of Communications, Canada, under Department of Supply and Services Contract No. OSU 76-00064. Permission by the Department of Communications to publish the results of this study is acknowledged. Support was also provided by the National Research Council of Canada, under Grant No. A-4140.

REFERENCES

Adamo, R.C. and J.E. Nanevicz (1976), Spacecraft-charging studies of voltage breakdown processes on spacecraft thermal control mirrors, in Spacecraft Charging by Magnetospheric Plasmas, A. Rosen (Ed.), Progress in Astronautics and Aeronautics, 47, 225-235.

Balmain, K.G. (1973), Charging of spacecraft materials simulated in a scanning electron microscope, Electronics Letters, 9, No. 23 544-546.

Balmain, K.G., M. Orszag, and P. Kremer (1976), Surface discharges on spacecraft dielectrics in a scanning electron microscope, in Spacecraft Charging by Magnetospheric Plasmas, A. Rosen (Ed.), Progress in Astronautics and Aeronautics, 47, 213-223.

De Forest, S.E. (1972), Spacecraft charging at synchronous orbit, J. Geophys. Res., 77, 651-659.

Fredricks, R.W., and Scarf, F.L. (1973), Observations of spacecraft charging effects in energetic plasma regions, in Photon and Particle Interactions with Surfaces in Space, Reidel, Dordrecht-Holland, 277-308.

Gross, B. (1957), Irradiation effects in borosilicate glass, Phys. Rev., 107, 368-373.

Gross, B. (1958), Irradiation effects in Plexiglas, J. Polymer Sci., 27, 135-143.

Hoffmaster, D.K. and J.M. Sellen Jr. (1976), Spacecraft material response to geosynchronous substorm conditions, in Spacecraft Charging in Magnetospheric Plasmas, A. Rosen (Ed.), Progress in Astronautics and Aeronautics, 47, 185-211.

Meulenberg, A., Jr. (1976), Evidence for a new discharge mechanism for dielectrics in a plasma, in Spacecraft Charging by Magnetospheric Plasmas, A. Rosen (Ed.), Progress in Astronautics and Aeronautics, 47, 237-246.

Rosen, A. (1975), Large discharges and arcs on spacecraft, Astronautics and Aeronautics, June, 36-44.

Rosen, A.(1976 a), Spacecraft charging: Environment-induced anomalies, J. Spacecraft and Rockets, 13, 129-136.

Rosen, A., Ed. (1976 b), Spacecraft Charging by Magnetospheric Plasmas, Progress in Astronautics and Aeronautics, 47.

Rosen, A. (1976 c), Spacecraft charging by magnetospheric plasmas, IEEE Transactions on Nuclear Science, NS-23, 1762-1768.

Stevens, N.J., R.R. Lovell, and J.V. Gore (1976), Spacecraft-charging investigation for the CTS project, in Spacecraft Charging by Magnetospheric Plasmas, A. Rosen (Ed.), Progress in Astronautics and Aeronautics, 47, 263-275.

A PROPOSED MECHANISM FOR THE INITIATION AND PROPAGATION OF DIELECTRIC SURFACE DISCHARGES

G. T. Inouye and J. M. Sellen, Jr.
TRW Defense and Space Systems Group
Redondo Beach, California

INTRODUCTION

This technical brief will examine the possible mechanisms under which negatively charged dielectric surfaces may release the stored charge on these surfaces. To be specifically relevant to the problems of differentially charged spacecraft in space, attention here will be restricted to the surface charge on very thin (less than .1 mm) polymeric films whose outer (space exposed) face is dielectric material and whose rear face contains a metallized layer which is electrically connected to the spacecraft ground. Such dielectric/metal films correspond, thus, to the thermal control blanket materials commonly used on spacecraft.

The specific concern of this technical brief is that condition in which the outer face of the dielectric film is charged to large negative potentials (in excess of 10 kilovolts) with respect to the rear face metal film. Such conditions of severe differential chargeup are believed to be present on many spacecraft at geosynchronous orbits during periods of magnetic substorm activity. The accumulation of negative charge on non-sunlit dielectric surfaces can result in differential chargeup voltages across the dielectric film in the range from 10 to 20 kilovolts if sufficient photoelectron release from spacecraft conducting surfaces can occur as a result of solar ultraviolet absorption on those conducting surfaces.

If sufficient differential chargeup across the dielectric films is present, discharge of the stored charge and the stored energy across the dielectric surface can result. These discharges will be termed dielectric-to-metal arcs to distinguish the (dielectric) cathode source of the discharge electrons. In space, the actual course of the arc electron currents may be from the dielectric surface to spacecraft metal surfaces (and thence to the rear face metallized film) or from the dielectric surface to the space plasma (with the subsequent motion of a space plasma electron from space to the exposed conducting portions of the spacecraft, and, again, to the rear face metallized films). The principal concerns of such surface discharges will be the total charge contained in the charge relocation transient and the peak levels of charge relocation. These total charges and maximum relocation currents can affect the operation of spacecraft systems and, if sufficiently large, can degrade the conduction properties of the ground straps from the spacecraft ground to the rear face metallized films.

INITIATION AND PROPAGATION OF THE DIELECTRIC SURFACE DISCHARGE

It is commonly observed in laboratory experiments of dielectric films under conditions of severe differential chargeup (voltage across the dielectric films in excess of 10 kilovolts and electric stress levels in excess of 1 megavolt per centimeter) that comparatively large areas of the film can be discharged. The explanation of these dielectric "clean-offs" can be approached from either one of two widely divergent viewpoints. In the first method of explanation, the discharge of the dielectric surface initiates at a single point and then propagates across the dielectric surface with the majority of the charge release at any given point on the surface occurring as the discharge wave propagates past the given point. Another possible direction of explanation for the surface clean-off is that, while the discharge may initiate at a single point, the charge release from the dielectric surface takes place at the same time over the bulk of the surface area being discharged. The explanation to be used in this technical brief is that discharge initiates at a single point on the surface and then propagates as a discharge wave. The principal areas of interest relating to such postulated processes are the mechanism of propagation, the speeds of propagation, the allowed extent of the propagation wave before extinction, and the effectiveness of the dielectric surface clean-off as the wave progresses past any given point on the surface.

Figure 1 illustrates the potential (as a function of position) along the exterior of a dielectric surface for a surface at high level negative chargeup, and indicates the sudden realteration of that potential under an assumed breakdown of the dielectric. The breakdown

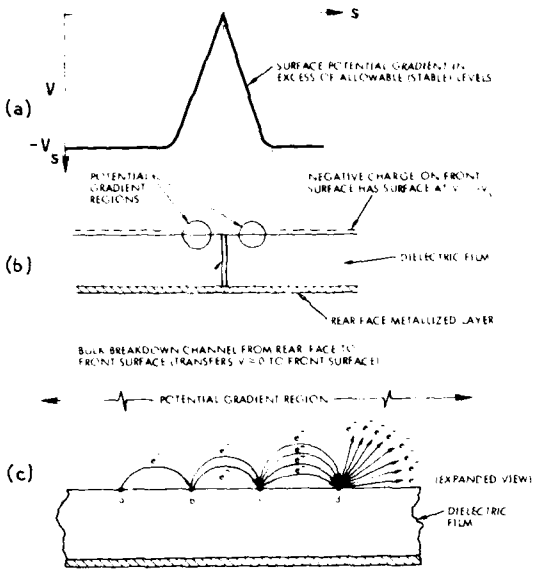


Fig. 1. Dielectric bulk breakdown, formation of a surface potential gradient region and secondary electron multiplication in the potential gradient region of a severely stressed dielectric film.

assumed here is a point breakdown through the bulk of the dielectric, but, it should be emphasized, that similar overall dielectric surface potential gradients may be expected if the breakdown is a surface breakdown (of limited extent) at the edge of the dielectric. In either instance, the surface of the dielectric now is the site of surface potential gradients which may range as high as the previously existing bulk potential gradients (i.e., at levels in excess of 1 megavolt/cm).

It is commonly known that dielectric surfaces cannot sustain surface potential gradients significantly above levels of $\sim 10^4$ volts/cm and Fig. 1 illustrates the action of a very small element of such a stressed surface under such levels of electric field. In the process illustrated there, an electron leaving the surface at point a is accelerated in the surface potential gradient field and strikes the surface at point b. If the electron energy gain over the distance from a to b exceeds the level at which secondary electron emission is in excess of unity, then the charge release at point b will exceed the entering charge at that point and a growing quantity of charge will continue across the surface from point b to point c, etc. as illustrated in Fig. 1.

The basic mechanism in Fig. 1 can result in a propagating discharge wave if the action of the charge release is, somehow, to preserve the conditions that severe surface potential gradients continue to exist on the surface. If, on the other hand, the release of charge on the surface causes a diminution of surface electric field, then the discharge wave might be expected to extinguish. Under many types of physical behavior, a point initiated disturbance which attempts to propagate into a broader spatial region will suffer a diminution of its driving force because of the continued increase in the size of the disturbance front and, hence, will have only limited propagation. For the surface discharge wave (which is assumed here to be point initiated) to propagate over these broad regions of dielectric surface, will require some mechanism which somehow "steepens" the surface potential gradient wave and maintains steepness of the wave above critical levels as the wave progresses.

Figure 2 illustrates the required wave steepening action which allows the wave to continue to propagate into ever broadening regions without an extinction of its driving force. The explanation of the wave steepening is the secondary electron multiplication gain which causes the surface potential at the trailing part of the wave to advance relative to the head of the wave until the head of the wave is forced forward by the steepening potential gradient along the surface.

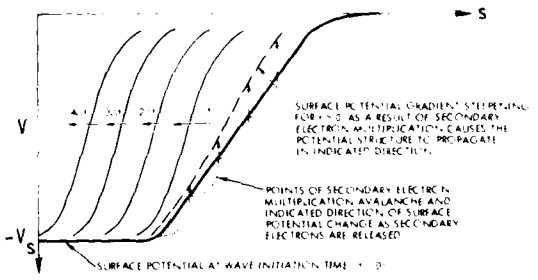


Fig. 2. Surface potential gradient steepening action as a result of secondary electron multiplication avalanching and indicated propagation of the surface potential wave structure.

ELECTRON CURRENT RELEASE LIMITATIONS IN DIELECTRIC SURFACE DISCHARGES

While the mechanisms illustrated in Fig. 2 would allow the discharge to propagate (and to propagate indefinitely provided only that

$$\vec{E}_s \cdot \vec{\lambda} > V_I, \text{ where } \vec{E}_s$$

is the surface electric field, $\vec{\lambda}$ is the electron hopping distance and V_I is the required electron acceleration voltage for the secondary electron emission ratio to exceed unity), there remain major unanswered questions over the disposition of the electron charge released from the surface through the secondary electron multiplication wave process. Initially (i.e., at the discharge outset), such electrons would be attracted to space (assumed here to be at a positive potential with respect to the dielectric surface) or to the exposed conducting portions of the spacecraft (also at positive potentials). It should

be noted, however, that as the area of discharged surface increases that, ultimately, many regions of the previously charged dielectric surface are now at positive potentials (as a result of the discharge) with respect to the surface potentials at the site of the still propagating surface discharge wave and, hence, may attract the electrons being released at the propagating wave site to initiate a "re-charge" process at the previously "cleaned-off" surface locations. This discharge/recharge action, if it occurs, would tend to limit the extent of final surface clean-off and it should not be expected that total surface clean-off would result.

The experience of various investigators has led many to conclude that the surface clean-offs were almost total. Experiments at TRW have ranged from very high levels of clean-off (nearing 100% within the accuracy of the measurement) to values more nearly 30%. Using dielectric surfaces in which the rear face metallized film has been segmented and with separate current leads to the separate rear face metal segments, some evidence has been found in the TRW measurements, that surface discharge clean-offs on one area of the dielectric may be re-depositing surface charge in other (previously discharged) areas of the dielectric. These particular measurements at TRW (which were undertaken primarily to determine the propagation speeds, if any, of surface discharge waves) did reveal some evidence of site-to-site charge relocation currents which can have implications relative to current recirculation loops in the spacecraft frame and in the several grounding loops which may be in place for a single, extended area, dielectric film and rear facing metallized layer.

If electrons from the discharge surface, and upon being released from that surface, must be drawn to space, then space charge limitations between the areas of electron release and space can limit the total current transported to space at any one instant. In the electron flow to space the effective current carrying configuration is that of a spherical diode in which the dielectric surface is the cathode and the space is the anode. The perveance of such diodes is typically of "unit" perveance levels (i.e., at 10^{-6} amperes per volt to the three halves power), and, hence, for a 10 kilovolt differential potential between the dielectric surface and space, electron release currents as large as the ampere level might be expected to flow before space charge limitations set in. This assumes, of course, that the potential of the underlying metallic spacecraft structure is unchanged during this emission process. If the replacement or recollected plasma electron currents to the underlying metallic surface are insufficient to maintain its potential, the space charge limited currents will be even smaller. It is also assumed that the secondary electron multiplication process on the dielectric surface is capable of releasing amperes of electrons and that the overall flow process is space charge limited, rather than emission limited.

Under the assumed process of space charge limited flow from the dielectric surface to space and with the perveance limitations presented there, it would appear that the electron release currents from large area dielectric surfaces would not materially exceed those of small area dielectric surfaces but merely that the discharge time of the large surfaces would be longer than those discharge times of smaller areas. This conclusion does not agree with the recent ESTEC[1,2] measurements in which both the magnitude of the surface electron release currents and the duration of the current release pulse were observed to increase as the extent of the charged surface area increased. It should be noted, however, that the magnitudes of released electron currents did not scale linearly with surface area but were at a somewhat reduced rate (increases by a factor of 7 in $I_{discharge}$ were observed for surface area increases in the order of 20) and it may be that the surface discharge waves propagating out over the larger dielectric surface have, because of their greater dispersion, a higher perveance with respect to space than the more compressed "cathode" location sites of the smaller dielectric samples used in these measurements.

Another possible limitation, in principle, in the electron release current is the velocity of the discharge wave propagation across the dielectric surface. If it is assumed that the passage of the discharge wave over a dielectric surface element causes a diminution of ΔV in the chargeup voltage across the film, the total release current by a discharge wave moving across a surface is

$$I_e = \dot{A}\,\hat{C}\,\Delta V \qquad (1)$$

where $\dot{A}$ is the area of surface discharged per unit time and $\hat{C}$ is the capacitance of the dielectric film/metallized film per unit area. If the discharge wave front is burning across a front width of W and if the wave propagation speed is v_{dw}, then

$$\dot{A} = Wv_{dw} \qquad (2)$$

and

$$I_e = Wv_{dw}\,\hat{C}\,\Delta V \qquad (3)$$

For capacitance values of 30 picofarads/cm^2, a differential voltage change of 10 kilovolts as a result of discharge wave passage (perhaps a high side estimate), then

$$I_e(\text{amperes}) = W(\text{cm})v_{dw}(\text{cm/sec})(3\times10^{-7}) \qquad (3')$$

and for a wave front length of 10 centimeters and a propagation wave velocity of 10^6 cm/sec,

$$I_e \sim 3 \text{ amperes}, \qquad (3'')$$

a typical clean-off current release observed for

small sample dielectrics. If, on the other hand, W ~ 100 centimeters and $v_{dw} = 10^7$ cm/sec, I_e rises to levels of several hundred amperes, as has been observed in the ESTEC experiments.

In searching for possible limitations to the electron release current, three such limitations may be noted. The first of these is that

$$\Delta V < V_{chargeup} \qquad (4)$$

and this limitation has been taken into account in the calculations above. A second limitation is that the width of the "burning" front of the discharge wave must be somehow related to the scale size of the sample and to the ability of the wave to propagate from an initially small discharge into the larger scale discharge. If the dielectric sample is a square of length, L, along each side

$$W < 4L \qquad (5)$$

by geometrical considerations alone. The final limitation is in the velocity of the propagating wave. Because the discharge in this postulated behavior is created by secondary electron multiplication

$$v_{dw} < \bar{v}_{se} \qquad (6)$$

where $\bar{v}_{se}$ is the average velocity of a secondary electron in the avalanche process. For secondary electron energies at an average value of several hundred electron volts,

$$v_{se} \sim 3 \times 10^8 \text{ cm/sec} \qquad (7)$$

If all of these upper bound limitations are utilized in Eq. (3) then a maximum release current (neglecting space charge effects and the resultant perveance limitations) is

$$I_e < 3 \times 10^4 \text{ amperes} \qquad (8)$$

for L = 100 cm and $v_{se} = 3 \times 10^8$ cm/sec. It is obvious that such large currents do not, in practice, take place so that other limiting mechanisms must be in effect.

The use of $v_{dw} < v_{se}$ as a limitation is certainly an upper bound to the discharge wave velocity. In point of fact, the pulse length for discharge currents from both small and medium sized dielectric samples tend to indicate that discharge wave propagation speeds must be in range from 10^6 to 10^7 cm/sec and, hence, are considerably less than v_{se}. The limitation on discharge wave propagation speed is probably limited to the rate of charge extraction at the head of the discharge wave rather than to secondary electron velocity in the wave mid-point. The assumed potential configuration in which the dielectric surface is at negative potential and the spacecraft frame is at the space plasma potential (infinity) is a valid starting point configuration. The release of negative charge from the outer face of the dielectric may or may not result in a positive excursion in spacecraft frame potential. If the released electrons move into the space plasma, then the spacecraft frame will move to positive potentials with respect to the space plasma and will extract electrons from that plasma to complete the current circuit. If, on the other hand, the electrons from the dielectric outer surface move to exposed conducting elements of the spacecraft frame, then only minor changes of the spacecraft frame potential occurs.

The degree to which these limitations are firm and the degree to which such limitations may be lifted through configurational changes remains to be determined. However, the present evidence is that peak discharge currents observed in the larger sample ESTEC measurements exceeded levels of 100 amperes, and, because there may be even larger currents than those observed which can flow from dielectric cathodes to positive potential locations or regions if those locations are more closely spaced to the dielectric cathode location, the examination of such surface discharge currents remains as an important (and presently uncompleted) task. What is apparent at present is that the surface clean-off mechanism can propagate over larger areas, and, in the larger area propagation case, can release large currents (in excess of 100 amperes) in these larger geometry situations. Charge-recharge currents of this order of magnitude can also be postulated between various discharging areas of dielectric and previously discharged areas of the dielectric.

SUMMARY

This technical brief has advanced a proposed mechanism for the initiation and propagation of dielectric surface discharges. The predicted behavior of dielectric surfaces under the mechanisms proposed here would be consistent with the presently known patterns of behavior of such surfaces. It should be emphasized, however, that the experimental data base in this surface chargeup and discharge area is still comparatively limited and that considerably more complicated behavior may, in fact, be present than is contained in this simplified discharge model. Because of the magnitudes of the peak discharge currents thus far observed and because of the comparatively large areas of dielectric which may (at least partially) clean off during such discharges, it would appear that a continued effort should be made to more thoroughly understand this discharge process.

REFERENCES

1. Bogus, K. P., "Investigation of a CTS Solar Cell Test Patch Under Simulated Geometric Substorm Charging Conditions," Proceedings of the Spacecraft Charging Technology Conference, AFGL-TR-77-0051, NASA TMX-73537, Feb. 24, 1977.

2. Private communication from K. P. Bogus, July 8, 1977, material to be published.

REDUCTION OF SPACECRAFT CHARGING USING HIGHLY EMISSIVE SURFACE MATERIALS

A. G. Rubin, P. L. Rothwell and G. K. Yates
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

ABSTRACT

Recently it has been found that electrical charging of spacecraft to high potentials takes place on satellites at high altitudes, in particular at synchronous orbit. There is evidence that arcing occurs, which degrades thermal control surfaces and damages electronics. Spacecraft charging is discussed by S. DeForest[1] and A. Rosen.[2]

It is not at present possible to eliminate spacecraft charging by simple expedients such as a metallic spacecraft surface, less complicated geometries or better Faraday cages because of constraints of weight, thermal control and satellite operational requirements.

The purpose of the present paper is to examine the possibility of preventing spacecraft from charging to high potentials by utilizing surface materials with a high secondary emission coefficient, such as teflon and quartz. A plasma simulation code developed by the authors is run for representative cases. It is shown that spacecraft charging is prevented over a wider range of ambient plasma temperatures for more emissive materials.

I. INTRODUCTION

It is now well established that spacecraft can attain very high voltages relative to the ambient environment. The design of spacecraft in terms of geometries and surface materials properties determines the resulting differential charging between various locations on the spacecraft. For a more recent summary we refer the interested reader to the Proceedings of the Spacecraft Charging Technology Conference.[3]

As Knott[4] previously showed, materials with high secondary emission yield tend to reduce the maximum potentials to which spacecraft charge. With the advent of computer simulation techniques and further measurements of the secondary emission coefficients of polymer materials, it is now possible to carry out a more detailed study. It is shown here, not only that secondary emission reduces the average spacecraft potential, but more precisely, this effect comes about because of the interplay between the distribution of incident particle energies and the secondary emission yield curve of the surface materials. This effect can be used to advantage to reduce the magnitude of spacecraft charging.

For materials with secondary emission coefficient greater than unity, we show that charging is due only to the high-energy tail of the distribution of incident particles. Given this result, we show that charging can be mitigated by employing materials whose secondary emission remains greater than unity to high incident energies. The magnitude of charging is calculated for several materials and for a range of incident energies to show in detail what potentials can be expected.

II. APPROACH

The objective of the computer model is to realistically simulate plasma-spacecraft interactions.[5,6] It presently treats time-dependent plasma phenomena in the limit of spherical symmetry. Although future plans anticipate the incorporation of a realistic three-dimensional spacecraft geometry, understanding of the spherically symmetric limit uniquely identifies plasma effects.

We use what is commonly called a "particle pusher" model. That is, one approximates the actual plasma by a number of "computer" particles whose charge,

position, angular momentum and velocity are tracked in time. By properly weighting these computer particles, reasonable statistics can be obtained near the spacecraft.[5] Appropriate particle distributions are generated either by a Monte Carlo technique or by a systematic loading of the velocity and spatial intervals ("quiet start").[6]

For purposes of comparison we have initially restricted ourselves to Maxwellian distributions. However, both methods (Monte Carlo and Quiet Start) can easily be extended to any distribution that can be numerically integrated. Once the neutral plasma is created, the computer tracks the particles and recalculates the potential at each time step. Particles that hit or are emitted from the spacecraft are taken into account as well as those that enter and exit the sheath boundary. Thus, the computerized sheath structure dynamically evolves in analogy with the physical situation.

This computer code treats secondaries by emitting a properly weighted electron for each electron incident on the surface. The energy of the emitted electron is consistent with a Maxwellian distribution with a temperature of 2.5 eV. Backscattering is taken into account by also reemitting the incident electron with an energy consistent with that shown by Sternglass. The amount of backscattering is determined by a random number generator that allows the total backscattering to be in agreement with experimental data.

The secondary emission yield is given by[7]

$$\delta/\delta_m = e^2 \frac{E}{E_m} \exp\left[-2\left(E/E_m\right)^{1/2}\right] \qquad (1)$$

where E is the incident energy and δ is the secondary yield at E. The parameter e is Euler's constant and the parameter δ_m is the maximum yield at E_m.

Figure 1 shows the secondary emission yield for various polymers. It is seen that kapton, for example, has a low yield while teflon has a high yield ($\delta \geq 1$) up to 1.75 keV.

III. RESULTS AND CONCLUSIONS

Figure 2 shows our results for two different yield curves. The top curves are the yields as calculated from equation (1). The bottom curves represent the surface potential as a function of the ambient electron temperature. The points on the bottom curves were obtained by simulating the presence of the spacecraft in a neutral plasma at t = o. The surface potential then changes in a time-dependent manner until entering and exiting surface fluxes are in equilibrium. This is the steady-state condition and is represented by the points shown. The number of points is restricted by the length of computer time needed to reach the steady-state.

Note from Figure 2 that an increase in δ_m from 1.5 to 2.5 increases the energy range for low spacecraft potentials by at least a factor of two. The higher value of δ_m is consistent with the emission properties of teflon.[8,9] Materials with lower secondary emission[10] will have a dip in the surface voltage versus energy profile. (Note the left-hand side of Figure 2). This dip arises because $\delta \ll 1$ at low energies and the potential will be[5] $\sim 2.5\ kT_e$ depending on the ratio T_e/T_i. As more secondaries are emitted at higher incident energies, they eventually dominate over the ambient current and causes the spacecraft to go slightly positive. This creates the dip. The spacecraft potential is restricted on the positive side due to the low energy ($\sim$2.5 eV) of the emitted secondaries. Any positive increase in the floating potential increases the returning secondary current to the spacecraft and, hence, prevents further rise in the potential.

Figure 2 is without photoemission and, therefore, represents a spacecraft in the eclipse when most charging events take place.[1] Previous simulation results with photoemission showed that photo currents that are much less than the ambient current will cause large swings in the spacecraft surface potential as the spacecraft comes into sunlight. The same effect is seen here with secondary emission replacing the photo current. Small increases in the secondary emission current can cause large changes in the surface potentials.

The right-hand side of Figure 2 is consistent with the secondary emission properties of quartz and teflon. Figure 1 shows that kapton, which is used for the external coating of thermal blankets, has an especially low secondary emission yield. The present results indicate that other more emissive materials should be considered as possible replacements in order to inhibit spacecraft charging over a wider range of

environmental conditions.

REFERENCES

1. DeForest, S.E., Spacecraft Charging at Synchronous Orbit, J. Geophys. Res., 77, 651, 1972.

2. Rosen, Alan, "Large Discharge and Arcs on Spacecraft", Astronautics and Aeronautics, June 1975, pp. 36-44.

3. Pike, C.P. and R.R. Lovell (editors), "Proceedings of the Spacecraft Charging Technology Conference", AFGL-TR-77-0051, NASA TMX-73537, AFGL, Hanscom AFB, MA 01731.

4. Knott, K., "The Equilibrium Potential of a Magnetospheric Satellite in an Eclipse Situation", Planet. Space Sci., 1972, pp. 1137-1146.

5. Rothwell, P.L., A.G. Rubin, A.L. Pavel and L. Katz in "Spacecraft Charging by Magnetospheric Plasmas", Progress in Astronautics and Aeronautics, Vol 47, AIAA, N.Y. and MIT Press, Cambridge, MA (1976).

6. Rothwell, P.L., A.G. Rubin and G.K. Yates, p. 389, Ref 3.

7. Sternglass, E.J., Westinghouse Research Lab., Scientific Paper No. 1772 (1954).

8. Willis, R.F. and D.K. Skinner, "Secondary Electron Emission Yield Behaviour of Polymers", Solid State Communications, 13, 685 (1973).

9. Matshevich, T.L., Fiz. Tverd. Tela., Akad. Nauk. SSSR 1, 277 (1959) (in Russian).

10. Hachenberg O. and W. Brauer, "Secondary Electron Emission from Solids", in Advances in Electronics and Electron Physics, XI, p. 413-499, Academic Press, N.Y. 1959.

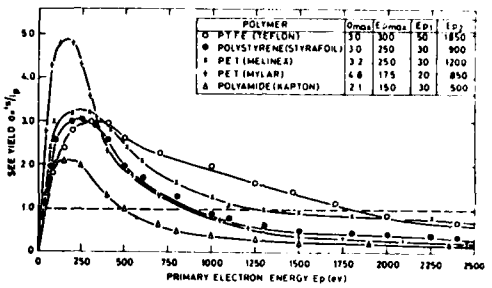


Figure 1. Secondary emission yield for various polymers as taken from Figure 1 of Reference 8.

EXPECTED SATELLITE POTENTIAL VS. THE AMBIENT ELECTRON TEMPERATURE FOR VARIOUS SURFACE MATERIALS

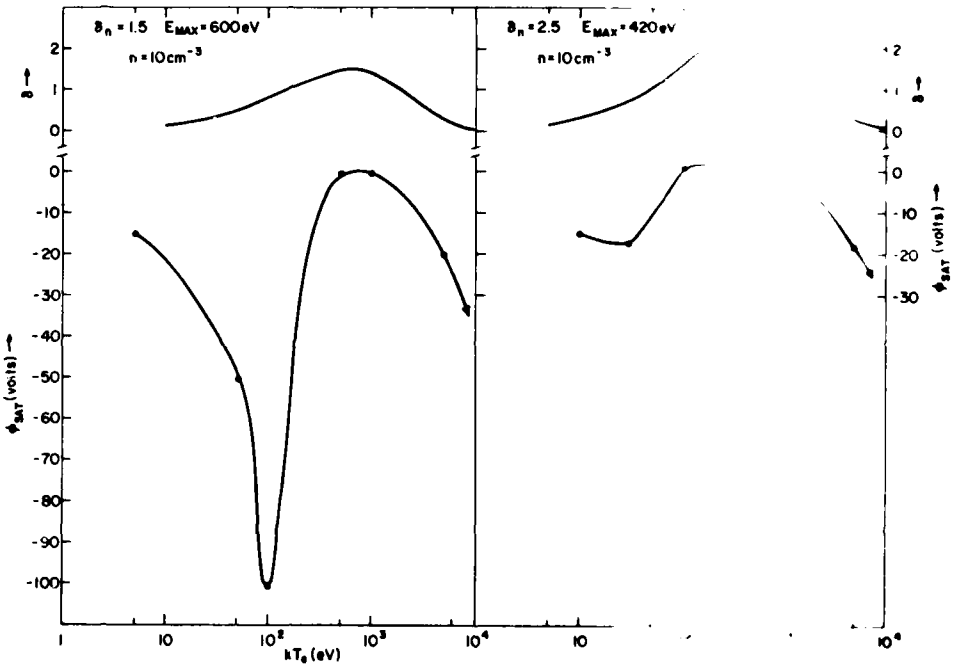


Figure 2. Expected satellite voltages vs the ambient plasma temperature for two separate cases of secondary emission density is taken as 10 particles/cc for all computer runs. The righthand side is for $\delta_m = 1.5$, $E_m = 600$ eV, $E_m = 420$ eV. The top curves describe the secondary emission electron energy. The bottom curves represent the satellite

A PLASMA BRIDGE NEUTRALIZER FOR THE NEUTRALIZATION OF DIFFERENTIALLY CHARGED SPACECRAFT SURFACES*

G. K. Komatsu and J. M. Sellen, Jr.

Systems Group Research Staff
TRW Defense and Space Systems Group
Redondo Beach, California 90278

INTRODUCTION

charge-up of spacecraft surfaces in
etic substorms and at geosynchronous
itudes is now a generally accepted
for the electrical equilibration of
ft under such conditions. A particu-
ere form of spacecraft charge-up,
cifically, differential charge-up,
hen the metallic frame of the space-
photoemissively clamped to the ambient
asma potential, while the dielectric
rfaces of the spacecraft are charged
ly to high potentials by the deposition
etic substorm electrons.

spacecraft which include an ion
in the system complement, it is of
to examine the possibility that charged
from the ion thruster can be used to
effects of these differential charge-
pacecraft surfaces. A specifically
ng possibility is that ions from the
lume of the plasma bridge discharge
er can be used to neutralize the de-
ubstorm electrons on spacecraft
c (insulating) surfaces.

not easily possible to duplicate, in
which occur for electrically isola
ft in the very dilute, but very ener-
arged particle environments of the
substorms. It is possible, however,
ne certain aspects of charged particle
n the plume of the plasma discharge
er to negatively charged surfaces,
these drainage current measurements
he charged particle flow behavior for
in space. It should be pointed out
pecific emphasis here is upon charged
diffusion in the presence of very dilute
plasma, such as those found at geo-
ous altitudes. For the comparatively dense (and comparatively nonenergetic) ambient plasmas encountered in the lower regions of the ionosphere, spacecraft charge-up is not of concern and the diffusion of charged particles from the thruster to spacecraft surfaces is considered to be significantly reduced relative to those diffusion levels encountered in the very dilute, geosynchronous region, plasmas.

This technical brief describes a series of measurements and the charged particle transport from the plasma discharge neutralizer of an ion thruster to various surfaces at differing distances and at varying potentials with respect to the thruster.

EXPERIMENT SETUP

In the initial planning of the experimental tests, the principal requirement of the experimental array was that it should be capable of detecting relatively small drainage currents of positively charged ions to various surfaces at variable negative bias potentials. The notion that the positive ion current flow would be at small levels follows generally from the space charge and [illegible] limitations on such flows. [illegible] space charge limitations in current flows in diodes cause [illegible] current density to diminish as the inverse square of the distance from the particle source to the particle collector. Because these distances can be comparatively large for spacecraft with ion thrusters, the experimental apparatus might anticipate that the currents over these large separation distances would be greatly reduced. A second major factor in the reduction of the current is the high mass of the Hg^+ ions in the discharge neutralizer plume (noting that the space charge limited current is inversely proportional to the square root of the charged particle mass). Another requirement of the experimental array is that it should be capable of detecting charged particle flows to several different surfaces which, if possible, should be at various distances from the thruster and its neutralizer.

*This [illegible]
[illegible] ork supported under Contract NAS
[illegible] for NASA-Lewis Research Center.

BEST AVAILABLE COPY

Figure 1 illustrates features of the experimental array in these current drainage tests. The ion thruster was installed in the 5 by 10 foot test chamber as illustrated. The ion thruster sputter shield was in place in view of anticipated future spacecraft operation of the thruster with such a sputter shield in place. In the present experimental array the sputter shield was electrically isolated from the thruster and could be separately biased for charged particle extraction with appropriate measurement of that drainage current.

Also present in the experimental array in Fig. 1 was an electrically isolated cylindrical electrode, the upper shroud, the lower shroud, and thrust beam collector. The upper shroud is electrically isolated for drainage current measurements as is the combined lower shroud and thrust beam collector.

DRAINAGE CURRENT MEASUREMENTS

During test, drainage currents were measured as a function of bias potential to: the cylinder, the upper shroud, and the thruster sputter shield. The thruster neutralizer discharge was established at nominal current and voltage on the discharge keeper electrode.

Measurement of Drainage Current to the Cylinder

When a negative bias potential is applied to the cylinder, a current of ions can be observed at this electrode. Figure 2 illustrates the ion current arriving at this electrode as a function of bias potential.

There are several features of the data in Figure 2 which are important. The first of these is the surprisingly high value of ion drainage current. For only small levels of bias (-10 volts), ion currents at the 40 microampere level can be observed. Such currents are clearly not consistent with a model of the ion current flow in which a physically small plasma cloud is the source of ions which emerge from this cloud to traverse the bulk of the separation distance from the thruster neutralizer to the electrode in a unipolar flow. The large values of ion current at low bias potentials, rather, indicate that the source of the ions is an extensive and dilute plasma cloud whose dimentions greatly exceed the dimensions of the thruster neutralizer proper and that ions moving from the neutralizer to the cylindrical electrode have the major

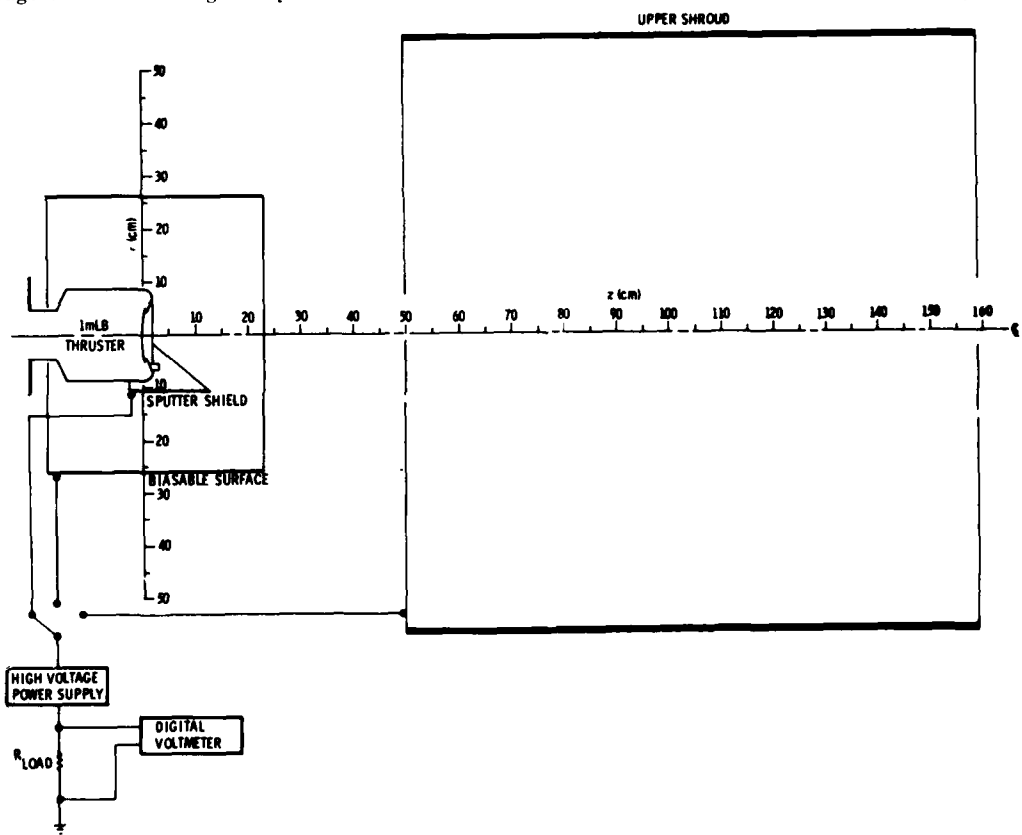


Figure 1. Overall Test Configuration for the Neutralization of Differentially Charged Surface Tests

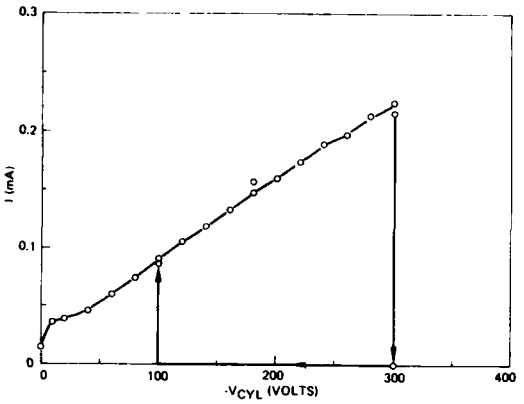


Figure 2. Ion Drainage Current to the Cylinder

portion of their traverse in plasma so that the space charge limitations of unipolar flow are not present.

The continued growth of the ion drainage current to levels of several hundred microamperes is also considered as an important feature. Because the deposition current densities in magnetic substorms are only at levels of a few nanoamperes per square centimeter, even spacecraft surface areas as large as 1 square meter may have electron deposition currents of only a few tens of microamperes. Ion drainage currents of several hundred microamperes, thus (as observed at ~300 volts of negative bias potential on the cylinder), could act to discharge differentially charged spacecraft surfaces of many square meters in area.

A third feature of importance in the data in Figure 2 is the abrupt cutoff in the ion drainage current to the cylinder for bias voltages more negative than 300 volts. As noted, the ion current falls abruptly to zero and remains at this low level for either positive or negative voltage motion of the extraction electrode. For more negative bias voltages, drainage currents remain small, even for bias potentials at several kilovolts (approximately -5 kV) of bias potential. For less negative bias voltages, drainage currents remain small until the cylinder potential approaches approximately -100 volts at which point the drainage current abruptly rises to the levels originally observed (for the I - V curve for the $0 < V_{CYL} < -300$ volt regime initially described.

The observed behavior suggests several possible processes described in the following hypothesized interaction. In the first regime, a dilute and extensive plasma cloud is hypothesized and ions diffusing outward in this cloud eventually move to the cylinder at only comparatively modest extraction potentials because most of the ion traversal is carried out in plasma (in the presence of electrons which mitigate the effects of ion space charge). As the cylinder bias is carried to more negative potentials, ion extraction current increases. At some upper bound limit on ion current, however, the plasma cloud is dissipated, collapsing to a much smaller physical size. In this more compact plasma cloud condition, the ions must move large distances without accompanying electrons and would appear at very low current levels as was originally anticipated. After the collapse of the plasma cloud volume, relaxations of the bias voltage do not act to restore the ambipolar diffusion until considerable relaxation of this bias has been exercised. At this lower bias point the ambipolar diffusion is once again active and the plasma volume once again enlarges so that significant ion current transport to the cylinder can take place at even modest cylinder bias voltages.

The hypothesized behavior above does not explain the method by which the plasma cloud collapse is initiated. One possible line of explanation could be an ion-electron instability which destabilizes the plasma at some upper end ion current passage. A second line of explanation can be that only a limited ion generation rate exists in the plasma plume and that the extraction of the entirety of this generated current to the cylinder causes the dissipation and collapse of the plasma. Either method of explanation remains speculative, however, in view of the present comparatively limited experimental evidence.

As a final treatment of the data in Figure 2, the I - V curve is restated in Figure 3 in log-log form and using, as the voltage variable, the term $V_K - V_{CYL}$ where V_K is the potential of the neutralizer keeper (which is, presumably, the potential at which the

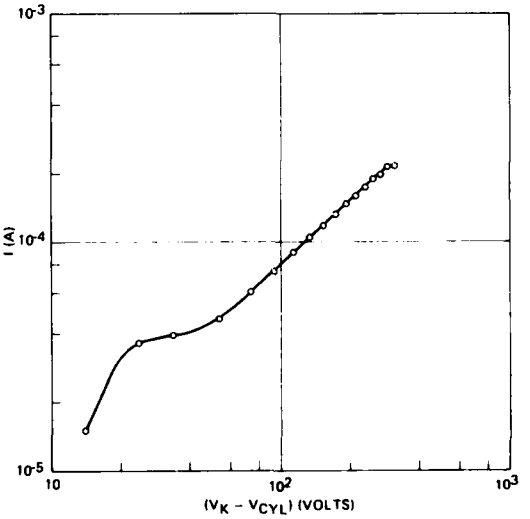


Figure 3. Ion Drainage Current to the Cylinder, Logarithmic Presentation

dilute plasma cloud forms) and V_{CYL} is the cylinder bias. Over the larger part of the I - V characteristic, the relationship is $I_{drainage} \alpha k (V_K - V_{CYL})$. The implications of these data are not completely understood at present.

Measurements of Drainage Current to the Upper Shroud

The upper shroud of the testing array is illustrated in Figure 1 and is a liquid nitrogen cooled liner for the 5 by 10 foot system. It is electrically isolated from the remaining shroud and the thrust beam collector and can be separately biased for ion current drainage from the neutralizer plasma plume.

In the measurements discussed above, this shroud was held at V = 0 while the electrical bias of the cylinder was varied. In the measurements described here, the cylinder is maintained at V = 0 while the shroud bias voltage is varied. For both this and the preceding data, the thruster sputter shield was held at V = 0.

Figure 4 illustrates the ion current to the upper shroud as the shroud bias is varied. The general behavior of the drainage current to this element is similar to that observed for ion flow to the cylinder. For comparatively small voltages, the shroud ion drainage reaches levels of ~200 microamperes. Further increases in shroud voltage, beyond ~-300 volts, cause an abrupt collapse of the observed ion flow which remains small, even for subsequent relaxations of the bias voltage, until the bias approaches -100 volts at which point the ion flow is re-established at the several hundred microampere level. The slope of the I - V curve is less than for the biased cylinder case.

The data in Figure 4 may be considered even more surprising in some respects than for the observed ion flow to the cylinder, in view of the generally larger distances of ion transport to

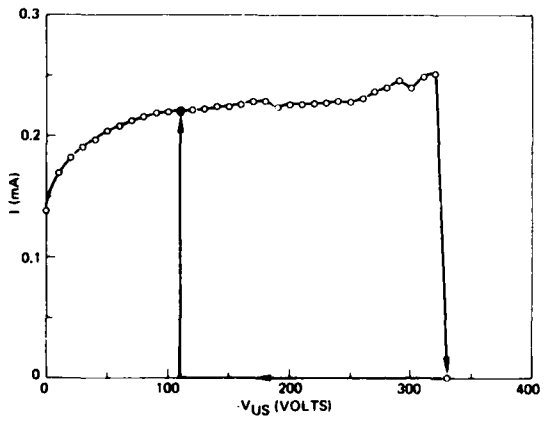


Figure 4. Ion Drainage Current to Upper Shroud

move from the thruster neutralizer to this downstream electrode. If the ion transport is, indeed, the result of the formation of a large and dilute plasma cloud emanating at the thruster neutralizer, then the dimensions of this cloud, from the standpoint of effective ion transport, must be of the order of 1 or more meters.

The abrupt cutoff of the ion drainage at ~240 microamperes is similar to the current level at cutoff for ion flow to the cylinder. In the above discussions of mechanisms for an implied plasma cloud collapse, it was considered that ion extraction beyond some level could exhaust the total ion production rate of the thruster neutralizer causing the plasma cloud to dissipate. If this mechanism is, indeed, present, the similarity of the cutoff levels in the two cases of ion flow would tend to be indicative of such a process.

Measurements of Drainage Current to the Thruster Sputter Shield

Figure 5 illustrates the ion current to the sputter shield as a function of sputter shield bias voltage. The behavior exhibited there is strongly similar to that behavior shown for ion flow to the cylinder and is relatively similar to the ion flow observed at the shroud. The hysteresis-loop aspects of the ion flow are present, and cutoff drainage level for hypothesized plasma cloud collapse is, again, at ~300 microamperes. As noted earlier, this similarity in ion current cutoff levels tends to support an exhaustion hypothesis for a plasma cloud collapse.

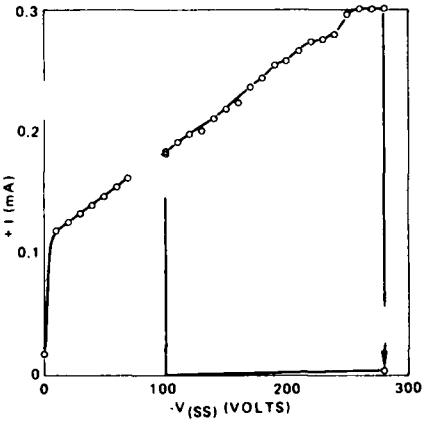


Figure 5. Ion Drainage Current to Sputter Shield

The thruster sputter shield was also biased positively to cause an extraction of electrons. Figure 6 illustrates these data. Electron current cutoff and plasma cloud collapse would not be expected for this condition and this level of current drainage and was not observed.

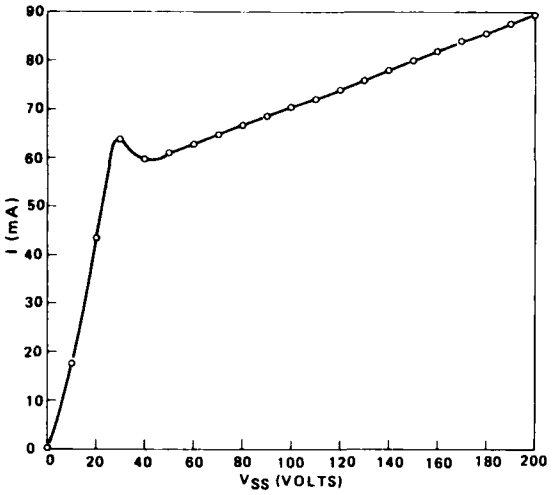


Figure 6. Electron Drainage Current to Sputter Shield

IMPLICATIONS OF THE OBSERVED ION CURRENT DRAINAGE BEHAVIOR

The principal feature of the data in Figures 2 through 5 is the comparative ease in which ion currents, at levels from tens to hundreds of microamperes, can be drawn from the thruster neutralizer plasma plume to relatively remote surface locations. If similar circumstances are present in space, then the flow of ions to negatively charged spacecraft surfaces should proceed at sufficient rates to match, essentially, the electron deposition levels on those surfaces. It is not positively determined, however, that all of the electrical equilibration features for spacecraft in energetic dilute charged particle environments in space can be simulated in laboratory facilities, and circumstances can be envisioned in which the various electric fields to and from various spacecraft surfaces could act to inhibit as well as to encourage the ion flow from the thruster neutralizer.

The observed phenomena of ion flow cutoff for extraction currents above a particular level (~300 microamperes) should not restrict possible applications of the plasma discharge neutralizer as a differential surface-charge neutralizer, at least for geosynchronous substorm conditions. As noted earlier, the electron depositions in such storms are at $\sim 10^{-9}$ A/cm^2 levels, and the 300 μA ion current production rate would represent a substorm electron deposition over an approximate area of 105 cm^2 (10 m^2). It would also appear possible to increase this ion current cutoff limit by suitable modification of the neutralizer.

SPACECRAFT CHARGING ON ATS-6

B. Johnson, J. Quinn, and S. DeForest
Physics Department, University of California, San Diego
La Jolla, California 92093

ABSTRACT

The UCSD Auroral Particles Experiment on the geosynchronous ATS-6 satellite has been used in a statistical study of spacecraft charging. Charging events and ambient plasma characteristics were studied for six 10-day periods during the first year of operation.

Events of several hundred volts negative occurred in the local nightside region on approximately half the days studied. On the majority of the days for which the spacecraft was eclipsed, potentials in excess of negative one thousand volts were observed. Effects due to differential charging were also observed. A correlative study has been undertaken to determine relationships between the plasma properties and the occurrence of charging.

INTRODUCTION

Since the launch of the ATS-6 satellite into a geostationary orbit in June of 1974, the UCSD Auroral Particles Experiment has collected an enormous wealth of data. Over 40 days of data were analyzed in an effort to obtain a reduced set of plasma parameters that would in some sense describe the environment at geostationary orbit.

It has been well established that spacecraft of varying configurations can frequently charge negatively to hundreds and sometimes thousands of volts (Pike and Lovell 1977, DeForest 1972, 1973; Reasoner et al 1975). The details of this charging are not totally understood, but are certainly related to the surrounding plasma. Thus, the plasma parameters were used to study the correlations between the plasma characteristics they represent and spacecraft charging.

The resulting data base can be made available to anyone interested. This paper will outline the contents of this data base and present the initial results of the study.

DESCRIPTION OF EXPERIMENT AND DATA

The UCSD Auroral Particles Experiment consists of five particle detectors. There are two rotating heads each containing a positive ion and electron detector. One head rotates in the north-south plane, while the other rotates in the east-west plane. The fifth particle detector is a fixed ion detector pointed eastward in the direction of the spacecraft motion. The rotating heads have a 220° range. Each detector can collect particles at 64 energy steps, ranging from 0 eV to 81 KeV, with the capability of dwelling at one particular energy step or scanning through all 64 steps in 16 seconds. The resolution of the particle analyzers is such that $\Delta E/E$ is approximately 20%. The angular resolution is approximately 2.5° by 6.4° for a flat spectrum. (A more detailed description of this instrument package is given in Mauk and McIlwain, 1975.) The data base was derived from 47 days of ATS-6 data, chosen for the quality and time of year. The days consist of a 20-day period from July 5 to July 24, 1974, a 10-day period from October 23 to November 1, 1974 and an 8-day period from February 10 to February 17, 1976. In addition, a 9-day period from September 23 to October 1, 1974 during which eclipse occurred was studied. On these 9 days, the satellite is eclipsed by the earth at local midnight, and charging events on the order of -1,000 volts are common. Data presented in this paper was taken in 1974 at 94° W. longitude.

Data from the north-south particle detector was studied with two purposes. The first purpose was to determine the potential of the spacecraft relative to the ambient plasma. The second purpose was to determine plasma parameters that would describe in some simplified sense the ambient plasma. A distribution function can be found from the data which when integrated from 0 to 81 KeV gives the four plasma parameters density, particle flux, energy density, and energy flux for the ions and electrons over this energy range.

Several points must be made concerning these integrals. The plasma parameters were obtained by integrating moments of the distribution function over velocity space, $d^3v = v^2 \sin\theta dv d\theta d\phi$. The angular dependence of the distribution function was assumed to be isotropic. This assumption is thought to be good for detector look directions making angles greater than 40^o with the local magnetic field (Mauk and McIlwain, 1975). The detected particles were assumed to be protons and electrons. It is known that other ions are present, but they are usually of lower abundance. Furthermore, to study the effects of the plasma on spacecraft charging, the effects of the spacecraft must be removed in doing the integrals. The plasma energies detected by the analyzers must be corrected to represent the true ambient plasma. For the ions, a negative spacecraft potential makes the particles arrive at the detector with a higher energy. Likewise, the electrons' energies are lowered. After correcting for the potential, the ion integral ranges from 0 eV to 81 KeV - $e\phi$, where ϕ is the magnitude of the potential and e is the fundamental unit of charge. This does not strongly affect the ions, but for electrons the corrected integrals range from $e\phi$ to 81 KeV + $e\phi$. Since most of the particle density is concentrated in the lower energies, the density integral may not be as representative of the plasma as the other three parameters which are weighted by higher powers of energy. For electrons, an additional correction must be made for secondaries and photoelectrons which are returned to the detector. Since these electrons are not representative of the ambient plasma, they must be excluded to prevent large densities from being incorrectly inferred. (For more details concerning the return of secondary electrons see Whipple, 1976.) After removal of the returned secondaries, the electron integrals range from $e\phi + E_R$ to 81 KeV + $e\phi$, where E_R is the maximum energy of the returned secondaries. No attempt was made to extrapolate the distribution function to energy ranges outside of the regime covered by the UCSD instrument.

Having the plasma parameters of density, particle flux, energy density, and energy flux, a procedure for relating these values to Maxwellian densities and temperatures was developed. The procedure is to find the two Maxwellian distribution functions which when integrated from zero to infinity yield the same four plasma parameters. The two Maxwellians are specified by densities n_1 and n_2, and temperatures T_1 and T_2. The values n_1, n_2, T_1, and T_2 are equivalent to the original parameters, but there is no guarantee that the Maxwellian distribution functions accurately represent the actual plasma distribution function. The reasons for this are that the Maxwellians were integrated to infinity instead of 81 KeV, and that the actual distribution function is not always well approximated by a sum of two Maxwellians. However, n_1, n_2, T_1, and T_2 may closely represent the ambient plasma under certain conditions, and a study of this is in progress.

DATA AND RESULTS

Data taken from hours 0-12 UT of July 22, 1974 is shown in spectrogram form in Figure 1. The upper half of the figure is electron data and the lower half is ion data. Intensity of the image varies logarithmically with particle count rate. A light region indicates high flux. Note that the energy scale for the ions is inverted. (For further description of the spectrogram, see DeForest and McIlwain, 1971.)

A typical charging event is shown between 8:00 and 10:30. An injection of hot electrons can be seen at 7:35. As the spacecraft charges negatively, low energy ions are accelerated to its surface, leaving an absence of detected ions with energy below the spacecraft potential. This is seen as a dark region in the spectrogram between 8:00 and 10:30, with a maximum potential of -340 volts being reached at 8:47. The bright rim around the charging region is due to cold ions being accelerated into the detector.

Effects of differential charging are manifested in the enhancement of the detected electron fluxes at energies somewhat below the spacecraft potential. These fluxes are secondary electrons which are returned to the spacecraft by an electric potential which is not radially monotonic (Whipple 1976).

Figure 2 shows the probability of negative charging events of magnitude greater than 10 volts (dark line) and greater than 100 volts (light line) at different local times, for 20 non-eclipse days. The occurrence of charging is seen to be localized in the nightside region. Two reasons for this localization are that substorms provide high electron fluxes in this region and that the plasma density tends to be low.

Figure 3 shows the frequency of occurrence of electron fluxes above 1.5×10^9 $(cm^2\text{-sec})^{-1}$ versus local time.

Unfortunately it is very difficult to obtain density measurements of an ion population with a temperature on the order of a few electron volts. When the spacecraft is not charged (or worse yet slightly positively charged), the detected flux from such a population is very small (or nonexistent), making density measurements inaccurate (or impossible). When the spacecraft is negatively charged the cold ions provide a high particle flux at the energy of the spacecraft potential, accounting for the bright rim seen on the spectrogram. However, the ions in this spike are detected at the energy of the spacecraft

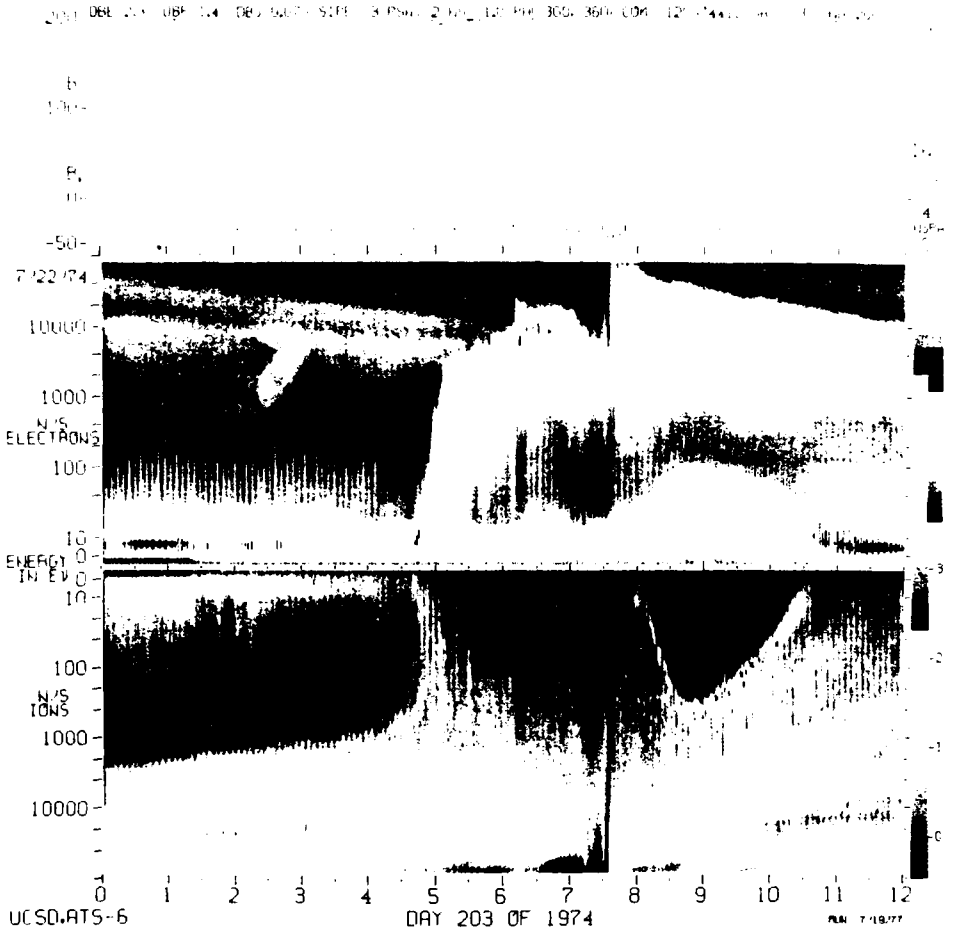


Figure 1. Data from July 22, 1974 in spectrogram form.

potential (hundreds of electron volts) while the necessary resolution for a density measurement is on the order of the particle temperature (around one electron volt).

Figure 4 shows the frequency of occurrence of ion density greater than 1 cm^{-3}. The probabilities were calculated using data taken when the spacecraft was uncharged within about 10 volts. It should be emphasized that the density measurements were made over an energy range of 0-81 KeV at the detector. A slightly positive spacecraft charge would eliminate detection of a cold ion plasma, and lead to an underestimate of the density.

Figure 5 is a plot of spacecraft potential versus electron energy density for a 20-day period. Each point is an average of 10 instrument scans corresponding to a time period of approximately 5 minutes. The data was taken when the detector look direction was within 20 degrees of perpendicular to the spacecraft. The content of this plot is clarified by recalling that for a Maxwellian distribution with density n and temperature T, the electron energy density is given by:

$$EED = \frac{3}{2} nkT \qquad (1)$$

Notice that for a given electron energy density there is a maximum potential to which the spacecraft will charge. It should be kept in mind that this plot contains many charging events, each with a different set of ambient plasma characteristics. A more orderly picture

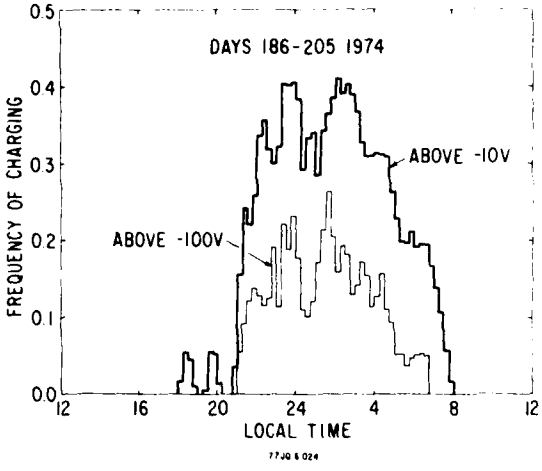


Figure 2. Distribution of spacecraft charging in local time.

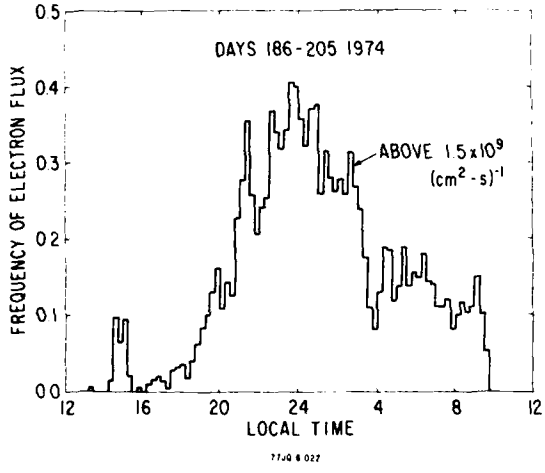


Figure 3. Distribution of electron flux in local time.

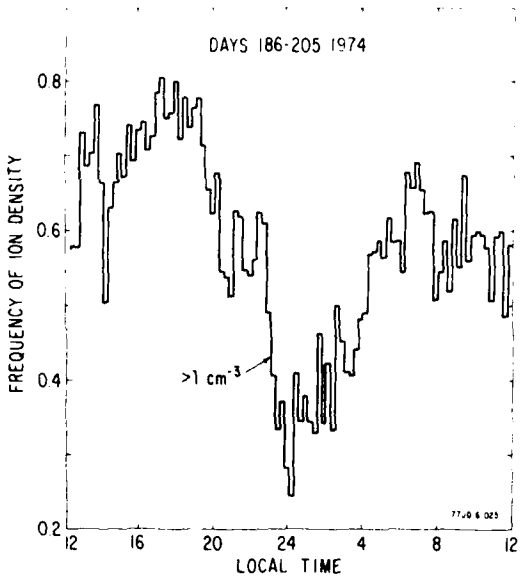


Figure 4. Distribution of ion density in local time.

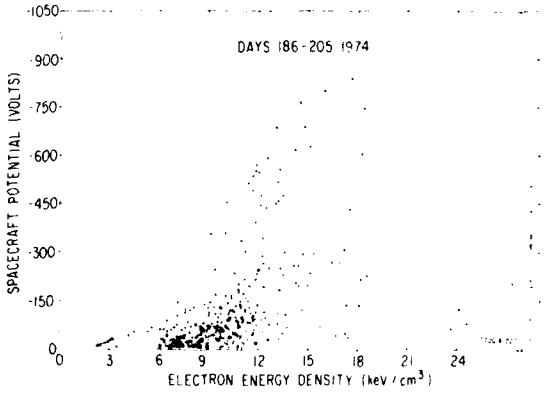


Figure 5. Correlation of spacecraft charging and electron energy density.

can be seen by studying a single event in greater detail.

The equilibrium potential of a spacecraft in the magnetospheric plasma is determined by a balance of several current sources.

$$J_e + J_{pe} + J_{se} + J_i + J_s = 0 \qquad (2)$$

J_e and J_i are the plasma electron and ion currents to the spacecraft. J_{se} and J_{pe} are the secondary and photoelectron currents and J_s is any current generated by the spacecraft, for example by firing an ion gun. For a discussion of controlling spacecraft charging by active means, see Olsen, Purvis, and Whipple (this conference). The dominant currents to an uncharged body are the plasma electron flux to the surface, and the photoelectron and secondary electron fluxes away from the surface. The plasma ion flux is relatively unimportant because of the slow thermal speed of ions compared to electrons. If a body develops a negative charge, cold ions which normally provide

a very small flux are drawn to the surface and provide a current which acts to reduce the charge.

Whipple (1965) has shown for a wide variety of conditions that the spacecraft potential is proportional to the electron temperature times a logarithmic function. Thus, it is useful to study the variation of potential as a function of electron temperature.

The temperature of a Maxwellian distribution is given by:

$$kT = EF/2PF \qquad (3)$$

where EF is the energy flux and PF is the particle flux. This same ratio may be used to derive a temperature for the non-Maxwellian electrons detected by ATS-6. Figure 6 is a plot of spacecraft potential versus this temperature for a 2-hour period of the charging event seen in Figure 1. The closed circles denote data from 8:00-9:00 UT, while the open circles are from 9:00-10:00 UT. During the first hour the potential rose to -340 volts at constant electron temperature, while in the course of the second hour the charge on the spacecraft fell approximately linearly with temperature.

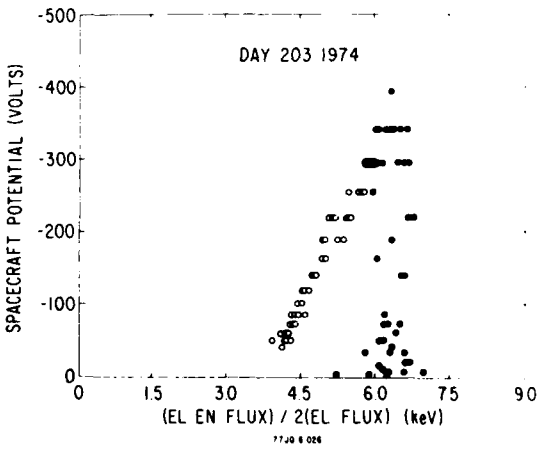


Figure 6. Dependence of spacecraft potential on electron "temperature".

A study of this and other charging events is proceeding with the goal of investigating several details of the charging process. The fact that the electron temperature was constant during the first hour of the charging event suggests that cold ion currents may be important. This behaviour seems consistent with a high density ion plasma providing a return current to the spacecraft and limiting the charge. The cold ion density presumably decreased with time, allowing the spacecraft potential to rise. Even in cases for which the electron fluxes dominate, the spacecraft charge is expected to depend strongly on the details of the spectrum. Figure 7 (Whipple 19) shows the total electron backscatter and secondary yield as a function of primary energy.

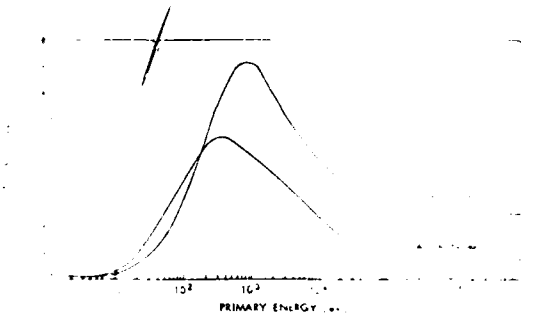


Figure 7. Total secondary electron yield.

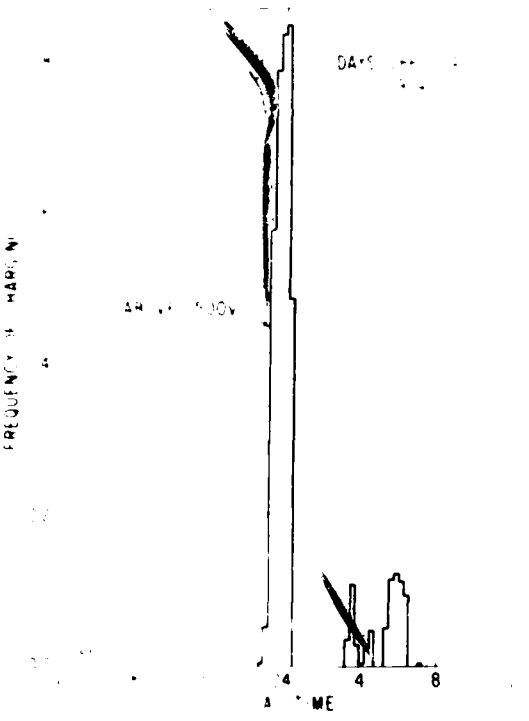


Figure 8. Distribution of spacecraft charging in local time for two separate days.

It is clear that for a given value of electron "temperature" defined by equation (3), a Maxwellian distribution would have a completely different net current (composed of the difference between primary and secondary fluxes) to the spacecraft than an electron distribution with two components, one hot and one cold. This effect may be quite important for charging at synchronous orbit.

The largest charging events occur when the spacecraft is eclipsed, removing the photoelectron current. Figure 8 shows the probability for charging events larger than -500 volts for 9 days during which the spacecraft was eclipsed.

ACKNOWLEDGEMENTS

The authors are grateful to Captain H. Garrett, B. Mauk, Professor C. McIlwain, C. Pike, and E. Whipple for many helpful comments and suggestions. We also thank L. McDaniel and G. Peters for their assistance with computer programming and C. McPhaden for the preparation of spectrograms. This work was supported by Air Force Contract AF F19628-76-C-0014 and NASA Grant NGL 05-005-007.

REFERENCES

DeForest, S. E., and McIlwain, C. E., Plasma clouds in the magnetosphere, J. Geophys. Res., 76, 3587, 1971.

DeForest, S. E., Spacecraft charging at synchronous orbit, J. Geophys. Res., 77, 4, 651, 1972.

DeForest, S. E., Electrostatic potentials developed by ATS-5, Photon and Particle Interactions with Surfaces in Space, Grard,R.J.L,

Pike, C. P., and R. R. Lovell, eds., Proc. Spacecraft Charging Technology Conference, NASA TMX-73537, AFGL-TR-77-0051, 1977.

Reasoner, D. L., Lennartsson, Walter, and Chappell, C. R., Relationship between ATS-6 spacecraft-charging occurrences and warm plasma encounters, Spacecraft Charging by Magnetospheric Plasmas, Rosen, A., ed. AIAA, New York, N.Y.

Mauk, B. H. and McIlwain, C. E., ATS-6 UCSD Auroral Particles Experiments, IEEE Trans. Aerospace and Electronic Systems, 11, 6, 1125, 1975.

Olsen, R. C., Purvis, C., and Whipple, E. C., Active modification of ATS-5 and ATS-6 spacecraft potentials, Proc. of Conf. for Effect of the Ionosphere on Space and Terrestrial Systems, Washington, D.C., 1977.

Whipple, E. C., The equilibrium electric potential of a body in the upper atmosphere and in interplanetary space, Ph.D. thesis, George Washington University, Washington, D.C., 1965; also published as NASA Tech. Note X-615-65-296, 1965.

Whipple, E. C., Observation of photoelectrons and secondary electrons reflected from a potential barrier in the vicinity of ATS-6, J. Geophys. Res., 81, 4, 715, 1976.

ACTIVE MODIFICATION OF ATS-5 AND ATS-6 SPACECRAFT POTENTIALS

R. C. Olsen and E. C. Whipple
University of California, San Diego
La Jolla, California 92093

Carolyn K. Purvis
NASA/Lewis Research Center
Cleveland, Ohio 44135

ABSTRACT

Instruments on ATS-5 and ATS-6 have been operated to study the phenomena of spacecraft charging. The ion engine neutralizers were operated in attempts to modify the spacecraft potentials. The neutralizer on ATS-5 is a hot filament emitting electrons, while the neutralizer on ATS-6 is a low energy plasma bridge. The diagnostics used were the UCSD particle detectors, counting electrons and protons as a function of energy. Operations were conducted in daylight and eclipse.

Preliminary analysis of some of these data has been carried out with the following results: (1) electron emission ($E \leq 10$ volts) does not perturb the status of a low potential ($|\phi| \leq 50$ volts) satellite by more than 50 volts (the ATS-5 low energy limit); (2) emission of a low energy plasma ($E \leq 10$ volts) does not change low potentials ($|\phi| \leq 5$ volts) by more than a few volts (ATS-6 low energy resolution); (3) when ATS-6 enters eclipse in the presence of a high energy plasma (10 kilovolt), the neutralizer suppresses any rise in $|\phi|$ (within a few volts resolution); (4) when the electron emitter on ATS-5 is operated, it serves to reduce negative potentials from thousands to hundreds of volts.

INTRODUCTION

A considerable amount of evidence has been accumulating over the last few years that anomalous behavior of spacecraft instrumentation is associated with electrostatic charging of spacecraft (Rosen, 1976). Charging effects appear to be particularly severe at synchronous orbit. Because of the importance of the synchronous altitude for a wide variety of spacecraft missions, this phenomenon has become a matter of concern to the aerospace community.

For example, McPherson et al. (1976) have shown that the occurrence of anomalies in synchronous spacecraft behavior correlates with local time and is most likely between local midnight and 0600. Reasoner et al. (1976) have shown that charging events on ATS-6 where the spacecraft potential goes more negative than -50 volts has the same correlation with local time. The failure of an entire Air Force satellite has been attributed to a spacecraft charging event.

Large spacecraft potentials have been observed on both ATS-5 and ATS-6. The largest potentials have occurred during eclipse conditions when photoelectrons are no longer emitted by the spacecraft, and during magnetic storms when the cool ambient plasma has disappeared. The largest potential observed so far was -19,000 volts on ATS-6 during an eclipse, but potentials as large as -2,200 volts have been observed in sunlight.

The recent analysis of low energy particle data on ATS-6 has shown that the local potential distribution about the satellite is dominated by the presence of differential charging of the spacecraft (Whipple, 1976a, b). Different portions of the spacecraft surface which are insulated from the main body can charge to very different potentials because of shadowing effects, anisotropic particle distributions, and non-homogeneous surface properties. As a result, large electrostatic fields can exist in the spacecraft vicinity. Possible consequences could be discharges across surfaces with associated material damage and electromagnetic interference. The UCSD experiment on ATS-6

This work was performed by UCSD for NASA/Lewis Research Center and GSFV under contract NAS5-23481.

has observed large spikes of accelerated electrons whose origin has been definitely traced to differentially charged portions of the spacecraft.

The packages of experiments on ATS-5 and ATS-6 have the capability not only of studying these effects, but also of performing active experiments by operating the ion engine instruments. These instruments have been operated in a coordinated program. Preliminary analysis indicated that during at least some of these operations the electrostatic configuration of the spacecraft was changed (Goldstein and DeForest, 1976).

Applied Technology Satellites 5 and 6 are both in geo-synchronous orbits, and carry particle detectors capable of measuring electron and proton fluxes as functions of energy and time (also angle on ATS-6). The spacecraft and detectors are compared in Table 1. There is a major qualitative difference between the two spacecraft in that ATS-5 is basically a homogeneous cylinder, while ATS-6 is a large umbrella-like structure with widely varying materials. See figures 1 and 2.

Table 1. Comparison of Spacecraft and Systems

	ATS-5	ATS-6
Characteristic size	2 m	10 m
Stabilization	Spin (axis parallel to earth's)	3-Axis
Outer surface	Mostly quartz (good insulator)	Quartz, kapton, paint, aluminum (mixed insulator and conductor)
Ion engine neutralizer	Thermal emission (electrons only)	Discharge plasma
Neutralizer placement	Recessed: 2.5 cm	Outboard: 17 cm
UCSD detectors	Body mounted (50 eV - 50 keV)	Rotating (1 eV - 80 keV)

Both spacecraft carry cesium ion thrusters. The thruster on ATS-6 was operated several times in 1974, and some results from these operations will be presented. Due to the spin-up of ATS-5, the thruster has not been used. More importantly for purposes of electrostatic control, each thruster carries a separate neutralizer. On ATS-5, this consists of a hot wire filament operating at less than 10 volts. The ATS-6 neutralizer is a small plasma bridge which emits a low energy ($\leq$10 volts) cesium plasma. Figures 3 and 4 display the two ion engines.

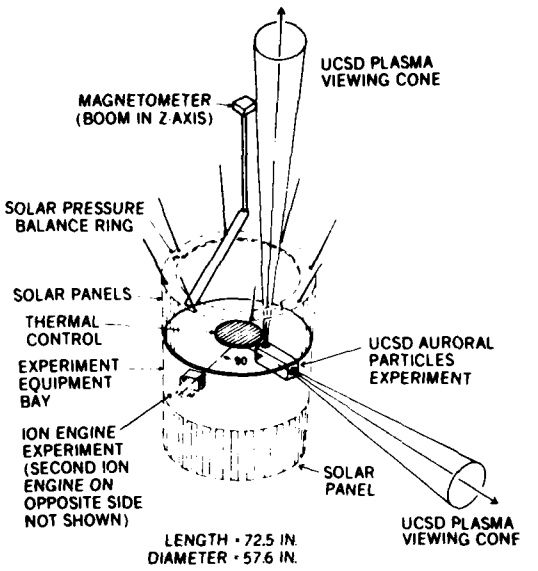


Figure 1. ATS-5: Detectors and Ion Engines

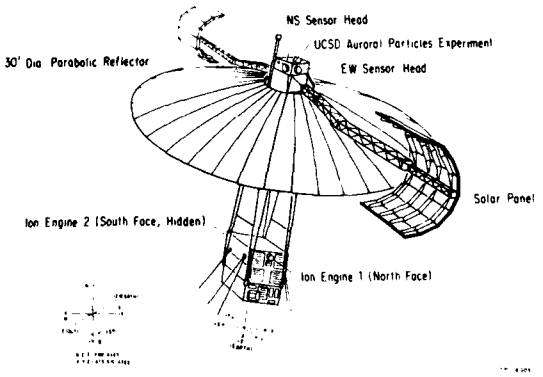


Figure 2. ATS-6: Detectors and Ion Engines

In this paper we will show a typical eclipse charging event on the ATS-6 satellite. This will be followed by the results of operating the ion engines on ATS-6 in daylight, and the results of operating the plasma bridge neutralizer in daylight. Moving back into eclipse conditions, we will give the results of operating the ATS-5 electron-emitting filament and the ATS-6 plasma bridge.

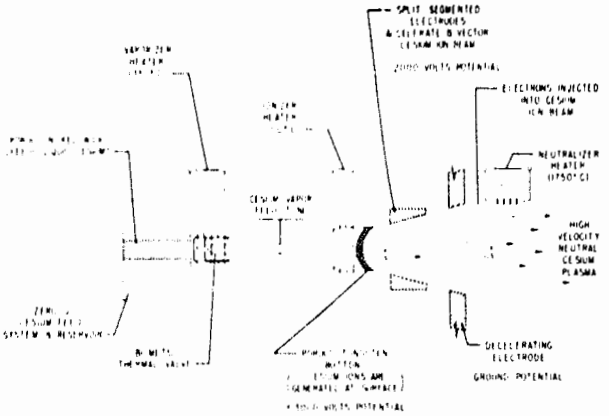


Figure 3. ATS-5: Thruster

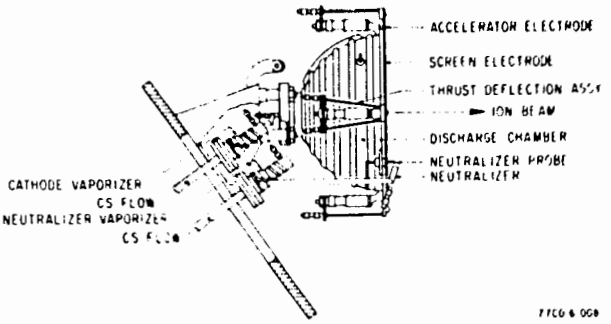
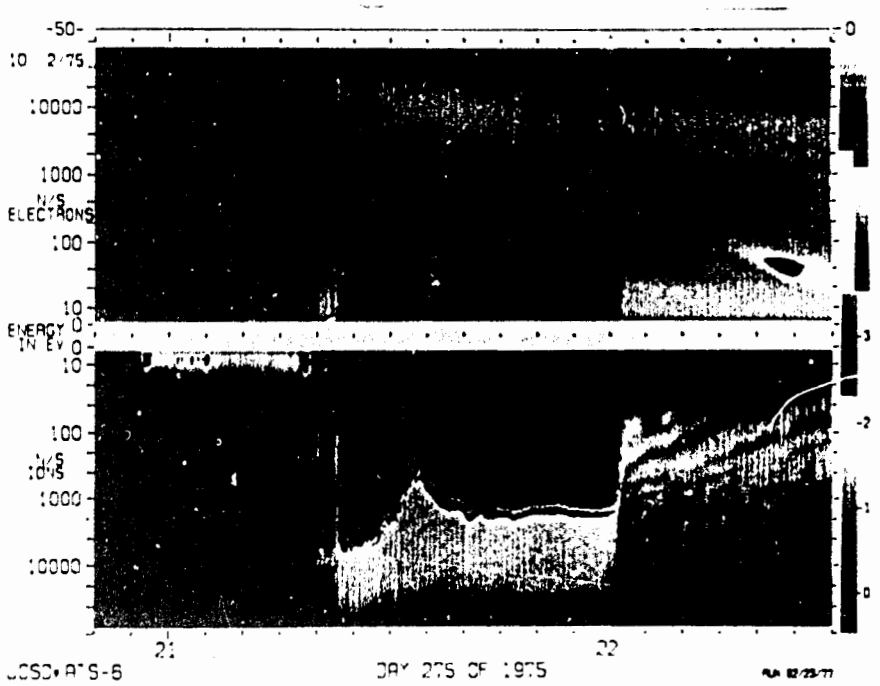


Figure 4. ATS-6: Thruster



Spectrogram 1. ATS-6: Eclipse with Injection of Hot Plasma; 10/2/75

ECLIPSE CHARGING

The primary qualitative means of data presentation is the spectrogram. Spectrogram 1 displays a typical charging event in eclipse, and will serve as an introduction to our data format. Both spectra are plots of count rate, which is proportional to the gray scale of the spectrogram. The other two axes are time and energy. Time is linear along the horizontal axis, while the energy scale is a combination of a linear scale at low energies and a log scale at higher energies. The electron spectra is on top, and the ion spectra on the bottom, usually with an inverted energy scale. In this spectrogram, the spacecraft is initially in sunlight, with an absolute potential below 10 volts. Upon entering eclipse at 20:55, the low energy band of electrons disappears, and a low energy band of ions appears. This may be interpreted as the disappearance of the photoelectrons and a small negative shift in potential. At 21:22, an injection of high energy plasma occurs. This plasma is the normal magnetosphere injection, not a result of any thruster operation. It shows up in our detectors as a large increase in particle count rates, particularly at high energies. This is displayed

in the
gray.
about
absenc
and th
(almo
the ho
indica
adjace
leaver
to sca

the AT
the ion
result
brief
taken
differ
reduc
tion ir
electr
tralize
and ag
4:35 w
ber of
differ
abunda

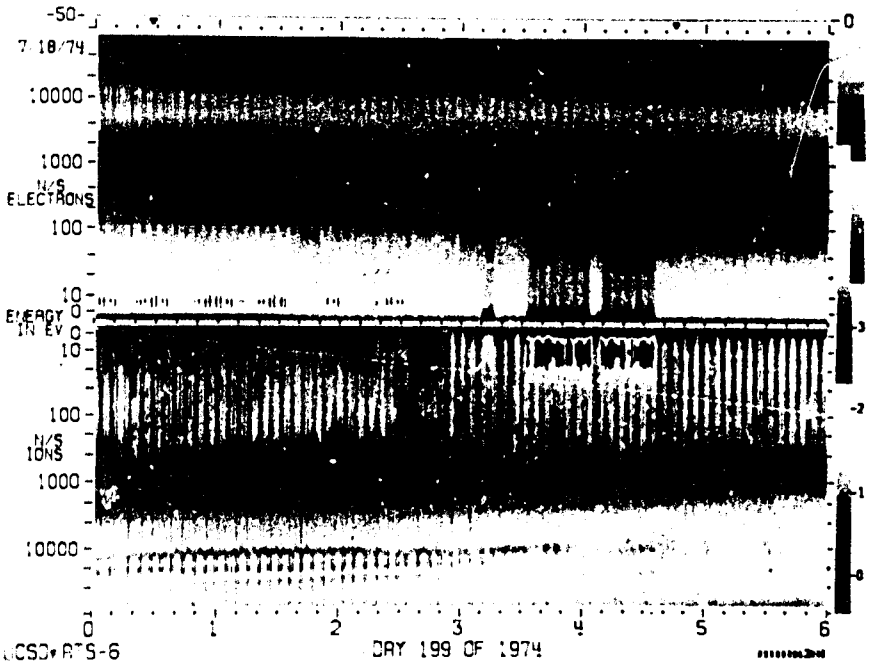
pectrogram as a much lighter shade of
n addition, the spacecraft charges up to
000 volts. This can be seen by the
of ions at energies below that value,
sudden transition to high count rates
white shade). The dark band marking
dary is an overflow of the gray scale
ng very high count rates compared to
energy channels. The spacecraft
clipse at 22:00, and drops in potential
where between 0 and -40 volts.

ION ENGINE OPERATIONS

here were two successful operations of
-6 ion engines. The initial operation of
thruster occurred on July 18, 1974. The
are displayed in Spectrogram 2. In a
eration, the spacecraft was apparently
a slightly negative potential, with any
tial charging of the spacecraft greatly
. This can be seen in the great reduc-
ntensity of the broad band of secondary
ns (0-100 volts) at 3:10 when the neu-
goes into full operation for 5 minutes,
in from 3:31 to 4:03 and from 4:10 to
en the thruster was on. The great num-
low energy ions could either be due to a
nt potential on the spacecraft or to the
nce of cesium ions.

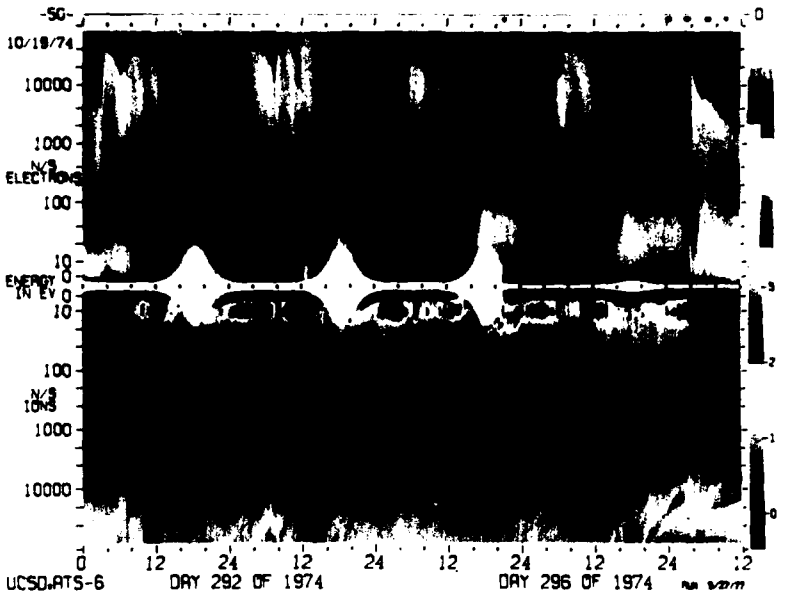
The second test ran for approximately 92 hours. It is displayed in Spectrogram 3. This spectrogram shows data from hour zero on day 292 to hour 12 on day 296. The engine ignited at 8:01 UT on day 292 and a low energy band of ions immediately appears. This band continues with small variations throughout the test. (The white indentations occurring between hours 12 and 24 of the first three days are temperature dependent effects and may be ignored.) This low energy band of ions may be attributed to cesium ions from the engine or ambient ions brought in by a shift in the potential. A negative drop of a few volts in the potential is consistent with both of these explanations.

There are two large injections of energetic particles, the first beginning at approximately 3:00 UT on day 293, the second at about 6:00 UT on day 295. Two injections of slightly lower energy particles occur on days 294 and 296. Charged particle injections have caused ATS-6 to fall to potentials of several hundred volts below ground in daylight. The engine may therefore be compensating for the change in the particle flux. It does hold the satellite within a few volts of ground throughout the test.



Spectrogram 2. ATS-6: Ion Engine Operation; 7/18/74

BEST AVAILABLE COPY



Spectrogram 3. ATS-6: Ion Engine Operation; 100 Hours; 10/19/74

NEUTRALIZER OPERATIONS: DAYLIGHT

Two sequences of daylight operations of the plasma bridge neutralizer occurred on ATS-6. Spectrogram 4 shows two of the neutralizer operations from August 20, 1976. The effects of particle emission can be seen at hours 8 and 10 in the low energy bands of ions appearing at these times. If these low energy bands are ambient ions, we are probably seeing a slight negative shift in the potential, bringing into view previously unseen particles. If so, we are seeing the same behavior as in some eclipses. Another possibility is that we are detecting particles from the neutralizer. Comparison of these data to similar eclipse data may help resolve the question of ions returning to the spacecraft.

On November 14, 1976, the operations at 11:35 and 13:35 seem to have taken the spacecraft from -100 volts to within a few volts of ground. This effect is displayed in Spectrogram 5. The reduction in secondary electrons implies the differential charge on the satellite has been reduced.

NEUTRALIZER OPERATIONS: ECLIPSE

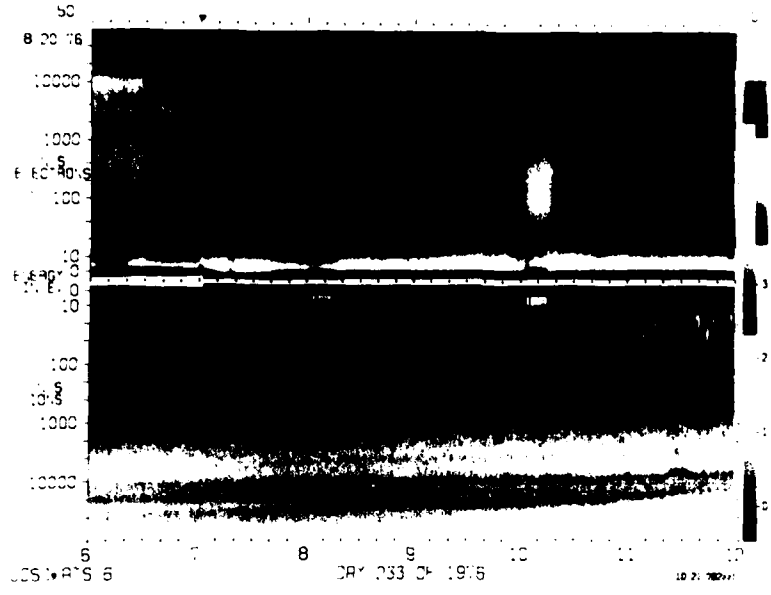
ATS-5

The hot filament neutralizer on ATS-5 was operated regularly in eclipse for several year. In the great bulk of ATS-5 spectrograms, it is apparent that switching on the neutralizer (the electron emitting filament) does drive a negatively charged satellite back towards lower potential. (See Figure 5.) (Taken from Purvis, 1976.) How well it succeeds is not clear, partly because of the lack of data in the 0-50 volt region.

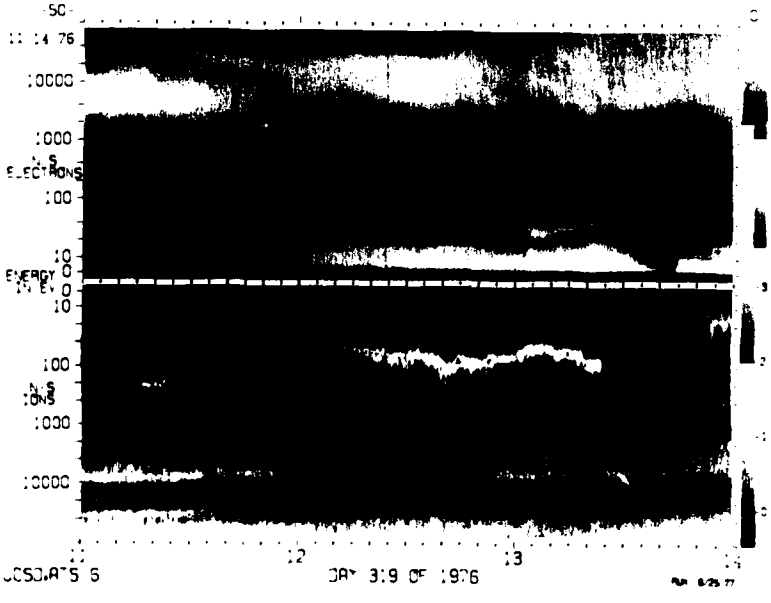
A data printout for a neutralizer operation on September 20, 1974, was reduced to graphical form in Figure 6. Spectrogram 6 displays the same data. The transitions at 6:31:30 and 6:36 correspond to neutralizer on and off commands. There is an immediate drop in the magnitude of the potential when the neutralizer is switched on. The speed of the transition is beyond the time resolution of ATS-5 (the energy scan takes about 20 seconds); therefore it is not known how close to zero volts (or the detectors' 50 volt lower limit) the spacecraft is driven. The potential then rises to an equilibrium with an apparent exponential behavior.

ATS-6

Finally, neutralizer operations have been conducted on ATS-6 in eclipse. In the first tests, the neutralizer was ignited before the spacecraft entered eclipse. This ignition was difficult to observe in the spectrograms. Generally, ignition happened at low potentials and



Spectrogram 4. ATS-6: Neutralizer Operation in Daylight; 8/20/76



Spectrogram 5. ATS-6: Neutralizer Operation in Daylight; 11/14/76

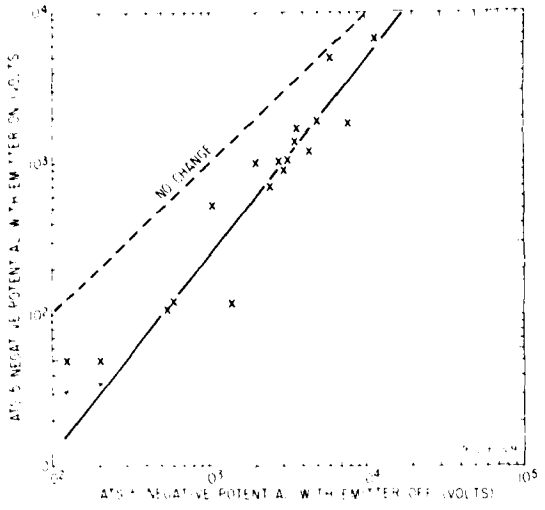


Figure 5. ATS-5: Effect of Electron Emitter on Spacecraft Potentials

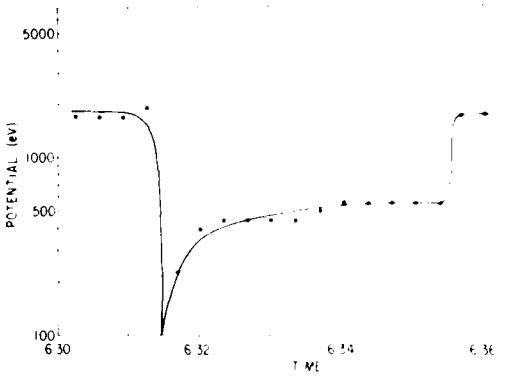


Figure 6. ATS-5: Potential during Eclipse/ Neutralizer Operation: 9/20/74



Spectrogram 6. ATS-5: Neutralized Operation in Eclipse: 9/20/74

the change in data was small. The main effect was the disappearance of the normal low energy band of electrons.

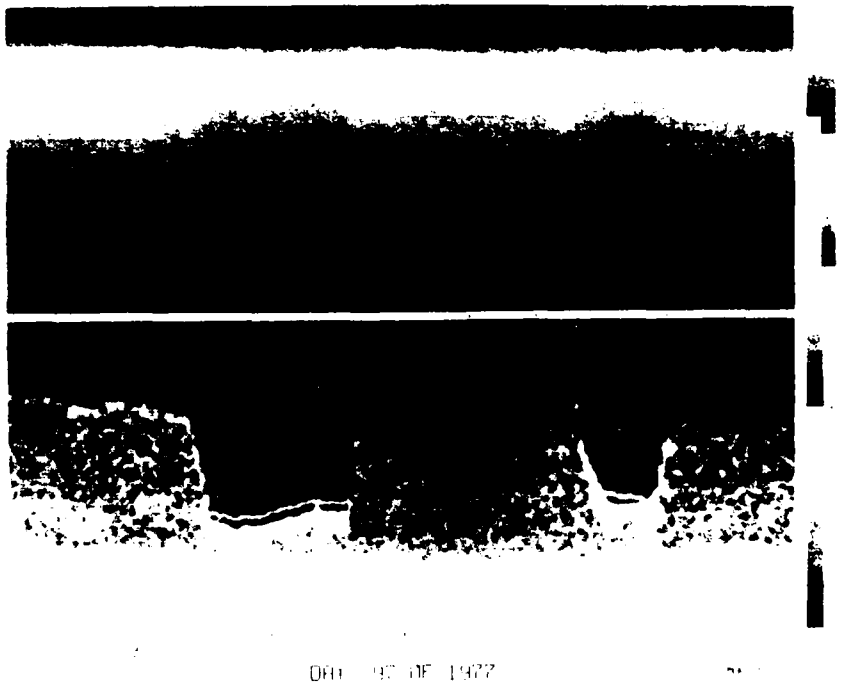
Movement of the spacecraft into eclipse caused no perceptible change in the particle data.

Our most spectacular results occurred in the spring of 1977. On April 7, 1977 the neutralizer was switched on after the spacecraft entered eclipse. Spectrogram 7 shows that the spacecraft was already charged to about -50 volts when it entered eclipse, and upon entering eclipse, charged to about -3000 volts. At 9:15 the neutralizer ignites, and the spacecraft is driven to within 40 volts of ground. The absence of a low energy plasma prevents us from saying the spacecraft was grounded, but the transition is impressive. As with ATS-5 and its neutralizer operations, the spacecraft returns to its previous status when the neutralizer is switched off.

SUMMARY

For ATS-5, we have several years of data involving electron emission to study. There are indications that electron emission is an effective means of modifying the potential of a negatively charged satellite, but it does not always ground the spacecraft to the ambient plasma.

Eight experiments were conducted with the ion engine neutralizer on ATS-6 in 1976, and seven were done in the first 100 days of 1977. Parts of this and earlier data have been reduced and given preliminary analysis. As with eclipses, the effect on the spacecraft depends on the particle environment. The ion engine seems to completely dominate all other particle sources, bringing the spacecraft within a few volts of ground and holding it there. Experiments with the smaller plasma bridge (the neutralizer) are not as overwhelming to the spacecraft but seem to hold it near ground



Spectrogram 7. Neutralizer Operation in Eclipse, Day 97/77

before, during, and after the transition into eclipse. When the spacecraft was in sunlight, operation of the neutralizer reduced the differential charge on the satellite.

REFERENCES

Goldstein, R. and S. E. DeForest, Active Control of spacecraft potentials at geosynchronous orbit, SPACECRAFT CHARGING BY MAGNETOSPHERIC PLASMAS, Progress in Astronautics and Aeronautics, 47, Rosen, ed.; AIAA, MIT Press, 1976.

McPherson, D. A., D. P. Cauffman and W. Schober, Spacecraft charging at high altitudes - The SCATHA satellite program, SPACECRAFT CHARGING BY MAGNETOSPHERIC PLASMA, Progress in Astronautics and Aeronautics, 47, Rosen, ed.; AIAA, MIT Press, 1976

Purvis, C. K., R. O. Bartlett and S. E. DeForest, Active Control of Spacecraft Charging on ATS-5 and ATS-6, Spacecraft Charging Technology Conference, Colorado Springs, Colorado, October 27-29, 1976.

Rosen, A., Spacecraft charging: Environmental induced anomalies, J. Spacecraft and Rockets, 13, 129, 1976.

Whipple, E. C., Jr., Theory of the spherically symmetric photoelectron sheath: A thick sheath approximation and comparison with the ATS-6 observation of a potential barrier. J. Geophys. Res., 81, 601, 1976a.

Whipple, E. C., Jr., Observation of photoelectrons and secondary electrons reflected from a potential barrier in the vicinity of ATS-6, J. Geophys. Res., 81, 715, 1976b.

ELECTRICAL INTERFERENCE TO SATELLITE SUBSYSTEMS RESULTING FROM SPACECRAFT CHARGING

Robert R. Shaw
Department of Physics and Astronomy
The University of Iowa
Iowa City, Iowa 52242

INTRODUCTION

Advanced electronics technology has been extensively used in synchronous orbiting satellites for both research and practical applications such as weather forecasting and communications. It is necessary that such spacecraft operate reliably in the space environment.

Recently, an aspect of the environment at synchronous altitude that can lead to serious operational problems has been identified. DeForest has reported that ATS-5 and ATS-6 charge to voltages as great as several thousand volts during magnetospheric substorms.[1,2] Satellite charging is caused by charged-particle currents from the surrounding plasma collected by the satellite (thermal electrons and ions) and charged-particle currents emitted by the satellite to the surrounding plasma (photoelectrons, secondary electrons, and backscattered electrons).

It is believed that satellite surfaces charge to the highest voltages during magnetospheric substorms. The magnitude of the voltage to which a satellite charges is roughly equal to the thermal energy of the dominant current constituent. During substorms, the dominant constituent is energetic plasma sheet electrons that are injected from the geomagnetic tail to synchronous orbit altitudes.[3] These electrons have thermal energies of several thousand electron volts, and, accordingly, satellite surfaces may charge negatively to voltages as great as ten thousand volts.

If a satellite has dielectric surfaces, those surfaces can charge to voltages that are large with respect to the frame of the satellite. If this voltage exceeds the breakdown voltage of the dielectric, an electrical discharge occurs.

It has been suggested that electromagnetic fields radiated from discharges, or direct-current surges from the discharges, may cause a spacecraft to operate with unexpected behavior that may be difficult to anticipate from ground testing.[4,5,6] Such anomalies, believed to be correlated with periods of enhanced substorm activity, have been observed on numerous spacecraft from several different missions flown at synchronous altitudes. One satellite failure was associated with unusual magnetospheric activity at synchronous altitudes.[6,7] In addition, ground-based laboratory tests have shown that discharges cause physical damage to thermal-control materials that are charged to voltages sufficient to generate electrical breakdowns.[8] Satellite charging can also cause the re-attraction of ionized contaminants that are emitted from a spacecraft.[9] As a consequence, satellite charging may contribute to deterioration of dielectric surfaces used for passive thermal control causing parts of the spacecraft to operate at undesirable temperatures.[10]

Recently, an experiment was flown on a synchronous orbiter to detect the presence of satellite charging, to detect the occurrence of electrical discharges caused by differential charging of the dielectric surfaces, and to detect coincidences between electrical discharges and the anomalous responses of several other spacecraft subsystems, which had been observed on previous missions of the same type of spacecraft. An example of data collected from this experiment is shown in Figure 1.

The interpretation of data from this experiment has been discussed in detail elsewhere; therefore, the major findings of this investigation will only be summarized briefly in this report.[10] Data collected by the current probe, shown in Figure 1, shows a spacecraft charging event (presumably caused by the injection of energetic plasma to synchronous altitudes during a magnetospheric substorm) beginning at approximately 23.7 hours local time when the spacecraft is eclipsed by the earth. The charging event is indicated by a small decrease in the total current density (measured across the probe area) during eclipse. Following eclipse the current density is decreased significantly from the value of 10 to 20 μamps m^{-2} typical of photoelectron emission from the probe usually observed at these orbital locations. This indicates

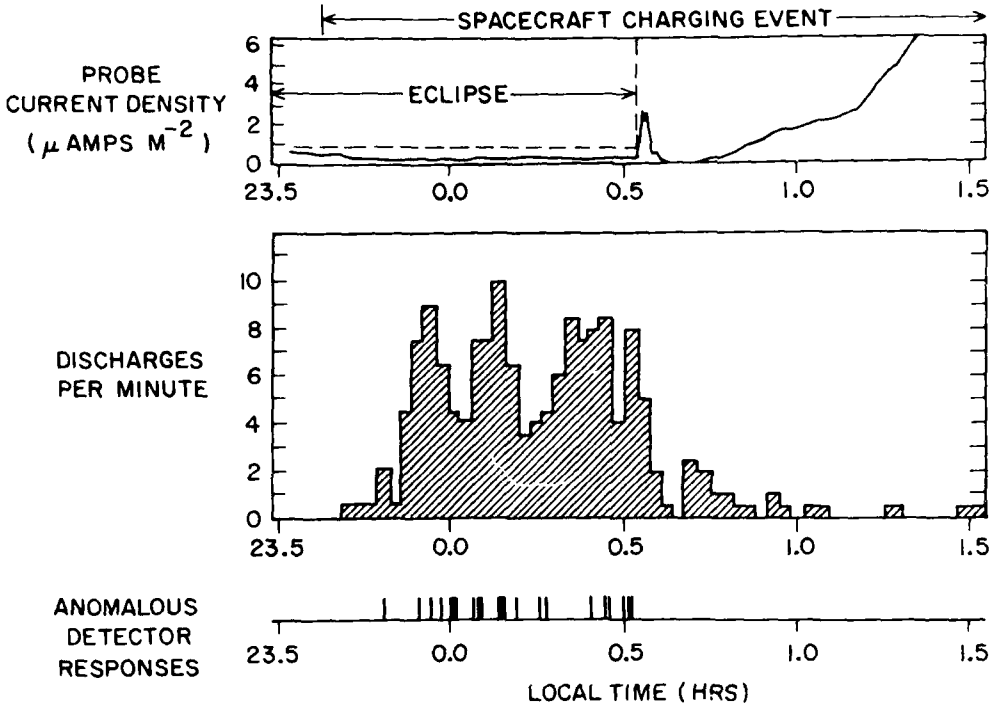


Figure 1. Satellite-charging experiment data collected during a period with twenty-five anomalous responses of a detector on another spacecraft subsystem. Each of these anomalous responses occurred coincidently with a discharge observed by the charging experiment.

that the spacecraft is immersed in energetic plasma electrons that can cause charging of spacecraft dielectric surfaces.

Near the beginning of the charging event the pulse counter begins to detect electrical discharges occurring on these dielectric surfaces as previously described by Cauffman and Shaw.[11] Rates of four to ten per minute are observed during this event, and an anomalous response of a detector is also observed on another subsystem during this event. Twenty-five anomalous detector responses were observed on this pass, and each of the twenty-five responses was observed coincidently with the detection of a discharge by the pulse counter experiment. (The pulse counter is sampled once per second and the detector is sampled more rapidly so that this coincidence is definitely not caused by overlapping data storage in the spacecraft data system.)

The principal findings from analysis of data collected from this experiment were that discharges occur at rates averaging about 20 per hour, and that two different types of anomalous behavior of the spacecraft's electronics subsystems do occur coincidently with discharges detected by the instrumentation and are, therefore, presumed to be caused by these discharges.[10] These anomalies occur infrequently in orbit; hence, it takes a specific discharge at a specific location on the spacecraft to generate such an anomaly because many discharges occur for which no anomalies are observed.

This paper reports some results of a study carried out to explain how the energy from electrical discharges generated by satellite charging couples into spacecraft electronics systems. Particular emphasis was placed on developing an understanding of the specific types of coupling mechanisms causing the anomalies on the type of spacecraft on which the charging experiment was flown, for example, the anomalous detector responses shown in Figure 1. This paper

reviews briefly the relatively simple coupling mechanism that was proposed to explain how discharges couple energy to this detector located well within the structure of the spacecraft. The following sections describe this mechanism and the engineering tests that were conducted to verify it.

DESCRIPTION OF ANOMALOUS DETECTOR RESPONSE

One type of orbital anomaly investigated in this study is the anomalous response of a detector located well within the structure of the spacecraft. Figure 2 schematically shows the configuration of this detector, the spacecraft structure, and the associated signal processing electronics.

The detector is mounted on a structural member the mountings of which are electrically isolated from the frame of the spacecraft for thermal purposes. This structural member passes through the spacecraft skin and is attached to a passive thermal radiator that is covered with dielectric surfaces. The detector is mounted on this structural member inside the spacecraft, and it is connected to a preamplifier through a long coaxial cable as shown schematically in Figure 2. The structural member is grounded to the frame of the spacecraft through the shield of the coaxial cable.

This subsystem has been observed to have responses in orbit that cannot be due to a normal stimulus of the detector, because the characteristics of the response differ significantly from the response to its usual stimulus. These responses are, therefore, easily identified and ignored by ground processing equipment. These anomalous responses have been observed to occur in coincidence with discharges observed by the satellite-charging experiment flown on this spacecraft[10] (see also Figure 1 of this paper).

When an anomalous response is observed, the preamplifier is observed to have an output waveform similar to that shown schematically in the lower panel of Figure 2. The preamplifier is observed to "ring" near the center frequency of its band pass and the resultant oscillation is damped as shown in Figure 2. The anomalous response is characterized by the ratio between the first and second peaks in this oscillation, P_1/P_2, and the time interval, ΔT, between the times at which these peaks occurred.

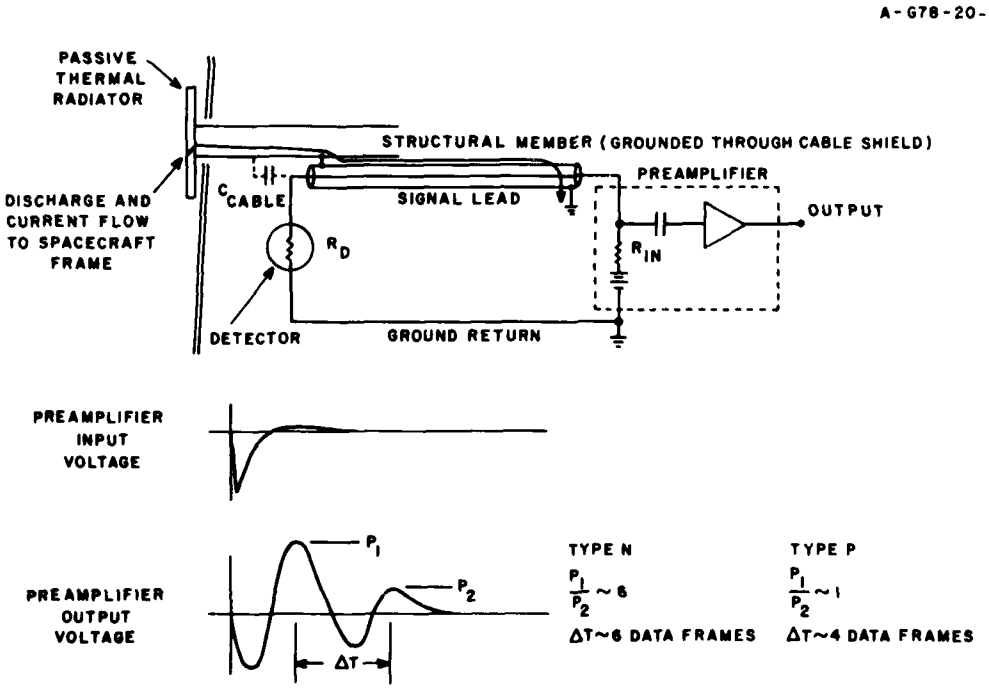


Figure 2. Mechanism for generating anomalous detector responses from discharge generated currents.

The anomalous responses fall into two distinct groups, or types, based on their observed characteristics as shown in Figure 3. Responses that we call Type P anomalies have values of P_1/P_2 equal to about 1.15 and ΔT equal to about 3.75 data frames. (1 data frame is equal to one complete sample of the detector.) The second category of anomalous responses, which we call Type N anomalies, have values of P_1/P_2 equal to 5.90 and ΔT equal to 5.75 data frames. (The separation between the three apparent subgroups of Type N responses shown in Figure 3 is caused by the digitizing step size in the data processing equipment.)

Any mechanism that uses electrical discharges as an energy source to generate these anomalies must couple sufficient energy into the electronics to account for the amplitude, P_1, and also for the spectral characteristics, P_1/P_2 and ΔT, that are observed in orbit. This study has identified such a mechanism.

Figure 2 shows how a Type N anomaly is generated. An electrical discharge can occur from the charged surface of the passive thermal radiator to the supporting structural member if the voltage between the dielectric surface and the metallic support exceeds the breakdown voltage of the dielectric. The discharge represents an injection of electrons to the structural member, and these electrons constitute a current that flows through the cable shield, through the ground clamp at the preamplifier, and into the spacecraft frame. (The assembly is electrically insulated from the frame, except through the cable shield clamped to the signal processing electronics housing.)

Because the cable shield and the connection at the ground clamp has a finite dc resistance, a voltage pulse exists on the supporting structure and the cable shield. This voltage pulse is coupled into the detector signal lead because of the capacitance of the signal lead to the cable shield. Thus, a voltage pulse exists at the preamplifier input and the resulting output is observed as an anomalous response of the subsystem.

This mechanism accounts for all observed features of the anomalous detector responses. Also, the characteristics of both Type N and Type P anomalies can be explained by this generating mechanism as shown in Figure 4. Voltage pulses of both positive and negative polarity can be produced at the preamplifier input for discharges occurring either on or next to the thermal radiator. A discharge occurring on the assembly produces a negative voltage pulse at the preamplifier, and a discharge occurring next to the radiator produces a positive voltage pulse because of ground return currents that are induced by the discharge. (See Figure 4.) For this case, the radiator acts like a short-range antenna, and the coupling strength should decrease roughly as the inverse cube of the distance of the discharge from the assembly.[11] This behavior is observed in orbit. The Type N responses have amplitudes that are constant to within a factor of 2 or 3. The amplitudes of Type P anomalies vary several orders of magnitude, however, because of the variation in coupling strength with the location of the discharge.

An engineering simulation was carried out to experimentally determine the response of this system to the currents injected by the discharges. The purpose of this simulation was to compare the responses generated in the test with the anomalous orbital responses to check the validity of this generation mechanism. The remainder of this paper consists of a description of these tests and the results obtained.

ENGINEERING TESTS

Because the bandpass of the preamplifiers is at low frequencies an equivalent circuit of the configuration shown in Figure 2 can be drawn as shown in the upper half of Figure 5. This circuit contains the essential parameters that determine the amplitude of the signal that is coupled into the preamplifier by the discharge current injected from the dielectric surfaces to the supporting structural member.

The discharge current was represented by a double exponential current source with a near zero rise time and a long decay time. Laboratory experiments with several types of dielectric materials suggest that the rise times of these currents are probably near 10^{-7} seconds and that the decay times may vary from 10^{-7} seconds to 10^{-4} seconds depending on the type of material and the physical configuration used.[12] Some theoretical considerations suggest that rise times of vacuum breakdowns may be as short at 10^{-8} seconds.[13]

The other factors influencing the amplitude of the signal coupled into the preamplifier are the resistance between the structural member and the spacecraft frame (labeled R_{SH} in Figure 5), the cable capacitance (C_{CABLE}), and the parallel combination of the detector resistance (R_D) and the preamplifier input resistance (R_{IN}).

This equivalent circuit was modeled in the laboratory as shown in the lower half of Figure 5. An exponentially decaying current source was generated by a differentiating the output of a pulse generator with a variable pulse rise time as short as 10^{-8} seconds. To simulate the cable capacity, a cable of the type and length used on the spacecraft was used to couple the signal to a preamplifier. A resistor, R_D, was used to simulate the detector resistance.

Several pulse decay times were selected, and a comparison was made between the laboratory model and the spacecraft configuration on a development test model of the spacecraft. The results of this comparison indicated that the

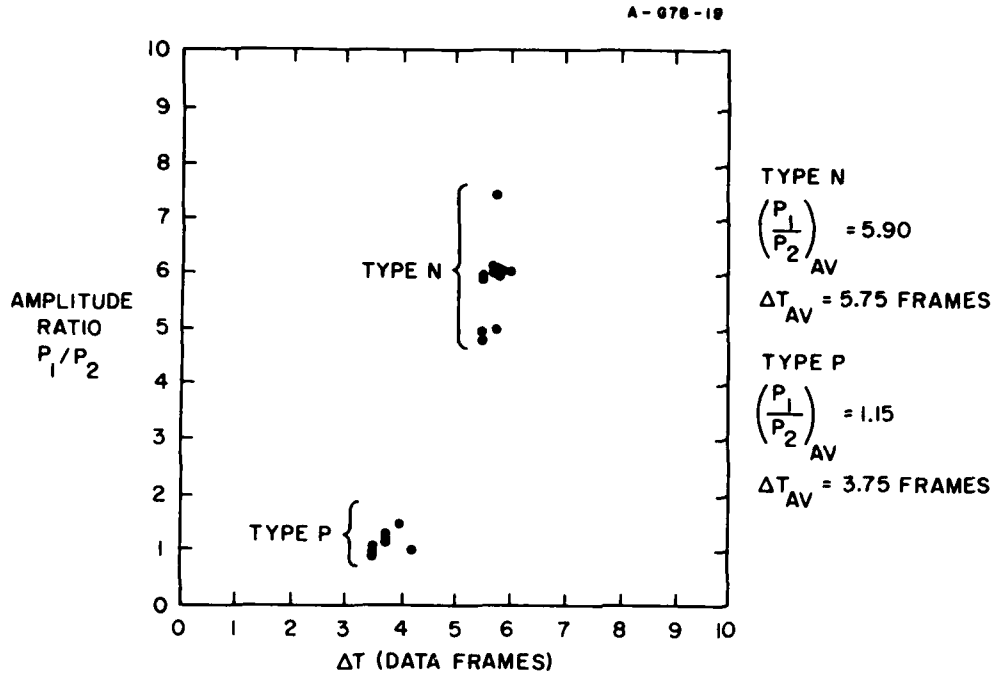
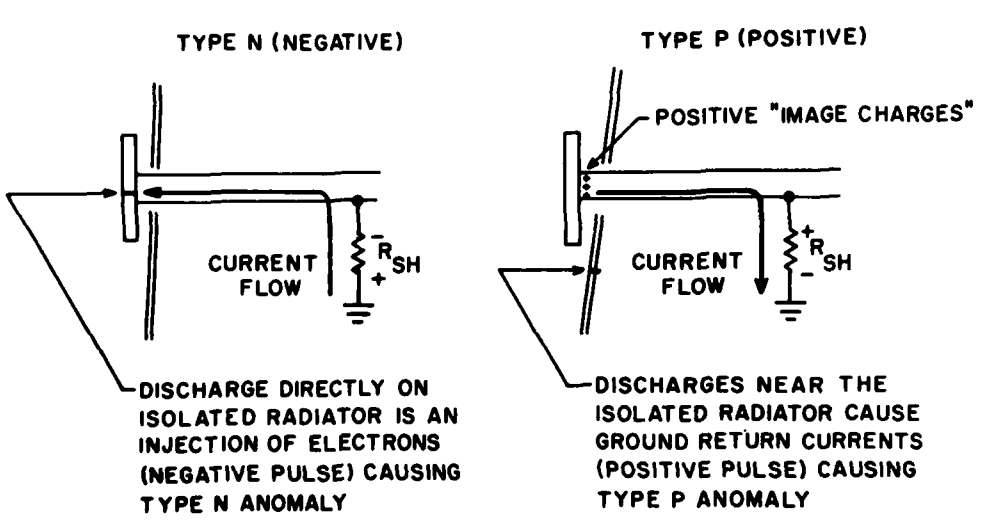


Figure 3. Anomalous detector responses occur in two distinct groups, which we call Type N and Type P. Apparent separation of Type N anomalies is caused by digitizing step size in signal processing electronics.



Negative voltage pulses (causing Type N anomalies) are caused by direct injection of [illegible] metallic support for the passive thermal radiator. Type P anomalies are caused by [illegible] resulting from ground return currents caused by discharges occurring near the [illegible]. All external surfaces shown are dielectric surfaces.

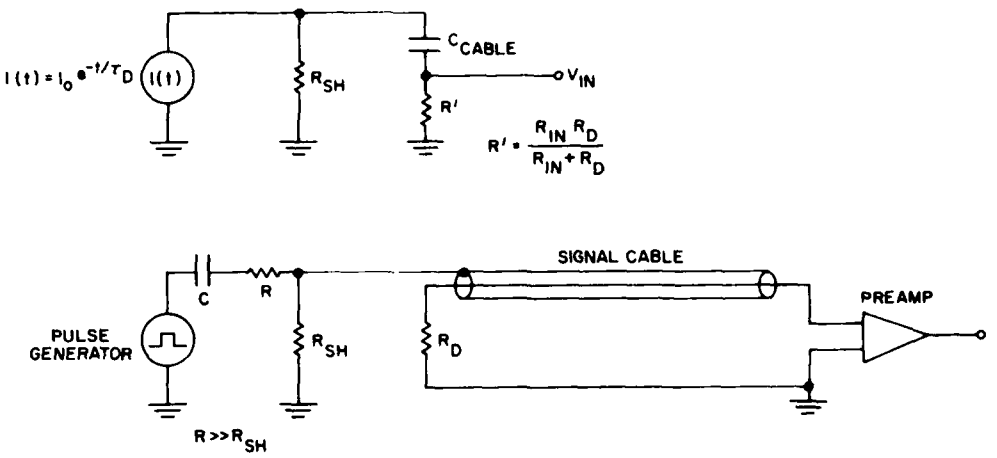


Figure 5. Equivalent circuit model and laboratory test configuration for modeling the detector subsystem response to discharges.

model was an accurate representation of the system response to within about twenty percent. This accuracy was judged to be adequate for the purposes of this test.

The results of the pulse injection tests are summarized in Figure 6. The left hand panels of Figure 6 show the amplitude ratio P_1/P_2 and time delay ΔT for a negative pulse with five different pulse decay times between 10^{-5} seconds and 10^{-2} seconds. The horizontal straight lines show the average value of P_1/P_2 and ΔT for the Type N anomaly.

Two features of Figure 6 are worth noting. First, the test results are in agreement with the values of P_1/P_2 and ΔT recorded on orbit for the Type N anomaly. (The precision of the test measurements is denoted by the vertical error bars.) Second, the system has the same response for any pulse shorter than about 1 millisecond. This occurs because the pulse is short enough so that its frequency content is nearly flat in the bandpass of the preamplifier filter. The response observed for this pulse (or any shorter pulse) is the impulse response of the system, and it is independent of the exact shape of the pulse.

To experimentally check this a square pulse 10 microseconds in width was used as an injected current to the simulated detector assembly. The response to the square pulse, labeled "impulse" response in Figure 6, was indistinguishable from those of the shorter exponentially decaying current pulses. Decreasing the width of the square pulse and increasing the amplitude by a factor of ten generated the same response as the 10-microsecond-wide pulse.

The right hand panels of Figure 6 show the test results for P_1/P_2 and ΔT for a positive pulse. These results are also in agreement with the corresponding values for P_1/P_2 and ΔT for the Type P anomalies.

These tests have shown that the generating mechanism described herein produces the spectral characteristics observed in orbit for these anomalies. In addition, this mechanism must predict reasonable values for the intensity of the preamplifier response.

The intensity cannot be calculated exactly because there is an important parameter that is not known. The total amount of charge participating in the discharges in flight has not been measured. Knowledge of the total charge is necessary to calculate the peak voltage of the preamplifier response. The observed peak voltage can be used, however, to calculate the total charge required to produce the anomalous response. This value can then be compared with other measurements made in laboratory experiments.

Table 1 lists typical discharge parameters that have been measured in laboratory experiments with similar dielectric surfaces at the Stanford Research Institute[14] and the parameters required to produce the anomalous detector responses. Approximately 6 μcoul of charge must participate in a discharge to produce the values of P_1 for these anomalous responses.

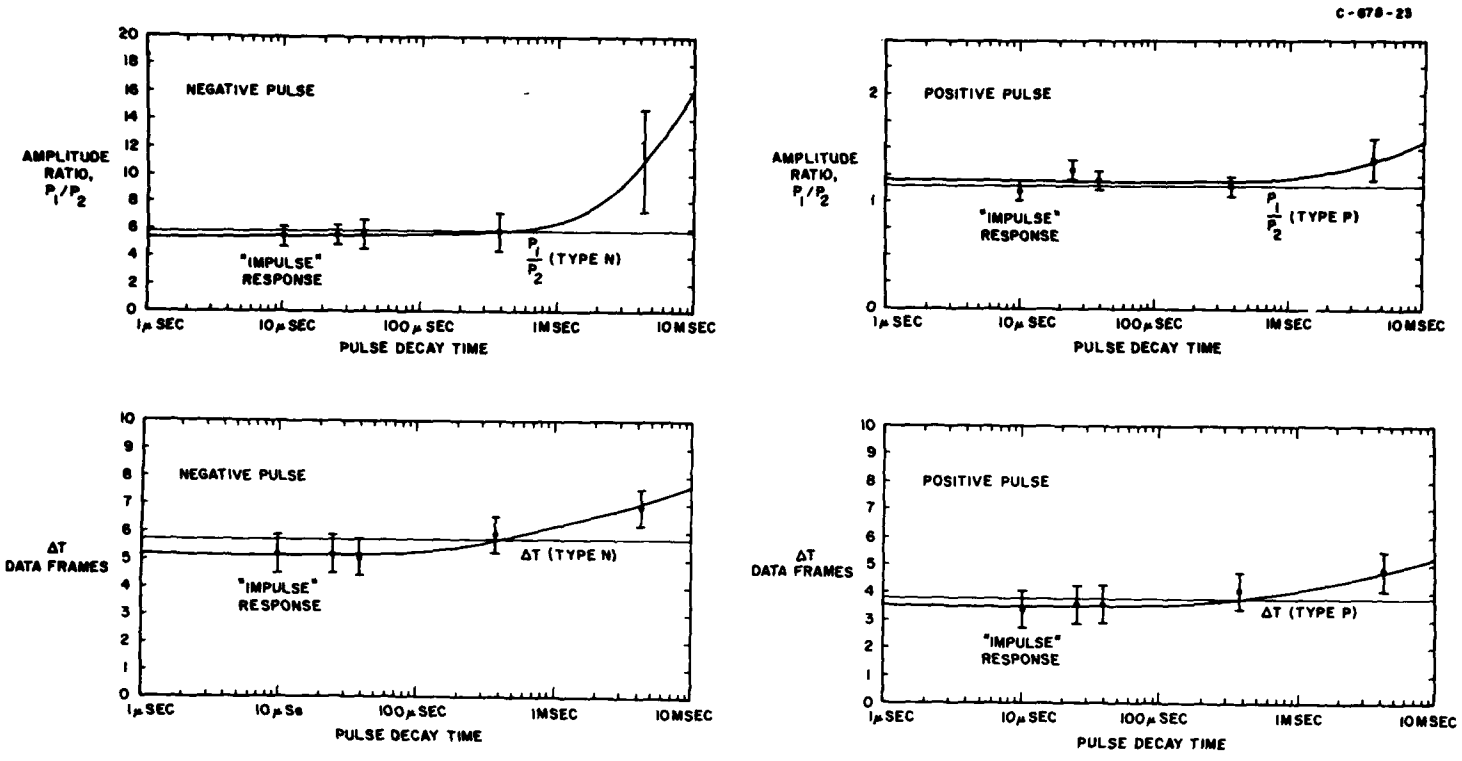


Figure 6. Test responses for negative and positive voltage pulses are in agreement with the orbital characteristics of Type N and Type P anomalies for discharges with decay times less than about 1 msec.

TABLE 1 COMPARISON OF CALCULATED DISCHARGE PARAMETERS WITH LABORATORY MEASUREMENTS

Characteristic	Probable Value	Determination
Total Charge	~ 6 μcoul	Coupling analysis
	0.2 - 2.0 μcoul	SRI laboratory experiments
Pulse Length	Less than ~ 1 ms	Comparison of pulse injection tests and orbital data
	<1 μs - 40 μs	SRI laboratory experiments

This value is reasonable when compared with laboratory measurements, as shown in Table 1. It is possible that slightly greater amounts of charge participate in a typical discharge in the orbital environment. This may be caused by contaminant films that could be present on the dielectric surfaces. Such films might increase the charge mobility parallel to the surface, resulting in slightly larger discharges. In any case, the amount of charge needed to cause the anomalous responses is in agreement with that typically measured in laboratory experiments.

SUMMARY

A physically reasonable mechanism for explaining the coupling between one type of spacecraft electronics subsystem and electrical discharges has been presented. A laboratory simulation of the coupling to the signal processing electronics has shown quantitative agreement between the test results and the anomalous detector responses that have been observed in orbit. The results of this study and experimental flight data collected by a satellite-charging experiment flown on this spacecraft confirm that these anomalies are likely generated by direct current surges from arc discharges.

REFERENCES

1. S. E. DeForest, "Spacecraft Charging at Synchronous Orbit," J. Geophys. Res., 77, 651, 1972.

2. S. E. DeForest, Correlation of Plasma Measurements on ATS-5 and ATS-6, Special Session on Initial Scientific Results from ATS-6 presented at the Fall Annual Meeting of the American Geophysical Union, San Francisco, CA, 1974.

3. S. E. DeForest and C. E. McIlwain, "Plasma Clouds in the Magnetosphere," J. Geophys. Res., 76, 3587, 1971.

4. R. W. Fredricks and F. L. Scarf, "Observations of Spacecraft Charging Effects in Energetic Plasma Regions," Photon and Particle Interactions with Surfaces in Space, D. Reidel Publishing Co., Dordrecht, Holland, 277, 1972.

5. D. P. Cauffman, Correlation of Strobe Anomalies with Geophysical Parameters, ATM-74 (4409-04)-3, The Aerospace Corporation, El Segundo, CA, 1973.

6. A. Rosen, Spacecraft Charging: Environment Induced Anomalies, presented at the AIAA 13th Aerospace Sciences Meeting, Pasadena, CA, 1975.

7. H. Wynne, Operational Experience with Satellite Anomalies, presented at the AIAA 13th Annual Aerospace Sciences Meeting, Pasadena, CA, 1975.

8. J. E. Nanevicz, R. C. Adamo, and W. E. Scharfman, Satellite Lifetime Monitoring, SRI Project 2611, Stanford Research Institute, Menlo Park, CA, 1974.

9. D. P. Cauffman, Ionization and Attraction of Neutral Molecules to a Charged Spacecraft, TR-0074 (9260-09)-1, The Aerospace Corporation, El Segundo, CA, 1973.

10. R. R. Shaw, J. E. Nanevicz, and R. C. Adamo, "Observations of Electrical Discharges Caused by Differential Satellite Charging," Spacecraft Charging by Magnetospheric Plasmas, Progress in Astronautics and Aeronautics, Vol. 47, American Institute of Aeronautics and Astronautics, New York, NY, 1976.

11. D. P. Cauffman and R. R. Shaw, "Transient Currents Generated by Electrical Discharges," Space Science Instrumentation, 1, 125, 1975.

12. D. P. Cauffman, Recommendations Concerning Spacecraft Charging in the Magnetosphere, Report No. SPL-33302, The Aerospace Corporation, El Segundo, CA, 1973.

13. C. Germain and F. Rohrbach, "High Voltage Breakdown in Vacuum," Vacuum. 18, 371, 1968.

14. R. C. Adamo, Stanford Research Institute, Menlo Park, CA, private communication, 1974.

BEST AVAILABLE COPY

VOYAGER SPACECRAFT ELECTROSTATIC DISCHARGE IMMUNITY VERIFICATION TESTS†

G. T. Inouye,* A. C. Whittlesey,** S. R. Ponamgi,**
B. D. Cooperstein,* and A. K. Thomas*
*TRW Defense and Space Systems Group, Redondo Beach, California
**Jet Propulsion Laboratory, Pasadena, California

INTRODUCTION

Current models of the Jovian magnetosphere indicate that the Voyager spacecraft will encounter a high energy plasma environment which may cause spacecraft charging problems. Anomolous behavior attributed to spacecraft charging have been observed on many geosynchronous satellites.[1,2] An accompanying paper[3] presents some calculations in which Voyager spacecraft potentials and differential potentials are estimated on the basis of several Jovian environmental models and a charging model of the spacecraft.

In view of the possible hazards to the Voyager mission a thorough examination of the spacecraft configuration from the viewpoint of spacecraft charging was instituted and many modifications were implemented:

- Comparison of Voyager grounding system with Pioneer 10 and 11 and other spacecraft, and tests of alternate grounding techniques. A number of grounding changes were incorporated.
- Examination of all exterior surfaces for possible chargeup.
- Elimination of nearly all dielectric exterior surfaces.
- Modification of circuits to eliminate susceptibility to arc discharges.

As a final measure, a spacecraft charging immunity verification test was performed, and appropriate actions were taken to remedy the remaining problem areas indicated by the test results.

Initially, the Voyager project performed some simple tests and analyses which indicated that the then current design could pose electrostatic discharge (ESD) problems in the vicinity of Jupiter. As a result of these tests and analyses, the following program was implemented to assess the magnitude of the problem and to correct those problems which presented a hazard to the spacecraft:

- Identify ESD sources.
- Analytically model the spacecraft to highlight possible problem areas.
- Test and analytically study materials for ESD characteristics.
- Identify circuitry and locations on the spacecraft sensitive to ESD's.
- Identify areas in which shielding and grounding improvements may be made.
- Investigate alternate materials less prone to generate ESD's.
- Establish test methods and equipment to validate the analytical assessment of hazards.
- Make changes to the spacecraft as needed.

At the present time, analysis of spacecraft charging and arc discharging, and also techniques for immunity verification testing as well, are not as well established when compared to other aspects of spacecraft design such as for thermal or mechanical design. This paper emphasizes the test considerations and interactions with the analytic modeling, with some results presented in conclusion.

TEST APPROACH

The basic test philosophy adopted for the Voyager 1 flight spacecraft tests was that all of the tests would be performed in a manner such that none of the on-board equipment would be exposed to test stimulus levels which could be considered hazardous to its in-flight performance, while the maximum immunity verification information was obtained. The purpose of the test program was to determine if a problem existed and then verify the value of any space-

†This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract NAS7-100, sponsored by the National Aeronautics and Space Administration.

craft design changes. Test requirements included: 1) how to develop test methods (need to know source characteristics and then determine how to apply an equivalent stimulus to the spacecraft) and 2) how to monitor the spacecraft performance (with maximum test visibility but with minimum disturbance of the system due to grounding and additional test wire harnesses). Many other facets of the real-life situation had to be considered such as the unavailability of developed test equipment, the realities of possible spacecraft test configurations in view of schedule and manpower limitations, and the paucity of applicable information and prior experience within the scientific and engineering community in testing for the phenomena at hand. The schedule restrictions, in general, were the most limiting as may be deduced from the fact that the immunization effort was begun at the first of the year and the successful launch of both Voyager spacecraft occurred near the end of August, 1977.

Figure 1 shows the logic flow diagram for the Voyager electrostatic discharge immunity verification program. It was recognized that the actual spacecraft ESD tests were essential, but could comprise only a small part of a cost effective verification program. A significant portion of the effort depended on analytical work supported by available information and subsidiary tests. The analytic model used to consolidate subsidiary test data and studies was the Specification and Electromagnetic Compatibility Program (SEMCAP), a computerized model of the Voyager spacecraft already in use at the Jet Propulsion Laboratory (JPL).

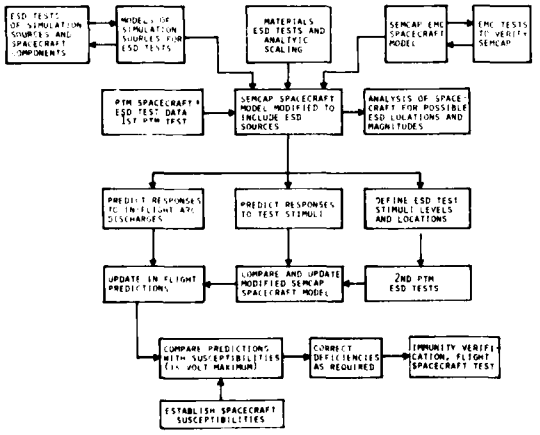


Fig. 1. Logic-Flow Diagram for Voyager ESD Immunity Verification

As implied by its name, SEMCAP was developed as an analytical tool to be used in the task of assuring the circuit-to-circuit electromagnetic compatibility (EMC) of all of the on-board electrical subsystems on a spacecraft. Basically, SEMCAP models all of the interbox harness cabling and all of the input and output interfaces for each box. The interaction of signals on each wire with every other wire is computed in terms of the physical configuration and terminating impedances. The multi-node set of electrical circuit equations are solved, and the computer program provides as outputs, the interference voltage levels on each interface wire integrated over the bandwidth of that circuit. In a typical EMC program the validity of SEMCAP is established on a statistical basis by checking only a selected fraction of all of the interfaces by tests.

A SEMCAP model which has been generated for EMC purposes for Voyager earlier in the program was modified to include arc discharge sources. Among the functions provided by the modified SEMCAP for the Voyager ESD immunity verification program were:

- Selection of diagnostic points and stimulus location
- Prediction of spacecraft responses to test stimuli
- Limitation of stimuli to benign levels
- Extrapolation of responses to those expected at other locations
- Prediction of spacecraft responses to in-flight arcs.

ARC DISCHARGE SIMULATION SOURCES

The two types of arc discharge simulation sources used in the Voyager tests are shown in Fig. 2. The radiated field arc source shown at the top of Fig. 2 was conceptually derived from Section 6.5.2.4.1, Electrostatic Discharge of MIL-STD-1541 (USAF). The coil used was a Transpark 400:1 automotive ignition coil. It was operated in parallel with a 2 μf capacitor which was discharged by a relay about once per second. Although radiated fields are not expected to occur in flight, this type of test stimulus was carried over from preceeding Proof Test Model (PTM) spacecraft tests to the flight spacecraft test to provide comparative data on the immunity improvements implemented. Tests with this type of stimulus also provided some of the data essential to establishing a measure of the accuracy of the SEMCAP model. The surface arc simulation source is shown in the lower part of Fig. 2. The aluminum foil is insulated from the surface to be tested with a 3-mil sheet of mylar. Its capacitance, determined by the test area, is charged to a predetermined voltage by manually adjusting the high voltage power supply through 500 megohm isolation resistors. For safety, the maximum applied potential was limited by preadjusted arc gaps in parallel with the power supply. It was recognized that unintentional coupling from the source to "target" circuits, including diagnostic equipment, could generate false data. For this reason the sources were battery operated in order to eliminate coupling into power lines.

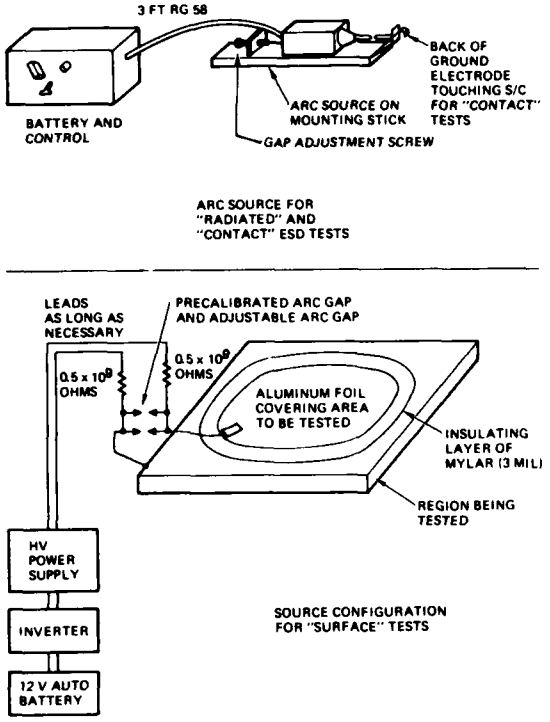


Fig. 2. Arc Simulation Sources

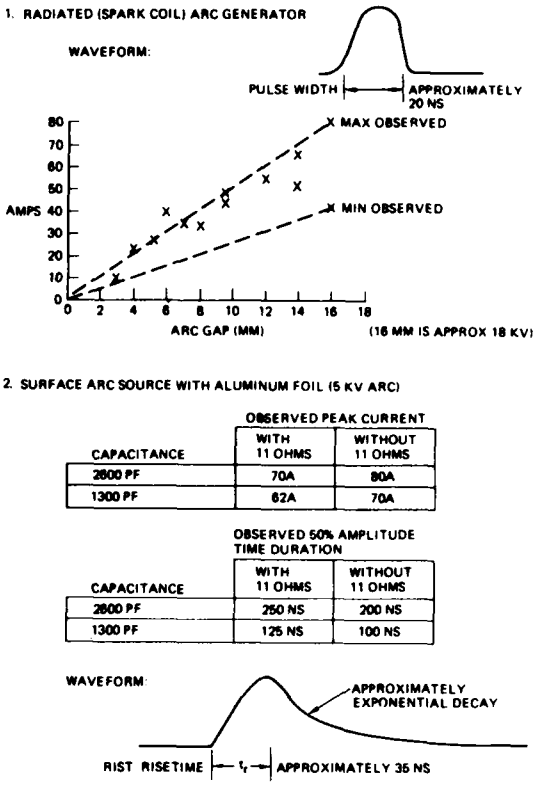


Fig. 3. Arc Simulation Source Calibration

In addition, the radiation and coupling from the supporting equipment, batteries and power supplies were minimized as best as permitted by available time constraints.

Figure 3 shows the calibration data for the sources. The spark coil arcing voltage is adjusted by means of the gap adjusting screw shown in Fig. 2. The relation between gap width and arcing voltage turned out to be nearly linear at about 1 kV/mm (actually 18 kV for 16 mm). The spark coil pulse shape as shown in Fig. 3 was approximately 20 ns in duration irrespective of arc breakdown voltage. The gap-width to peak current relation, a function of the coil self-capacitance (35 pf) and the external circuit inductance, was in the order of 50 amperes peak at 14 mm or 15 kV. The surface arc source was triggered at 5 kV for the test data shown in Fig. 3. The approximately 35 ns risetime observed was determined by the associated circuit inductance, and the 100-200 ns fall time increased as the area of aluminum foil used. As an aside, this simple test indicates that pulse duration scales with discharge area, contrasting with the hypothesis that pulse amplitude scales with area.

MODELING OF ARC DISCHARGES FOR SEMCAP

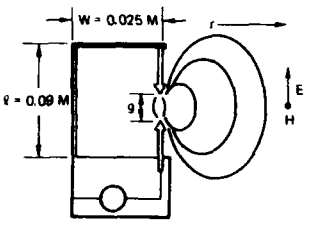
To modify SEMCAP to include arc discharge sources, it was necessary to model these in terms of equivalent voltage and current generators at the proper locations. Four generator models were developed to simulate radiated and conducted coupling effects of the electrostatic discharges. Due to time constraints, the simplest models that would provide approximate simulation to within about ± 10 db were developed.

The coil excited spark gap creates radiated electric and magnetic fields as shown in Fig. 4. These radiated fields were approximated by two models:

A. An induction field model consisting of quasi-static electric and magnetic fields proportional to the voltage and current of the source, respectively. These fields are short range and decay with distance as approximately r^{-3}.
B. A radiated field model representing the far field electromagnetic radiation of the loop antenna formed by the source. These fields are longer ranged and decay with distance as r^{-1}.

The local electric field representation approximates the quasi-static electric fields radiated from the dipole formed by the two electrodes. It has a frequency independent antenna

factor that is multiplied by the electrode voltage spectrum to obtain the local E-field spectrum. The other three antenna factors for these two models are taken from the near and far field terms for the radiation of a small magnetic dipole. It should be noted that the radiated E-field term is proportional to the arc current so that it utilizes the current parameters of the arc instead of its voltage parameters.



E AND H INDUCTION FIELDS PLUS RADIATED FIELDS

INDUCTION FIELD SOURCE MODEL

$E = V(\ell w) r^{-3}$

$H = I(\ell w/4\pi) r^{-3}$

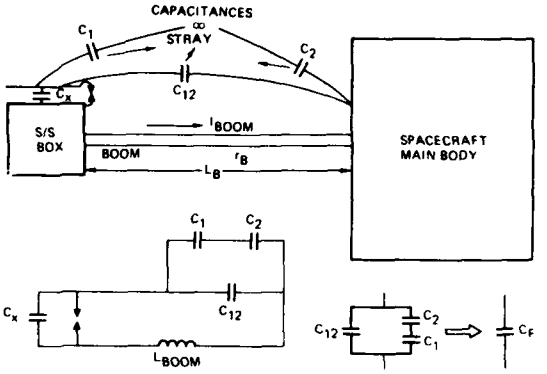
LOOP ANTENNA RADIATED FIELD MODEL

$E = I(377 \cdot 2\pi\ell \cdot w, \lambda^2) r^{-1}$

$H = I(2\pi\ell \cdot w, \lambda^2) r^{-1}$

Fig. 4. Models for Radiated Fields

Arc discharges of dielectric or insulated conductor surfaces create remote structure replacement currents as well as producing local capacitive and inductive coupling. Two surface discharge source models were required to simulate these effects. The structure replacement current effect is shown in Fig. 5 for a surface discharge on a box situated on one of the payload booms. The replacement current flows to restore potential equilibrium in the structure as the surface is discharged. The boom current will be a small fraction of the arc current and can be modeled as a filter which has a current transfer ratio equal to the ratio of impedances of the two paths. These impedances were approximated by a much simplified electrically equivalent model of the structure. The necessary equations are also shown in Fig. 5. The induced voltage on the boom was modeled as a second filter function multiplying the boom current. This filter has a gain equal to the impedance of the boom ground strap. The localized capacitive and inductive coupling effects of the surface ESD is most easily modeled by translating the surface into an approximately equivalent large diameter wire as shown in Fig. 6. This equivalent wire capacitive coupling is driven directly by the step voltage change of the discharged surface and the arc current drives the inductive coupling.



$$I_{BOOM} = I_{ARC}\left(\frac{X_{CX}}{X_{LB} + X_F + X_{CX}}\right)$$

$$= I_{ARC}\left(\frac{1/2\pi f C_X}{2\pi f L_B + 1/2\pi f C_F + 1/2\pi f C_X}\right)$$

$$= I_{ARC}\,(C_F/(C_F + C_X))\,(4\pi^2 f^2 C_{eq} L_B + 1)^{-1}$$

$$= I_{ARC} \cdot P_1(f)$$

$P(f)$ = LOW PASS (TYPE = 1)

G' = INBAND GAIN = $C_F/(C_F + C_X)$

$f_c = (2\pi\sqrt{L_B C_{eq}})^{-1}$

ORDER = 2

$C_{eq} = C_F C_X/(C_F + C_X)$

$V_{BOOM} = I_{BOOM} \cdot X_{LB}$

$V_B = I_{ARC} \cdot P_1(f) \cdot P_2(f)$

$P_2(f)$ = HIGH PASS (TYPE = 2)

$G_1'' = 2\pi L_B f_c$

$f_c = 10^9$ (ARBITRARILY HIGH)

ORDER = 1

L_B = INDUCTANCE OF BOOM GROUND STRAP

Fig. 5. Model for Boom Replacement Current

TEST CONFIGURATION AND TEST PROCEDURES

As noted previously, the possible test configurations were severely limited by schedule considerations as well as other considerations such as the unavailability of the radioisotope thermal generator (RTG) power source for this test. The most severe constraint was the desire to minimize test wire attachments to the spacecraft (to an oscilloscope) since testing indicated the extra wiring itself can act as an antenna and cause erroneous results. Also as noted previously, an objective of the test plan was to obtain the most crucial data in the most "efficient" manner. That is, to minimize the number of test locations and diagnostic points and to depend to a major extent on the analytical capabilities inherent in the use of the SEMCAP model.

Figure 7 shows the diagnostic setup used for the flight spacecraft tests as well as for the initial tests performed on the PTM spacecraft. As indicated in Fig. 7, four diagnostic test points were selected. These test points, made accessible with breakout connectors, were monitored differentially with high impedance probes on two oscilloscopes. Instead of bringing wires out from the spacecraft, long oscilloscope

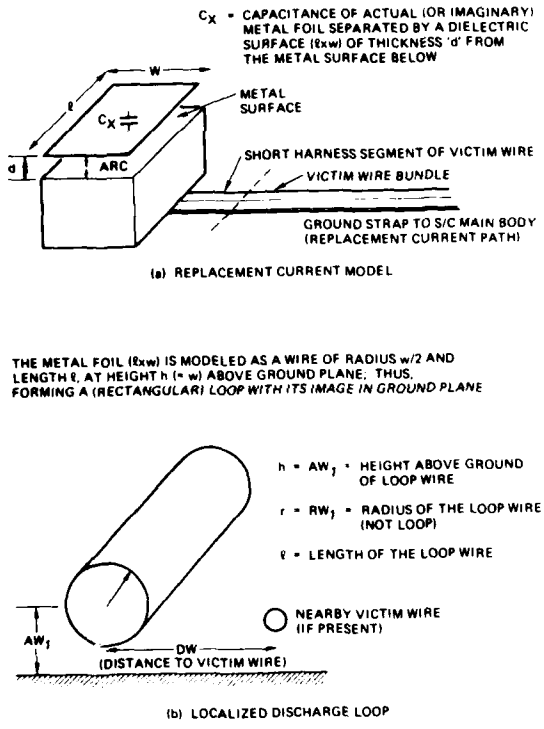


Fig. 6. Surface Discharge Model for Local Replacement Current Coupling

probes/leads were attached as close as possible to the desired monitor point, routed close to spacecraft structure for shielding from radiated fields and then brought out to the oscilloscopes which were located adjacent to the spacecraft. We had hoped to use battery and inverter power for the oscilloscopes but the oscilloscopes did not work on the inverters; ultra isolation transformers were used instead. The full-up spacecraft with the science boom deployed, as shown in the photograph of Fig. 8, was insulated from the floor to minimize unreal coupling effects. Power was brought in on cables from an external power supply, but the spacecraft was otherwise completely isolated electrically. The spacecraft systems were monitored via "air" using the spacecraft radio and telemetry link.

At each test location, after setting up the arc simulation source, the test data acquisition procedure was planned to increase the stimulation level from "very low" to a maximum of 3 db lower than benign and safe levels as determined from the earlier PTM spacecraft tests and from SEMCAP analyses. At each level, sufficient time was allowed monitoring personnel for each spacecraft subsystem as well as for each diagnostic point to verify that hazardous levels were not being approached, before going on to the next higher stimulation level. A sufficient number

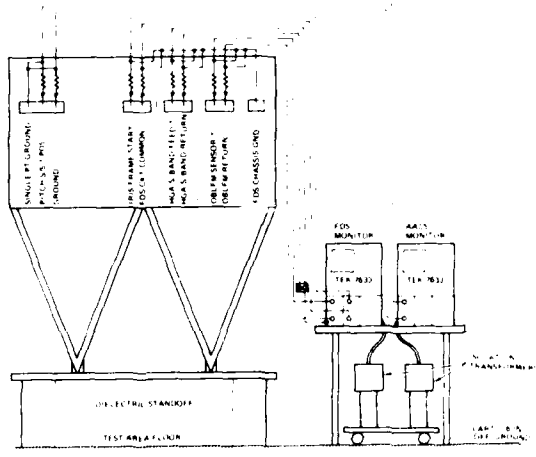


Fig. 7. Test Setup Showing Diagnostic Techniques

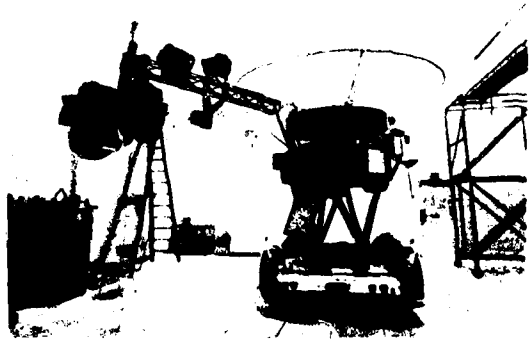


Fig. 8. Voyager I ESD Immunity Verification Test

of oscillographic traces for each diagnostic point were observed to assure that a representative photographic record was taken at each step.

TEST RESULTS AND EXTRAPOLATIONS TO IN-FLIGHT RESPONSES

The first ESD test of the PTM spacecraft (performed before ESD changes) demonstrated that the spacecraft was susceptible to arc discharges and showed susceptible areas. It also demonstrated the need for better source simulation, a more isolated spacecraft, and diagnostic monitors on some circuits. The second PTM ESD test (performed after many ESD fixes) showed great improvements in the spacecraft configuration, source simulation, refined our knowledge of susceptible spacecraft circuitry, and how to monitor spacecraft functional response.

The ESD test of the flight spacecraft (after further fixes) went very smoothly. The test results are shown in Table 1 along with the levels predicted by SEMCAP. The source parameters used in predicting the test results are shown in Table 2 (reflecting the arc parameters of the simulation source) and those for predicting in-flight responses are shown in Table 3 (reflecting test data and worst-case analytic scaling to spacecraft dimensions). A partial listing of the peak interference voltages are shown in Table 4 (sampled from a 12-source vs 68 victim interaction matrix compiled during the activity). As noted at the bottom of Table 1 the mean error between predicted and measured results is -6 dB and the standard deviation was 23 dB. These values are considerably different from those determined by previous error analyses of SEMCAP predictions for intra-system interference coupling. These other analyses resulted in a mean error of +9 dB, and the standard deviation was 10 dB. With more experience in modeling ESD's, it is expected that these errors will become smaller. Assuming that the above accuracy parameters also apply to the predicted in-flight responses, the spacecraft may be considered to be immune to arc discharges, since these were generally below 20 millivolts.

The predictions shown in Table 4 and similar computer runs were used to initiate numerous design changes, such as removing the Kapton (dielectric) thermal blanket on the MIRIS science instrument and replacing it with a conductive surface blanket material. The predictions also permitted the retention of the Brewster plate (a large dielectric area) on the basis that it created few if any significant transients on spacecraft circuitry.

SUMMARY AND CONCLUSIONS

The test philosophy, test simulation sources, test configuration and test results for the Voyager spacecraft arc discharge immunity verification have been presented. A major effort to

Table 1. Voyager ESD Test Results vs SEMCAP Predictions

LOCATION OF ARC	VOLTS DURING RADIATED TESTS							
	HGA		IRIS		OBLEM		SUN SENSOR	
	PRED	MEAS	PRED	MEAS	PRED	MEAS	PRED	MEAS
IRIS/ISS	0.9	1.2	2.0	0.8	0.09	1.0	0.08	2.5
LECP	1.5	0.8	6.0	0.45*	0.4	0.7	0.1	1.5
SUN/SENS	0.84	1.0	0.45	0.3*	0.9	0.4	4.0	1.6
MAG	0.4	0.4*	0.8	0.5*	3.6	1.2	0.4	0.5
FSS	4.9	4.0	0.96	1.5	2.1	1.4	0.002	2.0
BREWSTER PLATE MEAS: CONTACT PRED: RAD	2.2	4.0	1.5	4.0	0.9	3.3	0.54	12.2
	VOLTS DURING SURFACE TEST							
LECP	0.27	0.6	15.0	0.6*	0.04	0.6	0.017	0.8
BREWSTER PLATE	6.8	1.0	0.37	0.6	0.6	0.7	0.4	0.9
FSS	57.0	10.0	0.09	4.2	0.2	1.7	0.001	4.0

*BACKGROUND NOISE. NOISE DUE TO ARC UNNOTICEABLE

MEAN ERROR = -6.3 dB (UNDERPREDICTING) } NOT INCLUDING
STANDARD DEVIATION = 23.4 dB } *(STARRED) ENTRIES
EACH RECEPTOR MODELED WITH 75 MHz BANDWIDTH
PREDICTIONS FROM RUN OF 3/29/77

Table 2. Parameters of Test Arc Discharge Source Models

Arc Source	Type of Arc	Breakdown voltage V	Discharge current $I_{(A)}$	Risetime of discharge current t_r	Discharge current pulse width t_p	Replacement current path inductance $L_{(\mu h)}$	Main discharge capacitor C_x	Equiv stray capacitance C_f	I_{repl}/I_{arc} G'_I	$\frac{V_{repl\ volt}}{I_{arc}}$ G'_I	Corner frequency f_c	Looplength $\ell(m)$	Loopheight $h_{(m)}$	Wire radius $r_{(m)}$
IRIS/ISS	Radiated	12 kV	60	2 ns	20 ns	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Sun Sensor	Radiated	12 kV	60	2 ns	20 ns	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MAG Canister	Radiated	12 kV	60	2 ns	20 ns	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
FSS	Radiated	12 kV	60	2 ns	20 ns	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brewster Plate	Contact	12 kV	60	2 ns	20 ns	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
LECP	Surface	5 kV	25	10 ns	10 ns	1.8	56 pf	7 pf	0.12	1.9E4	35 MHz	0.3	0.15	0.07
Brewster Plate	Surface	5 kV	37.5	10 ns	20 ns	0.25	150 pf	12 pf	74E-3	1.5E3	95 MHz	0.25	0.15	0.07
FSS	Surface	5 kV	12	10 ns	10 ns	2.0	12 pf	13 pf	0.5	1.25E4	45 MHz	0.3	0.15	0.07

Table 3. Parameters of In-Flight Surface Arc Models

Assumes Arc Discharge Current Greater Than Measured "Replacement" Current by Factor $\frac{C_x + C_f}{C_f}$ (of test setup)

Arc Source	Breakdown Voltage V	Discharge or Arc Current I	Discharge Current Rise Time t_r	Discharge Current Pulse Width t_p	Replacement Current Path Inductance $L_{B(\mu h)}$	Main Discharge Capacitance C_x	Equivalent Stray Capacitance C_f	$G'_I = I_{repl}/I_{arc}$	$G''_I = V_{repl\ volt}/I_{arc}$	Corner Frequency f_c	Looplength ℓ	Loopheight $h_{(m)}$	Wire Radius $r_{(m)}$
MAG Cable	5 kV	20 A	10 ns	1.7 µs	2.3	0.05 µf	30 pf	5.0 E-4	82 E3	8 MHz	0.6 m	6 E-5	2.5 E-4
HGA Paint (OB)	1 kV	150 A	5 ns	3 µs	0.4	0.4 µf	500 pf	1.25 E-3	1.25 E3	15.9 MHz	1.8 m	2 E-4	5 E-5
Plume Shield (Sep Conn)	1 kV	16 A	20 ns	285 ns	1.5	4500 pf	0.2 pf	4 E-5	9.4 E3	290 MHz	0.2 m	5 E-5	2.5 E-5
FSS	7 kV	80 A	8 ns	80 ns	1.9	14 pf	53 pf	1.0	12 E3	34 MHz	0.35 m	0.025	5 E-5
HGA Paint (IB)	1 kV	150 A	5 ns	2.4 µs	0.4	0.3 µf	70 pf	2.3 E-4	1.25 E3	42 MHz	1.0 m	2 E-4	5 E-5
Plume Shield (RTG)	1 kV	16 A	20 ns	330 ns	0.8	5200 pf	2.5 pf	2.5 E-4	9.5 E3	83 MHz	0.17 m	5 E-5	2.5 E-5
RTG Oxide	3.5kV	925 A	20 ns	3.7 µs	2.8	0.34 uf	90 pf	2.5 E-4	1.1 E4	12 MHz	0.85 m	7.6 E-5	3.8 E-5
MIRIS KAPTON	1 kV	150 A	5 ns	26 ns	4	40 pf	0.032 pf	1.0	2.5 E4	12 MHz	0.24 m	5 E-5	2.5 E-5

G'_I = Replacement Curr/Arc Curr

G''_{II} = Repl Line Voltage/Arc Curr

Table 4. Revised SEMCAP Predictions of Interference (Volts) on Interface with RC Filters Due to In-Flight Arcs

RECEPTOR NAME	ARC SOURCE: MAG CABLE	HGA OUTBOARD FACE	FSS	HGA INBOARD FACE	PLUME SHIELD ON RTG STRUT	RTG OXIDE	KAPTON ON MIRIS
PLAYBACK DATA		---	---	---			---
CRS CMD WORD		---	---	---			0.044
PMS ADC BIT SYNC		0.001	---	0.001			---
PRA COMD WORD		---	---	---			---
LECP COMD WORD A		---	---	---			0.04
PPS COMD WORD		---	---	---			3.0
UVS MODE CONTROL		---	---	---			3.0
UVS SCIENCE DATA 1		---	---	---			3.0
MAG SAMPLE B		---	---	---		---	---
MAG OBLFM SENSOR TEMP		0.014	0.01	---	---	---	---
ISS-WA ADC START			---	---			35
ISS-WA ADC VID DATA							18
MAG IBHFM SENSOR TEMP	0.02		0.004	---	---	---	---
IRIS FRAME START		---	---	---			3.46
PITCH CRUISE S/SI POSN		---	---	---		---	---
HGA S-BAND FEED TEMP		---	0.23			---	---
RTG CASE TEMP		---	0.004			0.3	---

---DENOTES LESS THAN A MILLIVOLT, PREDICITONS FROM RUN OF 06/17/77
ARC SOURCE PARAMETERS PER TABLE 3.

achieve this immunity was implemented and the test results verify that it was achieved.

Testing of a spacecraft to verify its immunity to environmental charging and the resulting arc discharges is not an established procedure. The paucity of prior experience within the engineering as well as scientific community in understanding and in dealing with the phenomena, led to innovations in the test philosophy and the test implementation described in this paper. The approach of using analytical techniques to support and supplement the tests is very cost effective and leads to an efficient usage of in-line schedule time which is extremely expensive.

Much further work is required in developing techniques for spacecraft design as well as for arc discharge immunity verification testing. A way to simulate all of the arc discharge parameters of space during ground based testing is needed as well as improved analytical models. In addition to defining the environment, the characterization of spacecraft materials, how they charge and discharge, and how the arcs couple into spacecraft subsystems are the sort of basic information needed in order to achieve more cost effective spacecraft designs and to perform more realistic tests and analyses.

The Spacecraft Charging at High Altitudes (SCATHA) program[4] and the joint USAF/NASA Spacecraft Charging Investigation Program efforts are major undertakings which are currently funded. The work presented here underscores the urgency of the need for the results of these kinds of investigations.

ACKNOWLEDGMENTS

The work reported here involved extensive support and guidance from many areas of the Voyager team. Among those at JPL to whom we are indebted are J. B. Barengolta, A. J. Beck, W. F. Peer, and T. E. Gindorf of the Environmental Qualification Section, G. Cunningham and T. Risa of the System Design and Integration Section,

and W. S. Shipley of the Voyager Project Office. NASA/Lewis Research Center, particularly N. J. Stevens and F. D. Berkopec, provided consulting and ESD testing of Voyager components. The Boeing Aerospace Company also provided ESD data on Voyager parts under Contract 953957 CWO No. 10. The TRW authors were consultants to JPL under the leadership of A. Rosen under Contract No. 954684. We also acknowledge the work performed by Mrs. B. Benefield in coordinating the inputs from the various authors and in typing and producing the final manuscript.

REFERENCES

1. Rosen, A., "Spacecraft Charging: Environment Induced Anomalies," AIAA Paper 75-91, AIAA 13th Aerospace Sciences Meeting, January 20-22, 1975, Pasadena, Ca.; also *J. Spacecraft & Rockets, 13,* No. 3, March 1976, pp. 129-136.

2. Inouye, G. T., "Spacecraft Charging Anomalies on the DSCS II, Launch 2 Satellite," presented at Spacecraft Charging Technology Conference, U.S. Air Froce Academy, Colorado Springs, Colorado, Oct. 27-29, 1976; also published in the Proceedings, AFGL-TR-77-0051, 24 February 1977, NASA TMX-73537.

3. Sanders, N. L. and G. T. Inouye, "Voyager Spacecraft Charging Model Calculations," presented at the Second International Ionospheric Effects Symposium, IES '78, Jan. 24-26, 1978, Arlington, Va.

4. Shane, D. F., Lt. Col., USAF, "Space Test Program Spaceflight P78-2 (SCATHA)," presented at the Spacecraft Charging Technology Conference, U.S. Air Force Academy, Colorado Springs, Colo., Oct. 27-29, 1976; also published in the Proceedings, AFGL-TR-77-0051, 24 February 1977, NASA TMX-73537.

IONOSPHERIC EFFECTS OF A HIGH POWER SPACE-BORNE MICROWAVE BEAM*

L. H. Holway, Jr.
Raytheon Research Division, Waltham, MA 02154

and

A. H. Katz and G. Meltz+
Raytheon Company, Equipment Division, Wayland, MA 01778

ABSTRACT

A study has been made of the effect upon the ionosphere of microwave beams with intensities which would be produced by a satellite power system (SPS). The SPS captures energy from the sun and transmits the energy back to earth in a collimated microwave beam at 2.45 GHz for conversion to high voltage AC power for injection into the electrical power grid. The microwave field accelerates free electrons in the ionosphere which are then cooled by elastic and inelastic collisions with neutral molecules and also by the electron thermal conductivity along the magnetic field at high altitudes.

The rise in electron temperature, which depends not only upon the altitude and the intensity of the microwave beam but also upon the angle between the microwave beam and the magnetic field, can be several thousand degrees in the immediate vicinity of the beam. Nevertheless, the change, in the temperature of the neutral gases is negligible and no effect on chemical reactions has been identified, except that the higher electron temperature may reduce the recombination rate and increase the λ6300 optical emissions (nightglow). These effects were observed under comparable circumstances in the HF ionospheric modification experiments carried out at Platteville, Colo. In the microwave case, the effects will be localized in the vicinity of the beam so that there should not be a large effect upon the electromagnetic systems (communication/ control/ navigation/ positioning/ direction-finding/ radar/ etc.) which involve the properties of the ionosphere.

INTRODUCTION

A satellite power system which collects solar energy and transmits it back to earth by a microwave power transmission system (MPTS) could be a major practical application of space technology in the future, since it would generate electrical power on a large scale [1-4].

A series of investigations have been carried out to analyze critical aspects of the SPS. One area which must be thoroughly studied is the possibility of an environmental impact due to the microwave transmission through the ionosphere [4]. Based on theoretical models of ohmic heating, the magnitude of the ionospheric effects in both the F-layer and the D-layer has been calculated at 20, 80 and 320 mW/ cm^2 power densities at 2.45 GHz for both a southwest and a northeast U.S. site. These densities are representative of densities to be associated with a 5, 10 and 20 GW output SPS.

The study [4] of the ionospheric effects consisted of: (1) calculating the ionospheric changes that result from heating as a function of power density and time of day; and (2) identifying the effects of the heating on the power beam and on communications, navigation and other users of the ionosphere.

The analysis was carried out by dividing the ionosphere into low and high altitude regimes. At low altitudes (D- and E-regions), the electrons are cooled locally by inelastic collisions with the neutral atmosphere while, above 150 km, the most important effect is electron thermal conduction along the magnetic field.

One effect in the D region which, perhaps, might not have been anticipated is a sharp rise in electron temperature which occurs due to nonlinear effects when the intensity of the 2.45 GHz radiation increases from about 10 mW/ cm^2 to about 20 mW/ cm^2. Normally the major cooling effect for electrons in the D region is assumed to be the excitation of rotational modes of energy in diatomic oxygen and nitrogen. While this assumption is true for the "normal"

* Parts of this work were supported by NASA, MSFC, under Contract NAS8-31308.

+ Present address: United Technologies Research Center, East Hartford, Conn. 06108

BEST AVAILABLE COPY

power fluxes due to radar or radio transmissions, at the fluxes and electron temperatures that we are now considering, this cooling mode becomes "saturated". In fact, if only the rotational cooling were considered, the electron temperature would become unstable. Instead, a new cooling mechanism - the excitation of vibrational modes - becomes important and limits the electron temperature rise. The same effect has been described for the HF ionospheric modification experiments [5-7], where it would be expected to occur at much lower power fluxes (625 $\mu W/m^2$ at 5 MHz). These HF ionospheric modification experiments, carried out at Platteville, Colorado [8], have an S/f^2 ratio similar to the MPTS, where S is the power (Poynting) flux and f is the frequency. Since radio heating effects should be comparable when the ohmic loss is comparable, these experiments may provide important clues to the effects that the MPTS might have on the ionosphere.

D-REGION HEATING

At 2450 MHz, the microwave frequency is so high that only a minute fraction of the beam's power is absorbed in traversing the ionosphere. However, the ohmic heating per unit volume is

$$Q = \frac{e^2 \nu n}{m\omega^2} < E^2 > \qquad (1)$$

where n is the electron density, ν is electron collision frequency with neutral molecules and with ions, m and e are the mass and charge of the electron, and the brackets indicate that E^2 is averaged over an RF cycle. Since the ratio of S/ω^2 for the SPS is comparable to that produced in the ionospheric modification experiments at Platteville and Aericebo [8-9], the large body of information gathered in the modification experiments provides a good deal of insight into the effects which might be expected for the SPS.

In the daytime when there is considerable ionization in the D-region, the maximum absorption occurs at the altitude where $n\nu$ is largest, near 70 km. At these altitudes the effect of thermal conduction and motion of the plasma is negligible, and the rate of change of the electron temperature is

$$\frac{\partial T_e}{\partial t} = AS - \frac{2}{3}(U_1 + U_2 + U_3 + U_4) \qquad (2)$$

where the first term on the right hand side is derived from Eq. (1) and, for 2450 MHz transmission in the D-region,

$$A = 5.38 \times 10^{-13} n_2 T_e \qquad (3)$$

where n_2 is the sum of N_2 and O_2 densities in cm^{-3}, S is in mW/cm^2, and T_e is in °K. The U_i terms are cooling rates; for weak radio waves, only U_1, which is due to the excitation of rotational modes of energy in O_2 and N_2, is important. This rate is given by [5]

$$U_1 = 3.15 \times 10^{-10} n_2 T_e^{-1/2} \delta T_e \qquad (4)$$

where $\delta T_e = T_e - T_N$ where T_N is the ambient neutral gas temperature. For these weak intensities, T_e is close to T_n and Gurevich's linearization is valid, i.e., T_e can be replaced by T_n in Eqs. (3) and (4), in which case δT_e is proportional to S. However, when $T_e \gg T_n$, the linearization is obviously invalid. If only U_1 was considered, a solution of this nonlinear equation would indicate a temperature increasing unboundedly with time. Actually, according to a criterion formulated during an investigation of the HF case [5], the energy loss rate U_3 will control the cooling rate as soon as electromagnetic flux exceeds a critical power density, which is on the order of 15 mW/cm^2 for 2450 MHz frequency used by the MPTS. By numerically fitting a formula to Dalgarno's tabulated loss rate [11], we expressed this rate as

$$U_3 = 1.25 \times 10^{-21} n_2 (T_e^4 - T_N^4) \quad , \qquad (5)$$

over the temperature range from 1500-3500 °K.

The loss rate U_2, which describes fine-structure excitation in molecular oxygen, becomes important above 90 km. In principle, the loss rate U_4, due to elastic collisions, would be capable of controlling the runaway electron temperature in the absence of U_3, but, in fact, U_4 has a negligible effect upon the computations described in this paper.

The steady-state electron temperatures at 60, 100 and 110 km altitude are shown in Fig. 1 as a function of power flux. These electron temperatures were calculated from Eq. (2), with the formulas for U_1 and U_3 given by Eqs. (4) and (5) while the formulas for U_2 and U_4 are given elsewhere [5]. The dashed line shows the linear Gurevich approximation at 60 km; we see that the correct electron temperature is much higher than this approximation for fluxes greater than 10-15 mW/cm^2. Between 60 and 90 km, the steady-state temperature is almost independent of altitude; however, above 90 km, the fine-structure cooling rate, U_2, becomes important. The effect of U_2 is to reduce the temperature and increase the critical power, as is indicated by the curves for 100 and 110 km. This effect, due to the fine-structure cooling can be seen more dramatically in Fig. 2 where the steady-state temperatures are plotted as a function of altitude for specified values of the power flux.

The higher electron temperature increases the absorption by increasing the collision frequency. Although the absorption is still negligible for the microwaves themselves, Fig. 3 shows the increased

BEST AVAILABLE COPY

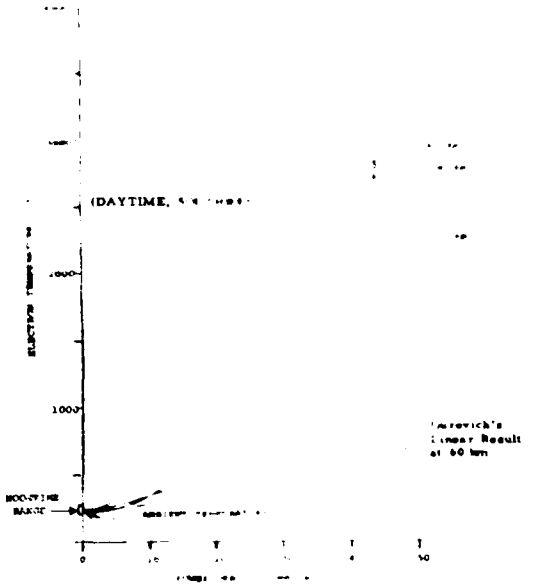


Fig. 1. Electron temperature at specified altitudes as a function of the incident energy flux.

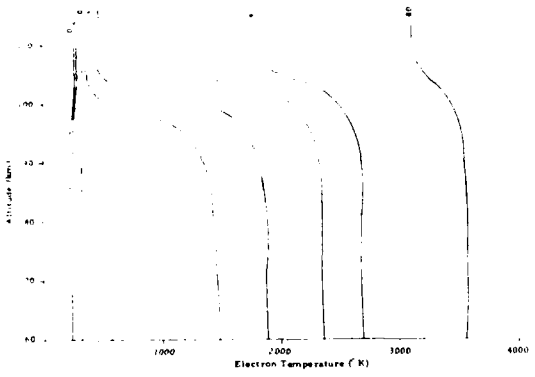


Fig. 2. Electron temperature for specified energy fluxes as a function of altitude. Intermediate fluxes, which exceed the critical flux at lower altitudes where the energy loss to atomic oxygen is negligible, are below the critical flux at altitudes where this cooling effect is important.

absorption that a 10 MHz radio wave would experience in passing through the disturbed volume. The solid curve shows the ambient absorption which, as shown by the dashed curve, is greatly increased after the electron temperature rises (which only requires times of the order of ms after the MPTS is turned on). The dot-dash curve shows a still

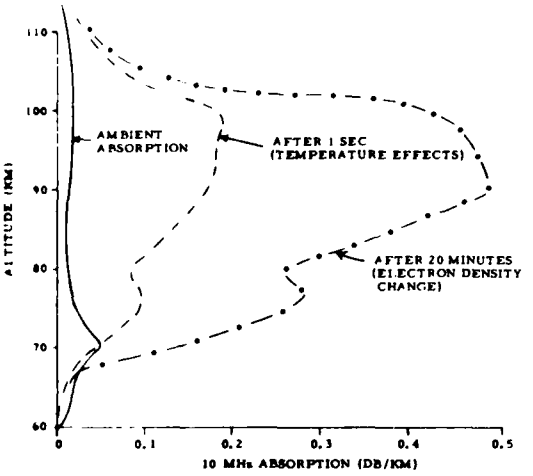


Fig. 3. Absorption coefficient per km for a 10 MHz ordinary wave passing through a region heated by a 20 mW/cm^2 MPTS. The dashed line includes only increased collision frequency due to increased temperature. The dot-dash line includes an increased electron density caused by the reduced recombination described by Eq. (6).

larger increase, requiring about 20 minutes to develop, which is due to an increase in electron density. This density increases slowly because the recombination coefficient decreases as the electron temperature rises. The recombination coefficient in these calculations was taken to be

$$\alpha = 5 \times 10^{-6} (180/T_e)\ \mathrm{cm}^3/\mathrm{sec} \qquad (6)$$

which is consistent with published recombination coefficients involving water cluster ions [7]. Although this value for the effective recombination coefficient gives a time-scale comparable to that observed at Platteville, the ion chemistry in the D-region is complex and requires further study.

HEATING EFFECTS IN THE F-REGION

For altitudes above about 115 km where the ion-neutral collision frequency, ν_{in}, is less than the (angular) ion cyclotron frequency ω_{Bi}, plasma motion is directed along the magnetic field (except for the $\vec{E} \times \vec{B}/B^2$ drift). Therefore, the plasma behavior can be treated as a one-dimensional problem along each magnetic flux tube; the segment of the flux tube which is heated will have a length equal to $D/\sin\alpha$ where D is the effective diameter of the beam and α is the angle between the flux line and the microwave beam. The geometry for a Southwest site (near White Sands, New Mexico) is shown in Fig. 4 where, assuming the satellite is directly above the equator, α

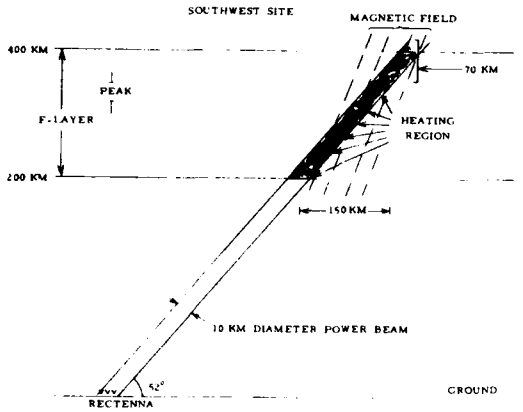


Fig. 4. A near-grazing angle between the power beam and the magnetic field at a Southwest site near White Sands, New Mexico.

is 7.6°. Results will also be given for a Northeast site near Bedford, Mass. where $\alpha = 31.5^{\circ}$. The calculations reported in this paper treat 3 system designs in which energy fluxes of 20, 80 and 320 mW/cm^2 were paired with beam diameters of 10, 7.1 and 5 km respectively.

The electron temperature is calculated by a computer program which solves one-dimensional plasma hydrodynamic equations which are described by Meltz and LeLevier [11]. The most important cooling effect is the thermal conduction in the electron gas which depends on the 5/2th power of the electron temperature. The coupled hydrodynamic equations also include pressure gradients and motion of the plasma, elastic energy transfer between electrons, ions and neutrals and inelastic transfer of the electron energy due to the rotational excitation of diatomic molecules and to fine-structure excitation in atomic oxygen. Although some energy is transferred to ions and neutral particles at high altitudes, much of the heat is conducted down to lower altitudes near 200 km where collisions with neutral particles are frequent enough to cool the electrons.

The electron temperatures are shown for the Northeast site at the peak of the F-region in Fig. 5, as a function of time after the MPTS is turned on, using a nighttime profile. These temperatures increase from their ambient value of 1200°K until they reach a steady-state after 20-30 seconds of heating.

The effect of continuing the heating for a longer time on the electron temperature and plasma density can be seen in Fig. 6. Although the temperature reaches a steady state in about 30 seconds, the density continues to decrease for a much longer time as

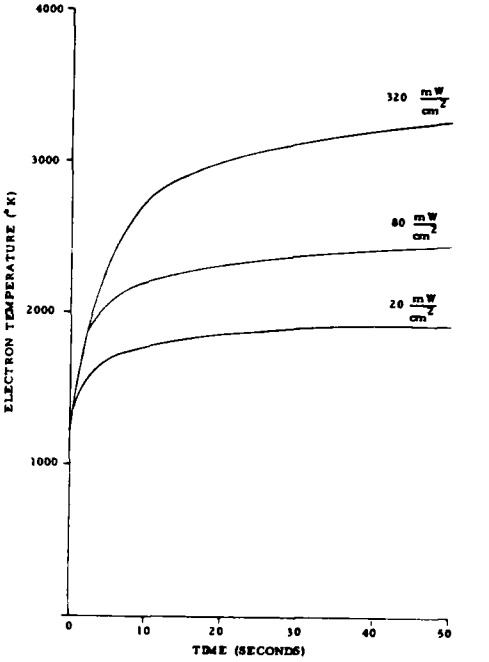


Fig. 5. Electron temperature as a function of time after turn-on at the peak of the F-region at the Northeast site.

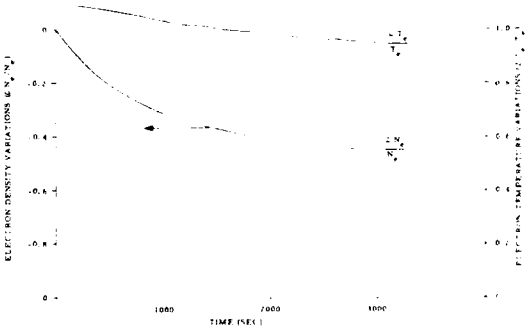


Fig. 6. Changes in electron density and temperature at night when the heating continued for longer times. Southwest site for 20 mW/cm^2 system. Actually the plasma will drift out of the beam after about 200 seconds.

the thermal expansion forces the plasma up the lines of force. These calculations overestimate the decrease in density because they ignore the $\vec{E} \times \vec{B}$ drift, which would cause a particular flux line to move out of the beam in about 200 seconds. After 200 seconds, Fig. 6 shows the plasma density is only about 8 percent below ambient, and a "wake" may be expected outside the beam since it will take time comparable to 200 seconds for ions to

drift back into the "hole". The plasma density changes shown here are due only to expansions and compressions and do not include alterations of the recombination coefficient due to electron temperature changes such as were included in our D-region calculations and which may also modify the high-altitude density changes.

The curves in Fig. 7 show the electron temperature as a function of altitude at both

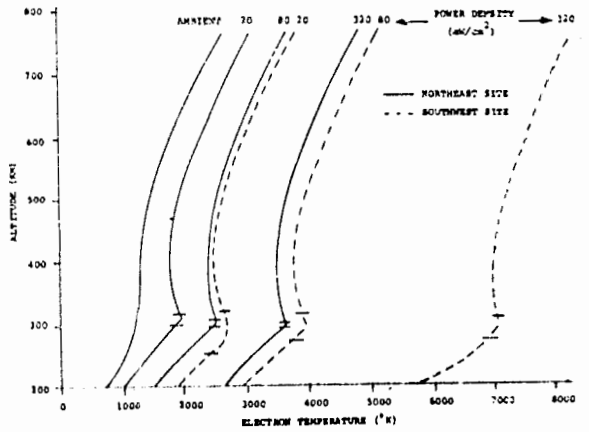


Fig. 7. Steady-state temperature at night as a function of altitude for the Northeast and Southwest sites in the ambient ionosphere and for the three system designs. Slashes indicate the altitude over which the line of force is heated.

the Northeast and Southwest sites for the 3 systems under design consideration. These temperatures were calculated at t = 200 seconds after turning on the beams, which corresponds to the length of time a given flux line remains in the beam and is sufficient time for the electron temperature to reach a steady state. Most of the heat is deposited near 300 km and much of it is then conducted downward to altitudes near 200 km where neutral particles are sufficiently numerous to cool the plasma. In these high altitude calculations, the loss of energy to vibrational levels in N_2 has been neglected although they are larger than rotational losses when the temperature exceeds 2000°K. However, although these losses may become noticeable for the highest power fluxes under consideration, the relatively low concentration of diatomic molecules in the F-region diminishes the importance of these losses compared to thermal conductivity losses, as pointed out earlier by Meltz and LeLevier [11].

The boundary conditions for the hydrodynamic equations have been discussed in [5] but we generally try to move the boundary conditions far enough away from the dissipation region that the precise boundary conditions are unimportant. For example, the electron temperature in Fig. 7 was assumed to equal the ambient temperature at the bottom boundary taken at 170 km. This boundary condition is only approximate; yet it causes little error in the electron temperatures above 200 km because most of the energy is absorbed by the neutral gases before it can be conducted down to the bottom boundary.

The electron temperature as it would be affected by the operation of the solar satellite in the daytime at the Northeast site is shown in Fig. 8. A comparison with Fig. 7 shows that

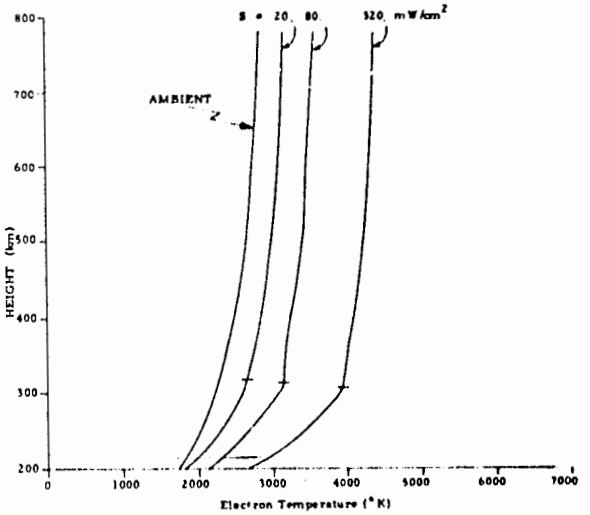


Fig. 8. Steady-state electron temperatures as they are affected at the Northeast site during the daytime. Calculations are for the magnetic field line undergoing maximum heating. The actual microwave irradiation for this line of force occurs near 300 km, over the narrow altitude range indicated by the slash marks.

the ambient electron temperature is considerably higher in the daytime than the nighttime, in part due to the fact that photoelectrons are born at relatively high energies. The additional increase caused by the microwave radiation is considerably smaller in the daytime, because the increased electron densities and higher electron temperatures increase the thermal conduction so that the thermal resistance between the dissipation region and the lower ionosphere is reduced.

These calculations have indicated large changes in electron temperature in the part of the ionosphere irradiated by the MPTS. However, these temperature changes are localized and involve only those magnetic lines of force which interest the beam. The calculations in this paper give the temperature rise along those lines of force which intersect the beam near the peak of the F-region because the temperature rise is greatest along these field lines.

The large increase in electron tempera- predicted for power fluxes of 80 mW/cm^2 greater is likely to cause a significant ge in the rates of several important c reactions. For example, an important ct of microwave heating of the ionosphere occur as a result of the vibrational and tronic excitation of the neutral compo-s. It is known that the excitation of reac-s will increase chemical activity through-the atmosphere. In particular, impact tation of the vibrational modes in mole-r nitrogen by hot thermal electrons can elerate the production of NO^+ ions through important reaction

$$O^+ + N_2 \rightarrow NO^+ + N \quad . \tag{7}$$

ne N_2 vibrational temperature T_{VIB} eeds 1200°K, then a rapid increase occurs he rate of Eq. (7), raising its value by an er of magnitude between 1000°K and 0°K. The increased production of NO^+ s supplants the usual slow creation of O_2^+ ich proceeds through the charge-exchange ocess

$$O^+ + O_2 \rightarrow O_2^+ + O \quad . \tag{8}$$

a result, electrons now recombine with O^+ ions as well, accelerating ionization ss by the reaction

$$NO^+ + e \rightarrow N + O \quad . \tag{9}$$

If the concentration of NO^+ increases due electron heating, a density depletion will cur. The effects may extend far beyond e beam, since the vibrational states of N_2 e metastable with very long lifetimes. At -layer heights, the principal loss of vibra-nally excited nitrogen N_2^* is diffusion. elow 200 km the atomic oxygen concentra-on is large enough to quench N_2^* at a rate at equals or exceeds the diffusion loss pro-ss. High velocity winds will transport the brationally excited species, N_2 and possi-y O_2, far from the region of microwave ating, leaving a wake of depletion ion ncentration. Schunk and Banks [12] estim-e the effective lifetime of N_2^* to be 5000 c. at altitudes above 200 km. If the wind is order 100 m sec^{-1}, then the excited N_2 olecules will travel approximately 4° of titude before being quenched by atomic xygen - a distance of nearly 450 km. This echanism has been put forth as one of the auses of the trough in electron concentration en near the equatorward edge of the auroral val.

F-REGION HEATING AT PLATTEVILLE

A typical power from a 5 MHz trans-itter at Platteville [8] was 2 MW with an ntenna gain of 17 dB. Such a transmitter roduces 88 μW/m^2 at 300 km which dissi-ates as much heat as 2.2 mW/cm^2 at 2450 MHz. Near the reflection height, the Airy "swelling" effect increases the power of the HF beam by more than an order of magnitude over a narrow height interval on the order of 1-2 km, giving power densities equivalent to power densities of 20-100 mW/cm^2 at 2450 MHz. Therefore, it seems reasonable to expect that power densities exceeding 20 mW/cm^2 of 2450 MHz radiation will produce effects comparable to the ionospheric modification experiments although some of the instabilities, such as the parametric decay instability, will not be excited when the heating frequency is far above the plasma frequency. The HF modification experiments caused enhanced airglow at 6300 A, triggered instabilities and produced field-aligned structures which scattered radiation in the UHF and VHF bands. Many effects were observed which were not anticipated theoretically, in large part because plasma instabilities involving nonlinear interactions which are more difficult to predict than the Ohmic heating results.

In the Platteville experiments, the self-focusing instability explained by Perkins and Valeo [13] produced field-aligned irregularities which scattered VHF and UHF signals propagating within the disturbed area with a direction perpendicular to the magnetic field. When a similar analysis was applied to the MPTS in [3] (Appendix B - "Self-Focusing Plasma Instabilities"), no instability modes were found for the crucial wavelengths less than 50 meters which could disrupt the MPTS control system. However, the threshold power decreases as the inverse fourth power of the instability wavelength. If infinitesimal modes exist in the ionosphere, with wave vectors within one degree of perpendicularity to both the beam direction and the magnetic field, the instability criterion indicated that they could be amplified for wavelengths greater than 200 meters with a 20 mW/cm^2 power flux. Self-focusing instabilities for plasma modes with effective wavelengths on the order of a kilometer have also been described by Drummond [14]. However, the instability theories are not fully believable because they assume infinitesimal perturbations, i.e., that electron temperature changes are much smaller than the ambient. This approach is not valid when Ohmic heating increases the electron temperature several times as we have found here. Some difficult analyses should be carried out which come to grips with the enormous temperature rises and the finite radius of the beam which allows the plasma to be partially pushed out of the beam by the expansion forces. Other complications include the electrodynamic drift which, because its intensity is a function of altitude, will tend to prevent the peaks and troughs of the plasma density in a potentially unstable mode from lining-up with the minima and maxima of the beam intensity over a large altitude range, as is necessary for amplifying a self-focusing mode. A thorough study is

BEST AVAILABLE COPY

necessary to assure us that any irregularities are large enough to be compensated for by the retrodirective phase-control system, that the disturbances are localized within the volume of the beam and the electron temperature excursions are comparable to those described in this paper.

EFFECTS OF IONOSPHERIC MODIFICATION ON OTHER USERS

The number of individual communication systems which potentially can be affected by ionospheric changes is too large to be adequately described here. A study was made by choosing specific examples which span the RF domain from 10 kHz to 6 GHz in generic systems such as navigation, communication and DoD radars [4]. For example, Table I lists three of the systems studied and shows the potential power beam ionospheric effects. If an experimental program for evaluating ionospheric effects is undertaken (and the large number of users which are likely to be affected by the ionospheric modification seem to indicate the advisability of an experimental verification program), a critical element will be the design and implementation of experiments which will quantitatively determine the risk that the MPTS ionospheric effects pose to other users.

TABLE 1

Example of Effects of Ionosphere Modification on Other Users

SYSTEM	POWER BEAM EFFECTS
Omega Navigation System (10.2 kHz)	Ionospheric changes produce sudden phase anomaly which decreases range accuracy of Omega.
HF Communications (3 to 30 MHz)	Disruption of HF communication likely which affects many users.
AF SATCOM; VHF Satellite-to-Aircraft Communication System	Generation of ionospheric irregularities could produce scintillations and/or multipath conditions.

The effects in Table I, even if they occur in a region localized with respect to the power beam, can potentially affect users at ranges upward to 2000 km from the interaction region. Another way of viewing this is that a user is potentially affected by the power beam in a region defined by the intersection of his local horizon and the ionosphere as shown in Fig. 9. For one power beam the probability of interaction might be small. However, with a large number of power beams, all located within the F-layer interaction region, the

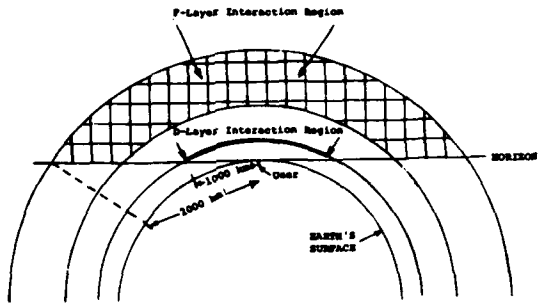


Fig. 9. F-layer and D-layer interaction region.

potential for user interference will increase significantly.

Ionospheric irregularities (in particular spread-F) could be caused by the self-focusing instability generated from the effects of Ohmic heating. During the Platteville heating experiment [8], ionospheric irregularities were observed in the 100 meter wavelength regime. The results of the Platteville heating are not directly related to microwave power beam effects because the Platteville heating is done at a frequency near the plasma frequency of the ionosphere. This increases the effective heating by at least an order of magnitude. However, the results did indicate the occurrence of irregularities at power densities equivalent to 20 mW/cm^2 at 2.45 GHz. Further, the occurrence of these irregularities was unexpected (i.e., not predicted) and it is risky at this stage of our understanding to predict the occurrence or non-occurrence of irregularities as a result of power beam heating at 2.45 GHz. This uncertainty provides a strong argument for undertaking an experimental program to determine if and to what extent ionospheric irregularities are produced.

DISCUSSION

The calculations reported in this paper indicate major changes in electron temperature in the vicinity of the power beam and moderate changes in electron density. While no serious effect upon potential users of the ionosphere has been identified due to this increase in the mean electron thermal energy, it is important to have experimental measurements of the effects produced by the power beam. In this regard, the Platteville and Aerecibo HF modification experiments are useful, but they were conducted at frequencies below or near the reflection frequencies. Therefore, one could come significantly closer to duplicating MPTS conditions by conducting modification experiments at 15-30 MHz, as discussed elsewhere [4, 15-16]. As the MPTS concept is further developed, it

will be important to study possible ionospheric effects both experimentally and theoretically.

ACKNOWLEDGEMENTS

The authors would like to thank W. E. Gordon, O. E. Maynard, F. W. Perkins and C. Wendell for advice and helpful discussions.

REFERENCES

1. Glaser, P.E., "Solar power from satellites," Physics Today 30, 30-38 (1977). Also "The potential of satellite solar power," Proc. of the IEEE 65, 1162-1176 (1977).

2. Brown, W.C., "The technology and application of free-space power transmission of microwave beams," Proc. IEEE 62, 11-25 (1974).

3. Raytheon Company, "Microwave Power Transmission System Studies," Vol. II, NASA CR-134886, ER-75-4368 (Dec. (1975).

4. Raytheon Company, "Space-Based Solar Power Conversion and Delivery System Study," Vol. III, Microwave Power Transmission Systems, NASA MSFC (March, 1977).

5. Holway, L.H., Jr. and G. Meltz, "Heating of the lower ionosphere by powerful radio waves," J. Geophys. Research 78, 8402-8408 (1973).

6. Holway, L.H., Jr. and G. Meltz, Modification of the lower ionosphere by high-power HF transmissions," Raytheon Research Division Technical Memorandum, November 1974 (also URSI Annual Meeting, October 1974).

7. Meltz, G., L.H. Holway, Jr. and N.M. Tomljanovich, "Ionospheric heating by powerful radio waves," Radio Science 9, 1049-64 (1974).

8. Special Issue of Radio Science on Ionospheric Modification, November 1974.

9. Gordon, W.E. and H.C. Carlson, Jr., "Aerecibo heating experiments," Radio Science 9, 1041-1048 (1974).

10. Dalgarno, A., "Collisions in the ionosphere," in Advances in Atomic and Molecular Physics," Vol. 4, D. R. Bates editor, 381-410, Academic Press, N.Y. (1968).

11. Meltz, G. and R. E. Lelevier "Heating of the F-region by deviative absorption of radio waves," J. Geophys. Research 75, 6406-6416 (1970).

12. Schunk, R.W. and P.M. Banks, "Auroral N_2 vibrational excitation and the electron density trough," Geophys. Res. Letters 2, 239 (1975).

13. Perkins, F.W. and E.J. Valeo, "Thermal self-focusing of electromagnetic waves in plasmas," Phys. Rev. Letts. 22, 1234-1237 (1974).

14. Drummond, J.E., "Thermal stability of earth's ionosphere under power transmitting satellites," IEEE Trans. on Plasma Science PS-4, 228 (1976).

15. Perkins, F.W. and R.G. Roble, "Ionospheric heating by radio waves: predictions for Aericebo and the satellite power station," J. Geophys. Res. (to be published).

16. Duncan, L.M. and W.E. Gordon, Final Report, "Ionosphere-microwave beam interaction study," NAS9-15212, DoD No. NA183TA (1977).

SOLAR POWER SATELLITE AND ITS INTERACTIONS WITH PLASMA AND THE IONOSPHERE

Henry Oman
Boeing Aerospace Company
Seattle, Washington

INTRODUCTION

A solar power satellite is the energy-converting portion of a system which injects into Earth-surface public utilities power generated from sunlight in a geosynchronous orbit, 35,693 km altitude. Other elements of the system are the microwave beam that transmits the power to Earth, and the receiving station with its array of antenna elements, rectifiers, and equipment for converting the collected power to high-voltage 60 Hz alternating current that is delivered to distributing utilities. A solar power satellite having solar cells in simple trough-type concentrators is shown in Figure 1, along with the other major system elements and their power losses.

To be practical, a solar power satellite must be large, in the order of 100 km^2. This is because the microwave beam should be sharp enough to focus its energy on a reasonable size receiving station, say 60 km^2 in area. An antenna generating such a beam would have to be about a kilometer in diameter. It can then transmit enough power to support on Earth a receiving station that delivers from 5 to 10 gigawatts (GW) of power. For reference, the larger nuclear power plants in operation in 1977 deliver about one GW.

Constructing a solar power satellite in low-Earth orbit, followed by self-powered electrically-thrusted transfer to its geosynchronous operating station seems to be the best approach. The alternate of assembly in geosynchronous orbit involves hoisting into low-Earth orbit huge quantities of propellants for hauling the satellite components into geosynchronous orbit. However, self-powered orbit transfer requires the generation of high voltage directly from the solar-cell array. About 1800 volts is the limit because at higher voltages current leaking through the plasma in low Earth orbit would steal too much power from the solar array.

In geosynchronous orbit, where there are only about 100 electrons per cm^3, the leakage current through the plasma will be insignificant, even when voltages up to 100 kV are generated in the solar array. However, ion engines used to control attitude and station location will generate charge-exchange plasma which can provide a path for leakage-current flow out of the solar cell array. The amount of power lost in this leakage current will depend on the ion-engine power and the engine-to-array spacing. Insulating the solar cells and their interconnectors may be necessary.

The microwave beam carrying power to the Earth receiving station may interact with the ionosphere. The extent of interaction will depend on the power transmission frequency and the power density in the beam. Power densities of 20 to 30 milliwatts per cm^2 are being considered. Most of the past work has been based on a 2.45 GHz frequency, but our latest analysis suggests that significant savings would result from the use of a higher frequency. For example a 10 GW receiving station might cost as much as $2.8 billion less if 5.8 GHz were substituted for 2.45 GHz.

SPACE ENVIRONMENT

The environment for the solar power satellite is summarized in Figure 2 where altitude is plotted in Earth radii, nautical miles, and kilometers horizontally on a logarithmic scale, but with the center of the Earth brought from minus infinity to the edge of the illustration.

The Earth's magnetic field does not directly affect the high voltage solar array of a solar power satellite, but it controls other phenomena that do affect the array. These phenomena vary by orders of magnitude in intensity as a result of solar activity induced changes in the Earth's magnetic field and particle arrival rates. For example, the ionospheric layers, of

Figure 1. Of the 17.3 GW Generated in Geosynchronous Orbit, 10 GW is Delivered to Distributing Utilities.

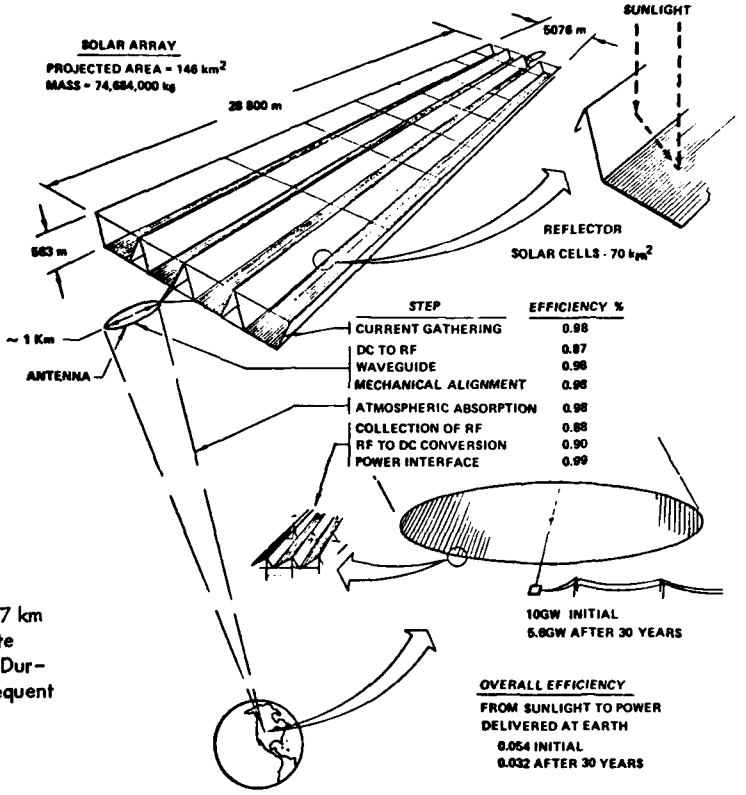
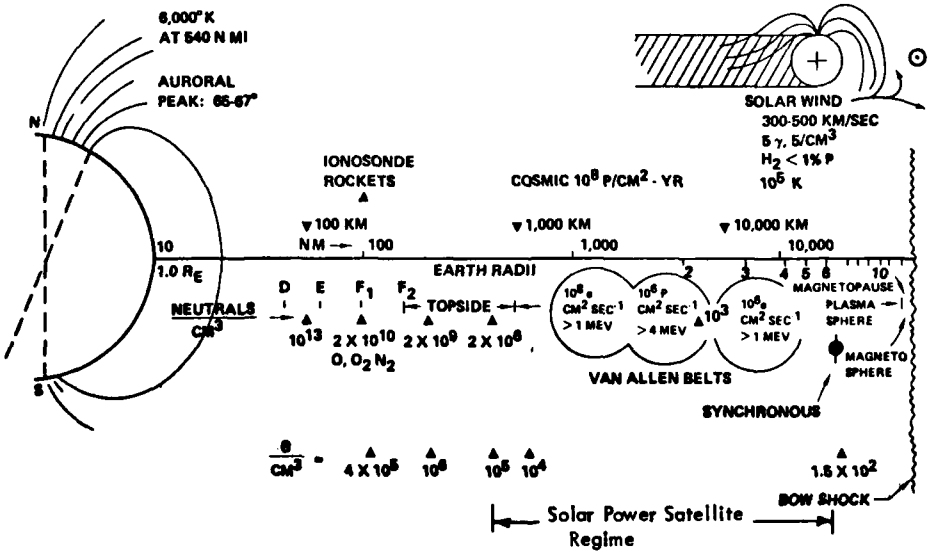


Figure 2. After Assembly at 477 km Altitude, a Solar Power Satellite Sees a Variety of Environments During Orbit Transfer and its Subsequent 30-Year Lifetime.



which the F_1 and F_2 are in the 100 nautical mile (NM) to synchronous-altitude operating regime, are affected not only by the magnetic fields, but also by the time of day and season of the year. During the day the ultraviolet in the sunlight ionizes the oxygen and nitrogen neutral atoms of the air, producing over 10^6 electrons per cm^3 (Figure 3). At night the recombination of electrons with ionized oxygen produces the air-glow. At synchronous altitude the normal electron count falls to 100 per cm^3, and 60 cubic kilometers must be swept to find a coulomb of charge. The thermal flux of electrons and the flux of ions swept out by the spacecraft are plotted as a function of altitude in Figure 4.

Solar cell degradation is most severe in the Van Allen trapped electron and proton belts, where up to 10^6 protons of greater than 4 MeV energy bombard every square centimeter of cell area each second. The energetic protons are particularly damaging because each proton of energy over 4 MeV can penetrate a 6-mil quartz cover and produce thousands of crystal defects before coming to rest. Compared with trapped protons the galactic cosmic rays (10^8 per sq cm per year) produce trivial damage in solar cells.

PINHOLE CURRENT COLLECTION

Current flow through pinholes in insulation is an important mechanism in the escape of power from high-voltage solar arrays through a plasma path. Kennerud measured the current flowing from plasma through pinholes while varying pinhole size, insulation type, area of electrode and surrounding insulation, shape of pinhole, and type of insulation adhesive (Reference 1). Environmental and electrical parameters that were varied are: plasma density, voltage level and polarity, length of plasma exposure, and background pressure.

In one of his tests he used 0.0127 cm Kapton insulation and measured current flow through the pinholes having hole diameters of 0.038-cm, 0.152-cm, and 0.381-cm. The Kapton was bonded to a stainless steel disc with conductive epoxy and mounted in a Teflon holder. This test holder exposed a 3.5-cm diameter disc (9.6 cm^2 area) of the Kapton surface with the hole in the center facing the ion beam.

The measured leakage current with the electrode positive is shown in Figure 5. At low voltages (< 40 volts) the collected plasma current strongly depends upon hole size and is nearly independent of voltage. In this voltage range the collected plasma current density of the largest two holes is of the same order of magnitude as the random electron current density calculated from Langmuir probe data. The relatively low plasma current density collected by the 0.015-inch diameter hole may be an insulation shielding effect; the hole diameter and depth, being nearly equal, make it a relatively "deep" hole whose walls limit the ability of the applied voltage to collect plasma electrons.

At intermediate voltages (+60 volts to +100 volts) the collected plasma currents start rising rapidly with increasing voltage while the effect of defect size diminishes. At high voltages (+1,000 volts to 20,000 volts) the currents for all three hole sizes are roughly equal and do not increase appreciably as voltage is increased.

The observed current saturation at high voltage was a surprise and at first was thought to be a limitation of the facility; e.g., perhaps the plasma source could provide no more current, or perhaps the limited size of the chamber walls restricted the ability of the holes' plasma sheath to collect current. To check these possibilities, the current collecting behavior of a 1.27-cm diameter sphere was measured under nearly the same plasma conditions with the sphere located close to where the pinhole had been. The result was that the sphere current never saturated and behaved according to theory, even though this current exceeded the pinhole saturation currents by a factor of 50. Thus it appears that the observed pinhole saturation is not a basic limitation of the plasma source or of the chamber size. This saturation is a repeatable phenomenon whose level depends primarily upon plasma density and area of insulation surrounding the pinhole.

Another surprise was the high level of the collected current at voltages above about 500 volts. Figure 6 shows that at high voltage the measured current for the smallest pinhole was several orders of magnitude larger than that calculated for a sphere having the same surface area as the pinhole. There is some evidence that this enhancement over what a sphere would collect is due to the presence of the insulation surface surrounding the pinhole. It had been suggested that one reason for this is that the voltage from portions of the electrode covered by Kapton is felt by plasma electrons and thereby enhance the pinhole current collection. To test this possibility the above tests were repeated with a Kapton-covered electrode in the same test holder without any pinhole. The region in front of the specimen was investigated

Figure 4 (right). Thermal Electron Flux Exceeds Thermal Ion Flux at Solar Array

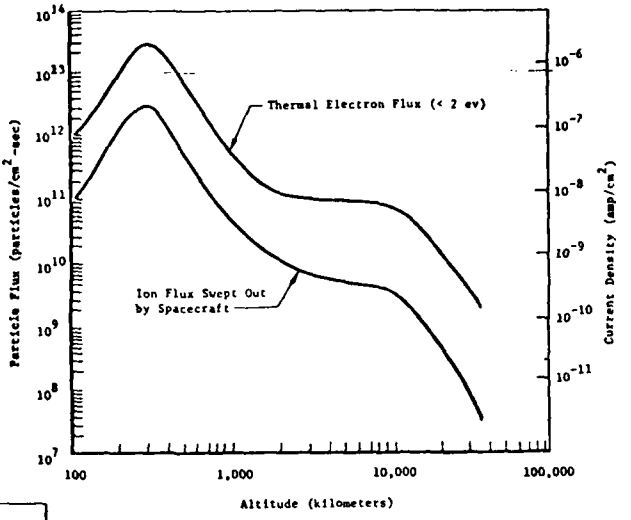
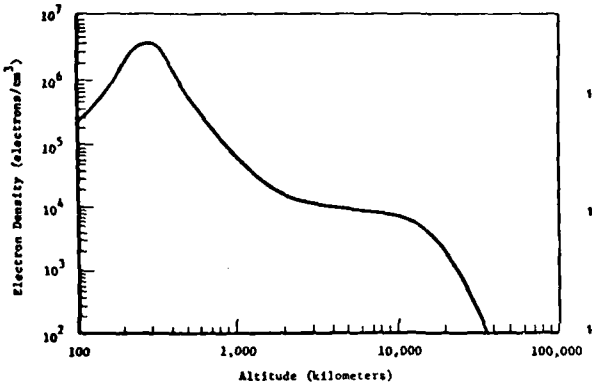


Figure 3 (below). Electron Density Varies by Orders of Magnitude with Altitude.



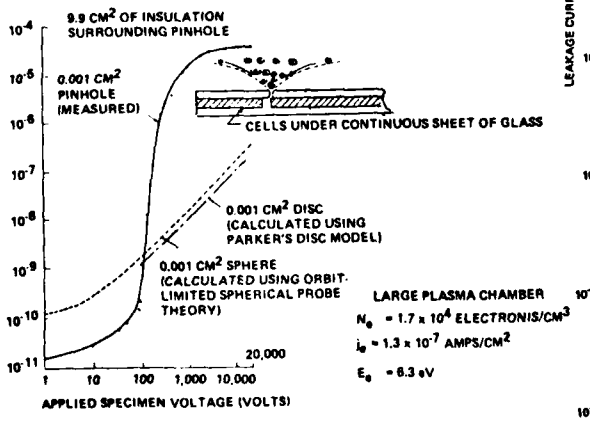


Figure 6. Area for Area, a Pinhole is a Better Current Collector than a Sphere if the Voltage Exceeds 100 Volts.

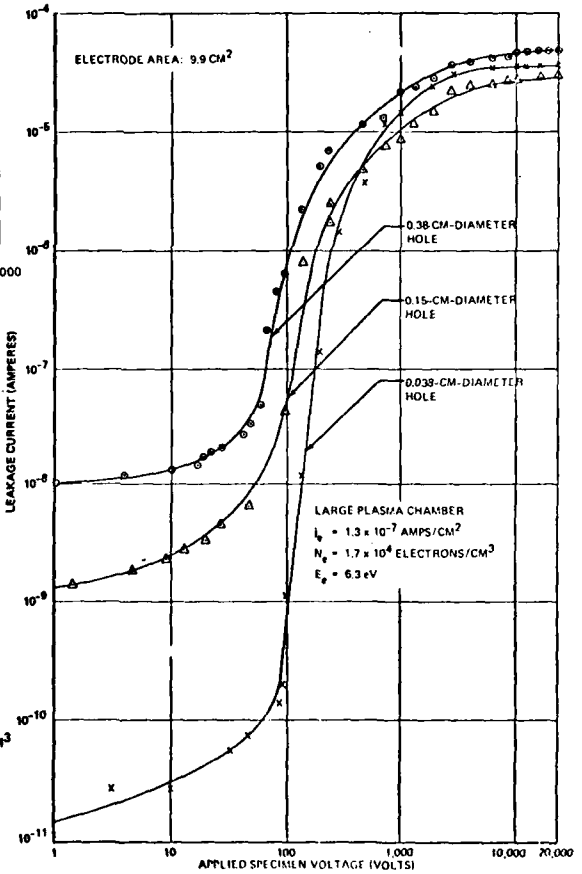


Figure 5. At Above 200 Volts the Hole Size in 125 μm Kapton Ceases to Significantly Affect Leakage Current.

with a hot wire probe. With the probe within 0.635-cm of the Kapton surface, no potential could be sensed. This result makes it seem unlikely that the electrode area under the insulation surrounding a pinhole contributes to the electron-collecting capability of the defect by virtue of electric fields extending through the insulation into the plasma.

POWER LOSS BY LEAKAGE THROUGH PLASMA

The space between 400 km altitude and the orbits of geosynchronous satellites contains neutral atoms, free electrons, positive ions, and high-energy charged particles. The high-energy particles, although damaging to solar cells and optical surfaces, are not numerous enough to carry a significant current. The free electrons, generated each morning when ultraviolet photons ionize neutral atoms, have energies of around one to two electron volts. This energy is dissipated in reactions with neutral atoms and ions increasing the temperature of the medium to the region of 500° to 2000° K. The temperature of an electron is related to its energy by Boltzmann's constant, 8.6171×10^{-5} eV per °K.

An electrically neutral gas containing free electrons and ions in equal numbers is called a plasma. A positively charged spherical electrode, say one cm in diameter, will collect electrons when inserted into a plasma. The volume in which electrons are influenced by the electrode, called a sheath, is much larger than the sphere. Some of the electrons will orbit around the electrode and escape back out of the sheath. Current collection is then said to be orbit-limited and is affected in a complex manner by the radius of the electrode, the voltage of the electrode, and the temperature and density of the free electrons.

The high-voltage solar-cell array for a solar power satellite looks more like a sheet electrode than like a spherical probe. For example, let us assume that 10 km^2 of a solar power satellite array is deployed to supply 1500-volt power for electric propulsion thrusters for raising the satellite from low-Earth orbit, say 500 km, to geosynchronous orbit. K.L. Kennerud has developed a method of analyzing the leakage current from such arrays (Reference 1) based on fundamental equations developed by I. Langmuir (Reference 2). Kennerud's technique converts the planar array into a sphere having the same area, and then he calculated the radius of the electron sheath surrounding the array. His experiments with small positively charged solar-cell panels correlated well with his predictions. With a negatively charged panel which collected ions, his experimental measurements did not correlate well with theoretical predictions, perhaps because the ion sheath extended to the chamber walls.

Using Langmuir's equations, we determined that at 500 km the electron sheath extends to a few meters above the plane of the solar cells, in the range of electron concentrations, electron temperatures, and array voltages of interest. The calculation of leakage current then simplifies into analyzing the rate at which electrons drift into an electron sheath having essentially the same area as the solar array. This electron current (J_r) is simply:

$$J_r = \frac{N_e \sqrt{E_e}}{3.7 \times 10^{11}} \quad \text{(in A/cm}^2\text{)}$$

where N_e = electron density, in electrons per cm^3

E_e = electron energy in eV

The calculated leakage currents from a 1500-volt array for several altitudes are shown in Table 1. The calculations were based on the electron densities in Figure 3 and the electron temperatures in Reference 3.

A flow of electrons from the plasma to the solar power satellite must be matched to an equivalent flow of electrons out of the solar power satellite. Otherwise the satellite will become negatively charged with respect to the plasma, and will cease attracting electrons. This flow of electrons away from the satellite is provided during orbit transfer by electron emitters which are installed for neutralizing the ions emitted by the thrusters. In geosynchronous orbit, where the satellite would be generating power, the electric thrusters would not always be in operation. Furthermore, in geosynchronous orbit the electron density is only about 100 per cm^3, so the power lost through plasma leakage, even at 44 kV, would be trivial.

A negatively charged solar array would attract ions rather than electrons. However, ions are less mobile than electrons, and the ion current would be much smaller than the electron current observed with a positively charged solar array. Thus, the positively charged array is the worst case.

EFFECT OF VOLTAGE, ELECTRON TEMPERATURE, AND ELECTRON DENSITY

It is interesting to note that in large solar arrays the voltage of the array does not significantly affect the leakage current. The array voltage affects only the thickness of the electron sheath which is small compared with its other dimensions. For example, with a 1500-volt, 10 km^2 array the sheath is only 2.6 m thick at 500 km altitude. Increasing the array voltage to 44 KV would increase the sheath thickness by only a few meters.

Sheath thickness is affected by electron density and temperature. For example, in geosynchronous orbit the electron density is only about 100 electrons per cm^3. A 44 KV solar array operating in geosynchronous orbit would have a sheath radius 15 percent larger than the radius of a sphere having the same area as the array. Contributing to this large r_o are the array voltage, the low electron density and hot electron temperature. However, the plasma leakage current in a solar array of a power satellite in geosynchronous orbit is limited, not by the electron supply, but by the inability of the heavy ions to move to the satellite to neutralize the charge deposited by collected electrons. If the electric propulsion thrusters are not operating, the leakage electron current would no longer be neutralized by the emitters used with the ion thrusters, and hence would be trivial.

SOLAR ARRAY POWER LOSS THROUGH CHARGE-EXCHANGE PLASMA

Charge-exchange plasma is generated in and downstream of an electric thruster when the beam ions interact with neutral atoms escaping from the thruster (Figure 7). The charge exchange plasma was first discovered to be a conducting medium in the path of electrons going from the thruster to the solar array during tests of the ATS-6 spacecraft in a large vacuum chamber (Reference 4). The spacecraft was biased at +15 volts relative to the thruster neutralizer, and substantial electron currents were observed flowing to the spacecraft.

The ions in the charge exchange plasma will be propelled by electric fields away from the ion beam and toward the back of the thruster (Figure 8). The positive ions in this plasma, not being readily absorbed by the solar array, constitute a minor part of the leakage current. The electrons, on the other hand, can funnel through holes in any solar array insulating surface, and can rob significant power from the solar array.

CHARGE EXCHANGE PLASMA GENERATION RATE

Kaufman and Isaacson in Reference 5 give the charge-exchange ion generation rate (N_{ce}) as

$$N_{ce} = \frac{7.62 \times 10^{33} J_b^2 (1 - \eta_u) \sigma_{ce} \sqrt{A}}{r_o \eta_u}$$

where the symbols represent the following quantities, with the values shown being those which might be applicable to our solar power satellite:

Symbol	Meaning	Possible Value, SPS
J_b	Beam current, amperes	64 kA per corner
η_u	Propellant utilization factor	0.892
σ_{ce}	Charge-exchange cross section, m^2	2.5×10^{-19} m^2 for argon
A	Atomic weight of propellant	39.94 for argon
r_o	Beam radius, meters	16.97

The values for the solar power satellite were derived as follows. Each argon thruster has an 80-ampere beam current, so the 800 thrusters at each corner would produce 64 kA of beam current. The argon propellant has an atomic weight of 39.94 atomic mass units and a charge-exchange cross section of 2.5×10^{-19} m^2 at 10^6 meters per second. The effective beam radius was assumed to be that of a circle having the same area as the total of 800 thrusters, each 120 cm in diameter.

Kaufman's equation gives a charge exchange generation rate of 3.52×10^{23} ions per second, the equivalent of 56.4 kA. However, if we assume that each of the neutral argon atoms is ionized only once, then the limiting rate of generation of ions for the charge-exchange plasma should correspond to the rate of release of neutral atoms by the thruster. The limiting charge-exchange current (I_{ce}) would then be

$$I_{ce} = I_b \left(\frac{1 - \eta_u}{\eta_u}\right) = 64 \text{ kA } \left(\frac{1 - 0.892}{0.892}\right) = 7.75 \text{ kA}$$

The difference between the 56.4 kA predicted with Kaufman's equation and the 7.75 kA of neutral atoms coming out of the thrusters might be explained as follows: Within the current-flow path extending from the thruster area to the solar array, the heavy argon ions might be considered as

TABLE 1. LEAKAGE CURRENT FROM POSITIVELY CHARGED SOLAR ARRAY

Array Altitude, Km	Electron Density, N_e Electrons/cm^3	Electron Temperature ^{0}K	Leakage Current nA/cm^2	Leakage Current Amperes per 1500 V String*	Power Loss, Percent of Generated
500	6×10^5	3,000	824.5	0.8494	7.72
700	2×10^5	3,000	274.8	0.2831	2.57
1,000	7×10^4	3,000	96.19	0.0990	0.90
2,000	2×10^4	3,200	28.38	0.0292	0.265
5,000	1×10^4	4,400	16.64	0.0171	0.156
10,000	8×10^3	5,400	14.75	0.0152	0.138
20,000	2×10^3	9,000	4.76	0.0049	0.044
30,000	1×10^2	13,600	0.29	0.0003	0

* The string is 0.404 m by 255 m, with an area of 133.02 m^2

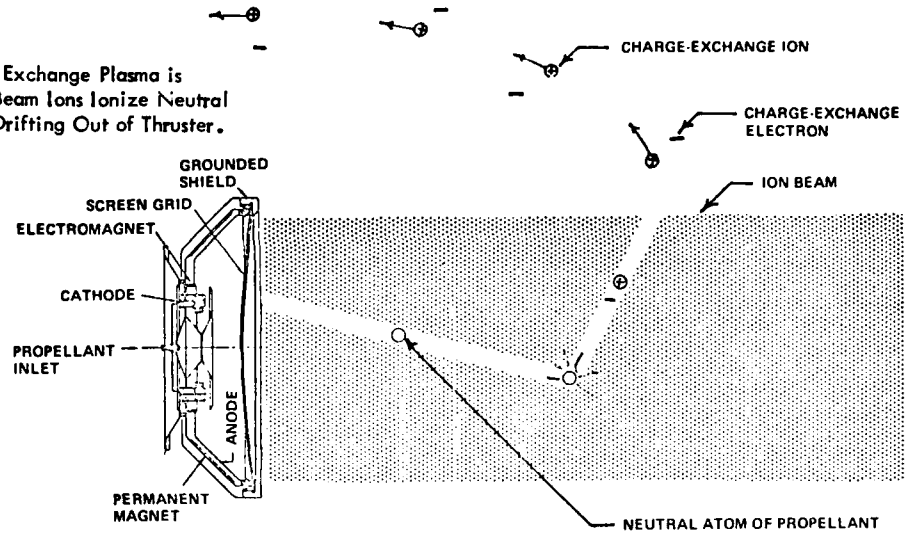


Figure 7. Charge Exchange Plasma is Generated When Beam Ions Ionize Neutral Propellant Atoms Drifting Out of Thruster.

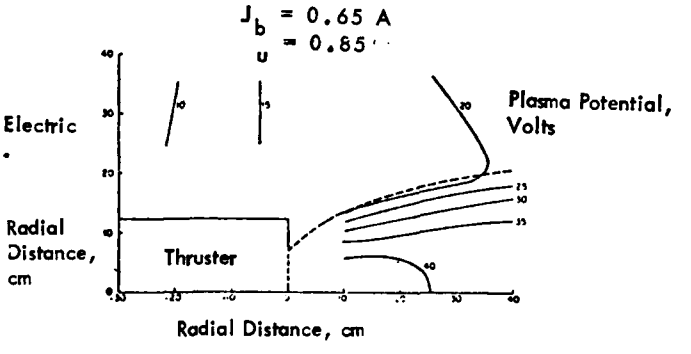


Figure 8. Ion Flow in Beam Generates Electric Field That Propels Charge Exchange Ions.

standing still relative to the much lighter electrons. There is always a one-to-one ratio of ions to electrons in this current-flow path, thus preserving the neutral character of plasma. However, the electrons are flowing at high velocity from the neutralizers to the solar array, and each electron doesn't stay long in the plasma. This is analagous to the action in a thermionic power-generating diode, where relatively few cesium ions neutralize the space charge around the cathode to promote a copious flow of electrons to the anode. In a sense, the argon ions form a conduit for expeditious transfer of electrons from the neutralizers at the thrusters to the solar array.

LIMITING POWER LOSS THROUGH CHARGE-EXCHANGE PLASMA

Methods of avoiding the solar array loss in output caused by charge-exchange plasma conduction are shown in Figure 9. Using the cone shield does not reduce losses, according to Reference 5. The cone does move the apparent source of plasma generation downstream and further from the solar array. However, it also increases the density of the neutral atoms, and the net result is an increase in the leakage current from the solar array.

Insulating the solar array would eliminate the leakage-current loss were it not for the pinholes. Kennerud showed that within the plasma sheath surrounding the solar array electrons will funnel into a pinhole from a large volume of the plasma (Reference 1). If the electron current is great enough, the pinholes will enlarge as the surrounding insulation sublimates away.

A third method of controlling leakage current going through the charge-exchange plasma is to collect the electrons with an anode before they can get into the solar array. Kaufman dismisses this approach with the note that is consumes too much power. However, most of the electrons can be collected on a 20-volt anode. Using Kaufman's generation rate of charge-exchange plasma gives a current of 56.4 kA, which at 20 volts represents 1.13 MW or 2.2 percent of the 52 MW consumed by the 800 thrusters.

If the charge-exchange plasma generation is limited by the supply of neutral argon atoms released by the thrusters, then the collecting plate will carry only 7.75 kA, which at 20 volts represents 155 kW or 0.3 percent of the power supplied to the thrusters. The anode could be a thin sheet of metal connected to the thruster structure through a 20-volt power supply.

If nothing else works, then the ion thrusters could be spaced away from the solar array. We have not yet calculated how far this spacing must be.

CONCLUSIONS

Generation of power at high voltage, around 40 kV, is advantageous on a solar power satellite. For example, power for the rf amplifiers would be carried by buses from solar cells as far as 10 km away, and even at 40 kV the current at the rotary joint between the array and antenna will be around 200 kA. Bus weight is reduced as voltage is raised. Also, generating power at 40 kV, a good input voltage for high-power klystron rf power amplifiers, avoids the need for heavy power conversion equipment.

Leakage of current through plasma can constitute a significant power loss from a high-voltage solar array. For example, at 300 km altitude a 2 kV array can barely generate enough power to feed the plasma losses. We believe that 477 km is a good altitude for assembling the solar power satellites, and that supplying power at 1.8 kV to the thrusters for orbit transfer is appropriate. Even at this voltage the I^2R losses in the power buses will be significant, but we need only about one-fourth of the satellite solar array for powering the thrusters.

In geosynchronous orbit where the satellite generates power and delivers it to Earth with a microwave beam, and the electron density is only around 100 per cm^3, plasma leakage current will be trivial if ion thrusters are not operating.

Leakage current through charge-exchange plasma, which is generated whenever ion engines are operated for orbit transfer or station keeping, is a phenomenon that is real but not yet fully understood. Leakage current through charge-exchange plasma might be reduced by shields, spacing, or ion collection at low voltage, but much more work needs to be done before we can be sure.

REFERENCES

1. Kennerud, K.L., "High Voltage Solar Array Experiments," NASA Lewis document CR-121280, March 1974.

2. Langmuir, I, and Blodgett, K.B., "Currents Limited by Space Charge Between Concentric Spheres," Physical Review 23, p. 49 (1924).

3. Doherty, W.R., and Wilkinson, M.C., "High Voltage Solar Array Space Environment Model," Boeing document D2-121333-1 (1970).

4. Worlock, R., Trump, G., Sellen, J.M. Jr., and Kemp, R.F., AIAA Paper No. 73-1101 (1973).

5. Kaufman, H.R., and Isaacson, G.C., "The Interactions of Solar Arrays with Electric Thrusters," AIAA Paper No. 76-105.

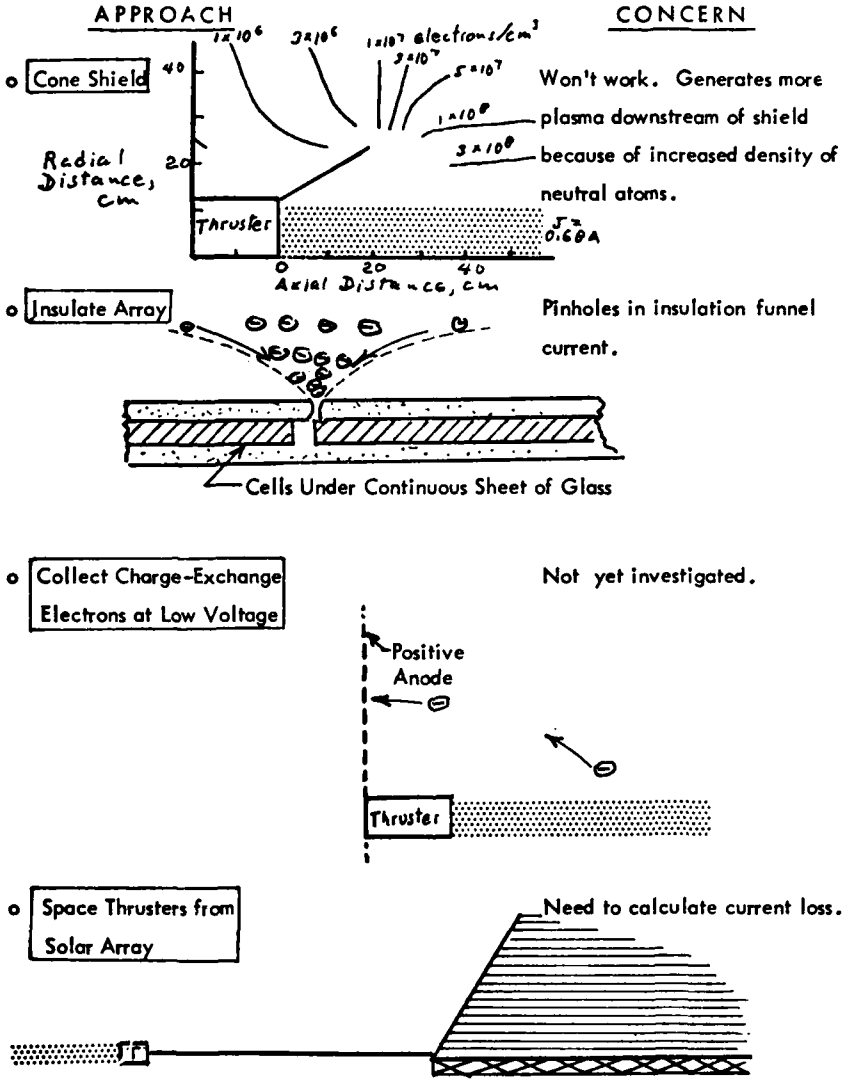


FIGURE 9. Ways of Reducting Power Loss Through Charge-Exchange Plasma.

RADIO PROPAGATION AND SOLAR CYCLE 21

Paul E. Argo and Jay R. Hill
EM Propagation Division, Naval Ocean Systems Center
San Diego CA 92152

INTRODUCTION

In October 1957, the smoothed sunspot number (SSN) reached 209, the highest value in recorded history. Several months earlier, the FCC had to warn all 2-50 MHz users that they should expect signal disruption, after it was learned that U.S. Highway Patrol vehicles were responding to European Police orders and South American taxi cabs were unable to decipher instructions from Florida dispatchers (George Jacobs, CQ Editor, private communication).

Although this was 1957 and solar cycle 19, there are indications that similar effects will occur in 1981 during cycle 21. Since 1961 most predictions have indicated abnormally low solar activity for the remainder of this century. Initial predictions based on unsophisticated techniques to fit to the sunspot record (Figure 1) were shown incorrect when cycle 20 reached 126, more than a factor of two above the predicted upper limit of 50. More sophisticated techniques still predicted that the sunspot cycle will not rise above 50. However, using the most up to date spectral analysis method available (Paul, 1972), Hill (1977) determined that the next cycle might climb above even the 1957 high. The cycle began close to the December 1976 prediction of Hill (rather than the September 1977 prediction of the others) and is rising even faster than the Hill prediction (Figure 2).

Because the ionosphere undergoes long term variations that follow the solar cycle and these variations cause corresponding changes in HF/VHF propagation, it is important that the communication system community recognize that the emerging solar cycle may behave differently then previously expected. Thus, the impact these differences can have on their systems must be evaluated. If cycle 21 reaches Hill's prediction then subsequent ionospheric effects (i.e. increased sporadic E, etc.) could cause compromising of military communication system by the opening of unexpected propagation paths (Crepeau, 1974). This paper, therefore, covers the ionospheric effects of both solar cycles 19 (SSN 206) and 20 (SSN 126) and how systems can be affected.

SOLAR ACTIVITY VARIATIONS

Solar activity, as measured by sunspot number, has been occasionally observed telescopically since 1610 and carefully observed since 1749. It appears to have an eleven year cycle which has been recently identified as a 22 year magnetic period with every other eleven year cycle having a reversed average solar magnetic field. Since the amplitude and frequency of the cycle varies slowly with time, it has become a point of interest to determine if sunspot activity variation is regular and constant over extended periods of time. If in fact it is regular, then it should be possible to make predictions based upon historical observations.

The sunspot record from 1749 to 1975 has been analyzed (Hill, 1977) using a high resolution spectral analysis to determine not only the frequencies of the periodicities, but also the amplitude and phase of each line. Since a number of closely spaced lines are resolved, it is possible to project backwards in time and form a comparison with the extended record as determined by Eddy. The analysis shows that discrete lines do exist in the sunspot record and a beat between several closely spaced lines produces a long (1000 year) period fluctuation in solar activity. The analysis is performed with 2724 monthly sunspot values given by Vitinskii (1956) and by *Solar Geophysical Data.* The algebraic sign of alternate cycles is negative to conform to an assumed 22 year magnetic reversal cycle.

Several methods for analyzing spectra have been applied to the sunspot record by the authors already mentioned. Cole used fast Fourier transform (FFT) to analyze the autocorrelation function and the solar cycle phase data of Schove and obtained the power spectral density as a smooth continuous function of frequency. Also both Cohen et.al. and Currie used the maximum entropy method (MEM) to find the smooth continuous power density function. Both the FFT and MEM methods estimate the frequency of a suspected spectral line at the point power. The method developed by A. Paul (1972) is a very accurate and sensitive method for determining frequencies, amplitudes and phases of unknown lines in a spectrum consisting of both broadband noise and discrete lines. Hill (1977) has used the Paul method to determine a Fourier Series fit to sunspot data. (See also the paper by Hill included in these proceedings.)

The summed Fourier Series corresponding to the sunspot model is plotted along with monthly data record in Figure 1. The series is evaluated beyond the data record at both ends to provide a forecast of future and an aft cast of past solar cycles. Note the striking agreement between the model and observations between 1720 and 1975. The fact that the summed series fails to fit the data as accurately in the first and last cycles as in the central cycles may be an artifact of the mathematical method or it may indicate changes in the physical processes of the sun, e.g., changes in the oscillation mode excitations.

The fit, Figures 1, 2 to the sunspot record between 1750 and 1975 is good enough to forecast a few years into the future. Using the zero crossing as an indicator, the model predicted this solar cycle would begin December 1976. [actual date was August 1976]. Several authors (Cole, 1973, Jose, 1965, Cohen and Lintz, 1974) have predicted a low sunspot activity level during the next cycle (number 21) with a peak near 50 to 60. Hill's analysis predicted a large peak between 130 and 200 during 1981 or 1982.

Although the present solar cycle has just begun, the SSN values are showing a cycle rise rate that is even faster than that predicted by Hill (Figure 2).

SOLAR CYCLE EFFECTS ON HF/VHF PROPAGATION

The ionosphere under long term variations that follow the solar sunspot activity (Figures 3, 4). The ionospheric variations cause corresponding changes in HF/VHF propagation. In this section, the cycle to cycle variations in ionospheric response to solar variation are examined and some of the pitfalls that the HF/VHF user can avoid are presented. It will be shown that commonly used ionospheric indices may not adequately describe the ionospheric variations unless used with care and that the effects are very dependent upon user and frequency. The 50-70 MHz band, which is usually groundwave or sporadic E supported, opens up to F-region mid latitude propagation during extremely high solar activity (SSN 190). The low VHF band (30-50 MHz) will be open to F-region mid-latitude propagation during medium to high solar activity during non-summer months. The top of the HF band (25-30 MHz) will also be open almost year round during the medium to high solar activity period. Incidentally, the 27 MHz citizens band channels will undergo much interference from skywave signals in a 4000 km radius during even a low solar cycle maxima (SSN 50).

A common method for quantizing the ionosphere is with the critical frequency of the F2 peak (foF2). Figure 4 shows the monthly mean noon foF2 values for Slough, England covering three solar cycles (1932-1962). The short period variations in Figure 4 are the seasonal variations indicated in Figure 3. Using the Slough data, one can see that the maximum foF2 value does vary from solar cycle peak to solar cycle peak (e.g., 13-15 MHz). When the monthly mean values of winter foF2 are plotted as function of the twelve month running average of sunspot number, a definite relationship can be seen (Figure 5). The usual interpretation (Davies, 1965) is that a "flattening" of the curve exists for SSN greater than 130 indicating saturation effects. If, however, the different cycles are separated (as suggested by J. W. Wright) (Figure 6), one sees immediately that each cycle has a linear (and different) foF2 vs. SSN relationship.

Notice that a SSN of 50, in a low cycle, does not give the same ionospheric response as a SSN of 50 in a very high cycle. In other words, the ionospheric response to SSN variations is predictable within each cycle and yet cannot be directly compared from cycle to cycle. This is important to the propagation forecaster (and system designer) because he should be including in his calculations the "best" predictions of solar cycle variations.

One of the problems involved in comparing solar activity to ionospheric variations is that SSN is probably not a good measure of the solar effects on the ionosphere. The reason is that the chief ionizing agent for the F-region of the ionosphere is the solar extreme ultraviolet (EUV) radiation. EUV has not been measured accurately and continuously for long periods of time, mainly because the measuring instruments must be satellite mounted. Solar physicists believe that the EUV flux varies in much the same manner as the sunspot number during a solar cycle, but the exact relationships have not been determined. In fact, the ionospheric measurements may lead to a more exact understanding of the solar process.

The oblique propagation of HF signals is affected by factors other than just the foF2 value. However, a useful measure of the propagation conditions is the highest frequency that can propagate over a given path, which is usually called the Maximum Observable Frequency (MOF). The MOF depends mainly upon the foF2, path length and the F layer reflection height. The relationship between MOF and foF2 is linear, while that with the F layer reflection height (being due to propagation geometry considerations) is less straight forward. In general, a lower reflection height yields a higher MOF while an increased reflection height has a decreased MOF. Also, a longer path length will have a higher MOF. For a given path length, the ration of the MOF to the foF2 is called the M-factor (i.e. MOF = foF2 x M-factor). Because it is a function of the F-layer height the M-factor for a path of 3000 km (called an M3000) is often used to characterize the layer height.

For a plane (flat) ionosphere, the M-factor increases as the path length increases and so at an infinitely long path length, the MOF would be infinite. In the curved earth case, there is a geometric limit on the M-factor for any specified distance.

During solar cycle variations, the F-layer height (h') varies, being lowest low sunspot cycle and rising with sunspot number (see Figure 3). From the previous paragraph, we see the effect on the MOF of this height variation is in the opposite direction as that of increasing foF2. Since h' increases with SSN, we see that M3000 decreases, canceling some of the increase in MOF due to the increase in foF2. In fact, White Sands, N. M. data (foF2, M3000) indicates that the differences between different solar cycle peak for MOF (3000) are small (5-25%) compared to the large minimum to maximum solar cycle SSN variations (80%). Figure 7 shows the variations in the maximum MOF supportable during the different solar conditions. The maximum MOF can be derived from the foF2 and (M3000) by following the M-factor curve from the 3000 km distance to the 0° elevation (launch) angle. This maximum MOF is the geometrically limited MOF. The MUF (3000) relates to the user working medium distances (across the U. S., for example). The MOF (max) is most relevant to transoceanic propagation with path lengths greater than 4500 km. In fact, during November 1958 while the peak MOF (3000) is ~43 MHz. The peak MOF (max) is ~47 MHz.

These MOF values are monthly medians, indicating that for half the days of the month, a frequency greater than 47 MHz was propagating. In fact, for 25% of the days in the month of November 1958, frequencies greater than 50 MHz were propagating over the long distance paths. This indicates that mid latitude 50 MHz propagation is not "standard" even during the peak of a high solar cycle.

Figure 7 shows the seasonal variations in the MOF for two solar cycle peaks and minima between. The seasonal variations indicated in Figures 3, 4 are evident and the solar cycle effects on HF propagation are clear. 50 MHz is available only on the best days in the winter during extremely high solar cycle (SSN 190) while 30 MHz is available for most of the year (excluding summer) for solar conditions other than the sunspot minima. During the summer even 20 MHz became unusable.

These propagation assessments are strictly for mid-latitude propagation paths. For paths crossing the equator, the results must be modified. Specifically, the ionospheric tilts in the equatorial region cause the signal to travel much further (Figure 8) with an M factor that is increased by a factor of 1.5. This means during conditions when a 45 MHz signal propagates at mid-latitudes, 68 MHz will be propagating transequatorially. In addition, the ionospheric responses to the solar cycle variations are distinctly latitude dependent (Figure 9). The high latitude variations (Kiruna, Sweden and Akita, Japan) are similar to the mid-latitude variations. The equatorial variations (Rarotonga and Kinshasa) show no such simple pattern and propagation prediction through the equator would prove extremely difficultly.

Measurements of scintillation activity (250 MHz) at Guam were taken (by NELC) over four years of declining sunspot numbers (Paulson and Hopkins, 1977). Figure 10 shows their finding of a direct dependence between SSN and scintillation activity. If the prediction of maximum SSN of 200 is realized, then a direct extrapolation indicates that six hours of intense scintillation every night might be expected. Note, however, that there may be cycle to cycle variations in this dependence, as we found in the foF2 vs. SSN relationship.

THE SOLAR CYCLE 21 AND THE USER

The MOF variations, as discussed, have different impacts on different type of users. The Navy, who usually transmits to distances less than 3000 km, will be basically unaffected by solar cycle peak values. However, the covert systems using frequencies of 25–30 MHz (expecting this to be above the maximum MOF) may find that for large portions of the day, they are compromising their covertness. Marine and Army users, in the lower VHF band (30–100 MHz), will have a large number of long skip propagation occurrances during high SSN.

All the above may never be aware of the problem, however, because their bands are minimally filled and interference won't be a factor; rather, they will be compromising security.

CBers, because of the large number of users in the bands (27 MHz), will have severe interference problems during the winter of any medium to high solar cycle. This interference will be of heterodyning nature - not individual signal receptions. Even a low solar cycle is likely to cause CB to go skywave often during the winter months.

CONCLUSION

We have reviewed the effects of solar activity on the ionosphere and HF/VHF propagation. The peaks of solar cycles allow much different propagation characteristics than the low of a solar cycle and the peak to peak variations may be significant under certain conditions. For example, in 1957, mid-latitude propagation at 50 MHz was fairly common while in 1968, the occurrances were rare. Previous estimates of the upcoming solar cycle have been that it will be much lower than the last cycle (1968). If this is true, 50 MHz will not propagate skywave and even 30 MHz may remain ground wave dominated for large parts of the year. Our analysis, on the other hand, indicates that the next cycle of solar activity may soar *above* the 1957 high, in which case, the 50+ MHz frequencies will be skywave for a record period of time. A more cautious estimate still places the next level of activity well above the 1968 peak and so frequent skywave propagation periods at 50+ MHz should be expected.

REFERENCES

Cohen, T. J. and P. R. Lintz, Nature 250, 398-400 (1974)
Cole, T. W., Solar Phys. 30, 103-110 (1973)
Crepeau, P. J., NRL Report 2873 (1974)
Currie, R. G., Astro Phys. Space Sci. 20, 509-518 (1973)
Davies K., Ionospheric Radio Propagation, NBS Monograph 80 (1965)
Eddy, J. A., Science 192, 1189-1202 (1976)
Hill, J. R., Nature 266, 151-153 (1977)
Jose, P. D., Astro J. 70, 193-200 (1965)
Paul, A. K., Math. Comput. 26, 437-447 (1972)
Paulson, M. R. and R. U. F. Hopkins, NOSC/TR 113 (1977)
Wright, J. W., Nature 194, 461 (1962)

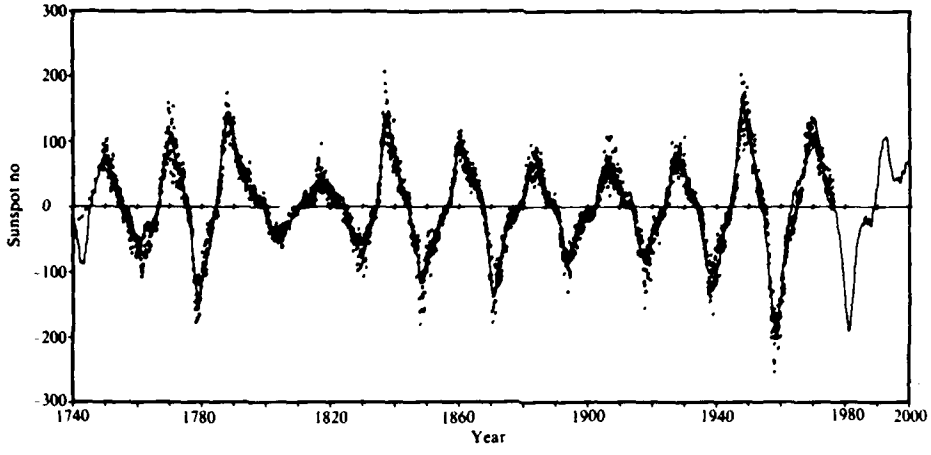


Figure 1. Observed monthly sunspot numbers (·) and Fourier series model (smooth curve)(from Hill, 1977)

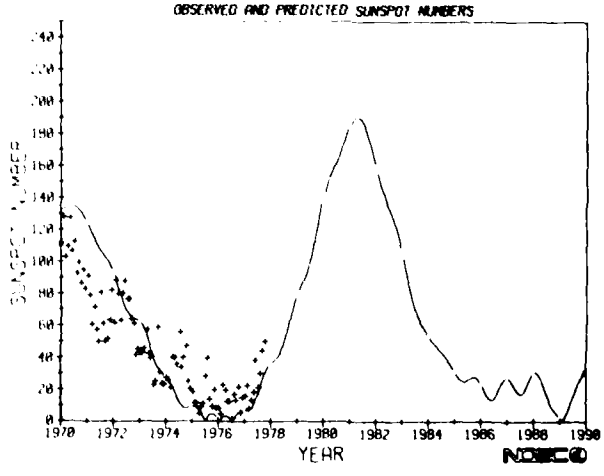


Figure 2 Absolute value of observed monthly sunspot numbers and Fourier Series Model (smooth curve) on an expanded scale

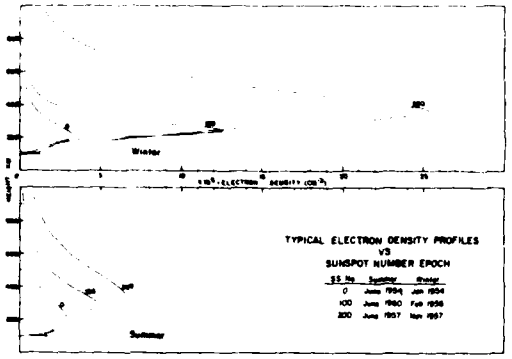


Figure 3 Solar cycle and seasonal variations in electron density profiles (after J. W. Wright, 1962)

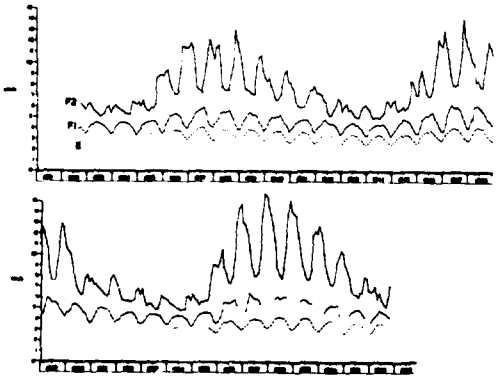


Figure 4 Monthly mean noon critical frequencies (foF2) at Slough, England for solar cycles 17-19. (from Davies, 1965)

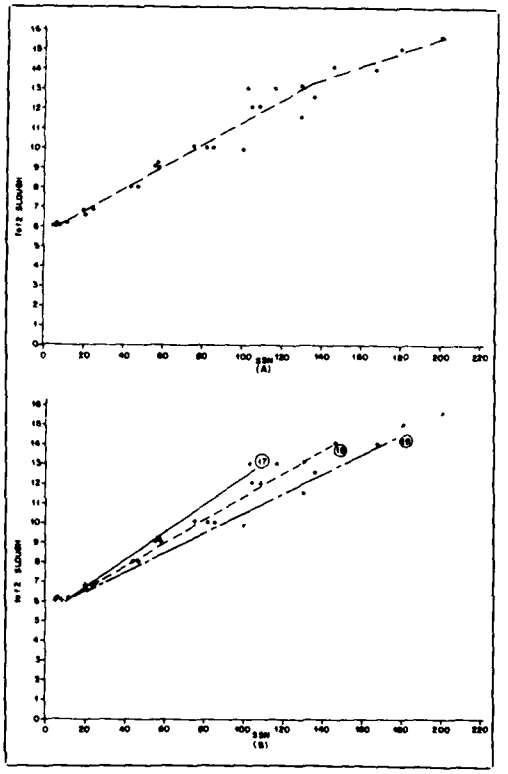


Figure 5, 6 Plotting foF2 versus SSN from Figure 4. Figure 5 is entire data, and shows flattening of the curve at SSN 130. If the cycles are plotted seperately (Figure 6), each cycle has substantially straight lines

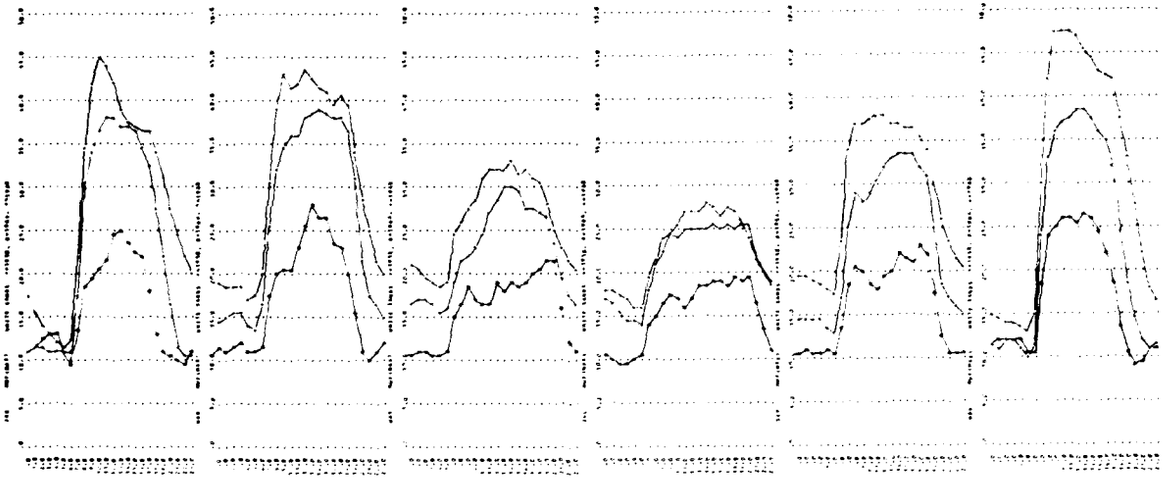
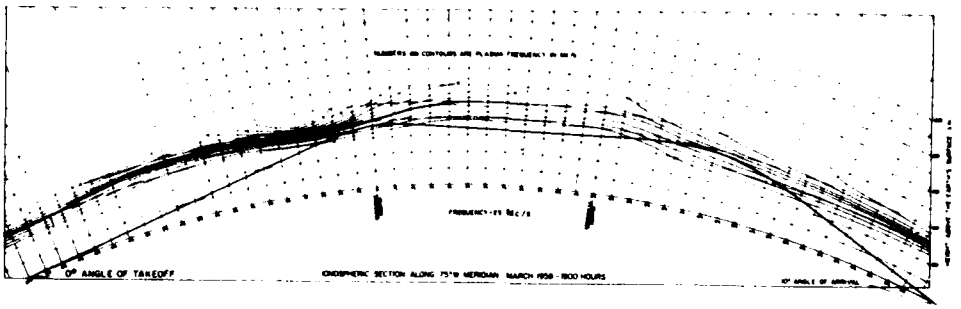


Figure 7 Solar cycle and seasonal variations in midlatitude HF propagation (MOF). 1958 was the peak of solar cycle 19 (SSN 206), 1968 was the peak of solar cycle 20 (SSN 126), while 1964 was the solar minimum inbetween.



Transequatorial ray path.

Figure 8 Notice the effect of the ionospheric tilts, which cause the pathlength to be significantly longer. (from Davies, 1965)

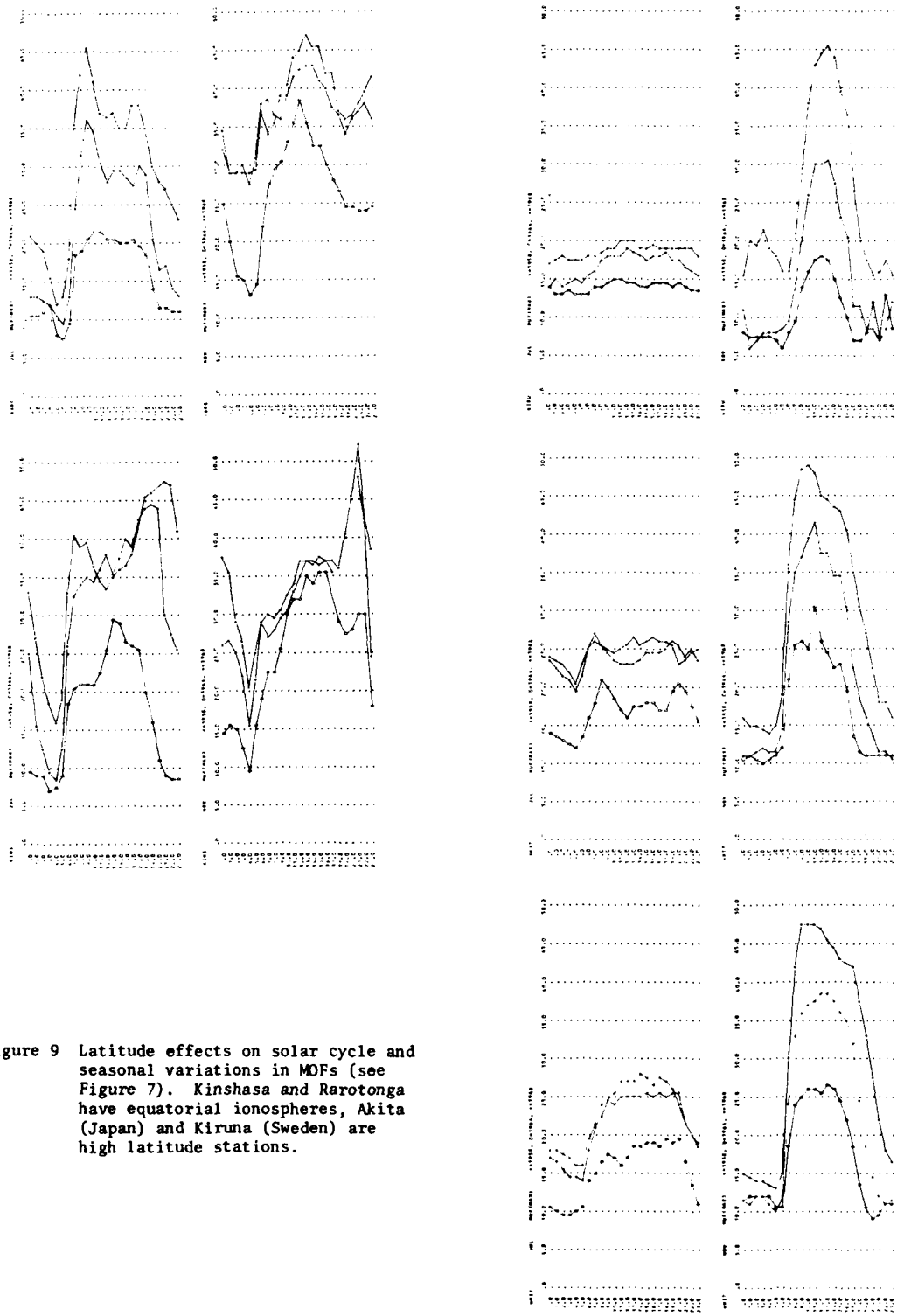


Figure 9 Latitude effects on solar cycle and seasonal variations in MOFs (see Figure 7). *Kinshasa* and *Rarotonga* have equatorial ionospheres, Akita (Japan) and Kiruna (Sweden) are high latitude stations.

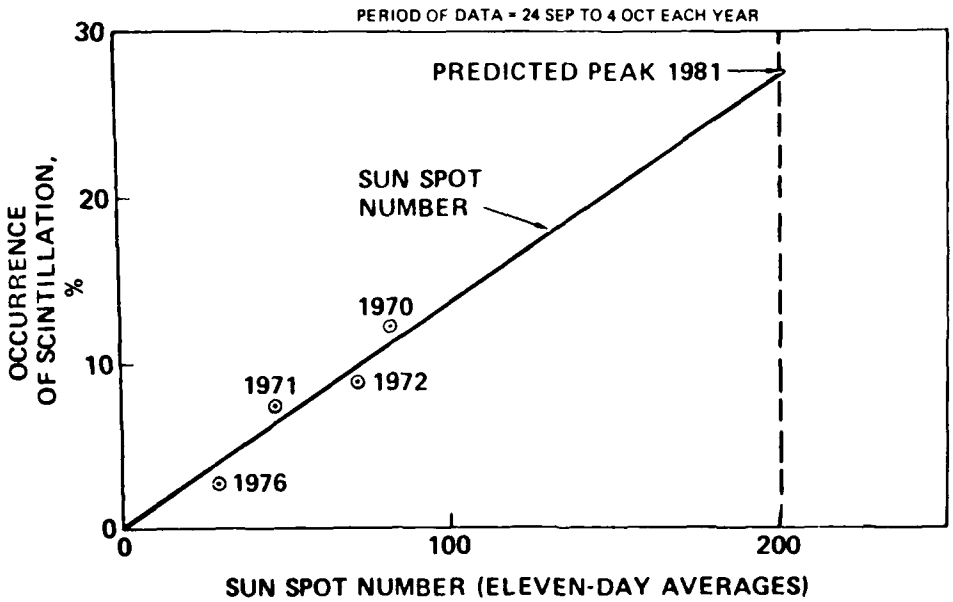


Figure 10 Dependence of scintillation occurrence on solar activity.

REAL TIME PROPAGATION ASSESSMENT: INITIAL TEST RESULTS

Ilan J. Rothmuller
Naval Ocean Systems Center
San Diego, CA 92152

INTRODUCTION

The operational effectiveness of a wide variety of Navy electromagnetic systems depends on the state of the propagation channel or environment. Here, the ionosphere is considered the propagation environment. Real-time propagation assessment is a technique which can be used to optimize system performance and minimize detrimental effects of an unknown or disturbed environment. Optimization may be achieved, for example, by choice of an optimum frequency to communicate in a specific geometric configuration. Real-time propagation assessment also provides warning of environmental disturbances which can interrupt communications, render surveillance systems useless, and cause gross inaccuracies in navigation systems.

A real-time environmental prediction and assessment system (EPAS) is an integration of three elements. These are: (1) a variety of solar/geophysical data sources; (2) a facility which collects, processes and selectively disseminates these data to; (3) a propagation assessment terminal where real-time, regionalized and tailored "products" are issued.

This paper describes an assessment terminal, the model development, and the results of testing the terminal in an operational environment.

PROPHET: A REAL TIME ASSESSMENT TERMINAL

The Naval Ocean Systems Center (NOSC) has developed a real-time propagation assessment terminal called PROPHET (a pseudo-acronym for "propagation forecasting terminal"). The objective of the NOSC effort was to develop a terminal to be used by operational personnel. PROPHET products were designed to be easily understood and applied. To achieve this objective, models were developed which translate solar/geophysical effects into specific system effects. The PROPHET terminal consists of a stand alone minicomputer with an interactive graphic display and hard copy unit. An example of a PROPHET product is shown in Figure 1 illustrating the concept of regionalized tailored assessments. The graphic display shows the area of concern for the communications controller. For the Naval Communication Station (NAVCOMMSTA) at Stockton, CA (near San Francisco, CA) this is the eastern Pacific Ocean. The map shows the current position of ships and (on the left) the maximum useable frequency (MUF), lowest useable frequency (LUF), and frequency of optimum transmission (FOT) for

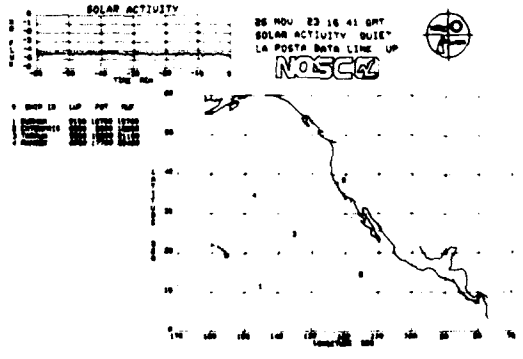


Figure 1. PROPHET display showing HF propagation characteristics between the Naval Communication Station (asterisk at 121°W, 38°N) and the ships Durham, Enterprise, Tarawa and Ranger.

each ship. The communications controller uses this information to guide his frequency selection. The current and past hour satellite measured solar x-ray flux, a measure of solar activity, is also displayed. When the x-ray flux rises to levels such that propagation is affected, PROPHET warns the circuit control officer and produces updated frequency advice. Such a warning not only alerts communication personnel that it is an environmental disturbance causing an outage rather than equipment failure, but also allows some flexibility to alleviate problems.

A very important PROPHET capability is the ability to produce ray traces which depict the path of the electromagnetic wave fronts. Figure 2 is an example of a 3000 km path. A number of important propagation features are evident from

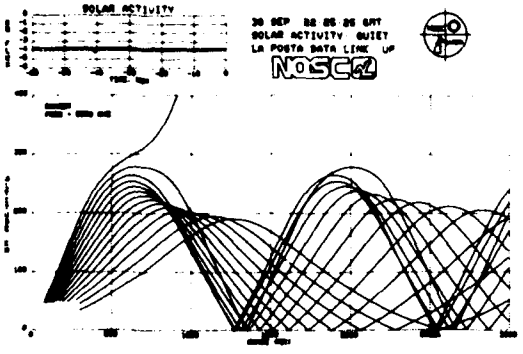


Figure 2. PROPHET raytrace display.

this presentation. No communication is possible for ranges out to approximately 1200 km. This so-called skip zone may be of advantage if an unwanted receiver is located within this range. Focusing, i.e. several rays concentrating in a small area, occurs between 1230-1330 km and, because of different wave travel times along the rays, signal degradation through interference may be expected. A similar interference must be expected at larger ranges (2500-2700 km) where rays reflected once from the ionosphere and waves reflected twice (after being reflected from the ocean's surface) are superimposed. Another application of the ray trace picture is the selection of specific antennas having launch angles favoring desired rays and suppressing undesired rays for a particular optimum coverage situation.

PROPHET MODEL DEVELOPMENT

An important aspect of the models developed for use in PROPHET is that they are tailored to a specific system application and that the "output" is in easily used terms (e.g. LUF). The NOSC model development approach is illustrated by the following chart:

Cause

Solar Effluents

Environment

Ionospheric Reaction

System Effects

HF: Comm. Outage - SWF
VLF: Nav. Error - PCA, SPA
VHF: Comm. Outage - Scintill.

For the systems considered, the sun is assumed to be the primal environmental driver.

Three approaches can be taken to develop models assessing the propagation environment: phenomenological, statistical, and semi empirical. The phenomenological is based on an understanding of the physical processes relating cause and effect. The statistical approach is based on synoptic studies and assumes that predictions can be made based on historic data. The semi-empirical method (presently used most often by the NOSC modeling group) combines knowledge of the basic physical causes with observational data to generate simplified models which obviate modeling the complicated processes between the basic driving cause and final effect.

An example of such a model is the effect of a solar x-ray flare on HF systems. The physics of the process is that x-rays from a solar flare create excess ionization causing increased absorption to signals in the lower portion of the HF spectrum. If severe enough no communication is possible and a short wave fade (SWF) results. To show this effect the flare case will be contrasted with the normal case.

Figure 3 illustrates the diurnal propagation characteristics for the HF band for a path from Guam to Hawaii under undisturbed ionospheric conditions. Frequency is plotted versus a 24 hour time period. Vertical deflections for the traces in the frequency-time plane indicate that communications over this path are possible; no deflections signify the contrary. *The lowest observed frequency* (LOF) follows very closely the secant of the midpath solar zenith angle. During the daylight hours, increased ionization in the D-region (50-90 km altitude) results in increased absorption for the HF band. Frequencies above the maximum observed frequency (MOF) penetrate the ionosphere and are not reflected back to the desired receiver. Figure 4 shows propagation conditions for the same path during a period

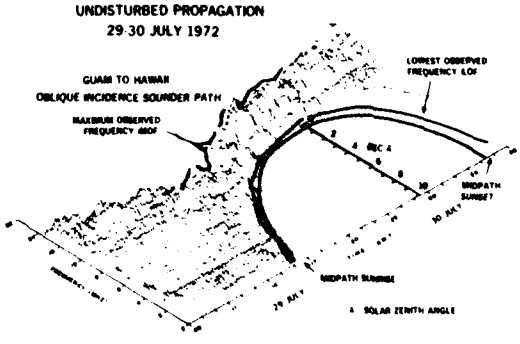


Figure 3. Diurnal HF propagation characteristics for the path from Guam to Hawaii.

when the sun emitted two bursts of x-ray radiation. The 1-8Å solar x-ray flux (ergs/cm^2/sec) peaks at 1840 and 2100 GMT making the entire HF spectrum unusable causing a SWF.

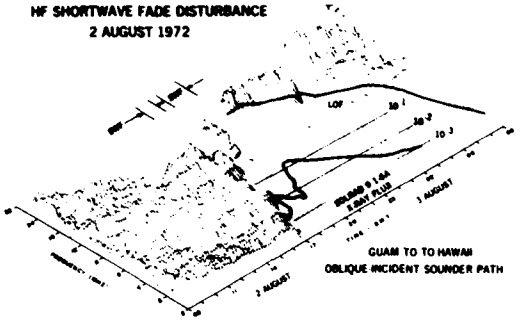
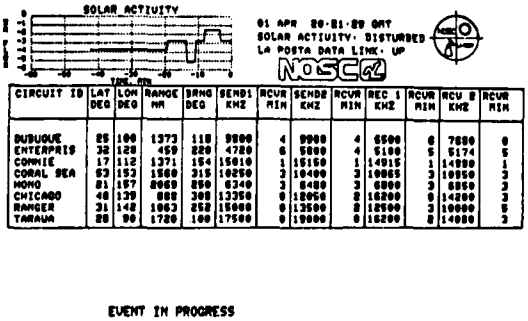


Figure 4. Propagation characteristics for the HF band during a solar flare.

Using the observed relationship between x-ray flux and LOF a model was developed which translates, in real-time, satellite obtained x-ray data directly into LOF, thus informing a communications officer of the available HF spectrum. To assure that propagation assessments are useful, this kind of tailoring to systems is mandatory.

This semi-empirical approach has yielded several other models relating space environmental data to systems effects. In addition to calculating the LOF as a function of x-ray flux, a model has been developed that predicts the duration of a flare (thus SWF) once the peak has been reached. This model uses the flare risetime and peak amplitude to estimate the decay rate. The model adjusts itself in case of multiple flare events and updates the decay estimate by tracking the actual decay rate.



CIRCUIT ID	LAT DEG	LON DEG	RANGE NM	BRNG DEG	E_{MB1} KHZ	RCVR MIN	E_{MB2} KHZ	RCVR MIN	REC 1 KHZ	RCVR MIN	RCV 2 KHZ	RCVR MIN
DUBUQUE	85	100	1373	118	9900	4	9900	4	6500	0	7850	0
ENTERPRIS	32	128	459	220	4720	0	5800	4	5190	5	5174	5
CONNIE	17	112	1371	154	15010	1	15150	1	14915	1	14980	1
CORAL SEA	53	153	1560	315	10250	3	10400	3	10065	3	10950	3
MONO	81	157	2069	250	6340	3	6480	3	6800	3	6850	3
CHICAGO	48	139	888	308	13350	0	13050	0	16200	0	14200	3
RANGER	31	148	1063	252	15000	0	13500	0	12900	3	10000	5
YAMAHA	20	90	1720	100	17500	0	19000	0	16200	0	14000	3

Figure 5. PROPHET display of circuits affected by a solar flare.

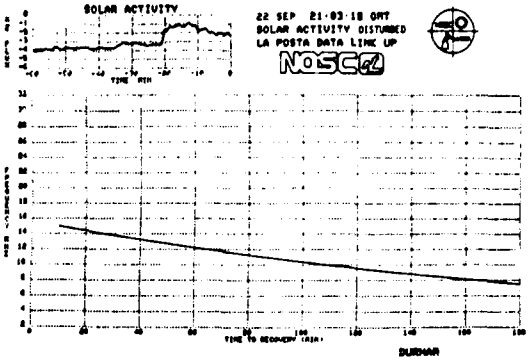


Figure 6. PROPHET display predicting time to recovery from a solar flare for the ship Durham.

Figures 5 and 6 show how this flare or fade duration forecast model is realized in PROPHET. Subsequent to a solar flare PROPHET issues a warning and displays a list of all active circuits with predicted recovery times (under columns headed "RCVR MIN") (Figure 5). This prediction can be examined on a circuit by circuit basis as shown in Figure 6. In this example, the MUF is 15 MHz where the predicted fade is only 10 minutes. This type of display shows not only how to shorten the effect of a SWF but also allows the recognition that the SWF was caused by a solar flare rather than an equipment malfunction.

The effects of other solar/geophysical disturbances have also been modeled. Errors in the Omega VLF navigation system can be caused by solar proton events (SPE) and also by x-ray flares. To account for the navigation errors caused by an SPE (these also cause polar cap absorption or PCA at HF), Argo (1975) has developed a model that converts satellite measured proton flux (E>10 MeV) into correction factors used with normal Omega correction tables. Absorption models for transpolar HF and polar region VHF/UHF satellite communications during PCAs have also been developed (Argo and Hill, 1976).

To perform predictions of HF propagation characteristics on a daily basis a model developed by the Air Force is used. This is an hourly update to a statistically based model - the updating done by refitting to data obtained from an HF ionospheric sounder network. To make longer range predictions, NOSC has developed a model also based on a statistical model.

Another statistical model adapted for Navy use, is the scintillation depth of fading model (LaBahn, 1974). This is an adaptation of the model developed by Fremouw and Rino (1973) and modified by Pope (1974).

Present work is aimed at modeling the effects of geomagnetic storms on a variety of systems and a phenomenological basis for a scintillation model.

TEST RESULTS

NOSC conducted a field test of the PROPHET assessment terminal at the Naval Communication Station, Stockton, CA. The general test objectives were to test the utility (and reliability) of real-time, tailored and regionalized propagation assessments in an operational environment and to specify the needs and uses of various components of an environmental prediction and assessment system (EPAS). Figure 7 illustrates the conceptual information flow in an EPAS as described earlier. Figure 8 shows the information flow for the NOSC field test. A variety of solar/geophysical data flow into La Posta Astrogeophysical Observatory (LPAO) where the data were sorted and combined with local data. Selected data were sent on to the PROPHET terminal. Data sources included SOLRAD 11B, SMS-1, GOES-2, NOAA-5, as well as other data from SELDADS (Williams, 1976) and the Air Force Global Weather Central (AFGWC).

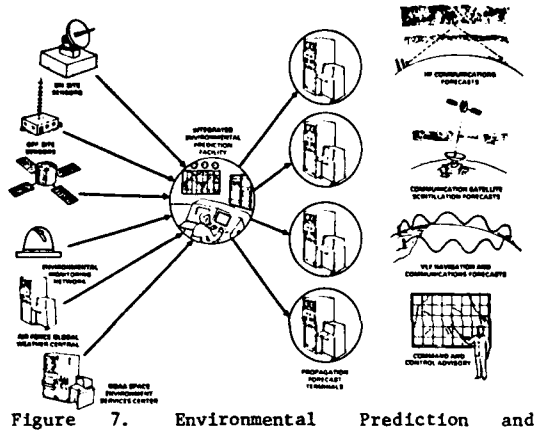


Figure 7. Environmental Prediction and Assessment System concept.

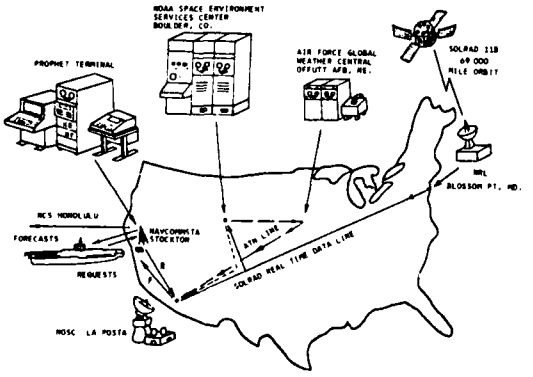


Figure 8. NOSC PROPHET field test information flow.

The field test was evaluated by analyzing the log tapes (which contain useage information as well as commentary by COMMSTA personnel) and questionnaires filled out by the users. Thus both qualitative and quantitative results were obtained.

Analysis to date has shown that: (1) COMMSTA personnel made frequency use of PROPHET - terminal access averaged once per nine minutes (during each 24 hour period); (2) using PROPHET predictions, QSYs (number of frequency shifts) and outages due to propagation were reduced by 15% with duration of outages reduced by 15%-20%; (3) COMMSTA personnel gave a qualitative general "usefulness" rating to PROPHET of eight on a scale of ten.

The ray tracing capability provided some unexpected returns. It not only was useful to give personnel an added (and previously unused) flexibility to perform antenna selection, but also was an educational tool providing insight to the operational personnel into the process of HF propagation. For example, according to records, a 6 MHz transmission from Stockton was not heard by a ship despite the fact that this frequency was within the MUF/LUF envelope. A PROPHET ray trace was called and showed the ship to be in the skip zone for the antenna in use. The ray trace was redone using an antenna with a different take off angle so that rays would reach the ship. The transmitting antenna was shifted and the frequency/antenna combination provided six hours of solid communications.

CONCLUSIONS

The initial field test in an operational environment of a propagation assessment terminal has shown that such information can be useful and will be used if it is tailored to the users needs. Feedback from Naval Communications Station users indicated that PROPHET has properly filled a need which had existed for a long time.

Further, this test has shown that a terminal such as PROPHET provides a useful vehicle for developers of models which account for the effects of the propagation environment on systems. Present capabilities are limited only by a lack of methods to reliably forecast ionospheric disturbances. Future work is aimed at removing those limitations.

REFERENCES

1. Argo, P. E., Modeling Omega PCA Phase Advances, Naval Electronics Laboratory Center Technical Report 1950, 1975.

2. Argo, P. E. and J. R. Hill, High Frequency Polar Cap Absorption Model: SOLRAD Application, Naval Electronics Laboratory Center Technical Note 3249, 1976.

3. Fremouw, E. J. and C. L. Rino, "Modeling of Transionospheric Radio Propagation," Radio Sci., 8, 213, 1973.

4. LaBahn, R. W., Development of a Scintillation Grid, Naval Electronics Laboratory Center Technical Note 2814, 1974.

5. Pope, J. H., "High Latitude Ionospheric Irregularity Model," Radio Sci., 9, 675, 1974.

6. Williams, D. J., SELDADS: An Operational Real-Time Solar-Terrestrial Environment Monitoring System, NOAA Technical Report ERL-357-SEL37, 1976.

IONOSPHERIC FORECASTING AT

AIR FORCE GLOBAL WEATHER CENTRAL

Charles S. von Flotow, Captain, U.S. Air Force
Headquarters Air Force Global Weather Central
Offutt AFB, NE. 68113

ABSTRACT

AFGWC uses worldwide data sources coupled with computer processing to generate data displays used for ionospheric forecasting. Several operational products fall into this area of forecasting: Radio propagation forecasts, radar propagation forecasts, and ionospheric specification. High frequency radio propagation forecasting is a two part process: the generation of monthly mean forecasts based on ITS-78 and preparation of real time updates to these monthly forecasts every six hours. Radar propagation forecasts are tailored for specific systems and are based on foF2 forecasts which are used to generate electron density profiles using the Damon-Hartranft ionospheric model. A four dimensional specification model is in limited use and is capable of using varied data inputs. Future areas of research for direct operational use would include dynamic ionospheric models with prediction capabilities, ionospheric storm behavior algorithms, the reaction of the ionosphere to geomagnetic disturbances, and improved ray tracing techniques.

INTRODUCTION

The Air Weather Service Space Environmental Support System (SESS) has the mission of providing environmental support to all Department of Defense agencies that operate in or use the space environment. Ionospheric forecasting and specification is a major effort of the Space Environmental Support Branch of Air Force Global Weather Central (AFGWC), Offutt AFB NE. One of the more important functions, and most time critical, is event alerts to systems that rely on, or are affected by, the ionosphere. This function is not forecasting, per se, but relies on observing some significant solar event and then forecasting the effects of that event. This is an important area but will not be discussed. This report will deal with another area of ionospheric forecasting, one of routine day-to-day interest.

Ionospheric forecasting has been developed to support and satisfy routine operational requirements. These requirements are generated by systems that depend on the ionosphere, or are affected by its irregularities. This report will describe the products and techniques used to generate these forecasts, as well as future needs. The first section outlines the data sources. High frequency (HF) radio forecasting methods are covered in section two. Section three will discuss ionospheric forecasting and the models used. Section four will provide an update on the development of the four dimensional ionosphere at AFGWC, and finally, in section five, a brief outline of problem areas and future requirements of the operational ionospheric forecaster will be presented.

I. DATA SOURCES

Ionospheric forecasting involves specifying the existing ionosphere and then forecasting the future state. Inputs from 40 to 50 worldwide ionosondes are used to specify the existing state (Table 1). Additional input is provided by 11 real-time polarimeters (Table 2). The ordering of these observations allows the analyst to detect changes in the mean field that will affect climatological forecasts. Specific uses will be described later, but some samples are shown in Figs. 1 & 2. One data source used for subjective input is the High Latitude Monitoring Stations (HLMS) which is a computerized network of riometers, magnetometers, and auroral radar in Alaska and Greenland. Other data sources include trapped radiation particle flux measurements by geostationary satellites, personal contact with communicators, and the solar activity observations. This real-time observing network of the Space Environmental Support System (SESS) was described by Snyder (1974) and remains essentially the same today.

II. HF FORECASTING

The first step in supporting HF frequency managers is to generate the monthly climatological forecast which can be modified by

TABLE 1. IONOSPHERIC SOUNDERS

OBSERVATORY NAME	LOCATION NORTH LATITUDE	LONGITUDE
MANILA	14	121 E
DJIBOUTI	12	43 E
OUGADOUGOU	12	1 W
TAIPEI	25	121 E
DAKAR	15	17 W
MAUI	21	156 W
KADENA	27	127 E
KOKOBUNJI	36	193 E
ASHKHABAD	38	58 E
VANDENBERG	35	121 W
ALMA ATA	43	77 E
AKITA	40	140 E
PATRICK	29	80 W
POITIERS	46	00 W
LANNION	48	03 E
KHABAROVSK	48	136 E
WAKANAI	45	142 E
BOULDER	40	105 W
PETROPAVLVSK	53	158 E
WALLOPS	37	75 W
KIEV	51	31 E
IRKUTSK	52	104 E
MOSCOW	55	37 E
DOURBES	50	5 E
LINDAU	51	10 E
SLOUGH	51	1 W
MAGADAN	59	151 E
OTTAWA	45	76 W
ST JOHNS	47	52 W
SVERDLOVSK	56	61 E
KENORA	50	97 W
TOMSK	56	85 E
GOOSE BAY	53	60 W
YAKUTSK	62	130 E
COLLEGE	65	147 W
CP SCHMIDT	69	179 W
KIRUNA	68	20 E
MURMANSK	69	33 E
TUNGUSKA	61	90 E
TROMSO	70	19 E
CHURCHILL	58	94 W
NARSSARSUAK	61	45 W
SALEKHARD	66	66 E
DIXON	73	80 E
KRENKEL	80	58 E
GODHAVEN	69	54 W
RESOLUTE BAY	75	95 W
QANAQ	76	68 W

TABLE 2. POLARIMETER SITES

LOCATION	LATITUDE	LONGITUDE	BEACON
ATHENS	N34	EØ15	NONE
GOOSE BAY	N48	WØ62	ATS5
OSAN	N35	E128	ETS2
PATRICK	N27	WØ8Ø	ATS5
RAMEY AFB	N17	WØ7Ø	SMS1
SAG HILL	N39	WØ71	ATS5
BOULDER	N37	W1Ø5	SMS1

TABLE 2. POLARIMETER SITES (Continued)

LOCATION	LATITUDE	LONGITUDE	BEACON
LA POSTA	N3Ø	W115	SMS1
TAIWAN	N23	E122	ETS2
PALEHUA	N2Ø	W157	ATS6
SHEMYA	N49	W179	ATS1

real-time updates. All support requests are sent to the Air Force Communications Service Frequency Management Office at Scott AFB where requests are tailored to specific user requirements. The current model used at AFGWC is the updated ITS-78 described by Barghausen (1967). This model (HFMUF) has been further modified by AFGWC for free format input and to accept additional data cards. These additional data cards are used to generate output products that are transmitted directly to the customer via AUTODIN (Automated Digital Network). The output allows the user to immediately determine maximum usable frequency, optimum frequency, and other parameters of interest to the HF communicator. Fig. 3 is a sample of a typical product and is tailored for a specific path and transmitter. The most significant improvement to HFMUF is the semi-automated method of processing these requirements. Cards are prepared by the Frequency Management Office of Air Force Communications Service (AFCS/DOYF) and sent via AUTODIN from AFCS at Scott directly to the AFGWC computer. The card images are stored on the computer where the forecaster edits and corrects any input errors by using demand CRT terminals. The modified ITS-78 program is run and the forecast generated and transmitted. AFGWC can receive, edit, and process these requests in less than one hour. However, since these request are generally for months or years in the future, they are routinely processed in a day or so.

Once the product is in the hands of the frequency manager, it is only useful for the mean ionsphere. Average errors on the order of 2 to 5 Mhz will result and this error will often exceed the usable frequency window during disturbed conditions (Rush, 1973). The frequency manager needs a real-time update to his monthly mean forecast and this is provided by AFGWC every 6 hours on a routine basis (Fig 4). This second step of HF forecasting is a complex problem depending on worldwide communication contacts and varied ordering of ionospheric sounder data for a solution.

The communicator contacts provide a valuable qualitative assessment on the state of the ionosphere that can be used for post analysis and forecasting. For example, observed two hour sunrise transition difficulties over long east-west equatorial paths in the eastern hemisphere would be used for forecasting similar problems in the western hemisphere later in the day. At other times, especially during disturbed conditions, polar communica-

	12L	13L	14L	15L	16L	17L	18L	19L	20L	21L	22L	23L
MANILA	7	6	3[illegible]	28	21	9	12	7	7	21	-7	-9
TAIPEI	-16	-16	-1[illegible]	1[illegible]	-2	-8	-29	-2[illegible]		-28	-27	-22
MAUI	-21	-21	-11	1	11	17	7	-6	-31	-32	-19	-1[illegible]
ASHKHA	8	19	9	-4	-1	7	-8	1[illegible]	17	25	86	111
VANDEN	-17	-9	9	2	3	16	21	5	-14	-19	-11	-2
PATRIC	-11	-6	1	-3	-9	[illegible]	5	21	-17	13	15	32
KHABAR								28	5[illegible]	2[illegible]	29	41
BOULDE	11	5	8	4	5	14	22	18	32	3	7	3
KIEV	16	-7	-7	-7	4	8	14	12	24	39	23	32
DOURBE	8	-4	8	7	20	13	16	3[illegible]	29	21	31	22
MAGADA			-6	-[illegible]	5	2	-3	14	22	29	19	17
SVERDL	3	-5	-1	3	9	13	6	3		16	-2	26
TOMSK	-3	0	-7	5	-2	1	8	1[illegible]	19	12	-9	3
GOOSE	-12	-15	-12	-9	1	3	15	25	2[illegible]	4	-2[illegible]	-9
YAKUTS	0	10	6	1	-1	-2	16	24	41	41	40	3[illegible]
COLLEG	3	-1	1	1	9	16	7	2[illegible]	44	57	85	53
MURMAN	8	16	2[illegible]	24	18	27	26	4	-4	-17	1[illegible]	15
NARSSA	-7	-5	-8				-18		7			
SALEKH	9	7	6	8			17	35	15	4	-22	-26
DIXON			15	27	3[illegible]							
KREUKE	37	4	17	19	49	57		72	17	17	44	64
RESOLU												-1[illegible]
GUAM[illegible]				11							3[illegible]	

FIG. 1

Percentage difference between observed Maximum Usable Frequency (MUF) and ITS-78 MUF for different ionospheric stations. These MUF deviations are for a 4000 Km path and are calculated from the M-factor and foF2 observations. Only last half of local day is shown.

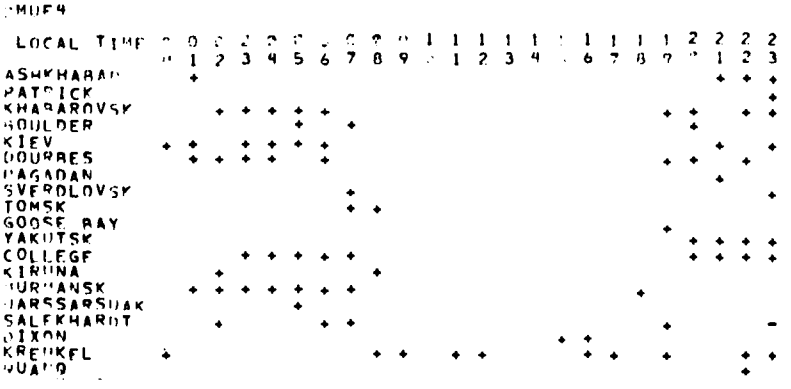


FIG. 2

Graphical representation of the same data used in figure 1. Here all MUF deviations greater than 25 percent are indicated by a "+" or "-". This display serves to filter out non-essential information and identify areas of significant deviations.

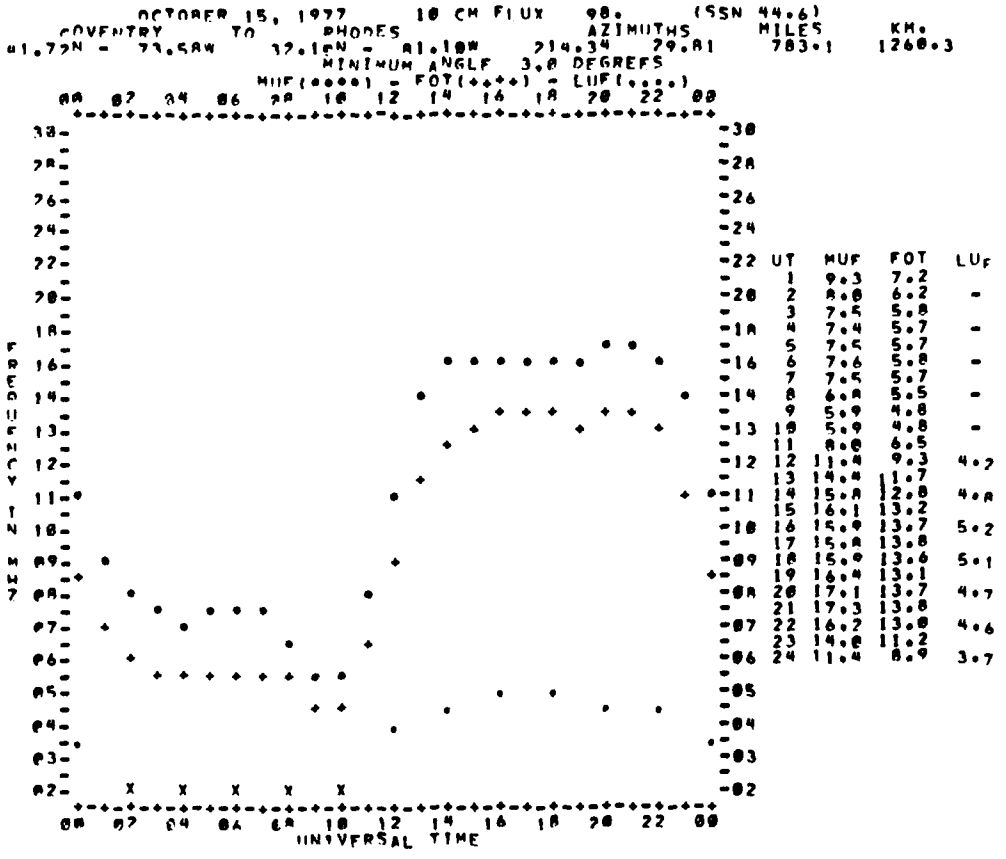


FIG. 3

Simple output of a MUF-FOT-LUF forecast for a given propagation path.

MUF = Maximum Usable Frequency FOT = Frequency of Optimum Transmission

LUF = Lowest Usable Frequency

tors may provide the only useful information because of the lack of other usable observations.

Objective data is required to make a numerical forecast and this is where the ionospheric sounder data plays an important role. Real-time updates to any model ionosphere will improve its accuracy; however, the limiting factor has always been the sparse data network. In fact, hourly observations are needed every 10-20 degrees of longitude and latitude to accurately specify the undisturbed ionosphere (Rush, 1972). AFGWC routinely receives real-time data from only 30 percent of the available stations. The remaining data are received from several hours to several days later. To compensate for this deficiency and enhance the data base, five-day means are used as a preliminary forecast and compared with the observed data from real-time sites. A variety of displays available to the forecaster allow the comparison of these data. Another example is shown in Fig 5. The key ingredients are the displays showing the difference between any observed parameter and the ITS-78 climatological value (Figs. 2,5). From these displays, one can generate an objective and quantitative forecast product that the communicator can use to adjust his climatological forecast. This system allows effective frequency management to a wide variety of users. To satisfy the needs of special users, specific tailor made products must be generated.

SUBJ: HF RADIO PROPAGATION REPORT
SPACE ENVIRONMENTAL SUPPORT BRANCH
AIR FORCE GLOBAL WEATHER CENTRAL
SECONDARY HF RADIO PROPAGATION REPORT ISSUED AT [illegible] OCT 77.
PART I. SUMMARY [illegible] TO [illegible] OCT 77/
FORECAST [illegible] TO [illegible] OCT 77.

	QUADRANT I	II	III	IV
REGION	[illegible] TO 90W	90W TO 180	180 TO 90E	90E TO 0
POLAR	U5/+4[illegible]	U5/+3[illegible]	U5/+25	U5/+20
AURORAL	N6/+35	N6/+25	N6/+25	N6/+25
MIDDLE	U7/+25	N7	N7	N7/+20
LOW	N7/+2[illegible]	N7	N7	N7/+20
EQUATORIAL	N6	N6/+2[illegible]	N7	N7

FIG. 4

Sample HF propagation forecast for first six hours of 10 Oct 1977. Forecasts are made for 20 geographic areas. The letter indicates average observed propagation during past six hours in the indicated region.

W = Poor U = Fair N = Normal

The first number indicates forecast conditions for the next six hours.

1 = Worst (none) 5 = Fair 9 = Best (excellent)

The number after the solidus is the MUF deviation forecast for the next six hours.

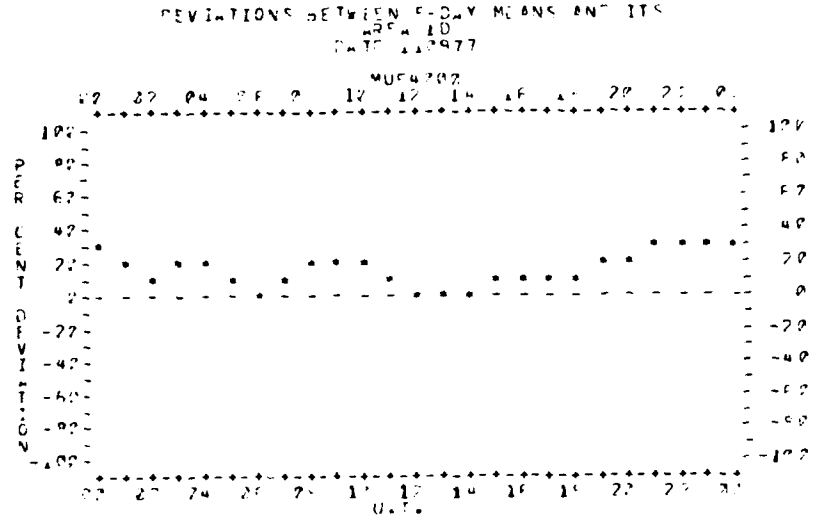


FIG. 5

Graph of 5-day means of observed MUF deviations which are averaged for all stations within a giver geographic area. This is used as the first guess at forecast MUF deviations. These values are modified using the real-time displays like those in Figs 1 and 2.

III. IONOSPHERIC FORECASTING

Some specialized transionospheric radar systems need detailed profiles of the ionosphere to correct returns for deviations induced by the ionosphere. Generally speaking, azimuth angles are affected by horizontal gradients, elevation angles by vertical gradients, and range by number of electrons along the ray path. One way to compensate for these errors is to use a forecast electron density profile or its equivalent and a suitable ray tracing program. This will correct errors in elevation and range, but will not affect azimuth appreciably unless horizontal gradients are accounted for. However, for most mid-latitude applications, horizontal gradients and azimuth errors can be neglected.

One current ionospheric model in use at

AFGWC meshes work by several people and produces tailored output depending on the desired use. This model uses the spectral analysis technique developed by Flattery (1974) coupled with the ITS-78 model ionosphere to generate gridded fields of the ionospheric parameters foF2 and M-factor which are then fed to the Damon-Hartranft Model (1970) to derive electron density profiles. These profiles are integrated over altitude to get Total Electron Content (TEC). Two improvements were made in 1975 which have not been previously reported. One improvement iterates on the sun spot number until the difference between the ITS-78 values of foF2 and observed foF2 are minimized. It is then assumed that the differences are also minimized over data sparse areas and the resulting field is more representative of existing conditions than one generated by the actual sun spot number. The second improvement incorporates empirically derived functions based on latitude, season, and time of day to increase or decrease the topside scale height of the electron density profile so that the derived TEC will match the average TEC observations. The major forecaster input is modifications to the 24 hour foF2 forecast generated by the modified ITS-78 model. The initial guess is five day means which will beat persistence more than 50 percent of the time. The forecaster will adjust the five day means to get the best profile for the intended use. The manual input reduces the RMS error significantly. For example, typical RMS errors for daily foF2 values generated by ITS are between 1.0 and 1.5 Mhz. The 5 day mean can reduce the error to less than 1.0 Mhz. With forecaster input the average RMS error is reduced to near 0.5 Mhz.

IV. USE OF THE 4-D IONSPHERIC MODEL

The four dimensional ionospheric model (4-D) was developed to overcome several deficiencies in the current production model. The Damon-Hartranft model is tied to foF2 observations, but new satellite data can provide much more information than ground-based sensors, especially on the structure of the topside ionosphere. Recognizing that future data sources may provide other inputs, the 4-D was designed to specify an accurate ionosphere based on many varied parameters. All data are reduced to plasma frequency and the ionosphere is represented as a set of coefficients that operate on a set of orthogonal functions. Each successive orthogonal function adds more and more detail to the plasma frequency profile. The 4-D can accurately specify an existing set of observations scattered in space and time and errors are reduced by increasing the number of input observations. The next step is to develop techniques to provide forecast fields as input to the 4-D.

Currently, techniques are under development that will display fields of any mission tailored parameters of interest. The forecaster adjusts these fields by delineating quadrilaterals and applying a multiplication factor to the "box". These boxes can be moved in the ionosphere so they remain over a fixed geographic location; or they are moved across

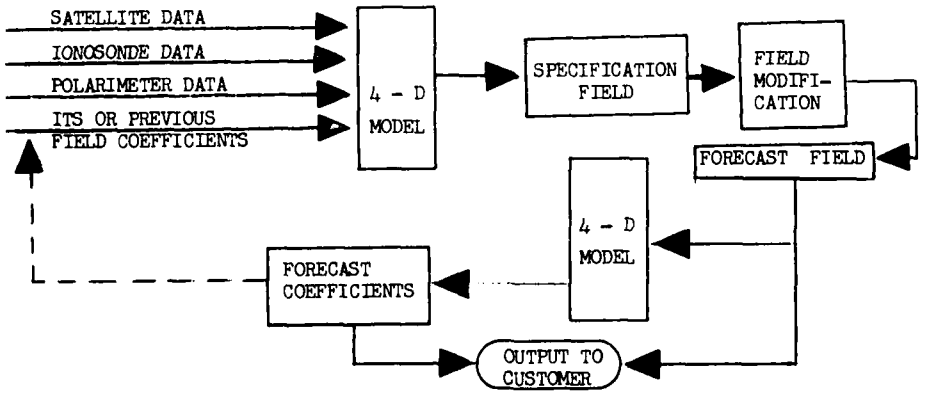


FIG. 6

Schematic Diagram showing the use of the four dimensional ionospheric model for generating forecast products.

the grid so they remain fixed in the ionosphere. Before the multiplying factors are applied to the forecast field they are smoothed in space and time with the rate of smoothing determined by the forecaster. This smoothing is done to eliminate sharp gradients that could be induced. The end result is that a series of enhancements or depressions of the mean field are generated that approximate the forecast conditions.

Displays that order the available data in different sets are an important factor in the forecasting process. By carefully monitoring the past behavior of the ionosphere over a particular site, some estimate can be made about potential behavior. These estimates are the forecast field that can be fed back to the 4-D for generating other parameters. Fig. 6 is a schematic of the entire process that is being developed. Note the ability to make recursive changes. This allows one to use the previous forecast as the first guess and update it with current observations or make additional minor changes.

Testing and evaluation of the 4-D is still in progress. Some satellite data which was expected to be available (topside passive sounder and electron density measurements at 800 Km) has not materialized due to data reduction difficulties. Further testing and implementation of satellite data into the 4-D will have to wait until the problems are corrected. When this testing is complete a flexible model will exist that can specify fairly accurately the observed state of the ionosphere. Now efforts must be directed at improving the forecasting abilities.

V. IONOSPHERIC FORECASTING IMPROVEMENTS

This section will discuss several areas where improvements in forecasting ability are desired. This does not mean that there has not been any work in these areas, or that the work already done is not usable. Investigators should consider operational applications of their work during development if they want to aid the forecaster. An elegant numerical model that takes excessive computer time is not operationally useful. The operational forecaster needs improved techniques in the area of ionospheric storm forecasting, scintillation forecasting, and ray tracing.

The techniques for specifying the ionosphere have been developed and can match the observed values. The development of dynamic models has just barely begun. For example, the best overall model for total electron content forecasting, as far statistics are concerned, is 5-day means. The only time the forecaster can consistently beat 5-day means is during ionospheric storms. Even then he has no better than an even chance because existing models can't match the observed behavior for all storms. Simple rules of thumb, based on available real-time observations, would be an enormous aid because the 5-day means could be adjusted according to the rule of thumb. Forecasts are dependent on a reasonable specification of different ionospheric parameters. A model now exists that can integrate these different observations. Techniques need to be developed that will specify the future state of these ionospheric parameters.

Any dynamic model development must satisfy known user requirements. Currently, products are required to specify electron density profiles, total electron content, and HF propagation reliability. Obviously, development in this area must be an improvement over the current techniques. One of the most needed techniques is to forecast the behavior of the electron density profile and TEC after geomagnetic storm commencement. AFGWC has a limited real-time (every 90 minutes) magnetometer network which can help pin-point storm onset. Additionally, geosynchronous satellites can indicate some substorms and major storm onsets by measuring differences in trapped radiation particle fluxes (Study in progress at AFGWC). The missing link is what parameter or observation can indicate when there is going to be a strong positive or negative phase and for how long and severe will these phases last. These are difficult questions and the answers are needed to satisfy operational requirements.

The largest area of concern is future requirements levied on the space environmental forecaster. One of these that is directly related to the ionosphere is scintillations of VHF, UHF, and higher frequencies. The current state of the art is best described by Aarrons (1975). There is a need for an operational forecast capability than can specify location, magnitude, and duration of scintillations. This will allow communications managers to circumvent outage periods caused by scintillations.

Finally, as transionospheric radar systems become more accurate, the errors induced by the ionosphere become more significant. Three dimensional ray-tracing models exists, but use excessive amounts of computer time. AFGWC has ray-tracing programs but they are not routinely used because of the excessive computations required. One of the requirements for the future would be a system that could use the 4-D coefficients for determining the path an electromagnetic ray would take through the ionosphere. Unfortunately, the constraints of time, accuracy, and reliability may be difficult to overcome.

SUMMARY

AFGWC uses a variety of real-time and near real-time data sources to satisfy operational needs with alerts forecasts, and specifica-

tions of the near-earth space environment. Limited ionospheric observations are coupled with climatological data from ITS-78 and are used to provide mission tailored products for HF frequency managers, radar operators and analysts, and Command and Control authorities. The four dimensional ionospheric model has put the state-of-the-art for ionospheric specification ahead of the forecasting ability. Future research directed at solving operational forecasting problems will have immediate benefits if the techniques are operational useful.

REFERENCES

1. Aarons, J. (1975), Global Morphology of Ionospheric Scintillations II. AFCRL-TR-75-0135.

2. Barghausen, Alfred F. etal., Predicting Long-Term Operational Parameters of High Frequency Sky-wave Telecommunications Systems. Institute for Telecommunications Sciences. Boulder, CO. May 1969, ESSA TR ERL 110-ITS-78.

3. Damon, T.D. & Hartranft, F.R. (1970), Ionospheric Electron Density Profile Model, Aerospace Environmental Support Center Technical Memorandum 70-3.

4. Flattery, T.W. and Davenport, G.R. (1977), "Four Dimensional Ionosphere Model", Paper presented at meeting of the U.S. National Committee of the Interational Union of Radio Science, Stanford University, 22-24 June 1977.

5. Flattery, T.W. and A.C. Ramsay (1975), "Derivation of Total Flectron Content for Real-Time Global Applications", Effect of the Ionosphere on Space Systems and Communications. Edited by Goodman, J.M.

6. Rush, Charles M. (1972), Improvements in Ionospheric Forecasting Capability, AFCRL-72-0138, AFCRL Environmental Research Paper No 387.

7. Rush, Charles M. & Gibbs, Joseph (1973), Predicting the day-to-day Variability of the Mid-Latitude Ionosphere For Application to HF Propagation Predictions. AFCRL-TR-73-0335., AF Surveys in Geophy #268.

8. Snyder, Arnold L. Jr. (1974), "Real-Time Magnetospheric and Ionospheric Monitoring"., Paper Presented at Sixth Conference on Aerospace & Aeronautical Meteorology, El Paso, TX. 12-14 Nov 74.

Prediction of Solar Proton Events at the Air Force Global Weather Central's Space Environmental Forecasting Facility

Edward W. Cliver, James A. Secan, Edward D. Beard, and James A. Manley
Headquarters Air Force Global Weather Central
Offutt AFB, NE 68113

ABSTRACT.

Large, favorably located, solar flares can emit low energy ($\backsim$ 5 Mev) protons which produce enhanced ionization in the earth's polar ionosphere. This enhanced ionization is responsible for the Polar Cap Absorption (PCA) phenomenon. Higher energy particles can adversely affect DOD satellite systems and pose a problem for astronauts engaged in extravehicular activity. Predicting the occurrence and intensity of proton events after a candidate particle flare has occurred is a priority concern of the Air Weather Service's space environmental forecasting facility. This paper discusses the evolving data sources, procedures, and techniques currently used for this purpose at the facility and their application to the recent energetic activity from McMath region 14943 (7-24 Sep 77). We suggest areas for further study and improvement.

1. INTRODUCTION

Energetic protons emitted by intense solar flares with energy $\backsim$ a few Mevs can disrupt high frequency command and control communications in the earth's polar regions for periods of days. Higher energy particles impinging on satellite sensors can render them temporarily inoperative. With the advent of the Space Shuttle in the late 1970's, our attention will once again focus on the problem of protecting men in space from hazardous solar flare radiations. Forecasting the occurrence and intensity of solar flare particle events for these diverse activities is an important facet of the mission of the Air Weather Service's space environmental forecasting facility located in the Air Force Global Weather Central at Offutt AFB, Nebraska.

With flare prediction still in its infancy, it is virtually impossible to forewarn customers of the sudden ionospheric disturbances (e.g. Short Wave Fade, Sudden Phase Anomaly, etc.) caused by flare soft x-radiation which arrives at earth concurrently with the visual light and radio emissions through which we normally "observe" the flare. Happily, this is not the case for the effects of flare protons in the earth environment. By exploiting the time differential between the observation of the flare and the arrival of particles, the solar forecaster is able to sift real time flare data reported over the Astrogeophysical Teletype Network (ATN), run quantitative prediction models, and notify customers in advance of all save the most energetic events. In Section 2 of the paper the various features of flares which are indicative of proton production will be discussed with the emphasis on those "yes-no" predictors which are best suited to the facility's real-time mission. Section 3 will deal with the proton prediction program developed by Smart and Shea (1977 a,b). This is the primary tool currently in use for predicting the size of events. In Section 4 the recent (September 7-24) energetic activity from McMath region 14943 will be discussed in terms of its proton emission; in this respect it was the most interesting and prolific region to appear on the sun since September of 1974. Based on this first extended operational test of the facility's proton prediction capability, during solar cycle 21, we offer some comments and suggestions in Section 5.

2. YES-NO PREDICTORS

With few exceptions, flares which produce protons at earth are significant flares but the converse is certainly not true. Other factors come into play. One of the first to be recognized (McCracken, 1959) was the longitudinal asymmetry in the location of particle producing (parent) flares on the sun, with the majority occurring in the western hemisphere. This is due to the well known "garden hose" effect of the interplanetary magnetic field. Because of the geometry of the interplanetary medium, no observable characteristic of large solar flares can be regarded as a sufficient condition for particles to arrive at earth. Over the years, however, statistical studies have identified several observables as necessary, "almost necessary", or, at least, favorable conditions for the production of high energy protons in solar flares, even though the actual mechanisms are not well understood. Ellison et al. (1961)

pointed out that all flares which produce ground level effects (GLE) are double ribbon flares, with the ribbons embedded in large sunspots. This also applies for the majority of proton flares, especially those which produce strong polar cap absorption events (PCA). The formation of loop prominences, after the flare, has also been suggested as evidence of the acceleration of protons to high energies (Bruzek, 1964). In terms of the facility's real-time mission, the most useful proton "signature" has been the U-shaped peak flux density radio spectrum (Castelli et al. 1967; Castelli and Guidice, 1971). O'Brien (1970) analyzed 54 visible hemisphere flares responsible for principal ($\geq$ 2.0 db) PCA events between 1952 and 1968 and concluded that the radio emission from 49 of them satisfied the "modified" U-shaped spectral criterion consisting of: high flux density burst at meter wavelengths, pronounced dip in the centimeter wavelengths, intensities approaching or $>$ 1000 solar flux units in the frequency ranges from 200 to 1000 MHZ and 8500 to 10000 MHZ, and a rise time $\geq$ 4 min in either frequency interval.

A parent flare feature which offers promise for the future as a yes-no predictor to be used in conjunction with the Castelli criterion is the Type II burst. Švestká and Fritzová-Švestková (1974) point out that previous poor correlations obtained between proton events and Type II bursts based on data from the 19th solar cycle were due to the low sensitivity of observation of both the phenomena under consideration. Figure 1, based on data compiled in the Catalog of Solar Particle Events, 1955-1969 (Švestká and Simon, 1975), shows how the correlation has improved with observing technology and expanded patrol hours.

The observability of both the Castelli U and the Type II burst proton signatures should improve significantly with the implementation of the Radio Interference Measuring Set (RIMS) at three circum-global Space Environmental Support System (SESS) observatories between 1977 and 1980. The RIMS consists of a fully automated system of eight discrete frequency solar radiometers operating at 245 MHZ, 410 MHZ, 606 MHZ, 1415 MHZ, 2695 MHZ, 4995 MHZ, 8800 MHZ, 15400 MHZ and a 25-75 MHZ swept frequency radio telescope. The antennas and receivers are linked to a mini-computer which monitors the instantaneous solar emission and transmits ATN reports of burst activity as it occurs. At present only the Saqamore Hill and Manila observatories monitor enough discrete frequencies to independently detect the occurrence of a U-shaped peak flux density spectrum. While the combined coverage of the sun by these two observatories is nearly complete throughout the year, events occasionally "fall through the cracks" as we shall see in Section 4. Also Saqamore Hill and Manila are the only observatories currently capable of detecting and reporting the occurrence of Type II bursts. Note added in proof: The Palehua observatory's RIMS became operational in the fall of 1977.

3. PREDICTING THE SIZE OF EVENTS.

Once a large, favorably located, flare has occurred and the observatory reports indicate it was accompanied by intense meter and microwave emission with a U-shaped spectrum and/or a strong Type II burst, the solar forecaster must predict the magnitude of the impending event. This can be done in a matter of minutes using the Smart-Shea prediction program which can be run on either of the two demand terminals in the facility. The minimum inputs to the program are the flare start time and heliographic position and an x-Ray or microwave burst energy used to predict the peak proton flux in a given energy channel (usually $>$ 5.2 Mev). To more completely specify an event, solar wind speeds may be input when available and a prediction of the proton spectral slope based on radio data for events with a U-shaped peak flux density spectrum (Bakshi and Barron, 1974, 1975) may also be entered. The output consists of predictions for riometer absorption and particle fluxes in energy ranges directly comparable to those observed by a number of different satellites such as SMS/GOES, VELA, or SOLRAD. The

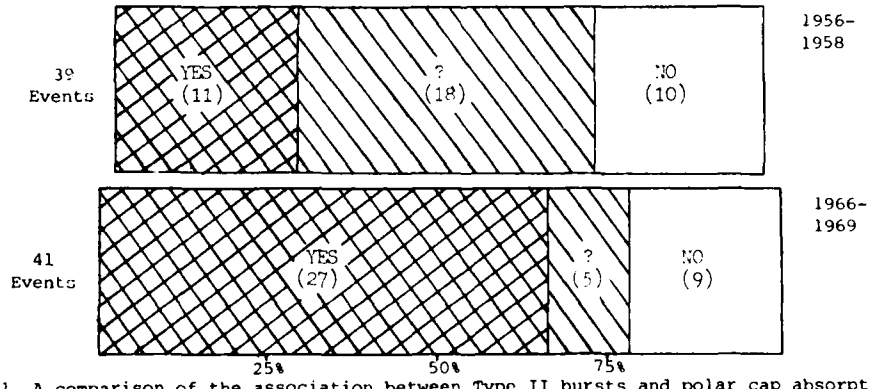


Figure 1 A comparison of the association between Type II bursts and polar cap absorption events for peak years during the 19th and 20th solar cycles. The associations in doubt are mainly due to a lack of dynamic spectrograph coverage during the parent flare.

program was designed to handle multiple flare events or overlapping particle injections and predictions are given for hourly intervals over the duration of the event(s). Since the predictions cover the entire spectral range of the accelerated protons, a wide variety of support requirements can be met.

While a discussion of the inner workings of the Smart-Shea program is beyond the scope and intent of this paper, a few words on the various electro-magnetic specifications of the peak proton flux (the "front-end" of the program) are in order before we discuss its actual application in Section 4. Historically, the algorithms relating x-Ray or microwave burst integrated or semi-integrated (rise) energy to proton fluxes have been derived using either (earth orbiting) satellite flux observations or riometer measurements since the earth system, after all, is the primary area of concern. In the Smart-Shea program, however, the front-end flux prediction is for the flux at 1 a.u. out on the Archimedes spiral connecting to the flare site. The flux predicted to arrive at earth is this value reduced by a coronal gradient of $e^{3\theta}$ where θ is the angle in radians between the heliographic longitude of the flare and the longitude of the foot point of the earth Archimedes spiral, both taken at the time of peak low energy flux at earth. Thus it has been necessary to modify the various prediction equations to establish relationships between the size of the x-Ray or microwave burst and the peak proton flux at 1 a.u. out on the spiral originating at the flare site. The result of this process for the various algorithms is given in Table 1. Note that the improvement in the correlation coefficient that one would hope to see following this normalization process (as affirmation of the validity of the particle propagation correction) is not apparent. In fact, most of the correlation coefficients have decreased slightly; those which were seriously degraded have been excluded from the program.

To completely determine the particle spectrum, it is necessary to specify a particle spectral slope as well as the peak flux in a given energy range in the front-end of the program. In the absence of an external specification of this slope by the forecaster, using the Bakshi-Barron algorithm, a default value (Van Hollebeke et al. 1975) is used.

4. MCMATH 14943

Now that we have had an overview of the tools and techniques at the forecaster's disposal, let us see how they worked in their first extended operational test of this solar cycle, the energetic activity occurring between September 7-24 in McMath region 14943. The Culgoora Observatory report on the morning of 08 Sep is worth repeating in its entirety as it appeared on the ATN because of the invigorating effect it had on those of us who had weathered the unusually long solar minimum.

> Good observing conditions. Spectacular limb activity has occurred all morning at NE10. On opening at 2124Z surges and thin suspended material were observed. At 2141Z, a large BSL (bright surge at the limb) occurred at SE25, reaching about 0.25 (solar) radius. The flare was certainly of brilliant intensity and was classified as a 1B. A surge occurred at 2239Z at NE10, followed at 2252Z by what appeared to be a second, and larger flare. This second event reached maximum intensity at about 2308Z but increased in size, eventually developing into a classical system of bright flare loops. Large short wave fades accompanied the flare activity. At meter wavelengths, a very strong Type II burst began at 2227Z, along with strong Type III activity. The Type II event was localized near the east limb near the equator and contained multiple fast changing lobes at 43, 80, and 160 MHZ. Flux at 43 MHZ was greater than 50,000 flux units and greater than 10,000 FU at 80 MHZ. A weak type IV event also began at 2340Z.

Thus region 14943 first made its presence known on the sun. The flare on the 7th was the first of four (possibly six) energetic events from the region which likely produced particles. Even though this flare was located slightly behind the limb, the proton signatures - Castelli U, Type II burst, and post flare loops plus ample soft x-Ray (class X2) and microwave (3100 Sfu at 4995 MHZ reported by Manila) emission - were all present. However, the extreme eastern position of the flare strongly mitigates against the arrival of protons at the earth and this is borne out by the observations of low and high energy protons at the earth given in Table 2a along with the Smart-Shea program predictions.

The 2B flare on 09 Sep at NO9E76 (1600UT maximum) was accompanied by strong (4700 Sfu at 8800 MHZ reported by Sagamore Hill) centimeter wavelength emission satisfying the Castelli criterion, an X2 1-8A$^{\circ}$ x-Ray burst and a major ($\geq$ 500 Sfu) Type II radio burst. Again, as seen in Table 2b, the eastern position of the flare ruled against the interception of significant numbers of flare particles at the earth.

The 2B flare (NO7W21) on 16 Sep at 2228UT did not have the classic U-shaped peak flux density radio spectrum associated with significant particle producing flares. Manila reported a burst maximum of 2400 Sfu at 606 MHZ and 1415 MHZ (2308UT maximum) with peak fluxes decreasing toward higher frequencies to 920 Sfu at 8800 MHZ. However, a class M6 soft x-Ray event was observed in addition to a significant ($<$ 500 Sfu) Type II radio burst. In addition region 14943 had already produced two events which had all the earmarks of proton production. This in itself is a favorable indicator of a large flare's potential for producing particles. (Of the 221 confirmed par-

TABLE I

Predictor(X)	Frequency	Algorithm	No. of Events	Correlation Coefficient	Correlation Coefficient Before Normalization Procedure
Integrated Radio Burst Energy (10^{-17}joules $m^{-2}Hz^{-1}$) Straka (1970)	606 MHZ	$Y=33.4\ X^{0.96}$	15	.780	.894
	1415	$Y=49.0\ X^{1.27}$	15	.704	.732
	2695	$Y=17.9\ X^{1.49}$	15	.702	.641
	4995	$Y=13.0\ X^{1.17}$	15	.573	.594
Castelli et al. (1973)	8800	$Y=12.3\cdot10^{28}X^{1.67}$	18	.816	.8?
Semi-Integrated Radio Burst Flux Density (joules $m^{-2}Hz^{-1}$) Newell (1972)	606 MHZ	$Y=894.4\ X^{-0.01}$	16	-.009	.241
	1415	$Y=28.8\cdot10^{18}X^{0.97}$	16	.508	.506
	2695	$Y=22.0\cdot10^{25}X^{1.40}$	16	.666	.881
	4995	$Y=15.6\cdot10^{19}X^{1.05}$	16	.535	.828
	8800	$Y=27.0\cdot10^{18}X^{1.01}$	16	.521	.772
Burst Energy (Sfu·min) Cliver (1976)	2.8 GHZ	$Y=0.0013\ X^{1.45}$	47	0.54	0.59
X-Ray Burst Energy (erg·min $cm^{-2}s^{-1}$) Kuck (1969) Kuck et al. (1971)	.5-4A°	$Y=201.7\ X^{0.99}$	21	.43	--
	1-8A°	$Y=30.7\ X^{1.33}$	28	.53	--
	1-16A°	$Y=16.7\ X^{0.51}$	20	.29	--

Microwave and x-Ray burst energies as predictors of the peak proton flux one a.u. out on the Archimedes spiral connecting to the flare site. For the three x-Ray and the 8800 MHZ integrated energy algorithms, Y is the > 10 Mev flux (protons $cm^{-2}s^{-1}ster^{-1}$). For all others, Y refers to the > 5.2 Mev flux. The units of X are given in the first column with the exception that the X units for the 8800 MHZ integrated energy equation are joules $m^{-2}Hz^{-1}$.

TABLE II (a,b)

Predictor (a) (1)(2)	30 MHZ Riometer Absorption (db)	18-38 Mev Flux (protons $cm^{-2}s^{-1}str^{-1}Mev^{-1}$) (3)	25 Mev Flux (protons $cm^{-2}s^{-1}str^{-1}$) (4)	Predictor (b)	30 MHZ Riometer Absorption (db)	18-38 Mev Flux (protons $cm^{-2}s^{-1}str^{-1}Mev^{-1}$)	25 Mev Flux (protons $cm^{-2}s^{-1}str^{-1}$)
07 Sep				09 Sep			
1415 SI(Mnl)	0.7	.007	--	1415 SI(Sag)	0.8	.002	--
2695 SI(Mnl)	1.1	.019	--	2695 SI(Sag)	0.5	.006	--
4995 SI(Mnl)	0.5	.005	--	4995 SI(Sag)	0.5	.005	--
8800 SI(Mnl)	0.4	.002	--	8800 SI(Sag)	0.5	.005	--
1415 I(Mnl)	0.5	.004	--	1415 I(Sag)	0.5	.006	--
2695 I(Mnl)	0.8	.011	--	2695 I(Sag)	0.8	.014	--
2800 I(Mnl)(6)	0.3	.002	--	2800 I(Sag)	0.3	.003	--
4995 I(Mnl)	0.7	.007	--	4995 I(Sag)	0.7	.010	--
8800 I(Mnl)	1.6	.042	--	8800 I(Sag)	4.7	.510	--
.5-4A°	0.3	.001	--	.5-4A°	0.2	--	--
1-8A°	0.3	.002	--	1-8A°	0.2	--	--
Observed	0.3	.7	--	Observed	0.5 (7) (8)	.08	--
Time of Maximum				Time of Maximum			
Predicted (5)	09 Sep 1700Z	09 Sep 1600Z	--	Predicted	11 Sep 0400Z	11 Sep 0300Z	--
Observed	09 Sep 1400Z	09 Sep 1600Z	--	Observed	11 Sep 2130Z	{10 Sep 1320Z 11 Sep 0040Z	--

Comparison of the Smart-Shea program predictions with observations for the energetic activity from McMath region 14943 (07-19 Sep 77). The numbers in parentheses refer to the notes.

Notes to Table II.

(1) S.I. = Semi-integrated flux density (rise energy); I = integrated flux density (total burst energy).

(2) Sag = Sagamore Hill; Mnl = Manila; Ath = Athens

(3) The predictions for this energy range are for the 19-29 Mev channel on the VELA satellites. The 18-38 Mev observations are from the SMS/GOES satellites.

(4) The > 25 Mev neutron monitor aboard the VELA satellites is an indirect sensor of high energy protons and its sensor is highly spectral dependent. We have considered it in this study because it serves as a good indicator of the onset times of high energy particle events.

(5) The predicted maximum time of a proton event in a given energy range is the same for all algorithms.

(6) The input observation for this predictor is the total burst energy at 2695 MHZ, 2800 MHZ, or 3000 MHZ. The nominal frequency of observation for the SESS observatories is 2695 MHZ.

TABLE II (c,d)

Predictor (c)	30 MHZ Riometer Absorption (db)	18-38 Mev Flux (protons $cm^{-2}s^{-1}str^{-1}Mev^{-1}$)	25 Mev Flux (protons $cm^{-2}s^{-1}str^{-1}$)	Predictor (d)	30 MHZ Riometer Absorption (db)	18-38 Mev Flux (protons $cm^{-2}s^{-1}str^{-1}Mev^{-1}$)	25 Mev Flux (protons $cm^{-2}s^{-1}str^{-1}$)
16 Sep				19 Sep (10)			
1415 SI(Mnl)	13.1	41.0	302.3	4995 SI(Ath)	4.4	8.6	103.4
2695 SI(Mnl)	16.1	62.0	463.5				
4995 SI(Mnl)	5.?	6.9	43.8				
8800 SI(Mnl)	4.0	3.8	28.6				
1415 I(Mnl)	6.7	11.0	79.6	1415 I(Sag)	4.1	7.6	95.5
2695 I(Mnl)	10.2	25.0	183.0	2695 I(Sag)	4.7	8.6	119.3
2800 I(Mnl)	3.7	3.4	25.5	2800 I(Sag)	2.5	2.7	35.8
4995 I(Mnl)	2.9	0.9	7.7	4995 I(Ath)	1.7 (11)	1.2	19.1
8800 I(Mnl)	4.8	5.5	40.6	4995 I(Sag)	2.7	3.3	43.0
				8800 I(Sag)	2.6	3.0	39.0
.5-4A°	1.0	0.2	2.8	.5-4A°	2.8	3.5	45.3
1-8A°	1.2	0.3	3.5	1-8A°	3.5	5.4	66.8
Observed	0.6 (7)	0.6	29.2	Observed	8.6 (7)(9)	5.0 (12)	251.3
Time of Maximum				Time of Maximum			
Predicted	17 Sep 0400Z	17 Sep 0300Z	17 Sep 0300Z	Predicted	19 Sep 0400Z	19 Sep 1200Z	19 Sep 1200Z
Observed	17 Sep 1215Z	17 Sep 0210Z	17 Sep 0020Z	*Observed*	19 Sep ~115Z	19 Sep 2115Z	19 Sep 1750Z

Comparison of the Smart-Shea *program predictions* with observations for the energetic activity from McMath region 14943 (07-19 Sep 77). The numbers in parentheses refer to the notes.

Notes to Table II (cont'd).

(7) The peak riometer absorption for each of these events may have been delayed (or masked) by the diurnal variation at Thule.

(8) The Thule 30 MHZ absorption readings taken 15 minutes before and 15 *minutes* after this observation were 0.0 db.

(9) This absorption value was a spike, rising out of a ~ 4.0 db background.

(10) Only partial measurements of the burst energy were available from Saqamore *Hill which did not* begin observing the event until after burst maximum (as reported by Athens). As such, the predictions will be lower than those made had the entire event been observed.

(11) Differences between predictions for an event which are based on the same algorithm but use observations from different observatories are caused by discrepancies in the energy measurement as discussed in *Section* 5.

(12) Gaps in the SMS/GOES data occurred from 0800-1400UT and from 1900-2100UT.

ticle events listed in the Catalog of Solar Particle Events, 1955-1969 which have definite or probable solar flare associations, 82 (37 percent) originated in regions which had already produced a particle event.) Finally, and equally as important, the region had by now rotated to the western hemisphere. The comparison of predictions and observations in Table 2c reveals large overprediction by all of the radio predictors for the low energy proton flux as measured by riometer absorption. At higher energies the results are somewhat better although the spread in the predictions is considerable. The initial increase in > 25 Mev protons was detected only 37 minutes after the peak of the microwave burst.

Early morning power problems at Sagamore Hill prevented the observation of all of the radio burst associated with the 3B flare (NO8W58) at 1038UT on 19 Sep. An X2 1-8A° event was observed by the SMS/GOES satellites and the partial microwave burst observations indicated large proton fluxes (Table 2d). For this event the time delay between the 5000 MHZ burst maximum (1036UT reported by Athens) and the onset of higher energy particles was 39 minutes. A 3N flare (NO9W66) observed by Culgoora beginning at 0257UT on the 20th may have been a contributing source of particles for the later stages of this event. A Type II burst was associated with this flare but other proton indicators were unimpressive. Manila reported ~ 50 Sfu peak flux densities at the higher microwave frequencies and an M4 event was observed by SMS/GOES in the 1-8A° band. As an interesting historical footnote, the first of these two flares occurred three years to the day after, and in almost the same position as, the proton flare of 19 Sep 1974.

At 0615UT on 24 Sep a > 25 Mev proton enhancement was detected by the Vela neutron monitor. These particles almost certainly emanated from region 14943 which was then two days behind the western limb. Culgoora reported a very energetic Type II/III burst combination with peak times at 0210UT and 0556UT, respectively, occurring in association with an importance 3 bright surge at the limb at about NW10. Manila reported a weak microwave burst (~ 20 Sfu at the higher frequencies) peaking at 0555UT and the 1-8A° x-Rays did not rise above class C level. It is likely that limb shielding was significant at these higher frequencies. Assuming that the protons were accelerated at the time of the Type II burst (Švestká and Fritzová-Švestková, 1974), the transit time to earth of the > 25 Mev protons was four hours.

5. COMMENTS AND SUGGESTIONS

Based on the experience gained from the activity in region 14943 we offer the following suggestions and comments:

(1) Our "yes-no" prediction capability should be significantly improved by the implementation of the RIMS. At present, as we saw for the event on 19 Sep, the peak flux density spectrum information is lost when either of the primary SESS radio observatories at Sagamore Hill or Manila is not observing. (For these events, even if other "yes-no" predictors indicate that a proton event is imminent, the particle spectral slope prediction based on the width of the U-shaped radio spectrum is lost and the default value of this important parameter must be used.) At lower frequencies, the 24 hour sweep frequency meter and dekameter wavelength burst patrol should prove a valuable supplement to the discrete frequency microwave patrol.

(2) While there is a good deal of scatter in the predictions of event size obtained from the Smart-Shea program depending on which "front-end" is used, the totality of inputs for each event from McMath 14943 generally gives a consensus answer in fair agreement with observation (Table 2). Operationally, however, sufficient time is not available to run the program using all inputs before the initial notification of customers is required. The protons simply arrive too fast. One of the prime challenges facing the facility during the coming cycle is the identification of those prediction algorithms which consistently give results that are in the ball park and the elimination of those which do not. At this point it is too early to comment.

(3) The prediction of the maximum times of proton fluxes in the various energy channels obtained from the Smart-Shea program appears to be quite reasonable considering the complexity of the problem and the uncertainties involved.

(4) As can be seen from the events on the 16th and 19th for favorably located flares, the onset of the high energy particle event at earth may occur before the microwave burst is terminated. For this reason the x-Ray rise energy or semi-integrated radio flux density should be used for the initial prediction of proton event size. For the semi-integrated radio flux algorithms the approximation

$$E_{rise} = 1/3\ S_p T_r \qquad (1)$$

where

E_{rise} = burst rise energy

S_p = burst flux density

and T_r = time from start to max,

based on the assumption that the burst flux density increases parabolically, is used. This assumption breaks down for bursts with long rise times and equation (1) will give too large a value for the rise energy. This in turn results in an over prediction of the proton flux as can be seen in Table 2c for the event on 16 Sep. One of the advantages of the RIMS will be that the rise energy can be accurately measured and reported directly. In addition, measurements of the total burst energy will also be automated and it will no longer be necessary to use rule of thumb approximations to determine this parameter.

(5) Finally, as we saw for the event on the 24th, forecasting behind-the-limb proton

events remains beyond the state of the art and will continue to present a challenge in the 21st solar cycle. Since limb shielding of the x-Ray, optical, and microwave emissions from these events can be severe, the longer frequency radiation offers the best chance of observation. Specifically, detection of strong Type II emission from a region beyond the western limb which had produced protons while on the disk should alert the forecaster to the possibility of the occurrence of a proton event.

REFERENCES

Bakshi, P. and Barron, W.: 1974, Spectral Correlations between Solar Flare Radio Bursts and Associated Proton Fluxes I, AFCRL TR 74-0508.

Bakshi, P. and Barron, W.: 1975, Spectral Correlations between Solar Flare Radio Bursts and Associated Proton Fluxes II, AFCRL TR 75-0579.

Bruzek, A.: 1964, Astrophys. J. 140, 746.

Castelli, J.P., Aarons, J., and Michael, G.L.: 1967, J. Geophys. Res. 72, 5491.

Castelli, J.P., Barron, W.R., and Aarons, J.: 1973, Solar Radio Activity in August 1972, AFCRL TR 73-0086.

Castelli, J.P. and Guidice, D.A.: 1971, Proceedings of COSPAR Symposium on Solar Particle Event of November 1969, p. 27.

Cliver, E.W.: 1976, Parent Flare Emission at 2.8GHz as a Predictor of the Peak Absorption of Polar-Cap Events, NELC TR 2015.

Ellison, M.A., McKenna, S.M.P., and Reid, J.H.: 1961, Dunsink Obs. Publ. 1, 5..

Kuck, G.A.: 1969, Predictions of Polar Cap Absorption Events, AFWL TN WLRTH 69-8.

Kuck, G.A., Davis, S.R., and Krause, G.J.: 1971, Prediction of Polar Cap Absorption Events, AFWL TR 21-1.

McCracken, K.G.: 1959, Nuovo Cim. (series 10), 13, 1081.

Newell, D.T.: 1972, Forecasting Peak Proton Flux and PCA Event Magnitude Using 'Flash Phase' Integrated Radio-Burst Flux Density, AFCRL TR 72-0543.

O'Brien, W.E.: 197., The Prediction of Solar Proton Events Based on Solar Radio Emissions, AFCRL TR 70-0425.

Smart, D.F. and Shea M.A.: 1977a, in Proceedings of 15th International Cosmic Ray Conference, Vol. 5, p. 131.

Smart, D.F. and Shea, M.A.: 1977b, Space Research 18, (in press).

Straka, R.M.: 1970, The Use of Solar Radio Bursts as Predictors of Proton Event Magnitude, AFCRL Space Forecasting Research Note 2.

Švestká, Z. and Fritzová-Švestková, L.: 1974, Solar Phys. 36, 417.

Švestká, Z. and Simon, P., ed: 1975, Catalog of Solar Particle Events, 1955-1969, Astrophysics and Space Science Library, D. Reidel Publishing Co., Dordrecht, Holland.

Van Hollebeke, M.A.I., Ma Sung, L.S., and McDonald, F.B.: 1975, Solar Phys., 41, 189.

REMOTE IONOSPHERIC MONITORING

Jurgen Buchau and William N. Hall

Air Force Geophysics Laboratory, Hanscom AFB, MA 01731

Bodo W. Reinisch and Sheryl Smith

University of Lowell Center for Atmospheric Research, Lowell, MA 01854

ABSTRACT

A program for the development of automatic real-time monitoring of the ionosphere at a remote observing location is described. The Digisonde 128, a digital sounder in routine operation since 1970, provides digital data suitable for online computer processing. Analytical methods for the detection of ionospheric echoes and the determination of their virtual height, amplitude, and range spread have been developed. For real time application, these methods have been implemented in hardware using microprocessors. The Geomonitor and Geomonitor Display System (GDS) provide the capacility to perform the monitoring at any desired site separated from the remote observing location. The Geomonitor, in addition to analyzing vertical and backscatter digital ionograms, digitizes magnetometer, riometer, radio polarimeter, and satellite scintillation data for transmission to the using site. Compression of the digital ionogram data and the other geophysical and ionospheric data permits transmission over a 600-1200 baud telecommunications link to the GDS. The GDS formats and displays the data for the remote observer and analyzes the data to call attention to disturbances. For each ionogram, lines are printed summarizing the frequency extent of the observed echoes and the variation with height or range of the echo amplitudes. As further data arrive, lines are added to the printouts, resulting in time histories, so-called "ionospheric characteristics" of layer heights, backscatter ranges and critical or top frequencies. In addition to showing regular diurnal variations, the onset, type, and severity of ionospheric disturbances can be determined from these characteristics. To aid in the interpretation of the characteristics, ionograms are reconstituted from the ionospheric data messages. These with displays of the other geophysical and ionospheric data create a comprehensive display for evaluation by the observer.

INTRODUCTION

The high latitude ionosphere, strongly controlled by auroral particle precipitation and especially disturbed during auroral substorms, affects a variety of USAF and DOD systems, as well as military and commercial communications links operating in this environment. To advise the various military users of prevailing conditions, the AF Air Weather Service (AWS) has monitored routinely the state of the local arctic ionosphere and the onset of disturbances from various ground sites as among others Goose Bay, Labrador, Eielson AFB, Alaska, Tromso, Norway. The reports are collected and evaluated by the AF Global Weather Central (AFGWC) and advisories and warnings are being disseminated routinely or on special request through the Space Environment Support System (SESS).

The monitoring of this environment has for a long time been accomplished by trained personnel through hourly reports from the various sites, using a limited selection of geophysical parameters. To automate the process of data analysis and data transmission especially at remote sites, we have developed the Geomonitor (Reinisch and Smith, 1976). This system extracts the essential information from the digital

ionograms, supplied by the Digisonde (Bibl et al, 1970; Bibl and Reinisch, 1978), a digital ionosonde used for vertical incidence and backscatter soundings. Together with properly formatted digitized data from other geophysical sensors messages are formed that can be sent to Air Force Global Weather Central or to other user sites as a final step to remotely monitor the environment of interest. This remote monitoring capability is of special interest at high latitudes, since a.) the auroral effects routinely and severely disturb the ionospheric environment making 24 hour coverage and fast issue of warnings essential, b.) rapdily rising manpower costs and the difficulty to find qualified personnel willing to work for extended periods at these remote, high latitude stations require automation as a cost effective solution.

GOOSE BAY OVAL MONITOR STATION

Requirements for better specifications of the high latitude ionospheric and auroral environment surfaced in the mid- to late sixties, as AF surveillance systems increased their accuracy requirements, the use of an Over-the-Horizon Backscatter System in these disturbed regions began to be contemplated and effects on trans-ionospheric propagation started to become a major concern. AFGL started a large airborne program to improve the understanding of the structure and dynamics of the high latitude ionosphere. Flights between 1967 and 1971 indicated that the auroral oval concept (Feldstein and Starkov, 1967) provided a good ordering frame for various ionospheric phenomena. Airborne investigations of initially the noon sector of the oval (Whalen et al, 1971) and then the more disturbed and complicated night sector (Buchau et al, 1972; Wagner and Pike, 1972) led to a unified picture of the structure of the high latitude environment (Gassmann, 1972) and suggested the feasibility of monitoring this environment from a few selected points. To illustrate the concept, we show in Figure 1 the auroral oval (shaded area) for average magnetic conditions and the location of Goose Bay in relation to the oval in hourly increments. Large scale ionospheric entities ordered by the auroral oval are: a. The FLIZ (F-layer irregularity zone, Pike, 1972) generally colocated with the oval belt, b. The midlatitude F-layer trough (Muldrew, 1965) found south of the equatorward boundary of the night sector of the oval, c. Auroral Es, colocated with the oval belt, d. Auroral E, a particle produced thick E-layer belt coinciding with the oval in the night sector and found just south of the oval in the noon sectors, and e. D region enhancements, during substorms within the night sector, of the oval, and traveling through sunrise around the oval into the noon sector of the oval, where they arrive in a region between 60° and 70° corrected geomagnetic (CG) latitude within .5 to 2 hours after the onset of the substorm (Driatskiy, 1968; Elkins, 1972) and f. The polar cap ionosphere within the oval belt, with auroral, E-layer and F-region phenomena different from those in the adjacent oval. The diameter of the auroral oval and the auroral activity within the oval belt and the polar cap are under strong control of the interplanetary and the earth magnetic field. The equatorward edge of the midnight sector of the oval is found at 70° (CG) latitude under quiet (Q=0) conditions and at 59° CG latitude under very disturbed (Q=8) conditions. All the described ionospheric regimes and phenomena change their locations and/or their intensities as the oval expands or contracts.

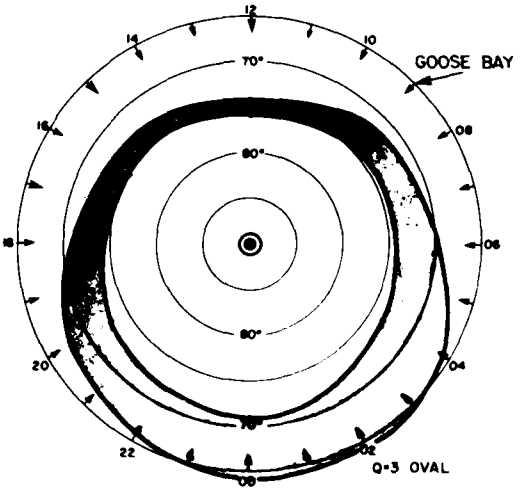


Figure 1 shows the auroral oval in a CG latitude-local time coordinate system. Goose Bay, located at 65° CG latitude, becomes an auroral oval station from 22:30 to 03:30 CGLT. During the other hours of darkness it is an F-layer trough station while in daytime it is between 500 to 1000 km to the south of the equatorward edge of the oval. While under the auroral oval, vertical incidence soundings, and riometer and magnetometer measurements allow the assessment of the state of the auroral ionosphere and the determination of the onset time and severity

of auroral disturbances. In the afternoon and evening hours, backscatter soundings directed towards the oval edge show strong irregularities which are associated with the aurora (Wagner and Pike, 1972) and thus permit the determination of the oval diameter many hours prior to the time, the oval moves over the station. In the early morning and pre-midday hours, Goose Bay is under the influence of the previously mentioned eastward drifting D-region enhancements associated with auroral substorms and monitoring of the existence of auroral disturbances in the midnight sector is possible for many hours.

Based on the evolving monitoring concepts the Air Force Geophysics Laboratory (AFGL) established in 1971 the Goose Bay Ionospheric Observatory. The station was equipped with riometers, magnetometers, satellite receivers for total electron content measurements, and a Digisonde 128 alternatingly connected to two separate transmitting antennas for vertical incidence and backscatter soundings. The Digisonde is a step frequency sounder with a 0.25 to 16 MHz frequency range. Phase coding and coherent integration are used to increase the signal-to-noise ratio. Envelope detection, i.e. power integration, is applied for backscatter soundings, since scatter type reflections or reflections from moving ionization fronts result in random or continuous (Doppler shift) phase changes, respectively, incompatible with a coherent integration scheme.

The two logarithmically compressed quadrature samples are integrated in 128 equidistantly spaced range bins (with selectable ΔZ) determining the amplitude with 6 bit resolution and the phase with 3 bits. The data are recorded on digital magnetic tape and, using an optically weighted font, are printed out on paper. Using available telecopier techniques the digital ionograms can be transmitted via telephone line to any other location for real time assessment of ionospheric conditions at Goose Bay.

Figure 2 shows in its left part a typical daytime ionogram obtained by the Goose Bay Digisonde. For each 1 MHz frequency band, shown between two frequency marks, 2160 characters of amplitude, phase and housekeeping (date-time-frequency, etc.) information is recorded on magnetic tape. The presently standardized 0-10 MHz ionogram requires the storage of 21600 characters. Transmission via telephone line has been accomplished requiring 2 seconds per frequency for a total of 3 minutes 20 seconds for the ionogram shown.

THE GEOMONITOR

In order to reduce the requirements for tape, prepare the data for automatic analysis and limit the amount of data to be transferred by data links to remote users, digital techniques of echo recognition were developed and implemented in the Geomonitor. Microprocessor controlled special electronic circuits achieve the high speed required for the simultaneous processing and display of the digital ionograms and the continuous data stream from various geophysical sensors.

For the processing of the ionograms initially the noise level on each frequency is established. Strong spread F or auroral Es echoes could lead to an overestimate of the noise level, thus suppressing weaker echoes. To circumvent this problem the noise level is determined separately for the lower and the upper 64 height bins by calculating the amplitude distributions for both ranges; the distribution which peaks at a lower amplitude is used to determine the noise level. The noise threshold is set at the positive half point of the distribution and only amplitudes larger than this level are considered as possible echoes. The noise threshold itself is recorded for reference.

The echo detection algorithm scans the 128 height bins on each frequency and detects up to six echoes, two from the E-region (<156 km) and four from the F-region. The selection of echoes is by average pulse amplitude, separately for the two regions. For two different echoes to be recognized as such they must be separated by a dip in amplitude down to the noise threshold. The echo spread is an important parameter in the assessment of the degree of disturbances, and it strongly influences the determination of ionospheric parameters such as foE and foF2. The Geomonitor determines and records, therefore, the spread of the main E and the main F echo.

The accurate heights of the leading edges of all six echoes is determined by sliding a standard pulse along an array of amplitudes with twice the resolution of the initial height steps. The exact virtual height of the echo is found at the position, where the average deviation (Bevington, 1969) between

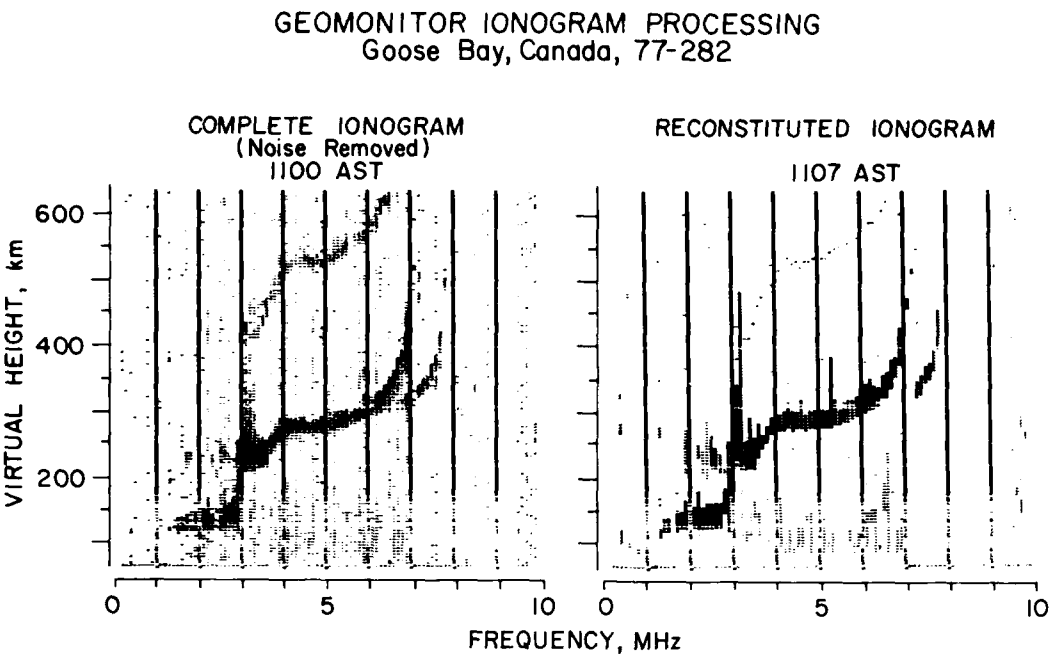


Figure 2. Geomonitor Ionogram Processing

standard pulse and (normalized) data is a minimum. The double-resolution array is generated by linear amplitude interpolation between two adjacent original height bins.

Finally, to investigate the dynamic height variations of the ionospheric layers, the Geomonitor calculates the so-called integrated heights. For each of the 128 height-bins the amplitudes are summed over all frequencies and normalized. This presentation was especially developed for use in the presentation of ionospheric characteristics to be discussed later.

The complement of data, with the addition of the gain setting on each frequency and the housekeeping information, for a standardized total of 128 frequencies, is collated or formatted into one digital record of 2340 characters, a reduction to about 10% of the original data. Evaluations have shown, that the chosen parameters and the number of echoes adequately depict an original ionogram for all routine and most special investigation purposes even under very disturbed conditions. The data are in a format perfectly suited for further processing, as the online determination of critical frequencies, the minimum frequency, layer heights and true height analysis. All these parameters can be determined using minimum computational effort. The development of the techniques to be applied is underway and will be concluded within the near future.

For verification of proper performance of the algorithms the collated ionograms described above are recast into the format of a regular Digisonde ionogram and presented on various displays. Shown in the right section of Figure 2 is such a "reconstituted" ionogram, taken 7 minutes after the one on the left, which is an unprocessed "cleaned" ionogram with amplitudes smaller than the noise level (see above) removed. The example is typical for the performance of the presently used algorithms. The high quality of the reconstituted ionogram is evident, minor problems as raggedness of the spread width will be corrected by refining the respective algorithms.

Sandwiched between the collated ionograms are the so-called "Geophysical Data", also **packaged** in a 2340 character long record representing 5 minutes of data. Analog signals from the three magnetometer channels, two riometers, satellite signals for Total Electron Content (TEC) and amplitude scintillation measurements are digitized and recorded. Additional analog

data up to a total of 15 can be added as required. Digitization rates presently used are 2 Hz for two scintillation channels and 0.1 Hz for the remaining 13 channels.

In its present operation the Geomonitor produces every five minutes one record of geophysical data and two records for two ionograms of various types (vertical, backscatter, coherent or power integration) as selected at the Digisonde. Reduced operation as, for example, quarter hourly ionograms only is readily possible. The data load of 3 x 2340 or 7020 characters/5 min can be handled by available 600-1200 baud telecommunications links.

DATA PRESENTATION

An important requirement for the analysis of multi-dimensional data under all conditions, but especially important for real time applications, is intelligent presentation that reveals the characteristics of the data set. In the case of ionogram observation, it is the original ionogram itself that gives the scientist the most detailed answer in regard to the momentary conditions of the ionosphere. It does not show, however, the time development of the ionospheric parameters, i.e. the diurnal variations and disturbances, unless a sequence of ionograms is studied simultaneously. Analog methods to present characteristic ionospheric parameters as a function of time, like the critical frequencies of E and F-region and the layer heights, were developed in the fifties (Nakata et al. 1953; Bibl, 1956). Since 1969 Digisondes have produced digital ionograms and we have developed computerized techniques to generate digital characteristics. Use of microprocessor technology in the Geomonitor has made it possible in Goose Bay to print out digital ionospheric characteristics in real time. To extend the usefulness of the display techniques to other users, a Geomonitor Display System which is close to completion, can take the Geomonitor data stream, as arriving via any suitable data link, and present these data in the form of characteristics, and as reconstituted ionograms at any remote site.

Figure 3 is a typical example of three selected characteristics. Shown on top are the integrated heights and ranges derived from backscatter ionograms, below, the F-region characteristic and at the very bottom the E-region characteristic of the vertical incidence ionograms. Each ionogram produces one line in each characteristic.

The height or range characteristic is produced by simply printing out the 128 characters available in each ionogram record, which result from the summation of amplitudes over all frequencies for each height bin separately. Since the ionogram traces are almost horizontal at the minimum heights of the respective layers, the summation results in large numbers at these heights. Similar properties of oblique, oval associated backscatter echoes, result in strong integrated amplitudes at the range of these echoes. Range or height changes with time are thus visible at a glance in this characteristic.

The F-layer characteristic is obtained by sequentially printing the largest of the four F-echoes amplitudes into the respective frequency bin. The presentation thus obtained is an amplitude and frequency vs. time history and, as time progresses, reflects the changes of upper and lower limit of the frequency band reflected from the F-layer. The lower limit is generally a function of occulation by the lower E-layer or of enhanced absorption. The upper limit and its time variation is in daytime a direct measure of the foF2 (the Digisonde uses circularly polarized antennas that suppress the extraordinary component). Spread-F conditions routinely observed at night at Goose Bay make the upper limit a more complicated parameter of the degree of F-region disturbance.

The E-layer characteristic shown in the lowest panel is obtained in an identical fashion by printing the larger amplitude of the two E-echoes into the respective frequency bin. The upper limit of the band of strong amplitudes is equal to foE in those time sectors, where the cosinusoidal time variation shows the solar zenith angle control of the E-ionization. At other times, sharp spikes or deviations from a smooth daytime E-ionization indicate the presence of Es with ftEs > foE.

The daily variations of foF2, foE, h'F and backscatter characteristics for the Goose Bay ionosphere are evident on these hard copy time histories. Comparison with variations of the previous days and with easily produced average curves allows the assessment of current trends as for example enhanced or depressed foF2. The range of auroral backscatter at a given hour in the afternoon or evening can be converted

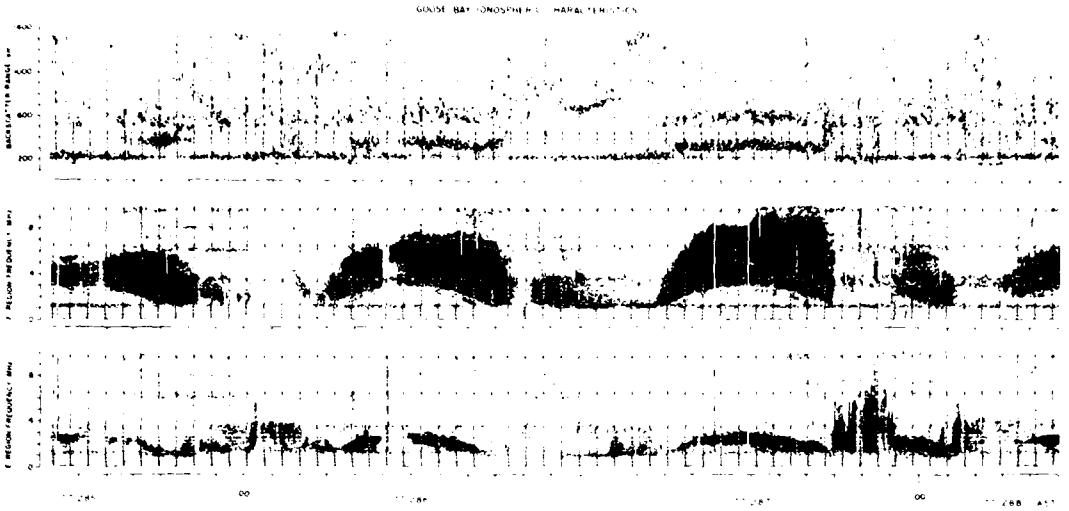


Figure 3. Goose Bay Ionospheric Characteristics

into a specific Q-value. Well established knowledge of oval continuity (Buchau et al, 1970) and the prediction of the circumpolar location of discrete auroral forms (Meng, 1977) from a locally established Q-value makes these backscatter range determinations a powerful measurement. The "anchor point" technique of using a one-point measurement to predict the whole oval has been developed by Gassmann (1973) and is at present successfully used at AFCWC with satellite auroral images providing the input parameter Q_E, the equatorward oval boundary. Of special importance is the ability, to interpret various types of signatures in the characteristics as evidence of auroral oval and ionospheric disturbances. Examples of such signatures are increased minimum frequency resulting from an increase in D-region electron density, the time history of such an increase, the existence of total absorption, sporadic E events, strong spread F occurrence, the sudden depression of foF2, and the early appearance or the rapid advance of backscatter fronts.

The parallel display of properly formatted Geophysical Data provides insight into the nature of the disturbance from different points of view and has to be made an integral part of the assessment of the event prior to issuing of an event report. As the disturbance grows in intensity, ionospheric characteristics become more difficult to interpret either due to the general lack of clear patterns found especially during magnetic storms or due to the strong increase in auroral absorption resulting in total lack of ionospheric echoes. Under these conditions the measurements most important to the preparation of advisories are the riometer and the magnetometer measurements.

The following discussion of the possible real time use of the Geomonitor and Display System in support of an Over-the-Horizon Backscatter Radar (OTH) System operating in this environment demonstrates the approach. The discussion also shows on examples the signatures of specific events and the value of the data compression and characteristic presentation to the investigation of ionospheric phenomena in general.

Let us consider the OTH radar located at a midlatitude location with Goose Bay within the surveillance area. (This is actually the situation of the planned 414L CONUS OTH-B Experimental Radar System, which is scheduled to start tests in 1980 and which will receive Goose Bay Ionospheric Data.) The latitude dependent ionospheric features associated with the auroral ionosphere such as the F-layer trough and the trough wall location, and the time dependent disturbances, such as absorption and sporadic E all have strong impacts on the propagation situation. Limitation of coverage, multipath to one target and azimuth/range errors due to the large gradients are but a few of the related problems.

Assume a station like GBIO with a Geomonitor connected by a communication link to the Geomonitor Display System located in the radar system operations

center. The GDS then provides a comprehensive ionospheric and geophysical data display for use by the radar's ionospheric forecaster. The forecaster assists the radar operators in the frequency management and in determining whether and how the radar is being affected by natural disturbances over part or all of the surveillance area. He also estimates the probability of the disturbance subsiding or continuing. Figure 4 shows a vertical incidence reconstituted ionogram from GBIO at 1000 local time on day 286. An absorption condition exists and only fragments of echoes from the usual daytime layers are seen. More information is available from the characteristics in Figure 3. They show no absorption on this day prior to the event onset and during the previous night. Such time history is indicative of a localized absorption event drifting around from a nightside auroral substorm, and it is likely to be of short duration. The ionospheric forecaster would advise the operators to this effect. Looking at the characteristics at later times that day, we see that the absorption did indeed diminish, disappearing completely in 2 or 3 hours. The ionogram (Figure 5) shortly afterward at 1307 local time on day 286 shows the expected daytime ionospheric layers indicating normal radar system operation.

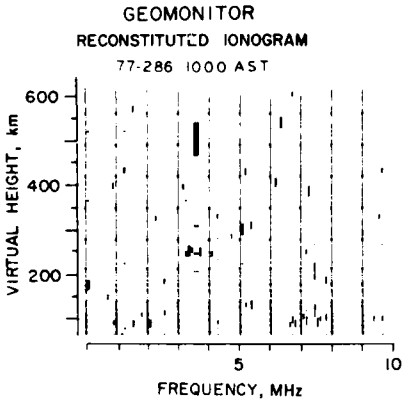


Figure 4. Geomonitor Reconstituted Ionogram 77-286 1000 AST

The vertical incidence ionogram at 0930 local time on day 292 is quite similar to Figure 4, showing complete absorption. The characteristics in Figure 6 show that the E layer echoes had been absorbed throughout the morning, and that absorption had occurred periodically during the previous night; a high level of disturbance is evident

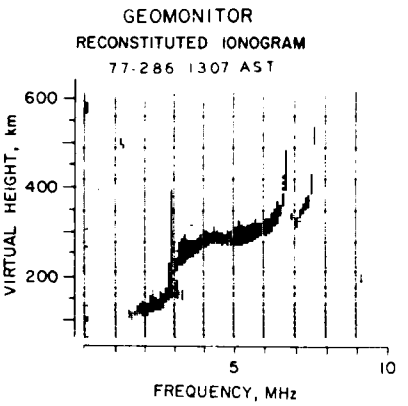


Figure 5. Geomonitor Reconstituted Ionogram 77-286 1307 AST

in the rapid changes in the Es top frequency. This suggests that disturbed conditions will continue with the resulting disturbing effects on the radar system. The ionospheric forecaster would advise the operator on expected future problems. This would permit enhancement of other available means of surveillance if extended outages could not be tolerated. The characteristics later in the day continue to show complete absorption as was inferred above. During short periods, when the absorption subsides, foF2 can be determined as 5.5 MHz, compared to 6.8 MHz on day 286 at the same time. This depression supports the assessment of a continuing disturbance, and aids the frequency management during times when propagation is possible.

The ionogram in Figure 7 shows a typical late afternoon ionosphere at 1637 AST on day 287 with an foF2 of 8.7 MHz. HF propagation would be by F-layer modes. Examination of the backscatter ionogram range characteristic shows that echoes which are associated with the auroral oval have been present as early as 1530 AST with decreasing range throughout the next hours. This indicates that the auroral oval was expanded and that the observatory and thus the OTH coverage area would rotate under the oval early in the evening. The ionogram at 1800 AST in Figure 8 verifies the prediction. The Es layer is well developed, the presence of an Es multiple indicates that absorption is not strong. Thus for selected azimuths, HF propagation would be by E layer modes only. Knowledge that the propagation has changed from F-layer to E-layer modes is impor-

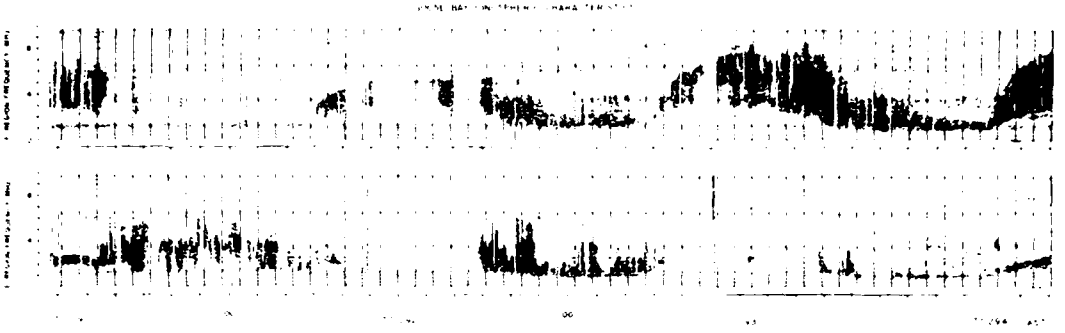


Figure 6. Goose Bay Ionospheric Characteristics, disturbed period

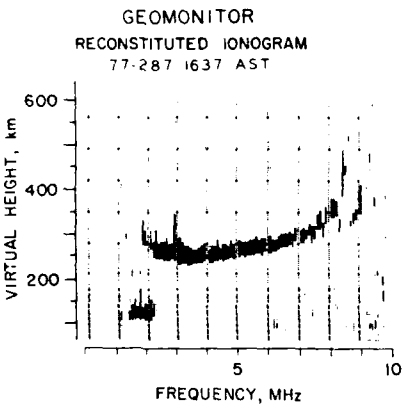


Figure 7. Geomonitor Reconstituted Ionogram 77-287 1637 AST

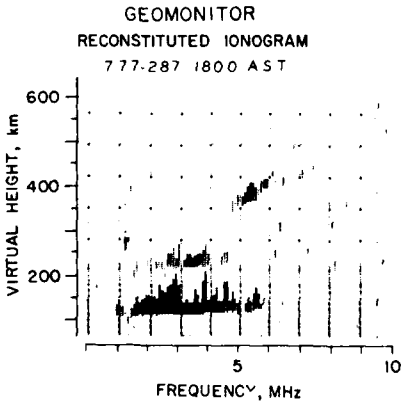


Figure 8. Geomonitor Reconstituted Ionogram 77-287 1800 AST

tant in interpreting the range of targets in the surveillance area. The characteristics show that the Es event continued for several hours. Figure 9 shows that the Es reached 10 MHz at 2052 AST allowing for the use of rather high surveillance frequencies but providing only short (<2200 km) range coverage in the general direction of Goose Bay.

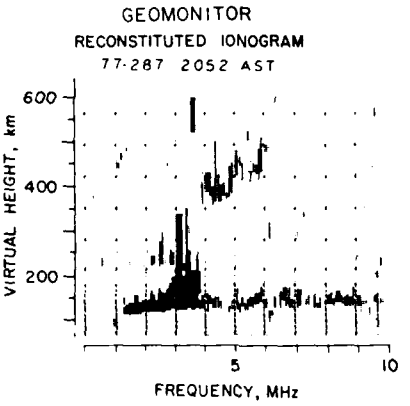


Figure 9. Geomonitor Reconstituted Ionogram 77-287 2052 AST

Another type of behavior of the ionospheric features associated with the auroral ovals occurred on the night of day 286. The backscatter range characteristic showed that the oval stayed well poleward of GBIO throughout the night. The ionogram at 2345 AST in Figure 10 showed the off-vertical echoes from the oval appearing in the vertical ionogram display. The foF2 at this time was approximately 2 MHz, indicating GBIO and thus a large part of the coverage area was in the midlati-

tude ionospheric trough. Knowledge that the auroral oval stayed poleward of Goose Bay and that GBIO was underneath the trough is informative in evaluating the ionosphere effecting the radar surveillance area equatorward of Goose Bay. The reconstituted ionogram at this time would be very important input to the coordinate conversion conducted by the radar to convert azimuth and range information into target ground coordinates, dependent on the reflecting height of the ray path.

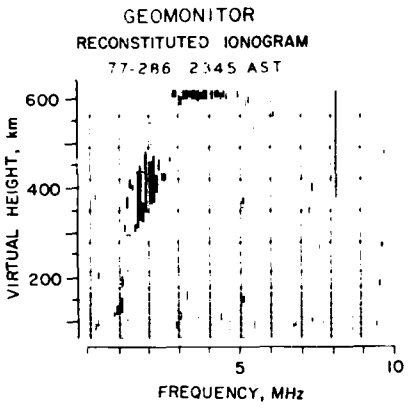


Figure 10. Geomonitor Reconstituted Ionogram 77-286 2345 AST

Notice the contrast in the size of the auroral oval on days 286 and 287. On 286 it remained poleward of GBIO all night, indicating a Q index value of 0. On 287 it arrived overhead at 1800, consistent with a Q index of 8.

SUMMARY

We have described an approach to the remote monitoring of the high latitude ionosphere, which is summarized in Figure 11. Digital vertical incidence and backscatter ionograms are reduced in the Geomonitor by extracting amplitude, height and spread of the ionospheric echoes. The algorithm works remarkably well even under disturbed conditions during an auroral event. Properly formatted digitized magnetometer, riometer and satellite propagation data and the processed ionospheric data are stored on tape, or for real time use, are transmitted over a suitable medium capacity data link to any desired user site. Here a Geomonitor Display System separates the data, presents them as characteristics or time histories of various ionospheric parameters and of the geophysical data for interpretation by an observer. The usefulness to an OTH system has been shown on selected samples. Quiet day curves for the various geophysical data and thresholds for minimum frequency and Es-events can be used in the Geomonitor and the Display System to initiate alarms when deviations of predetermined magnitude occur.

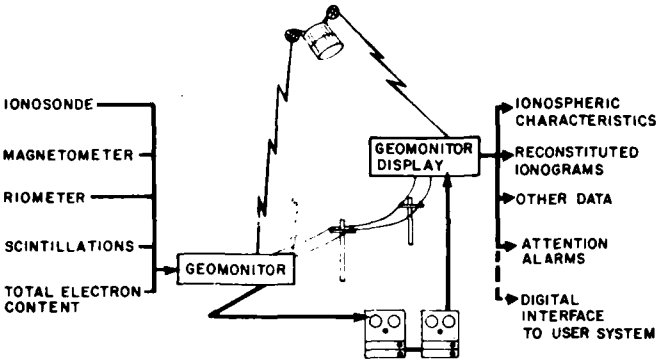


Figure 11. Schematic presentation of data requirements, data flow and display for remote oval monitoring.

The addition of two more monitor sites, one in the European Arctic and one in Alaska could provide inputs into AWS-SESS resulting in a 24 hour continuous coverage of all high latitude disturbances and the continuous determination of the location of the boundary between F-layer trough and auroral oval. Together with the sporadic DMSP satellite auroral images, complete and continuous understanding of the state of this important section of our environment can be achieved. The U. S. Army employs an instrument (Automatic Ionogram Collator) similar to the Geomonitor for their Digisonde operation in Fort Monmouth, N.J. (F. Gorman, private communication). This system could offer the possibility of a mid-latitude expansion of the arctic network.

ACKNOWLEDGEMENTS

The technical support by Messrs. J. B. Watrous and E. W. Jewell during the establishment of the Goose Bay Oval Monitoring facility is thankfully acknowledged.

REFERENCES

Bevington, P. R., 1956, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill.

Bibl, K., 1956, J. Atmos. Terr. Phys., 8, p. 295.

Bibl, K., J. A. Patenaude and B. W. Reinisch, 1970, Digital Integrating Goniometric Ionospheric Sounder Digisonde 128, Final Report, U. of Lowell, AFCRL-71-0002.

Bibl, K., B. W. Reinisch and S. Smith, 1976, Digital Ionospheric Sounding in Support of Arctic Research, Final Report, U. of Lowell, AFGL-TR-76-0037.

Bibl, K. and B. W. Reinisch, 1978, The Universal Digital Ionosonde, Radio Science (accepted July 1977).

Buchau, J., J. A. Whalen and S. I. Akasofu, 1970, On the Continuity of the Auroral Oval, J. Geophys. Res., 75, P. 7147.

Buchau, J., G. J. Gassmann, C. P. Pike, R. A. Wagner and J. A. Whalen, 1972, Precipitation Patterns in the Arctic Ionosphere Determined from Airborne Observations, Ann. Geophys., 28, 2, pp. 443-453.

Driatskiy, V. M., 1968, Diurnal Pattern of Auroral Absorption in the Auroral Zone, Geomagnetism and Aeronomy, VIII, 1, p. 33.

Elkins, T. J., 1972, A Model of Auroral Substorm Absorption, AFCRL-72-0413, Environmental Research Papers, No. 404.

Feldstein, Y. I. and G. V. Starkov, 1967, Dynamics of Auroral Belt and Polar Geomagnetic Disturbances, Planet. Space Sci., 15, pp. 209-229.

Gassmann, G. J., 1973, Analog Model 1972 of the Arctic Ionosphere, AFCRL-TR-73-0151, Air Force Surveys in Geophysics, No. 259.

Ming, C. I., R. H. Holzworth and S. I. Akasofu, 1977, Auroral Circle-Delineating the Polarward Boundary of the Quiet Auroral Belt, J. Geophys. Res., 82, 1, pp. 164-172.

Muldrew, D. B., 1965, F-Layer Ionization Troughs Deduced from Alouette Data, J. Geophys. Res., 70, pp. 2600-2635.

Nakata, Y., M. Kan and H. Kyeda, 1953, Rep. on Ion. Res., Japan, pp. 129-135.

Pike, C. P., 1971, Latitudinal Survey of the Daytime Polar F-Layer, J. Geophys. Res., 76, 31, pp. 7745-7753.

Reinisch, B. W. and S. Smith, 1976, Geomonitor, Digital Real Time Processor for Geophysical Data, AFGL Technical Report TR-76-0292.

Wagner, R. A. and C. P. Pike, 1972, A Discussion of Arctic Ionograms, AGARD-CP-97, AGARD Conference Proceedings "Radar Propagation in the Arctic" held at Lindau/Germany September 1971.

Whalen, J. A., J. Buchau and R. A. Wagner, 1971, Airborne Ionospheric and Optical Measurements of Noontime Aurora, J.A.T.P., 33, pp. 661-678.

ATTEMPTS TO PREDICT TROUGH/PLASMAPAUSE BOUNDARIES IN REAL TIME

M. Mendillo, C. Chacko, and F. Lynch
Department of Astronomy
Boston University
Boston, MA 02215

and

P. J. L. Wildman
Air Force Geophysics Laboratory
L. G. Hanscom AFB
Bedford, MA 01730

1. INTRODUCTION

The nighttime ionospheric trough found near L=4 marks the transition region between the mid-latitude and auroral ionospheres. The latitudinal gradients associated with the trough (particularly on the poleward side) are severe plasma boundaries which impact F-region supported communications systems. Past statistical relations have suggested a close connection between the trough's position (L_T) and the equatorial plasmapause location (L_{pp}). While these relationships generally use local time and the geomagnetic index Kp to specify conditions described, the relatively large uncertainties associated with the correlations have never really suggested that the regression equations could be used as predictors of actual boundary locations.

This paper describes two new approaches for specifying trough and plasmapause locations using satellite-borne auroral photographs and computer-simulated magnetospheric convection patterns. The availability of USAF/DMSP auroral photographs in near real time suggests the possibility of locating the poleward wall of the trough over an extensive geographic area by observing the equatorward edge of the continuous (diffuse) aurora. The concept is tested with case study events employing simultaneous DMSP and ISIS-2 satellite passes in conjunction with a ground-based ionosonde network and a meridianal chain of electron content observatories.

A second scheme tested involves computer modelling of equatorial plasmapause dynamics, initially in an attempt to reproduce satellite probe determinations of the trough/plasmapause location in the topside F-region (h≈800-3200 km). A magnetospheric electric field model keyed to the Kp index is used to compute plasmapause dynamics in the night sector as a way of specifying temporal changes in the L-value of the plasmapause during periods of increasing geomagnetic activity.

2. RELATIONSHIP BETWEEN THE AURORA AND THE TROUGH

The first part of this paper concerns an examination of the detailed relationships between the poleward wall of the F-region trough and the equatorward edge of the continuous (or diffuse aurora). Energetic particle precipitations are known to cause both features, and thus the question arises of whether or not detailed knowledge of the latitudinal pattern of one can be used to predict the detailed structure of the other. The notion of specifying optical auroral features from ionospheric data is not nearly as appealing as the reverse possibility, i.e., using a single optical photograph (obtained during a conventional DMSP or ISIS satellite pass) to specify the location of an important F-region structure over the extensive geographic area covered in the photograph. Since topside sounders or single-height, in-situ satellite probes provide the only routine way of

specifying trough features (along a given satellite track), the concept of using "optical diagnostics" from a single satellite pass to infer F-region structure over a region much more extensive than the sub-satellite track implies a capability broadly equivalent to the unrealistic scheme of having many simultaneous ionospheric sounding satellites pass through a given world region (see Figure 1).

In examining trough morphology, the poleward wall is generally found to be the best-defined feature of all trough characteristics (Pike 1976, Mendillo and Chacko, 1977). Thus, if a consistent relationship exists between the auroral boundary and the poleward wall, and DMSP-type satellites can fix the latitude of the wall over an extensive region, this key reference point can be used to anchor trough models giving latitudinal and height-dependent structure (Rycroft and Thomas, 1970; Feinblum and Horan, 1973; Halcrow, 1976; Mendillo and Chacko, 1977).

We tested this scheme by carrying out ten "case study" investigations using ISIS-2 topside sounder passes in December 1971 for which DMSP auroral photographs were available in near simultaneity. For all ten cases, the temporal separation between the ISIS and DMSP satellite passes fell within the 8 to 40 minute range. The ISIS data were used to determine the N_e latitude gradients (in CGL) at h_{max} near the midnight meridian. Since each ISIS pass crossed (or came very close to) the equatorward edge (ϕ_a) of the continuous aurora, the latitude separation ($\Delta\phi$) between the poleward wall (at ϕ_{pw}) and the aurora could be determined at the longitude or local time (2230-2400 MLT) of the ISIS pass. For ϕ_{pw} corresponding to the foot of the poleward wall, $\Delta\phi \simeq 4.7^{\circ} \pm 1.2^{\circ}$, while for ϕ_{pw} corresponding to the top of the poleward wall, $\Delta\phi = 2.4^{\circ} \pm 1.1^{\circ}$. Apart from the observation that the auroral boundary (via DMSP) is always found poleward of the trough boundary, no systematic relationship between $\Delta\phi$ and magnetic activity (using Kp or AE, for example) could be found. The number of cases examined (ten) was of course small, and the spectral range of the DMSP detector is not ideally suited for high resolution studies. We do, however, believe that the separations observed between the diffuse aurora (as monitored by DMSP) and the poleward wall of the main electron density trough are physically significant quantities.

Given the observed $\Delta\phi$ at the ISIS sub-satellite location, we were interested in determining the longitudinal (or local time) consistency of the separation over the region covered by the DMSP photograph, and indeed beyond. This is not a simple question to answer since no F-region monitoring technique can provide ISIS-type latitude resolution at points to the east and west of a given ISIS pass. We turned to the ground-based network of ionosonde sites (via WDC-A) and to the AFGL chain of total electron content (TEC) observing stations (courtesy of J. A. Klobuchar) for supporting data to estimate the spatial consistency of the trough's location away from the ISIS longitude. Figure 2 shows the network of 24 ionosondes and the 4 TEC sites used in the study.

Two of the case studies are presented in Figures 3 and 4. For the 06UT period on 9 December 1971 (Figure 3), the ISIS pass crosses the DMSP image at $\phi_a \simeq 68^{\circ}$ CGL, with the foot of the trough's poleward wall in foF2 at $\phi_{pw} \simeq 62^{\circ}$ CGL. The ionosonde stations immediately to the east show foF2 $\simeq$ 5MHz to the north of ϕ_a and foF2 $\simeq$ 2MHz equatorward of ϕ_{pw}. Looking further to the east, ionosondes poleward of ϕ_a give foF2 in the 5-6 MHz, again typical of poleward wall values, while sites equatorward of ϕ_{pw} show foF2 values more typical of the trough minimum (i.e., foF2 $\simeq$ 2-3 MHz). The TEC network covers North American, Atlantic Coast longitudes from 73°-53° CGL. The trough in TEC has its minimum near 60°, with 3×10^{12} el/cm^2 consistent with a 2MHz peak. In fixing the trough minimum at $\simeq$ 2MHz near 60°, no inconsistency is found within the DMSP field of view, and indeed as far to the east as the western European foF2 stations in the pre-dawn sector.

In Figure 4, the case of 0600UT on 21 December 1971 is depicted. The ISIS pass occurs several degrees to the west of the DMSP field of view. Once again, the foot of the poleward wall is nearly 5° equatorward of ϕ_a. In the North American sector, the ionosonde and TEC data support a local time consistency of $\Delta\phi \simeq 5^{\circ}$ within (and several hours to the east of) the DMSP field of view. A trough minimum of foF2 $\simeq$ 2MHz near 60° CGL may again be inferred from the satellite track to the west European sector.

In summary, a limited set of "case study" investigations points to the possibility of inferring real-time F-region trough morphology from DMSP auroral images, at least over the longitude span covered by a photograph,

and probably to several hours beyond it.

3. SIMULATION OF PLASMAPAUSE DYNAMICS

Most of the fundamental theoretical studies of the origin and behavior of the equatorial plasmapause (e.g., Nishida, 1966; Brice, 1967; Kavanagh et al., 1968) deal with the application of magnetospheric convection patterns originally suggested by Axford and Hines (1961) to the now classic plasmapause geometry reported by Carpenter (1966). The notion of a contracting plasmapause due to enhanced convection patterns during periods of increased geomagnetic activity is now a widely accepted view, and is, in fact, the scheme generally said to explain why statistical results for the plasmapause location (L_{pp}) vs. an index like Kp seem to provide reasonable descriptions of average behavior (Rycroft and Thomas, 1970; Kohnlein and Raitt, 1977). Yet, these simple statistical patterns of L_{pp} vs. Kp are not expected to provide an actual description of the time dependence of plasmapause changes during a specific storm's Kp-scenario. The real plasmapause cannot adjust instantaneously to a new configuration for each Kp step, and within a given Kp interval, the L_{pp} position must be a function of previous magnetic history.

In an attempt to overcome these two limitations, we decided to use a very simple model for magnetospheric convection keyed to magnetic activity to see if a time-dependent representation for plasmapause distortions in the night sector could be obtained. The convection model is limited to E x B motions in the equatorial plane, using a centered dipole magnetic field and a spatially uniform dawn-dusk magnetospheric electric field. For any given particle in the equatorial plane (specified by a LT and L-value), the instantaneous motion is given by $W = (ExB)/B^2$, where E is the resultant electric field arising from co-rotation (E_R, directed radially inward) and convection (E_{DD}, directed from dawn to dusk), and B is the dipole field strength. A "fluid" of particles is then followed by keeping track of all the individual motions using a 10-minute time step. The initial distribution of particles was taken to be circular, with a particle at each ½ L-value from L = 1.5 to 5.5, and spaced every 15 minutes of local time. A total of 864 individual points were therefore available to define subsequent transformations (see Figure 5a).

For a given dawn-dusk electric field, a unique plasmapause exists (Kavanagh et al., 1968; Chen, 1970); it defines the boundary between plasma particles which share in the co-rotational field associated with the Earth and those which are lost due to magnetospheric convection. The shape of the resultant plasmapause for such a case is the familiar "tear-drop" described by Chen (1970). Thus, our initial circular distribution of particles becomes distorted into a "tear-drop" configuration due to the E_{DD} field (see Figure 5b). It typically takes about 30 hours of simulation time for a completely stable configuration to develop, i.e., of the original 864 test particles 50-75% are trapped in motion about the Earth in a region defined by the appropriate theoretical "tear-drop" plasmapause associated with E_{DD}. Given this equilibrium plasmapause, we are interested in subsequent transformations caused by a stronger E_{DD} field.

If E_{DD} is increased, the new equilibrium plasmapause position is known in advance since it depends only on the new field strength (Chen, 1970). Our calculations show that it can take typically 8 hours for the new plasmapause to become stable on the nightside. Since geomagnetic activity changes occur on a much smaller time scale, a storm-time scenario for L_{pp} cannot be pictured as a collection of equilibrium configurations.

We decided to by-pass the equilibrium configuration problem by using a dynamic calculation scheme in which the fluid of convecting particles experiences a time-dependent E_{DD} field, and instantaneous L_{pp} (LT,t) locations are obtained by graphical inspection of the fluid's boundary. In an earlier work (Mendillo and Papagiannis, 1971), we developed a model for E_{DD} which showed a quadratic dependence upon Kp. We normalized this model to the "tear-drop plasmapause" which gave $L_{pp}(00LT) = 5$ for Kp = 2 (for which $E_{DD} = 0.14$mv/m). Using our quadratic relation between E_{DD} and Kp, a set of "best estimate" E_{DD}values were now available to describe magnetospheric convection/plasmapause scenarios keyed to observed Kp variations.

To illustrate the procedure, consider the storm period of 21-22 April, 1971. Prior to 1200 UT, magnetic activity was steady for nearly 9 hours, with $\overline{Kp} = 2$. From 1200 UT on, Kp (a 3-hour index) varied as follows: 3, 4, 5, 7, 5, 3, 2, . To follow the nightside L_{pp} variations associated with

this scenario, we first used our simulation program to generate the equilibrium ("tear-drop") particle distribution appropriate to the t=0 point of the scenario. Thus, the initial circular distribution of particles was allowed to evolve to the case: 12:00 UT on 21 April 1971, Kp = 2 (Figure 5b). The electric field model was then used to input new E_{DD} fields in concert with the observed Kp variation. Computer generated plots of the modified particle distributions were made at hourly intervals, and the L_{pp} boundary values for the LT sector in question were scaled from the plots (see Figure 6 for sample outputs 2, 9, 12, and 20 hours into the simulation). For this particular scenario, we examined the 2345 LT values in order to compare the model's predictions with ISIS-1 in-situ probe data for the equatorward edge of the total ion trough in the topside F-region (Wildman et al., 1976). The results are given in Figure 7, together with the "statistical predictions" of the L-value of the plasmapause using the relation given by Rycroft and Thomas (1970).

In examining the results, there are several features to note:
(1) The trough's equatorward edge (measured high in the F-region) is probably a better indicator of the plasmapause location during disturbed times than during quiet periods. Thus, at the t=0 point of the scenario, the difference between the data and the model is the largest.
(2) As the storm progresses (in this case, rather smoothly from Kp = 3→4→5→7), the steady contraction of the plasmapause is reproduced. The statistical results of Rycroft and Thomas (1970) appear as discontinuous jumps every 3 hours (the Kp interval); they also tend to agree with the observed behavior.
(3) Following the storm maximum, i.e., after 00LT on 22 April 1971, the Kp transition 5→3→2 causes a gradual recovery of the plasmapause to higher L-values. There is a definite "sluggishness" built into the model arising from the previously high Kp values. This ability to take previous magnetic history into account is in sharp contrast to the statistical predictions which slow instantaneous adjustments to higher L_{pp} values.

We have tested this procedure for other Kp scenarios and find the "sluggishness" of the model to be a major asset for both intensification effects (e.g., Kp = 1→5) and recovery patterns. In summary, then, a very simple physical model for a convection-dominated plasmapause can be used to simulate nightside dynamics effects in near real-time.

Acknowledgements

We wish to thank the World Data Center A in Boulder for providing the ionosonde values used in this study. Mr. Louis York's efforts with the convection program were greatly appreciated. TEC data were kindly provided by J. A. Klobuchar of AFGL, and we benefited by discussions with L. Snyder, J. Whelan, E. Weber and R. Sagalyn. This work was supported in part by AFGL contracts F-19628-75-C-0044 and F-19628-77-C-0019 to Boston University.

REFERENCES

Axford, W. I. and C. O. Hines, A Unifying Theory of High-Latitude Geophysical Phenomena and Geomagnetic Storms, Can. J. Phys., 39, 1433, 1961.

Brice, N. M., Bulk Motion of the Magnetosphere, J. Geophys. Res., 72, 5193, 1967.

Carpenter, D. L., Whistler Studies of the Plasmapause in the Magnetosphere, 1, Temporal Variations in the Position of the Knee and Some Evidence on Plasma Motions Near the Knee, J. Geophys. Res., 75, 693, 1966.

Chacko, C. and Michael Mendillo, Electron Density Enhancements in the F-Region Beneath the Magnetospheric Cusp, J. Geophys. Res. (November, 1977).

Chen, A. J., Thermal and Quasi-Energetic Plasma Flow in the Magnetosphere-A Theoretical Study, Ph. D. Dissertation, Rice Univ., Houston, Texas, 1970.

Feinblum, D. A. and R. J. Horan, HILON-A Model of the High Latitude Ionospheric F2 Layer and Statistics of Regular Ionospheric Effects at Ft. Churchill, 1968. Bell Labs., Murray Hill, N. J., 1973.

Halcrow, Barry W., F2 Peak Electron Densities in the Main Trough Region of the Ionosphere, Tech. Rept. PSU-IRL-IR-55, IOn. Res. Lab., Penn. State Univ., University Park, PA, 1976.

Kavanagh, Jr., L. D., Freeman, Jr., S. W. and A. J. Chen, Plasma Flow in the Magnetosphere, J. Geophys. Res., 73,

5511, 1968.

Kohnlein, W, and W. J. Raitt, Position of the Mid-Latitude Trough in the Topside Ionosphere As Deduced From ESRO.4 Observations, Planet. Space Sci., 25, 600, 1977.

Mendillo, Michael and C.C. Chacko, The Base Level Electron Density Trough, J. Geophys. Res. (November, 1977).

Nishida, A., Formation of Plasmapause, or Magnetospheric Plasma Knee, By the Combined Action of Magnetospheric Convection and Plasma Escape From the Tail, J. Geophys. Res., 71,5669, 1966.

Pike, C. P., An Analytical Model of the Main F-Layer Trough, AFGL-TR-76-0098, May, 1976.

Rycroft, M. J. and J. O. Thomas, The Magnetospheric Plasmapause and the Electron Density Trough at the Alouette I Orbit, Planet. Space Sci., 18, 65, 1970.

Wildman, P.J.L., Sagalyn, R.C. and M. Ahmed, Structure and Morphology of the Main Plasma Trough in the Topside Ionosphere, Proc. COSPAR Sym. Geophys. Use of Sat. Beacon Obs., M. Mendillo (ed), Boston University, Boston, MA, June 1976.

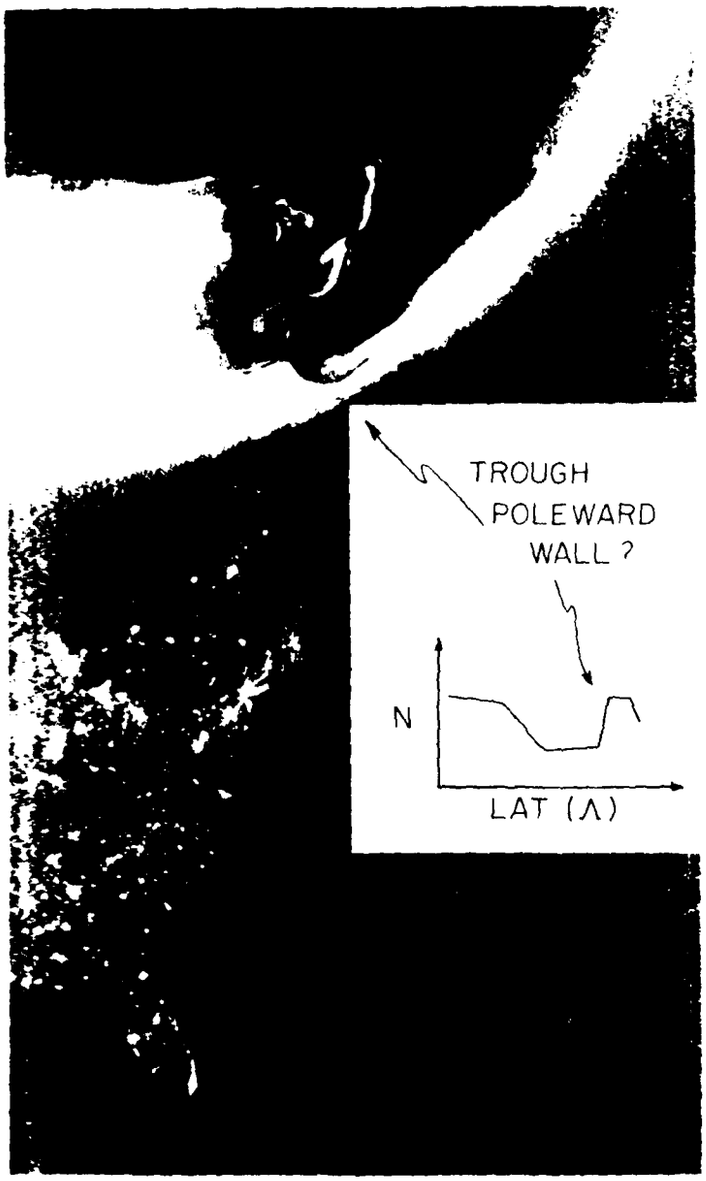


FIGURE 1. Sample DMSP auroral photograph illustrating possible relationship between the equatorward edge of the aurora and the poleward wall of the ionospheric trough.

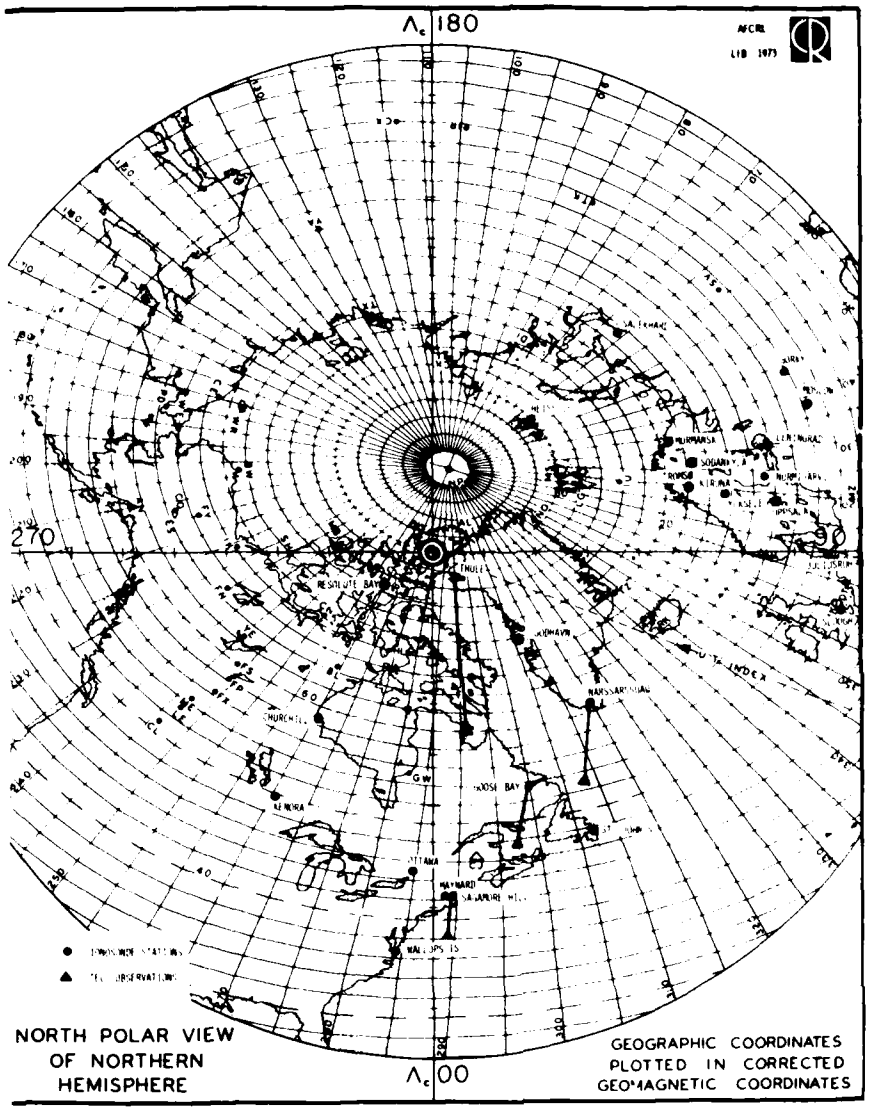


FIGURE 2. Polar view of northern hemisphere with geographic coordinates plotted in Corrected Geomagnetic Coordinates. Ionosonde stations (•) and TEC observing sites (▲) used in "case study" events are indicated. Corrected Geomagnetic Latitudes (CGL) appear as concentric circles on subsequent figures.

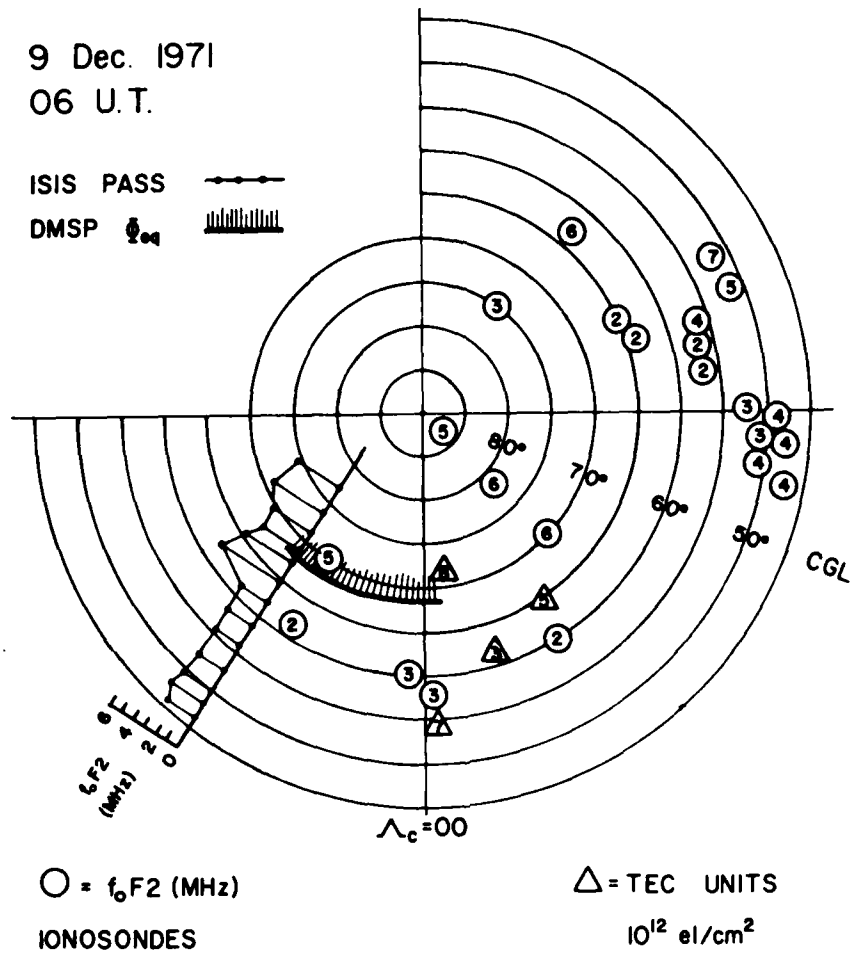


FIGURE 3. North polar view in CGL showing the location of near simultaneous ISIS and DMSf satellite data for 0600 UT on 9 December 1971. Ground-based foF2 values and TEC data at 0600 UT from the network shown in Figure 2 are presented using integer values.

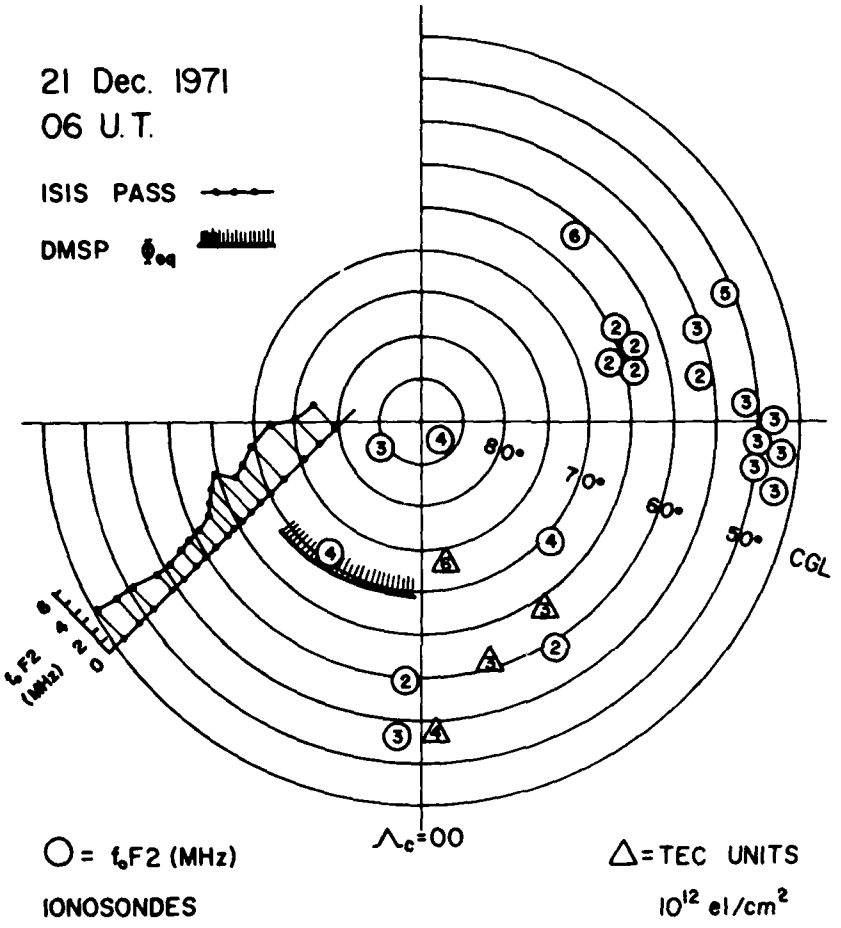


FIGURE 4. North polar view in CGL showing the location of near simultaneous ISIS and DMSP satellite data for 0600 UT on 21 December 1971. Ground-based foF2 values and TEC data at 0600 UT from the network shown in Figure 2 are presented using integer values.

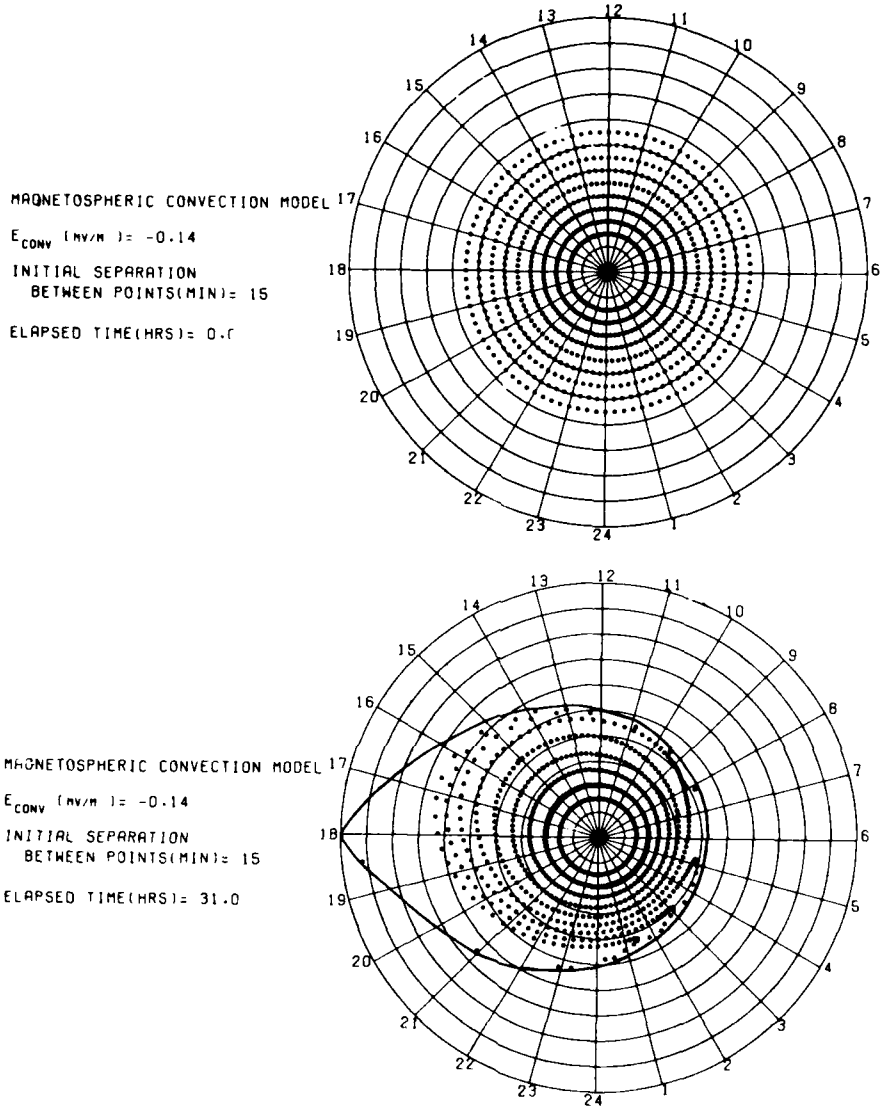


FIGURE 5. (a)Top. Magnetospheric convection model t=0 configuration for transition to a K_p = 2 equilibrium plasmapause (E_{DD} = 0.14 mv/m).

(b)Bottom. Resultant "tear drop" plasmapause for K_p = 2 after 31 hours simulation time. This equilibrium boundary represents the t=0 point for a K_p scenario beginning with K_p = 2 (e.g., at 1200 UT on 21 April 1971).

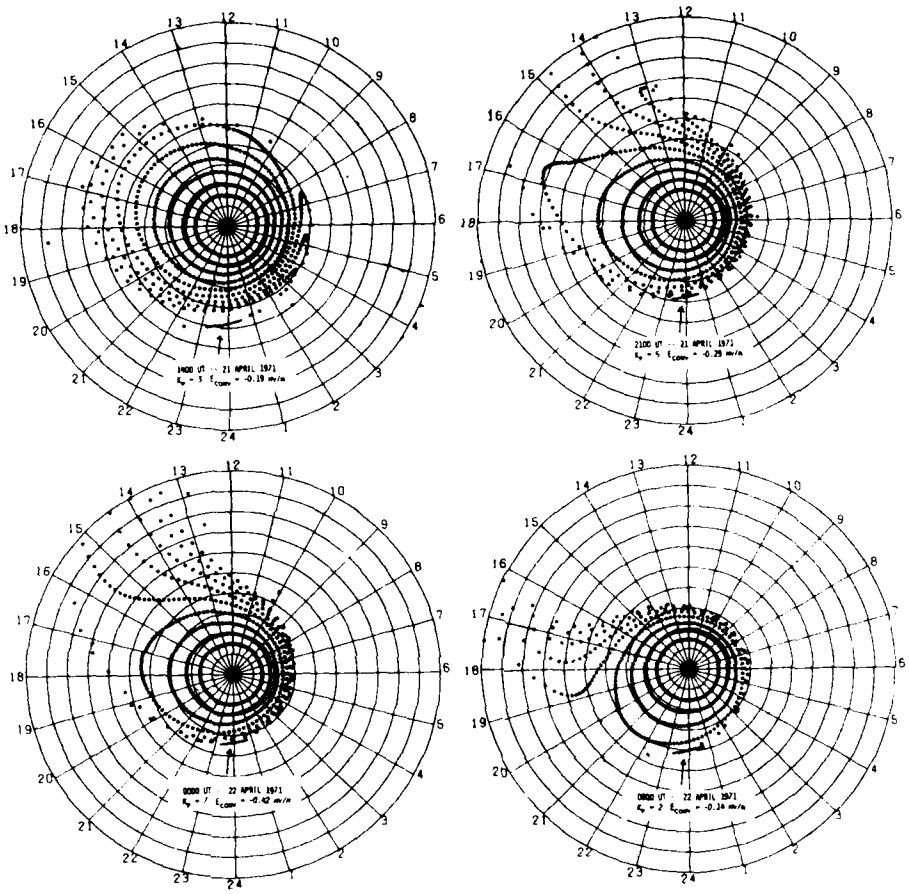


FIGURE 6. Four representative deformations of the plasmasphere depicted in Figure 5(b), at t=2, 9, 12, and 20 hours into the K_p scenario for 21-22 April 1971. The arrows indicate the L_{pp} boundary segments as defined by the particles at 23:45 LT.

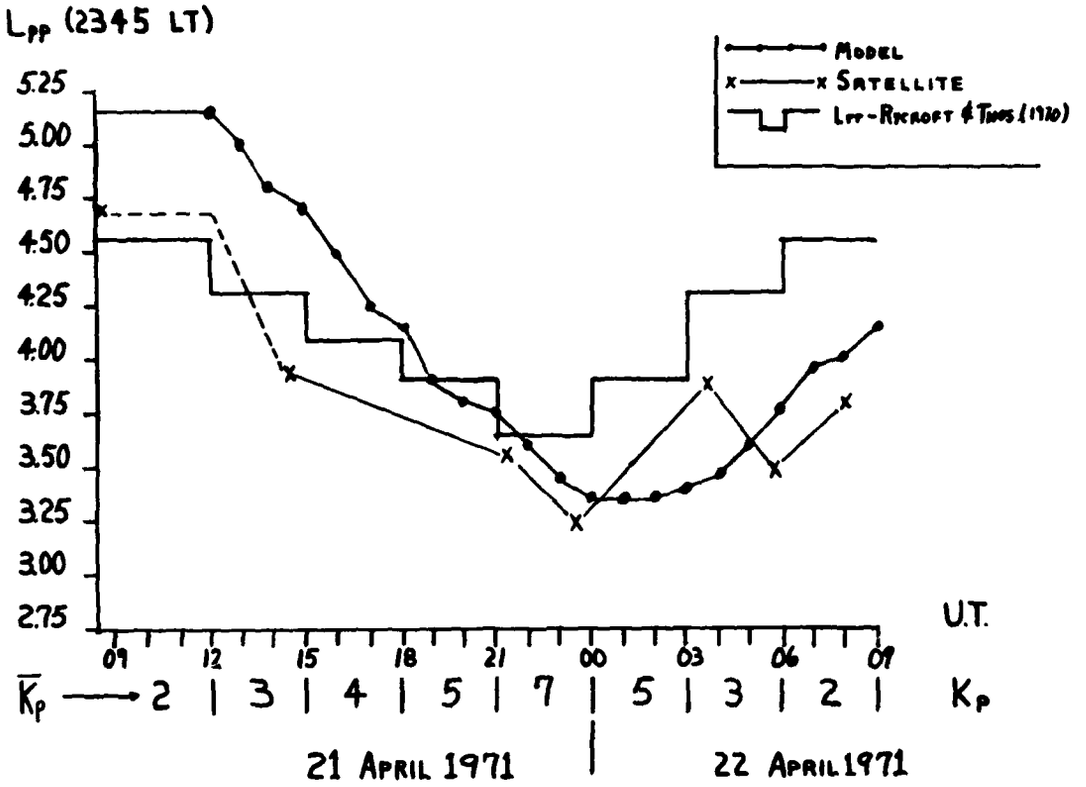


FIGURE 7. Overall Model Results for the K_p scenario of 21-22 April 1971, shown in comparison to ISIS probe data giving the equatorward edge of the F-region trough, and statistical results of Rycroft and Thomas (1970).

DIGITAL DATA TRANSMISSION OVER HF SKYWAVE CIRCUITS: PREDICTION AND EXPERIMENT

GERALD A. CLAPP
NAVAL OCEAN SYSTEMS CENTER
SAN DIEGO, CALIFORNIA

INTRODUCTION

In most areas of communications there is an increasing trend toward digital systems, and this is especially true for HF radio circuits. The traditional measure of HF circuit quality is signal-to-noise ratios. However, for digital systems the circuit bit error rate (BER) is often observed to be poorly correlated with signal-to-noise ratios because of the wide variety of noise and fading characteristics seen on HF channels.

One goal of this effort is to establish a base of information about the hourly and daily characteristics of multifrequency HF radio digital data links. The major characteristic of interest is the channel BER and its variation in time and frequency. The proper selection of HF frequencies is crucial if low error rates are to be provided to a communicator on a continuous basis. The second goal of this effort is to determine circuit performance when frequency selection is made by long-term frequency predictions. This information is then to be compared to the actual observed performance and any difference will be the result lack of real-time HF channel information. If the two goals are achieved then systems planners can estimate data throughput for proposed systems depending on the frequency management technique selected.

Many HF users routinely depend on frequency prediction techniques for guidance in selection present and future operating frequencies. Various computer programs are used to determine maximum usable frequencies (MUF) and optimums frequencies for traffic (FOT) as a function of path length, location, season, time of day, etc. Some predictions include expected signal-to-noise ratios (SNR) and circuit reliability or service probability for a specified grade of service. It is well established that these prediction techniques provide good forecasts of usable frequencies for monthly or bi-monthly averages.

The day-to-day HF user is well aware of varying propagation conditions on any time scale. If the user requires better frequency selection guidance than provided by long-term computer predictions then the alternative is to adopt real-time information such as provided by ionospheric sounders. Sounders provide real-time information about propagation over the path. A common stepped-frequency sounder will yield signal strength and multipath conditions for 1/20 octave frequency steps over the HF band once each minute.

Frequency prediction schemes such as DNC-14/NTP-6, Supp.1, will be compared with the HF digital data taken for this report. Measures of expected improvement by employing a sounder will be determined. Circuit availability using prediction techniques and sounding techniques will then be calculated and compared. Of significance for this analysis is that measured digital bit error rates will be used to characterize the quality of a frequency. This differs from other analysis which have usually used measured signal-to-noise ratios. The results are especially useful to the user since they do not require conversion from signal-to-noise ratios to bit error rates.

DATA

The baseline data consists of bit error rate (BER) measurements taken simultaneously on five frequencies. There are 528 hours of five-frequency BER data taken from December 1973 through 9 January 1974. The data consists of hourly values of the median BER and 90-percentile BER for each of the five frequencies. The 90-percentile number is the BER that is exceeded during 10% of the hourly sample; or equivalently, for 90% of the hour sample the measured BER is less than the given value of the BER.

From years of analyzing Naval ship/shore HF links, it has been determined that a median bit error rate of 10^{-2} corresponds to a TTY character error rate (CER) of from 3 to 4 percent. At 3 - 4% CER a teletype circuit is only marginally usable. A 90-percentile BER of 10^{-2} corresponds to a median bit error rate of about 2.5×10^{-3} and at this error rate teletype copy

would be rated as good-to-excellent.

The HF circuit was between Vallejo, CA, and San Diego, CA, a distance of 750 km or 470 miles. Transmitter power was 10 kw and both transmitting and receiving antennas were omni-directional and vertically polarized,, which resulted in about a 2 dB antenna pattern loss for the path length under test. The five simultaneously used frequencies were 2.7, 4.9, 6.4, 9.1 and 12.0 MHz. Bit error rates were measured on one narrow-band (110 Hz) frequency shift keyed (FSK) channel. There were generally from eight to twelve FSK channels in use in the transmitter tone package thus yielding a transmit power of approximately one kw for the data channel used for the test. Bit errors on the HF path were determined by comparison of the HF data stream with the original data stream that was provided over a trunk phone line. Phone line outage caused the loss of about six hours of data over the test period.

Little or no interference was present on any of the frequencies as each had been in continuous use several months before the beginning of the test period.

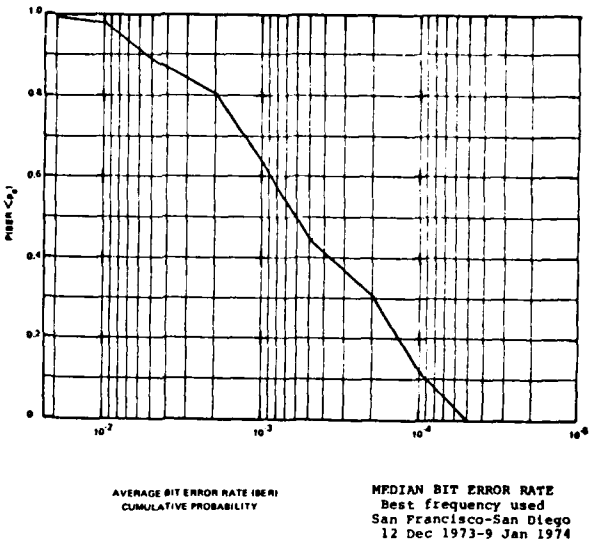


FIGURE 1

RESULTS

The basis for determining HF circuit performance will be results of the analysis of the five-frequency data. A request for frequencies for a tactical exercise, experiment, or other purpose is likely to result in the assignment of on the order of five usable frequencies. More may have been assigned but may be only marginally usable due to interference from other users on the authorized frequency. Additional authorized frequencies may be unusable because they are too far above or below the range of propagating frequencies for daytime or nighttime use.

The ultimate performance of a five-available-frequency circuit will be measured by selecting the best frequency out of the five frequencies at each hourly interval. The five frequencies were selected based on anticipated propagation conditions for the 470 mile path and should provide good 24-hour coverage for the path. This section presents the results that would be attained on the test path if the frequency were optimally selected each hour. Included in this section will be statistics of bit error rates, availability of grades of service, day/night differences in propagation, and the number of frequency changes required. Figure 1 shows the hourly median bit error rate for the entire test period. Every hour the error rate on each frequency was looked at and the frequency with the lowest error rate was selected. For only 3% of the entire test period was the hourly median BER at 10^{-2} or worse and at no time was it poorer than 2×10^{-2}. Thus there were no complete circuit outages and only three percent of the time would TTY traffic be rates as marginal. For fifty percent of the test period the median bit error rate was 6×10^{-4} or better. At this error rate only about 2 character errors per thousand or less would occur on a TTY channel. As can be seen, given a selection of five frequencies that are relatively free of interference excellent HF circuit quality was maintained. The problem for the HF communicator is to: (1) obtain a set of frequency allocations that are optimized for his path; and (2) know when to change frequencies and what frequency to change to.

To aid the HF communicator with his problems a number of computer propagation prediction programs have been written. Included in these is ESSA/ITS 78 which is the basis for DNC-14 (now superceded by NTP-6, Supp. 1) and the U.S. Army Skywave Propagation Predictions. These programs are very successful for the problem of specifying the range of frequency allocations that are optimum for specified path for a specified time. They also outline the diurnal frequency changes required to remain in operation as near to the predicted optimum working frequency (FOT) as possible. These propagation predictions are monthly averages and the expected statistical deviations from average; they cannot predict hour-to-hour or day-to-day variations. The significance of not being able to predict day-

to-day variations is shown in the following tables. Table 1 gives the number of days during the test period of 12 December 1973 - 9 January 1974, that each of the five frequencies had the lowest BER. This data is given for each hour of the day. For instance, the ESSA/ITS prediction for the circuit says that at 0000 GMT the predicted best frequency out of the five frequencies is 6.3 MHz. The test data shows that 6.3 MHz had the lowest error rate on only 11 days out of 29 days. On 9 occasions 4.9 MHz was the best frequency. As observed, at 0000 GMT, any of the five frequencies had a probability of being the best. The normal diurnal variation in propagation is clear from table 1; low frequencies are preferred during night and higher frequencies during the day. However, there clearly is no predictable frequency that is going to be best at any given hour of the day.

TABLE I

Number of Days Given Frequency has Lowest Bit Error Rate

Time Gmt	Frequency (mHz) 2.7	4.9	6.3	9.1	12.0
0000	4	9	11	3	2
0100	6	14	4	3	2
0200	1	17	1	3	2
0300	5	17		1	1
0400	9	10		4	2
0500	13	5		4	3
0600	8	10		4	3
0700	7	14		3	1
0800	6	18	2	1	1
0900	6	17			1
1000	6	18			
1100	7	17			
1200	11	12		1	
1300	13	10	1		
1400	10	17	2		
1500	4	5	19		
1600		1	10	9	
1700	1	2	8	11	
1800		2	3	15	
1900		3	2	17	2
2000	1	4	1	15	3
2100		4	3	17	4
2200		4	4	17	2
2300	1	8	10	13	1

A communications requirement may not necessitate finding the best frequency but one that meets a particular requirement. Table II illustrates the occurrence of meeting a requirement of the 90-percentile BER being better than 10^{-2}. Recall that this error rate would provide good-to-excellent teletype copy. It is evident from the table that there will generally be little problem finding a frequency that meets the requirement during the daytime hours. However, from 0800-1300, GMT (0000-0500 local) there is only at best a .78 probability that a given frequency will meet the 10^{-2} BER requirement. For only 17 hours out of the 528-hour test period (3.2% of the time) no frequency could meet the 10^{-2} requirement. However, all 17 hours had 90-percentile error rates of about 2×10^{-2} and would have provided at least marginal communications.

TABLE II

Fraction of Days the Given Frequency Has a 90-percentile Bit Error Rate $\leq 10^{-2}$

Time Gmt	Frequency (mHz) 2.7	4.9	6.3	9.1	12.0
0000	.59	1.00	.95	.27	.09
0100	.87	.91	.35	.22	.17
0200	.96	.96	.30	.26	.13
0300	.96	.83	.39	.17	.09
0400	1.00	.74	.30	.26	.17
0500	.95	.52	.17	.26	.22
0600	.87	.62	.17	.30	.22
0700	.83	.70	.17	.22	.17
0800	.70	.78	.26	.09	.04
0900	.52	.70	.09		
1000	.52	.78	.09		
1100	.61	.70	.04		
1200	.74	.52	.09	.04	.04
1300	.83	.48	.09	.04	.04
1400	.87	.70	.17		
1500	.82	1.00	.91		
1600	.29	.86	.95	.48	
1700	.11	.89	1.00	.79	
1800		.81	.86	.76	
1900		.95	1.00	.90	.14
2000	.05	.91	.91	.91	.27
2100	.09	.96	1.00	.96	.27
2200	.09	.87	.91	.83	.13
2300	.26	.96	.96	.78	.13

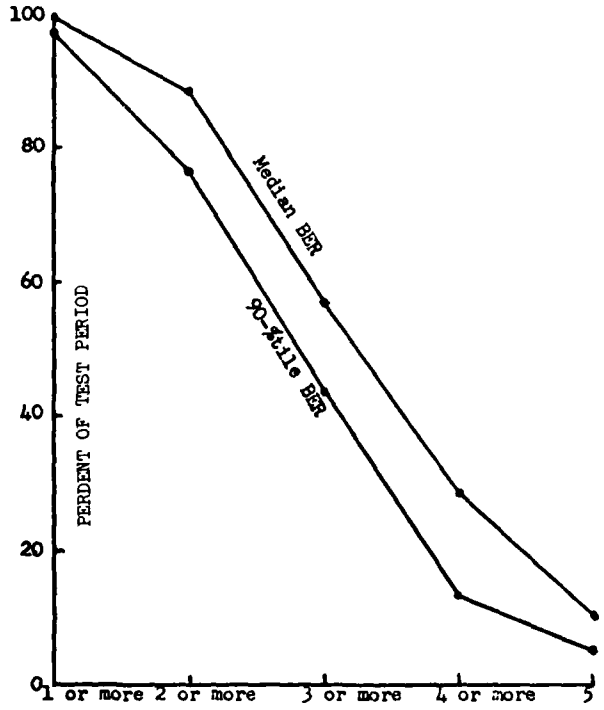
Frequency or time diversity used at these times would have put the observed BER to better than 10^{-2}. For 109 hours of the test period, only one frequency was available; for the balance of the test period (76.1% of the time) two or more frequencies exceed the 90-percentile BER limit of 10^{-2}. Figure 2 summarizes these results for the median and 90-percentile bit error rates.

Sunrise/sunset propagation difficulties were not evident from the test data. The data in table III shows that the periods of difficulty occurred primarily at or shortly after midnight local time. During the sunset hour, 0100 local time, usually two or more good frequencies were available. During the path's sunset hour, 1500 PST, there was always at least two frequencies available with BER better than 10^{-2}. The

TABLE III

Number of Days

Time Gmt	Number of Frequencies with 90-percentile BER $\leq 10^{-2}$ 0	1	2	3	4	5
0000			7	11	3	1
0100		2	11	3	1	4
0200		2	11	1	3	3
0300		5	10	5	1	2
0400		6	8	4	1	4
0500		11	6	1	1	4
0600	1	8	8	2	2	3
0700		11	6	3		3
0800	3	3	13	3		1
0900	4	9	10			
1000	3	9	10		1	
1100	4	9	10			
1200	1	13	7	1	1	
1300	1	14	5	2	1	
1400		7	12	4		
1500			6	16		
1600			8	8	4	
1700			5	13	1	
1800			3	16		
1900			2	16	2	
2000			2	13	6	
2100			3	11	8	
2200			3	16	1	2
2300			4	13	4	1
Total Hours	17	109	171	162	41	28



NUMBER OF FREQUENCIES AVAILALBE WITH BER $\leq 10^{-2}$

FIGURE 2

problem that occurs at sunrise/sunset transitions is an example of a HF communications problem; the problem to find the frequencies that are propagating. Within the one-hop sky-wave range, communications problem on HF is seldom due to the fact that there are no frequencies propagating.

The frequency change question is crucial; when to change frequencies and what frequency to change to. For the test path, computerized path predictions would indicate four to five frequency changes per day, occurring primarily at sunrise and sunset transitions. The test data shows a different pattern of frequency changes required to remain on the best frequency. At the end of each hour the five frequency test data was looked at to see if a better frequency could be found. A better frequency could be found 33.0% of the time, 67.0% of the time no frequency change was needed. If a frequency change was required, 73% of the time it was to the next adjacent available frequency, and 27% of the time the change was to at least two frequencies higher or lower than the current frequency. The average for the test period was 8 frequency changes per day, however, there was a large day-to-day variation in the number of required frequency changes as can be seen in table IV.

TABLE IV

No. of Freq. Changes Per Day	No. of Days
0	
3	
4	2
5	1
6	4
7	6
8	3
9	2
10	1
11	1
12	1
13	1
14	

Of perhaps more importance is the time of day that a frequency change would be required. Table V illustrates this for the test data. The result is that frequency changes may be necessary at any time of the day with only a slight preference for sunrise/sunset frequency changes. It is this type of day-to-day and hour-to-hour propagation variations that frequency predictions can not make an allowance for.

PERFORMANCE OF FREQUENCY SELECTION TECHNIQUES

In this section the accuracy of a number of frequency prediction techniques will be calculated from the test data. Each prediction method will be applied to the test data and the resulting circuit quality will be determined. These techniques are:

TABLE V

Time of Day (GMT)	No. of Days a Freq. Change was Required at Given Time
0000	10
0100	8
0200	5
0300	5
0400	8
0500	6
0600	4
0700	6
0800	8
0900	4
1000	7
1100	3
1200	8
1300	7
1400	6
1500	14
1600	11
1700	9
1800	9
1900	2
2000	5
2100	7
2200	8
2300	7

These techniques are:

(1) Sounder
(2) Idealized Computer Prediction
(3) Idealized Prediction with Modification
(4) ESSA/ITS 78 HF Prediction Program
(5) DNC-14/NTP-6, Supp. 1
(6) U.S. Army Skywave Propagation Charts

Sounder

The results by using this technique are obtained by selecting the best frequency each hour throughout the test period. The resulting observed bit error rates then sets the standard as the absolutely best performance available from the circuit. It is against this standard that other frequency prediction techniques will be comapred. An actual HF sounding system differs from the technique just described in that: (1) a sounder attempts to select the best frequency for current use rather than takes the best frequency for the previous hour; and (2) a sounder system can indicate that a better frequency is available at any time rather than at just the end of each hour. These two factors will tend to cancel and the resulting performance of a circuit being guided by a sounder system should be nearly as indicated in this section.

Idealized Computer Prediction

The test data was analyzed to determine for each given hour of the day the frequency that had the highest probability of being the frequency with the lowest bit error rate. The bit error rate distributions that result from using this technique is the upper limit that computerized frequency predictions will be able to achieve. Because of the large number of variables in a prediction program, many of which are imperfectly known or understood, it is unlikely that the Idealized Computer Prediction scheme can ever be written.

Idealized Computer Prediction

The calculated nighttime MUF was nearly identical with one of five assigned frequencies. This frequency was generally the best, but on occasions it was above the observed MUF; the modification consists of using the next lowest frequency during times of depressed nighttime MUF's.

ESSA/ITS 78 HF Frequency Prediction Program

"Predicting Long-Term Operational Parameters of High-Frequency Skywave Telecommunication Systems," is the title of the report that describes the use of the 'standard' computerized HF frequency selection program. This program in some form is used almost universally for HF frequency predictions.

DNC-14/NTP-6, Supp. 1

A condensation of the above computer program into tabular and chart form for use by the Navy. Designed as a bi-monthly publication for future frequency prediction for the Navy user.

U.S. Army Skywave Charts

A second condensation of the ESSA computer prediction program for Army use. The performance of each of the frequency selection techniques is given in table VI.

From the table VI data we see that the ESSA/ITS HF prediction program predicted the best frequency for digital data transmission only 45.6% of the time. However, a frequency with a 90-percentile less than 10^{-2} was predicted 83.9% of the time. The two tabular condensations of the ESSA/ITS program for military use are DNC-14 and the U.S. Army Skywave Prediction; both of these were observed to predict frequencies poorly. DNC-14/NTP-6, Supp.1 predicted use of a frequency that would have been unusable 39.2% of the time; and the Army book would have predicted an unusable frequency an astounding 53.4% of the time. For this analysis a usable frequency is defined as providing good teletype copy which occurs when 90 percentile bit error rates are about 10^{-2} or better.

Figure 3 shows the distribution of bit error rates that would have resulted had each of the six frequency prediction methods been used. Again, the very poor performance of DNC-14 and the Army method is obvious.

TABLE VI

PERFORMANCE OF FREQUENCY PREDICTION TECHNIQUES

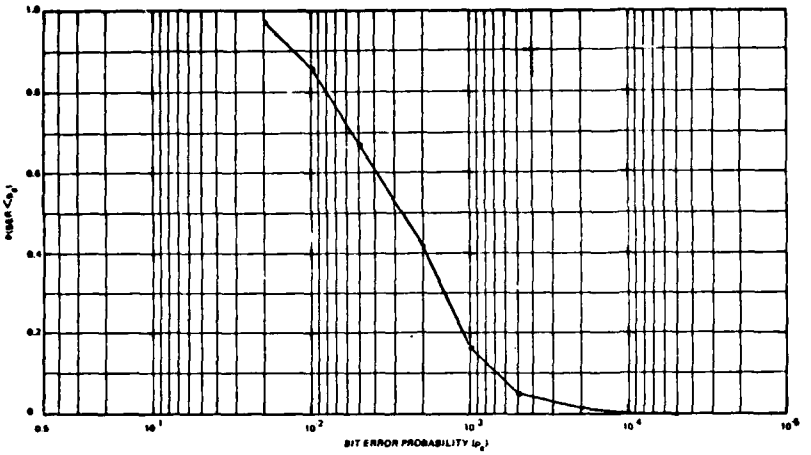
	% of time selected frequency was best observed frequency	% of time selected frequency had 90%-tile BER worse than 10^{-2}
1. Sounder	(100)	3.2
2. 'Ideal' Computer Prediction	57.8	18.3
3. 'Ideal' Prediction with MUF mod	71.9	5.8
4. ESSA/ITS Prediction Program	45.6	16.1
5. DNC-14/NTP-6, Supp. 1	39.9	39.2
6. U.S. Army Skywave Propagation Charts	31.9	53.4

The poor performance of the prediction methods cannot be related to unusual path location or abnormal propagation conditions. The geomagnetic index Kp during the month test period ranged from 5 to 35 an averaged about 20. No correlation was found to exist between the daily Kp and daily average BER. Sporadic-E days phenomena was observed during the early evening hours on four of the test days and was a minor source of test data variation. The poor performance of the predictions seems to be the result of hour-to-hour and day-to-day variations in propagation not accounted for in prediction techniques.

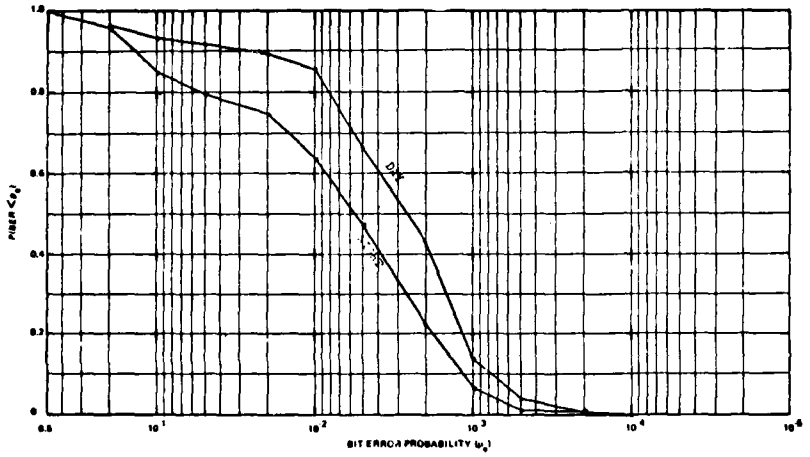
Other observations from the data show that the daily average BER ranged from 1.5×10^{-3} to 5.5×10^{-3} if the best frequency was selected each hour. These bit error rates were observed to be uncorrelated with the geomagnetic Kp index.

The hourly frequency with the lowest BER was compared with the predicted optimum frequency (FOT) for each hour of the test period. It was found that the FOT was the best estimate for the nighttime portion of the data and a value of 0.75 of the predicted FOT was best frequency for digital transmission during daytime hours. However, the standard deviation about the predicted FOT was 0.2 of the FOT. Twelve hours daily were labled nighttime and daytime; for this midwinter test period both the sunrise and sunset transition period was included in the daytime data. For the entire test period the nighttime BER was twice the daytime measured BER.

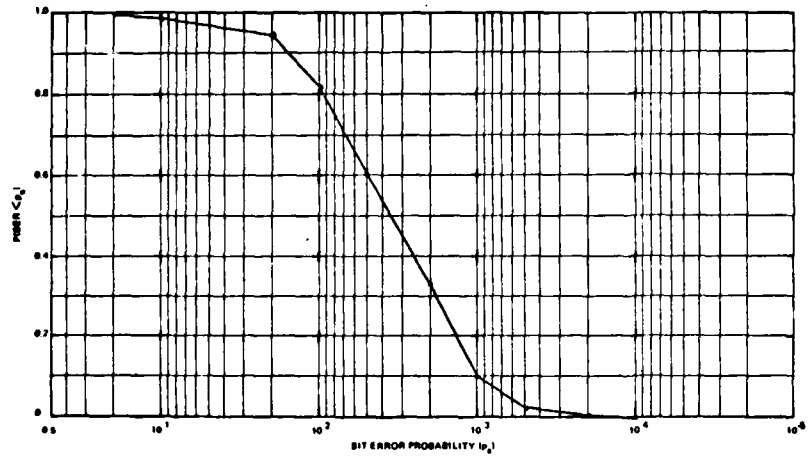
The data given here is for a system that employed FSK keying on narrow-shift channel at a rate of 75 baud. Sixteen such channels can be be placed in a normal 3 kHz voice sideband. If the power per data channel is kept constant the data here should represent performance of a 1200 baud FSK or PSK circuit. The digital error error performance of a 4-phase PSK, 2400 baud, modulation is anticipated to be only slightly less. If care is taken, a reasonable extrapolation to systems employing other power levels or antenna gains can be expected. Likewise, sufficient data exists to estimate performance if frequency or spaced-antenna diversity is employed. A future report will address these areas.



'SOUNDER' (Best of five frequencies, 528 hours of data)



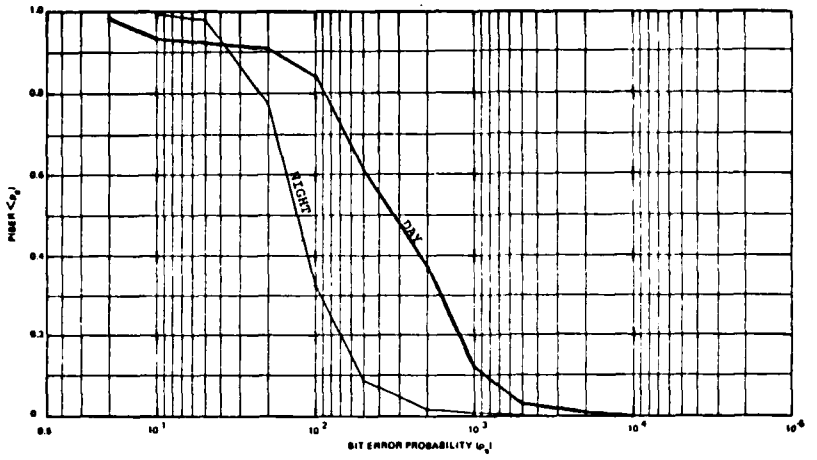
'IDEAL' COMPUTER PREDICTION



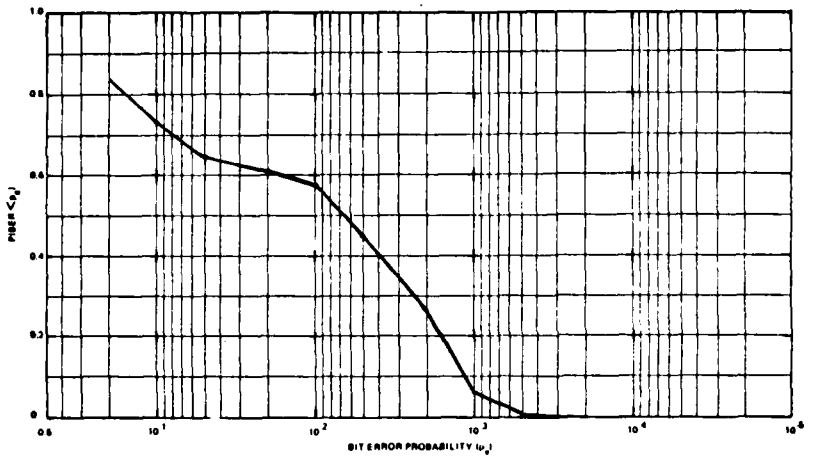
'IDEAL' COMPUTER PREDICTION WITH MUF MODIFICATION

HOURLY CUMULATIVE 90-PERCENTILE BER

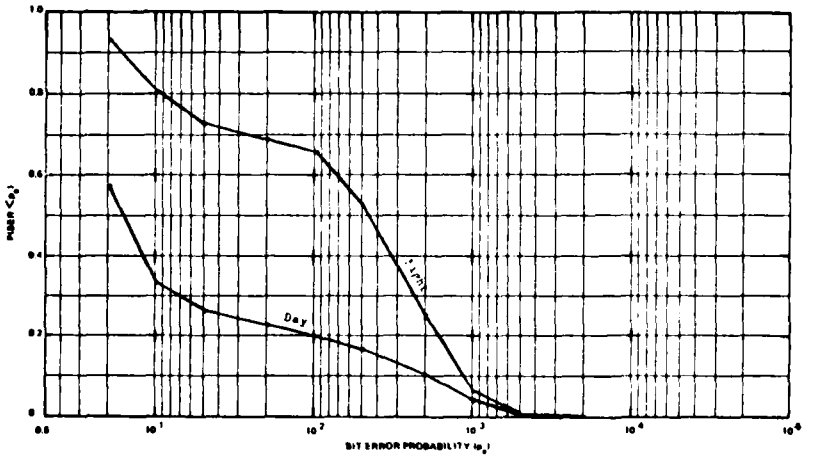
FIGURE 3
PREDICTION PERFORMANCE



ESSA/ITS Computer Prediction



DNC-14/NTP-6, Supp. 1



U.S. Army Skywave Prediction

HOURLY CUMULATIVE 90-PERCENTILE BER

FIGURE 3 (Cont.)
PREDICTION PERFORMANCE

ENVIRONMENTAL FORECASTING BY USING IONOSPHERE OBSERVATIONS ON ACOUSTIC-GRAVITY WAVES ASSOCIATED WITH SEVERE STORMS

R. J. Hung
The University of Alabama in Huntsville
Huntsville, Alabama 35807

and

R. E. Smith
NASA/Marshall Space Flight Center
Huntsville, Alabama 35812

ABSTRACT

Atmospheric acoustic-gravity waves associated with severe local thunderstorms, tornadoes and hurricanes have been extensively studied. Group ray tracing computations, based on acoustic-gravity waves observed by an ionospheric Doppler sounder array, indicate they are associated with severe storms. The study of wave-wave resonant interaction also suggests a possibility that tornadic storms are dynamical evolutions from thunderstorms. This study reveals the possibility of the development of an environmental forecasting system based on ionosphere observations of atmospheric acoustic-gravity waves associated with severe storms.

I. INTRODUCTION

The ionosphere is capable of sustaining a large number of wave phenomena, waves propagating downward from the magnetosphere, waves propagating upward from the neutral lower atmosphere, etc. Ionospheric-tropospheric interactions can also be studied using ionospheric observations of atmospheric waves propagating upward from the troposphere. These atmospheric waves are observed on ground-based ionospheric sounding records as perturbations in electron density.

Atmospheric acoustic-gravity waves studies can help to understand severe weather activity. The correlation between atmospheric acoustic-gravity waves and severe storms has been investigated during the past twenty years. Tepper (1950, 1954) proposed that the pressure jump lines effectively lifted the lowest layers of the atmosphere and appeared to initiate squall line development in convectively unstable air. Matsumoto and Akiyama (1969), Matsumoto and Tsuneoka (1969), and Matsumoto et al. (1967a, 1967b) contended that atmospheric acoustic-gravity waves were responsible for the pulsating tendency of winter and summer convective storms in Western Japan. Recently, Uccellini (1975) proposed that atmospheric acoustic-gravity waves are an important mechanism for triggering severe convective storms. He also suggested that the study of atmospheric acoustic-gravity waves could reveal much about the development of thunderstorms.

In recent years, experimental observations from CW Doppler soundings of the upper atmosphere show that there are ionospheric disturbances apparently associated with severe weather and thunderstorms (Baker and Davies, 1969; Davies and Jones, 1972; Prasad et al., 1975; Smith and Hung, 1975; Georges and Greene, 1975; Georges, 1976; Jones and Georges, 1976; Hung et al., 1978, etc.). These studies reveal that quasi-sinusoidal oscillations, with two harmonics of wave periods, 3 to 5 minutes and 6 to 9 minutes, are observed when thunderstorms with cloud tops (radar heights) in excess of about 12 Km are present within a radius of several hundred kilometers of the ionospheric reflection points.

During the extreme tornado outbreak of April 3, 1974, Hung et al. (1978) detected another two harmonics of wave periods, 11 to 15 minutes and 26 to 30 minutes in the CW Doppler array observational data. On this particular day, two hundred and thirty-six tornadoes occurred in the U.S. Among these tornadoes, 41 tornado touchdowns were identified by the National Severe Storms Forecast Center (hereafter NSSFC) within a 800 Km radius from Huntsville, Alabama during the time period 1700-2200 UT. These tornadoes were divided into five groups depending upon the touchdown times and geographical locations. A group ray path computation was used to trace the atmospheric gravity waves observed at ionospheric height in an effort to locate this probable source. Six ray tracing computations which covered the 41 tornado touchdowns within a 800 Km radius

from Huntsville, Alabama, showed that the acoustic-gravity waves were excited 1 to 3 hours before the touchdown times (Hung et al., 1978).

From the analysis of the extreme tornado outbreak of April 3, 1974, it appears as if the waves observed are a precursor phenomena due to the integrated effect of a group of tornadic storms which occurred approximately at the same time and in the same region. Atmospheric acoustic-gravity waves associated with isolated tornadic storms have also been analyzed with similar results.

Analyses have shown that atmospheric acoustic-gravity waves with wave periods of 20 to 24 minutes were detected when Hurricane Eloise was in the Gulf of Mexico (Hung and Smith, 1977a; 1977b). The computed location of the source of the waves was roughly where the hurricane would be located three hours after the waves were excited.

The transformation of atmospheric acoustic-gravity waves with wave periods, 3 to 5 minutes, associated with thunderstorm activity, to wave periods, 11 to 15 minutes, associated with tornadic storm activity, can be studied through wave-wave resonant interactions. The mechanism of development of tornadic storms from thunderstorms can be further studied by means of the resonant interaction.

II. ANALYSIS OF IONOSPHERE OBSERVATIONS

The ground-based ionosphere observation technique uses radio receivers located at a central site to monitor signals transmitted from three independent remote sites and reflected off the ionosphere approximately half way between the transmitter and receiver sites. When a wave perturbs the electron density profile in the ionosphere, the total phase path of the radio signal changes resulting in an instantaneous frequency shift in the radio signal.

The ionospheric Doppler sounder array system employed in this study consists of three sites with nine field transmitters operating at 4.0125, 4.759 and 5.734 MHz. These sites are located in the Tennessee Valley area. A detailed description of the array system has been given by Hung et al. (1978).

The observed data were subjected to both power spectral density analysis and cross correlation analysis. In some severe storm active days, the power spectral density analysis revealed more than one peak. When this occurred, Butterworth's digital filter (Otness, 1968) was applied to band pass the peaks. After using this digital filter, each peak corresponds to a wave excited by an individual source. Thus, we are able to detect different wave patterns associated with different sources during the same time period. The horizontal wave vectors and horizontal phase velocities of the disturbances are calculated from cross correlograms.

The azimuthal angle of wave propagation can be determined to an accuracy of ± 5°, and the horizontal phase velocity to ± 10 percent.

In the six cases during the extreme tornado outbreak of April 3, 1974, gravity waves with two harmonics of wave periods, 11 to 15 minutes and 26 to 30 minutes, horizontal wavelengths in the range 100 to 220 Km, and horizontal phase speeds in the range 90 to 220 m/sec were detected as being closely associated with tornadic storms (Hung et al., 1978 for details).

Results for the five isolated storms, two on November 20, 1973 and three on January 13, 1976 are shown in Table 1. A summary of these results shows that gravity waves with two harmonics of wave periods, 10 to 15 minutes and 25 to 30 minutes, horizontal wavelengths in the range 100 to 290 Km, and horizontal phase speeds in the ranges 100 to 220 m/sec were apparently associated with these isolated tornadic storms.

TABLE 1. PROPAGATION CHARACTERISTICS OF THE OBSERVED GRAVITY WAVES ASSOCIATED WITH ISOLATED TORNADIC STORMS

Observation Time (UT)	Wave Period (Min.)	Horizontal Wavelength (Km)	Azimuthal Angle of Wave Arrival (deg)	Horizontal Phase Speed (m/sec)
1900-1945, November 20, 1973	10.7	120	85	186.9
1945-2045, November 20, 1973	11.8	125	37	176.5
1930-2130, January 13, 1976	25.7	234	165	152
2030-2230, January 13, 1976	28.3	287	179.7	169.4
2100-2300, January 13, 1976	26.3	234	179.9	148.8

Two events during the time period of Hurricane Eloise, 0200-0340 UT and 0500-0730 UT, September 23, 1975, indicate that gravity waves with wave periods of 20 to 24 minutes, wavelengths from 220 to 300 Km, and horizontal phase speeds from 150 to 200 m/sec were present.

III. RAY TRACING COMPUTATIONS AND WAVE SOURCES

Theoretical discussions of atmospheric acoustic-gravity waves have been carried out by Bretherton (1966), Jones (1969), Cowling et al. (1971) and Bertin et al. (1975). These discussions indicate that the geometrical optics approximation is valid for the conditions discussed. The wave propagation is assumed to be locally plane so that the local dispersion relation of atmospheric acoustic-gravity waves, proposed by Hines (1960), is satisfied. Ray tracing, thus, can be carried out by following the group velocity direction in a wind-stratified model atmosphere.

The propagation of wave energy in a lossless transparent medium follows the direction of the group velocity (e.g., Yeh and Liu, 1972). This direction is termed the ray direction. In general, in an anisotropic medium the ray direction is different from that of the wave vector. The reverse ray tracing computation is the integration of the group velocity with respect to the time domain from the ionospheric reflection point back down to the tropopause using the wave period, wavelength, and azimuthal direction of wave propagation obtained from the observational data, the initial vertical wave vector computed from the dispersion relation, and appropriate atmospheric parameters. The effect of wind is taken into account by considering the time-space transformation given by the Galilean transformations of the displacement vector, time, Doppler-shifted wave frequency, and wave vector. The detailed description of the group ray tracing computation has been given in another article (Hung et al., 1978).

In this study, the neutral wind is treated as a constant in each slab of the atmosphere considered. The values of atmospheric parameters for each altitude are calculated from the U. S. Standard Atmosphere (1962), and profiles of the neutral wind are established by fairing in winds computed from the Kohl and King (1967) model above 100 Km altitude, with meteorological rocketsonde data from Cape Kennedy, Florida below 90 Km. Detailed treatment of wind profiles are given in Hung et al. (1978)

The reverse ray tracing is started at the ionospheric reflection height for the radio frequency wave used in this study and continued as long as the calculation is possible to a lower limit of 12 Km, an altitude equivalent to the tropopause. For the purpose of this study the geographic location of the point at which the calculation is terminated is referred to as the probable source of gravity waves. These computed sources are then compared with the actual observations from the computer printout of locations and times of tornado touchdowns supplied by NSSFC and/or the storm track of the hurricane supplied by the National Climatic Center.

Results of this comparison for the six events of April 3, 1974 indicate that the computed wave sources are near the locations where tornadoes touchdown from 1 to 3 hours later.

Results of computed probable sources of waves for the five isolated tornadic storms are given in Table 2. In summary, the waves appear to be excited near the location where tornadoes touchdown more than one hour later. The computed wave sources are within the ± 5° accuracy of the azimuthal angle of wave arrival. Figures 1, 2, 3, 4 and 5 show the geographical locations of computed sources of waves which were detected at ionospheric heights with the receivers at Huntsville, Alabama, at 1900-1945, November 20, 1973; 1945-2045 UT, November 20, 1973; 1930-2130 UT, January 13, 1976; 2030-2230 UT, January 13, 1976; and 2100-2300 UT, January 13, 1976, respectively.

In the case of gravity waves associated with hurricanes, analysis of two cases from Hurricane Eloise during the time periods, 0200-0340 UT and 0500-0730 UT, September 23, 1975 indicate that the computed sources were located roughly along the track of the storm, and more than three hours in advance of the location of the storm (Hung and Smith, 1977a; 1977b).

IV. MECHANISM OF RESONANT INTERACTIONS OF WAVES ASSOCIATED WITH STORMS

It is known that tornadoes are closely associated with severe thunderstorms (Davies-Jones and Kessler, 1974). Our observations of atmospheric acoustic-gravity waves associated with tornadoes indicate that waves are observed more than one hour before touchdown times. These results are apparently similar to the conclusion reached by Uccellini (1975) that in the case of severe convection, the acoustic-gravity waves are a precursor to the thunderstorms.

It has been shown by Georges (1973), Prasad et al. (1975), etc., that the ionospheric wave-like disturbances associated with thunderstorms were observed when thunderstorms with tops were in excess of about 12 Km, or when intense updrafts of convection had penetrated the tropopause. A similar idea was also proposed by Malkus (1960) and Saunders (1962), that the convective regions imbedded in the stratiform anvil of the thunderstorms are clearly the overshooting convective cells which penetrate the tropopause. By taking photographs from a U-2 airplane flying over the thunderstorms, Vonnegut et al. (1966) showed that convective overshooting turrets rose

above the anvil cloud. The top of the observed stratiform thundercloud anvil was above the tropopause (Vonnegut et al., 1966). Observations made by Shenk (1974) further strengthen these results.

TABLE 2. COMPARISON BETWEEN COMPUTED PROBABLE SOURCES OF WAVES ASSOCIATED WITH ISOLATED *TORNADIC STORMS AND ACTUAL TOUCHDOWNS OF TORNADOES*

Observation Time (UT)	Wave Traveling Time (Min)	Tornado Touchdown Time (UT)	Wave Excited Earlier Than Tornado Touchdown (Hours)
1900-1945, November 20, 1973	66	1935	2
1945-2045, November 20, 1973	65	2100	2
1930-2130, January 13, 1976	53	2015	2
2030-2230, January 13, 1976	55	2114	1
2100-2300, January 13, 1976	57	2150	1

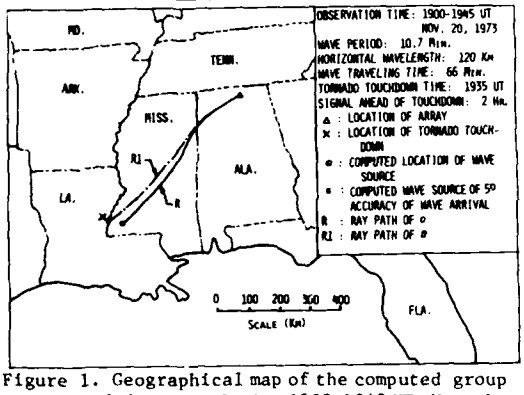


Figure 1. Geographical map of the computed group ray path of the waves during 1900-1945 UT, November 20, 1973, and the location of actual tornado touchdown.

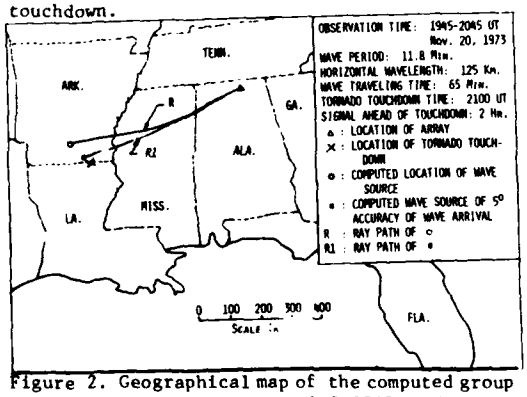


Figure 2. Geographical map of the computed group ray path of the waves during 1945-2045 UT, November 20, 1973, and the location of actual tornado touchdown.

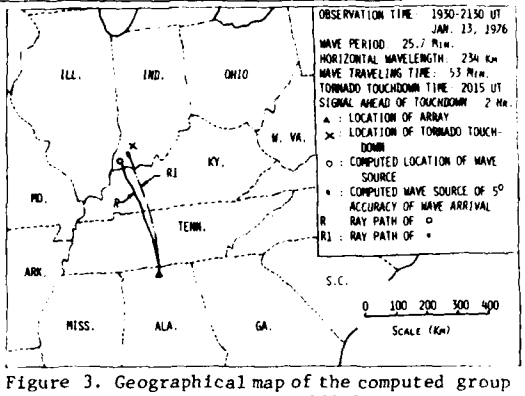


Figure 3. Geographical map of the computed group ray path of the waves during 1930-2130 UT, January 13, 1976, and the location of actual tornado touchdown.

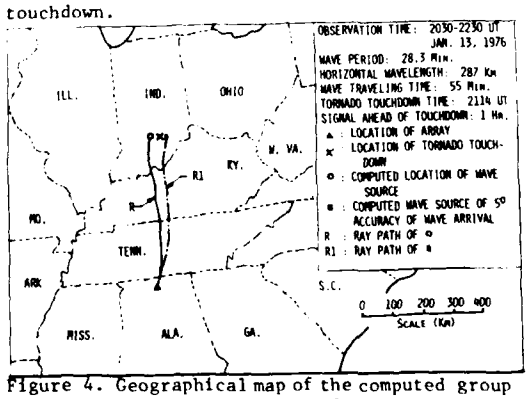


Figure 4. Geographical map of the computed group ray path of the waves during 2030-2230 UT, January 13, 1976, and the location of actual tornado touchdown.

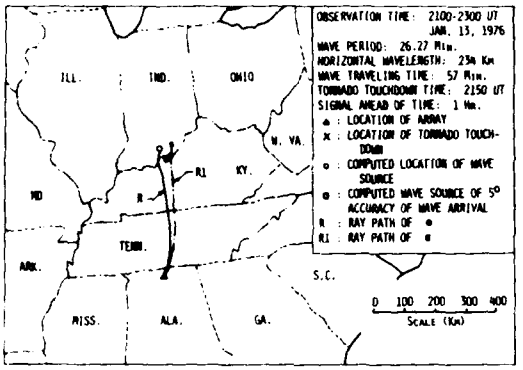


Figure 5. Geographical map of the computed group ray path of the waves during 2130-2300 UT, January 13, 1976, and the location of actual tornado touchdown.

The dynamical evolution of tornadoes from thunderstorms could be based on the wave-wave resonant interaction of two thunderstorm-induced waves becoming the tornado associated waves detected in the ionosphere. If the gravity waves associated with tornadoes are the results of the wave-wave resonant interaction of two waves associated with thunderstorms, the following conditions must be satisfied (Sagdeev and Galeev, 1969)

$$\omega_1(\underset{\sim}{k}_1) \pm \omega_2(\underset{\sim}{k}_2) = \omega_3(\underset{\sim}{k}_3) \tag{1}$$

where ω_1, ω_2 and ω_3 denote the wave-frequencies of thunderstorm 1, thunderstorm 2, and waves-associated with tornadoes, respectively; and $\underset{\sim}{k}_1$, $\underset{\sim}{k}_2$ and $\underset{\sim}{k}_3$ are the wave-vectors of thunderstorm 1, thunderstorm 2, and waves-associated with tornadoes, respectively. For the resonant-interaction to be valid, $\left|(\omega_1 - \omega_2)/\omega_1\right| \ll 1$, and the Taylor expansion is applicable to Equation (1). This leads to

$$\omega_1(\underset{\sim}{k}_1) = \omega_2(\underset{\sim}{k}_2) + \left(\frac{\partial\omega}{\partial\underset{\sim}{k}}\right) \cdot \Delta\underset{\sim}{k} \tag{2}$$

where

$$\Delta\underset{\sim}{k} = \underset{\sim}{k}_1 - \underset{\sim}{k}_2.$$

If we choose the negative sign in Equation (1), substitution of Equations (1) into (2) leads to:

$$\left(\frac{\partial\omega}{\partial\underset{\sim}{k}}\right) \cdot \Delta\underset{\sim}{k} = \omega_3(\underset{\sim}{k}_3) \tag{3}$$

By definition, $\omega_i = 2\pi/\tau_i$, where τ_i is the wave period of species i, and Equation (1) becomes

$$\frac{\tau_2 - \tau_1}{\tau_1 \tau_2} = \frac{1}{\tau_3} \tag{4}$$

which is the relation between the two waves associated with thunderstorms, τ_1 and τ_2, interacting resonantly which appears as the τ_3 detected at ionospheric height. For example, thunderstorm-induced waves, with a wave period of 3 minutes, interacting resonantly with another thunderstorm-induced wave with a wave period of $\tau_2 \simeq 3.896$ minutes generates a wave with a period of 13 minutes identical to those detected at ionospheric heights during periods with tornadoes (Davies and Jones, 1972; Hung and Smith, 1977a; Hung et al., 1978).

Again by definition $|\underset{\sim}{k}_i| = 2\pi/\lambda_i$, where λ_i is the wavelength of species i, Equation (3) leads to

$$\frac{\lambda_2 - \lambda_1}{\lambda_1 \lambda_2} \simeq \frac{1}{\tau_3} \left| \left(\frac{\partial\omega}{\partial\underset{\sim}{k}}\right) \right|^{-1} \tag{5}$$

which is another relationship between two thunderstorm-induced waves that can result in tornado-associated waves through wave-wave resonant interaction. To give a numerical example, let us consider thunderstorm-induced waves with a wavelength of 100 Km, resonantly interacting with another thunderstorm-induced wave with a wavelength $\lambda_2 \simeq 133$ km.

If $(\partial\omega/\partial\underset{\sim}{k}) = 500$ m/sec, a wave with a period of 13 minutes can be generated. This is a typical group velocity of the waves associated with thunderstorms. The wavelength of 133 Km is also within the range of our observations, 100 to 300 Km (Davies and Jones, 1972; Hung and Smith, 1977a; Hung et al., 1978).

Equations (4) and (5) give two criteria for any two thunderstorm-induced waves to generate the waves associated with tornadoes through resonant interaction. The two numerical examples seem reasonable based on our observations.

The wave-wave interaction condition, Equation (1), applies to a homogeneous medium. When the medium is inhomogeneous like the atmosphere, it requires additional conditions. If $\omega = \omega(\underset{\sim}{k}; t, \underset{\sim}{x})$ for a wave propagating in an inhomogeneous medium, the following conditions must be satisfied (Whitham, 1974)

$$\frac{\partial k_\alpha}{\partial t} + v_{g\beta} \frac{\partial k_\alpha}{\partial x_\beta} = -\frac{\partial\omega}{\partial x_\alpha} \tag{6}$$

and

$$\frac{\partial k_\alpha}{\partial x_\beta} - \frac{\partial k_\beta}{\partial x_\alpha} = 0 \tag{7}$$

where

$$V_{g\beta} \equiv \frac{\partial \omega}{\partial k_\beta} \equiv \frac{dx_\beta}{dt} .$$

Here, the subscripts of the Greek alphabet, α and β, stand for the vector components.

This study suggests a possible mechanism for the dynamical evolution of tornadic storms from severe thunderstorms based on wave-wave resonant interactions. This study also indicates that a tornado cannot develop in an isolated overshooting turret.

V. CONCLUSION

Ionospheric Doppler sounding observations during time periods with severe storm activity indicate that atmospheric acoustic-gravity waves are detectable as perturbations in electron density. Results based on six observed events from the extreme tornado outbreak of April 3, 1974, and five events of isolated tornadoes show that gravity waves associated with the severe storm system are excited more than one hour before tornado touchdowns. Results for gravity waves associated with hurricanes show the computed sources of the waves were located along the track of and more than three hours in advance of the location of the storm.

A study of wave-wave resonant interactions shows that there is a possibility that tornadoes evolve from severe thunderstorms through this mechanism. This study further indicates that a tornado cannot develop from an isolated overshooting convective turret.

It is suggested that further analysis of ionosphere observations of atmospheric acoustic-gravity waves may provide a technique for the early observation and possible short range predictions of severe storm activity.

ACKNOWLEDGMENT

The authors acknowledge the ray tracing computations made by T. Phan. One of the authors (RJH) also appreciates the support of the present research from NASA/Marshall Space Flight Center through Contract NAS8-31171 and the National Science Foundation/U.S. Army Research Office through Grant NSF/ATM75-15706.

REFERENCES

Baker, D. M., and K. Davies, J. Atmosph. Terr. Phys., 31, 1345, 1969.

Bertin, F., J. Testud, and L. Kersley, Planet. Space Sci., 23, 493, 1975.

Bretherton, F. P., Quart. J. Roy. Meteorol Soc., 92, 466, 1966.

Cowling, D. H., H. D. Webb, and K. C. Yeh, J. Geophys. Res., 76, 213, 1971.

Davies, K., and J. E. Jones, Space Research, 12, 1149, 1972.

Davies-Jones, R., and E. Kessler, Tornadoes, in Weather and Climate Modification, ed. by W. N. Ness, pp. 552-595, Wiley Interscience, New York, 1974.

Georges, T. M., Rev. Geophys. Space Phys., 11, 571, 1973.

Georges, T. M., Infrasound from Convective Storms, Part II: A Critique of Source Candidates, NOAA Tech. Rep. ERL 380-WPL49, pp. 59, 1976.

Georges, T. M., and G. E. Greene, J. Appl. Meteor., 14, 1303, 1975.

Hines, C. O., Can J. Phys., 38, 1441, 1960.

Hung, R. J., and R. E. Smith, Space Research, 17, 205, 1977a.

Hung, R. J., and R. E. Smith, Space Research, 17, 211, 1977b.

Hung, R. J., and R. E. Smith, Ray Tracing of Gravity Waves as a Possible Warning System for Tornadic Storms and Hurricanes, J. Appl. Meteor., 1978 (in press).

Hung, R. J., T. Phan, and R. E. Smith, Observation of Gravity Waves During the Extreme Tornado Outbreak of April 3, 1974, J. Atmos. Terr. Phys., 1978 (in press).

Hung, R. J., R. E. Smith, G. S. West and B. B. Henson, 9th Conference on Severe Local Storms, pp. 294, Am. Meteor. Soc., Boston, Mass., 1975.

Jones, W. L., J. Geophys. Res., 74, 2028, 1969.

Jones, R. M., and T. M. Georges, J. Acoustic Soc. America, 59, 765, 1976.

Kohl, H., and J. W. King, J. Atmos. Terr. Phys., 29, 1045, 1967.

Matsumoto, S., and T. Akiyama, J. Meteor. Soc. Japan, 47, 255, 1969.

Matsumoto, S., and Y. Tsuneoka, J. Meteor. Soc. Japan, 47, 267, 1969.

Matsumoto, S., K. Ninomiya, and T. Akiyama, J. Meteor. Soc. Japan, 45, 64, 1967a.

Matsumoto, S., K. Ninomiya, and T. Akiyama, J. Meteor. Soc. Japan, 45, 292, 1967b.

Otnes, R. K., Trans. IEEE Audio and Electro-Acoustics AU-16, 330, 1968.

Prasad, S. S., L. J. Schnack, and K. Davies, J. Atmos. Terr. Phys., 37, 1357, 1975.

Sagdeev, R. Z., and A. Galeev, Nonlinear Plasma Theory, W. A. Benjamin, Inc., New York, pp. 250, 1969.

Saunders, P. M., Tellus, 14, 177, 1962.

Shenk, W. E., J. Appl. Meteor., 13, 917, 1974.

Smith, R. E., and R. J. Hung, J. Appl. Meteor., 14, 1611, 1975.

Tepper, M., J. Meteor., 7, 1950.

Tepper, M., Pressure Jump Lines in Midwestern United States, January-August, 1951, Research Paper No. 37, U. S. Weather Bureau, Washington, D. C., pp. 70, 1954.

Uccellini, L. W., Mon. Wea. Rev., 103, 497, 1975.

United States Committee on Extension to the Standard Atmosphere, U. S. Standard Atmosphere 1962, U. S. Government Printing Office, Washington, D. C., pp. 278, 1962.

Vonnegut, B., C. B. Moore, R. P. Espinola, and H. H. Blau, Jr., J. Atmos. Sci., 23, 764, 1966.

Whitham, G. B., in Nonlinear Waves, ed. by S. Leibovich and A. R. Seebass, pp. 139-169, Cornell Univ. Press, 1974.

Yeh, K. C., and C. H. Liu, The Theory of Ionospheric Waves, Academic Press, New York, pp. 464, 1972.

LAGOPEDO - TWO F-REGION IONOSPHERIC DEPLETION EXPERIMENTS

Morris B. Pongratz, Gordon M. Smith,
C. Dexter Sutherland and John Zinn
University of California, Los Alamos Scientific Laboratory
Los Alamos, New Mexico

I. Introduction

At this Symposium on the effects of the ionosphere we would like to report on a very recent experiment wherein we turned the tables and produced an effect upon the ionosphere. The effect produced was a significant depletion of ionospheric plasma within a roughly spherical volume of approximately 100 km diameter. In cooperation with Sandia Laboratories, Albuquerque, Group J-10 of the Los Alamos Scientific Laboratory conducted two chemical release experiments into the F-layer ionosphere over the Hawaiian Islands in early September 1977. These experiments, codenamed Lagopedo (Spanish for ptarmigan), were directed to an investigation of the chemical kinetics which follow an injection of certain molecular species (H_2O, CO_2) into the F-layer. Previous work by Mendillo et al (1975) and Bernhardt (1976) had suggested that such an injection would greatly enhance the O^+ loss rate because molecular charge exchange with these artificial species proceeds at rates several orders of magnitude larger than the ambient, natural loss rates. Enhanced electron depletion would follow via the rapid dissociative recombination of electrons and molecular ions. Because the loss rates are much larger than production rates the net effect would be plasma depletion and a hole in the ionospheric plasma.

The questions about the hole we sought to answer were:

1. Would a plasma depletion hole form?
2. What are the spatial characteristics of the hole?
3. How long does it take to make the hole and how long does it persist?
4. What airglow emissions result from the hole formation?
5. Are the sides of the hole smooth or irregular?
6. What are the effects of a hole in the ionosphere?

The answers to these questions were functions of the characteristics of the expansion of the molecular vapors and the chemistry. To model the chemistry we needed to know the velocity and spatial distribution of the vapors as a function of time. We also needed to know whether or not the released gases clustered and froze, and if they did, what was the characteristic time for sublimation back to the vapor state?

The experiment was an excellent test of the techniques used in computer programs to model the chemical kinetics of the atmosphere, both during the natural, near equilibrium, variations and after perturbations of varying magnitudes. Before the experiment the emphasis in the chemical modeling was upon features which lent themselves to experimental observation such as airglow emission. Model predictions were also used as input for operational decisions such as release height and necessary ionospheric conditions for launch.

II. Theory of Plasma Depletions

We used the diffusive model of Bernhardt to characterize the expansion of the molecular vapors-H_2O and CO_2. In our theoretical calculations we assumed that none of the vapors clustered and froze.

In the F-region the dominant chemical reactions should be O^+ charge exchange

$$O^+ + H_2O \xrightarrow{k_1} H_2O^+ + O \qquad (1)$$

$$k_1 \sim 2.3 \times 10^{-9}\ cm^{-3}s^{-1}$$

$$O^+ + CO_2 \xrightarrow{k_2} O_2^+ + CO \qquad (2)$$

$$k_2 \sim 1.1 \times 10^{-9}\ cm^{-3}s^{-1}$$

followed rapidly by dissociative recombination.

$$H_2O^+ + e \xrightarrow{k_3} OH + H \qquad (3)$$

$$k_3 \sim 1.7 \times 10^{-7}\ cm^{-3}s^{-1}$$

$$O_2^+ + e \xrightarrow{k_4} O + O(^1D) \qquad (4)$$

$$k_4 \sim 1.89 \times 10^{-7}\ cm^{-3}s^{-1}$$

These reactions and species, as well as many others, were included in our 78-species, 1484-reaction chemistry code (Sutherland and Zinn, 1975). Our theoretical effort has not yet included charged particle transport along magnetic field lines.

The model calculations begin by selecting an altitude and starting the code off with CIRA 1965 values for the concentrations of the dominant neutral species for that altitude. Initially the ion and electron concentrations are zero. The model does include the effects of solar photons. Consequently as the model proceeds forward in time and the sun rises and sets an ionosphere forms self-consistent with the solar input and the 78-species and 1484-reactions. After several model days of self-consistently calculating an ionosphere we introduce the artificial Lagopedo release vapors. The model then calculates the temporal variations of the neutral, ionic and electron concentrations at various distances from the release point.

The calculations indicate a rapid depletion of O^+ and electrons. The concentrations of H_2O^+, H_3O^+ and O_2^+ rise following the release. By integrating along a column in space we were able to determine O (^1D) and O (^1S) column densities, and the 630.0 nm and 557.7 nm photon emission rates from these OI metastable states. We did not calculate expected levels of other airglow features such as OH band emissions which might result from reaction (3). We suspect that the excess energy resulting from reaction (3) could appear as hydrogen atoms with kinetic energy in excess of 5eV.

We have also compared the predictions of the many species, many reactions model with the predictions of a model with very simple chemistry. In the complex chemistry model reactions (1) and (3) are the dominant reactions, and because of the rapid rate constant for dissociative recombination in our simple model we assume that electrons are removed at a rate given by

$$\frac{d[e]}{dt} = -k_1 [e][H_2O]$$

where [e] is the electron concentration and $[H_2O]$ is the water vapor concentration.

With this one equation chemistry and Bernhardt's solution to the diffusion equation we can rapidly calculate the temporal and spatial effects of the release. Comparison of the results of the simple model shows that, although at early times the simple model predicts more electron removal, after about 100 s the two models agree in their electron concentration predictions. The simple model is useful for generating two dimensional pictures of the electron hole at many different times, but the complex model can be used to check theory with experiment on details such as airglow intensity or H_3O^+ concentration.

We are quite impressed by the fact that the calculated temporal and spatial concentration profiles are very sensitive to the characteristics of the expansion of the molecular vapors. Consequently significant improvement in the theory will come from better definition of the expansion characteristics.

III. Experiment Description

The two Lagopedo experiment payloads were lofted to altitude above the Kauai Test Facility, Hawaii, by Terrier-Sandhawk rocket systems. The total scientific payload mass was 200 kg, of which 88 kg was a mixture of the high explosives nitromethane and ammonium nitrate which yield only H_2O, CO_2 and N_2 in an ideal detonation. About 2500 moles of H_2O and 500 moles of CO_2 were produced and dispersed by the high explosive detonation.

On board the Sandhawk second stage (but spring separated from the high explosive package after motor burnout and prior to detonation) were four experiments designed to quantitatively measure the effects of the release. These instruments passed through the volume surrounding the detonation seconds after the event.

A thermal ion detector (TID), similar to that which detected ion dropout on the Buaro shaped charge barium injection experiment, was provided by Brian Whalen of the Canadian National Research Council at Ottawa. The TID measured ion flux with good directional but modest (1 in 16) mass resolution. To complement the TID measurements, Charlie Johnson and Gary Sjolander from Ed Szuszczewicz's group at NRL provided Bennett-type ion mass spectrometers with sensitivity in the 9-40 AMU range.

The two additional experiments contained in the Sandhawk section were designed to measure electron properties. Langmuir probes were flown by Dave Winningham of the University of Texas, Dallas, and measured electron concentration and temperature *in situ*. A. V. da Rosa and Paul Bernhardt of Stanford University flew dual frequency (137.35 and 412.05 MHZ), phase coherent beacon transmitters. Using ground based receivers and differential Doppler techniques, electron column densities were measured.

Because of the dynamic nature of the F-layer ionosphere, it was necessary to monitor electron concentration as a function of altitude in real time. Bill Wright of NOAA Boulder installed a Dynasonde on the island of Kauai. The Dynasonde is a state-of-the-art ionospheric sounder with computer control and data processing. During the countdown the Dynasonde provided minute by minute values of three parameters ($[e]_{max}$, h_{max} and thickness) used to describe the ionosphere. A necessary condition for launch was that calculations

indicate that a threshold number of electrons would be removed by the release effects. This number required that h_{max} not be far beyond rocket apogee, that the thickness parameter, or scale height, be greater than about 25 km, and that $[e]_{max}$ not be too low. The Dynasonde was also used to detect the hole created by the releases. It was able to monitor the late time existence and motion of the hole.

Jim Clynch of the University of Texas at Austin and AFGL personnel monitored total electron content between French Frigate Shoals, small islands northwest of Kauai, and the synchronous satellites ATS-3 and SMS-1 by observing the Faraday rotation of the VHF transmitter signals from these satellites. By fortunate coincidence the island ground station, the two ionospheric holes and the satellites were colinear. The effects of the depletions were variations of the amplitude and phase of the electromagnetic waves transmitted from satellites.

Walt Chesnut of SRI International operated a HF-radar on the island of Hawaii. The radar probed the release region in a geometry such that the radar line of sight was perpendicular to the geomagnetic field lines through the hole. With this geometry the radar could observe scattering from field aligned irregularities that were predicted by theories which treated the hole as a "negative barium cloud" which would be subject to the gradient drift instability because of the motion of neutrals across the electron density gradient.

Optical ground stations instrumented to observe the airglow resulting from the chemistry, and the expansion of the sunlit ice cloud were on the islands of Kauai and Maui. These ground stations were operated by LASL and EG&G personnel. The details of the instrumentation and results of the experiments will be discussed in the next section.

IV. Lagopedo Results

Necessary conditions regarding clear weather, the parameters of the ionosphere, functioning instrumentation and communications complicated the operational decisions. In satisfying the minimum requirements of all participating experimenters we had to make compromises with the ideal conditions for some experimenters. Nevertheless we were able to complete two (of two) quite succesful launches and releases. The release and ionospheric parameters of the Lagopedo Uno and Dos experiments are given in table I.

Although investigators from other institutions cooperated in providing a wide variety of diagnostics for Lagopedo experiments, in this report we shall limit the presentation of results to our own optical observations. However, we think it is appropriate to note that the observations our colleagues have reported to us are consistent with the theory of hole formation via the charge exchange-dissociative recombination mechanism proposed by Bernhardt.

Table I

	Uno	Dos
Release Parameters		
Event Time (UT)	0554:00 9/2/77	0604:10 9/12/77
Event Altitude	261.22 km	282.86 km
Event Latitude (geodetic)	22.279° N	22.262° N
Event Longitude	159.982° W	160.351° W
Released in Sunlight	Yes	No
Ionospheric Parameters		
Peak Electron Density	1.0×10^6 e/cm^3	3.4×10^5 e/cm^3
Height of Peak	275 km	323 km
F2 Layer Scale Height	42 km	65 km
Stability	Stable	Weak Spread F

On Lagopedo Uno the expanding cloud of ice crystals was observed with instrumentation atop Mt. Haleakala on Maui. For Lagopedo Dos which was not released in sunlight no ice cloud was observed. The early time ($t < E + 30s$) spectrograph data is dominated by the solar continuum spectrum reflected by the cloud. Intensified imaging cameras were equipped with 455.4 nm interference filters to measure the scattered sunlight. No natural airglow emission lines are in this region of the spectrum. The ice cloud was observed above background for about 100 s. The shape was generally spherical. The apparent (optical) diameter of the ice cloud as a function of time has been fit by a power law formula

$$\text{diameter} = (37\ \text{km})\ t^{0.42} \quad (t < 30s) .$$

The ice cloud diameter was over 100 km in less than 30 s!

Assuming that the ice cloud expands diffusively in a CIRA 1965 neutral atmosphere we find that the diffusion constant needed to explain the observations is

$$D = 7.7 \times 10^{10}\ \text{cm}^2/\text{s} .$$

This value of D is five times larger than the diffusion coefficient for water vapor calculated using Bernhardt's formula and a 1965 CIRA atmosphere.

Because ice cloud brightness decreased in time faster than the 1/t decrease in column density given by diffusion we assume that the ice clusters sublimated, and the molecules returned to the vapor state. We let the number of molecules per cluster, g, vary as

$$g = g_o\ e^{-t/\tau}$$

where g_o is the number in a cluster of initial radius, r_o. As the molecules sublimate the

cluster radius gets smaller and the cross-section for scattering sunlight, $\sigma \equiv \alpha\pi r^2$, decreases, and the ice cloud brightness decreases faster than the column density η. The value of τ necessary to fit the data is $\tau = 16.5s$. The absolute value of the brightness requires the relationship

$$\frac{\alpha x_o}{r_o} \gtrsim 0.1/\text{micron}$$

where x_o is the fraction of the H_2O molecules which form clusters and freeze.

We infer that $r_o < .3$ microns because the spectral measurements indicate increasing spectral radiance with decreasing wavelength. Because $\alpha \sim l$ we conclude that less than 3% of the vapor clustered and froze.

The airglow measurements are also consistent with a large hole diameter. The spectrographic measurements indicate enhancement of the OI emission lines at 557.7 nm and 630.0 nm. OH band emissions were not observed by either of two spectrographs or two intensified imaging cameras equipped with OH near IR interference filters.

The $O(^1D)$ emission to the ground state of OI provided dramatic confirmation of the predictions of the complex chemistry code. Because $O(^1D)$ is a metastable state its radiative lifetime is longer than 100 s. This, coupled with the fact that not all the $O(^1D)$ states are produced promptly at event, gives a theoretical temporal profile for 630.0 nm emission with a slow rise to a maximum level some 90 s after the release and then a similar decay back to pre-release radiance. The actual measured 630.0 nm radiance from both stations for both experiments does reach a maximum value some 90-120 s after the release. Naturally the signal to noise and the confidence level is quite high when one detonates the explosive and then observes peak signal some 90-120 s later!

The spatial profile and peak radiance of the airglow emission at 630.0 nm also support the temporal profile evidence that an ionospheric hole was formed by chemical effects. The emitting region expanded to overfill the fields-of-view of all the imaging instruments. This region was at least 125 km in diameter. The shape was generally spherical. The Lagopedo Uno peak radiance was about a kilorayleigh above background. The Lagopedo Dos peak radiance was about 300 Rayleighs above background.

V. Conclusions

Even though the experiments were performed only last September we are able to give preliminary answers to our questions. A plasma depletion hole in the ionosphere was formed via the predicted chemical processes. The hole is nearly isotropic and Gaussian in profile with a thickness at half depletion of 60 km. The hole is formed in less than 5 minutes and persists for nearly a half hour. Only atomic oxygen airglow was observed. No irregularities were observed along the sides of the hole. In the near future we anticipate publishing more complete and more detailed results of this first successful experiment directed towards producing a hole in the ionosphere.

References

M. Mendillo, G. S. Hawkins, and J. A. Klobuchar, "A Large-Scale Hole in the Ionosphere Caused by the Launch of Skylab," Science 187, 343 (1975).

P. A. Bernhardt, "The Response of the Ionosphere to the Injection of Chemically Reactive Vapors," Stanford Electronic Laboratories Technical Report No. 17, May 1975.

C. Dexter Sutherland and John Zinn, "Chemistry Computations for Irradiated Hot Air," Los Alamos Scientific Laboratory report LA-6055-MS, September 1975.

ION COMPOSITION MEASUREMENTS OF A CHEMICALLY MODIFIED IONOSPHERE - AN INITIAL REPORT

Gary W. Sjolander*, Charles Y. Johnson
Julian C. Holmes, and Edward P. Szuszczewicz

E. O. Hulburt Center for Space Research
Naval Research Laboratory
Washington, D.C. 20375

INTRODUCTION

On the evenings of 1 and 11 September 1977 a Terrier-Sandhawk rocket was launched from the Kauai Test Facility, Barking Sands, Hawaii. The launches were part of Project Lagopedo which was directed by the Los Alamos Scientific Laboratory to test ionospheric response to chemical injection. Each rocket carried a dual payload that consisted of a diagnostic instrument package and a nitromethane - ammonium nitrate explosive package.

The rockets were launched at 1950:54 HST, 1 Sept 1977 and 2000:09 HST, 11 Sept 1977. At an altitude of 90 km the explosive payload was spring separated from the rocket and moved ahead of the instrument section about 0.6 km prior to detonation (the event) in the F_2 region maximum. The event explosively released 53 kg H_2O, 27 kg N_2, and 11 kg CO_2 into the ionosphere. The instrument package then passed through the event region. The Naval Research Laboratory provided an ion mass spectrometer for each instrument section. It is the mass spectrometer results and a discussion of the modification chemistry that are the subjects of this initial report.

INSTRUMENTATION AND FLIGHT PATH

The NRL instrument was a 7-5 cycle Bennett Ion Mass Spectrometer (BIMS) that measured relative positive ion density in the 9 to 50 amu range (Johnson, 1960). Figure 1 shows the BIMS at the top of the instrument payload, looking forward, and mounted off center. Separation of the explosive payload uncovered the spectrometer to ambient ions.

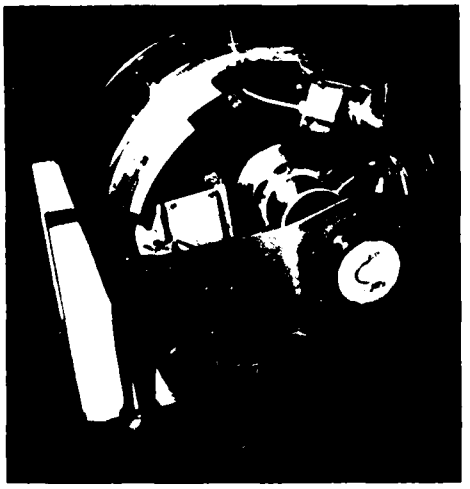


FIGURE 1

Bennett Ion Mass Spectrometer in instrument payload.

Figure 2 shows the flight trajectories of the two instrument sections relative to their positions at event time. The dashed line represents the 33 km radius from the event position. All times are referenced to launch. "Event" on figure 2 is to be interpreted as location of instrument package at event time.

---

*Physics Department
Johns Hopkins University
Baltimore, MD 21218

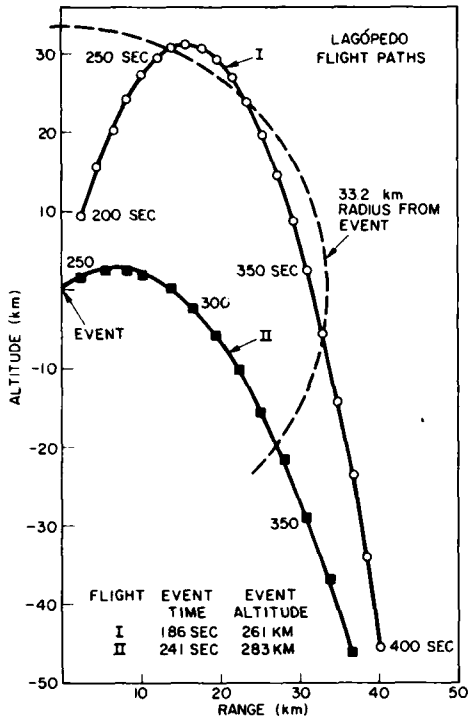


FIGURE 2

Instrument payload trajectories relative to coordinates at event time. The time between indicated position is 10 seconds.

The first rocket had a flight characteristic that allowed a time behavior study of the data apart from its spatial dependence. Notice, that as the payload moved away from the event point its trajectory maintained a radial distance of 33 ± 3 km from 240 to 360 seconds. Theoretical studies of an expanding gas into a tenuous atmosphere (Bernhardt, 1976) indicate that the gas cloud should be spherically symmetric for radial distances less than one scale height (50 km at event altitude) and that the expanding gas peaked in density and remained stable during the 240-360 second time period. This flight path and assumed cloud model form the basis for neutral density calculations of H_2O and CO_2 and for "first look" calculations of relative ion densities.

Flight II had no stationary radial position as did flight I. The second flight moved directly away from the event, and the resulting data are time-space variable throughout the entire trajectory.

IONOSPHERIC MODIFICATION CHEMISTRY

The depletion of the F_2 region charged particles is mainly through a two stage reaction process where O^+ is converted to a molecular ion which, in turn, dissociatively recombines with an electron removing an ion-electron pair from the ionosphere. In an undisturbed F_2 ionosphere O^+ is naturally reduced by reactions with atmospheric O_2 and N_2 (stage one) whose reaction rates are 2×10^{-11} and 1×10^{-12} $cm^3 sec^{-1}$ respectively. By using model values for N_2 and O_2 densities and the above reaction rates the O^+ lifetime in the F_2 region above 260 km is in the order of hours. Since dissociative recombination rates (stage two reaction) are in the order of 10^{-7} $cm^3 sec^{-1}$, it is the first stage, O^+ depletion, that limits the speed by which the natural ionosphere loses charge. The key to modifying the ionosphere is to change the rate by which O^+ is converted to a molecular ion.

The Lagopedo event changed the lifetime of a localized portion of the ionosphere. As the result of the injected gases the O^+ lifetime was reduced to milliseconds within 2 km of the event, and for distances tens of kilometers away the O^+ lifetime was reduced to minutes. Dissociative recombination, a fast reaction compared to normal ionospheric O^+ decay, became the rate limiting reaction in the modified ionosphere. The shortest electron lifetime possible in the near event region (less than 3 km) was 5 seconds; however, 50 to 100 second electron lifetimes were more typical for most of the region probed by the BIMS.

Figure 3 outlines the first order chemistry between the ionospheric O^+ and the injected gases. In the upper lefthand corner atomic O^+ is depicted as interacting with CO_2, N_2, or H_2O released at the event to produce (CO, O_2^+), (N, NO^+), or (O, H_2O^+) respectively. The formation of H_2O^+ is second order reaction of importance where H_2O density is high. The newly formed molecular ions then dissociatively recombine with electrons.

FLIGHT RESULTS

Figures 4-6 show the BIMS relative ion densities in arbitrary units as a function of time after launch for the two flights. Radial distance from the event is shown on a separate scale at the top of each figure. For flight I only, a scale factor of 1.2×10^4 was used to convert relative ion densities to

absolute densities for model calculations. No such scale factor has been worked out for the second flight; however, pre-event electron densities were determined to be 3 x 10^5 cm^{-3} (Bernhardt, 1977). Pre-event measurements of ambient ions were made above 205 km for the first flight and above 101 km on the second flight. Vehicle outgassing of the unpressurized instrument section is believed responsible for the delay in starting ambient ion measurements on the first flight. Vehicle outgassing of H_2O was seen in the mass spectra as H_2O^+ that resulted from H_2O charge exchange with ambient O^+. For both flights the pre-event H_2O^+ ion current was at 8% of the O^+ value. This observation was used to correct H_2O^+ data shown in figures 4 and 5. A similar correction has not been applied to flight II data shown in figure 6.

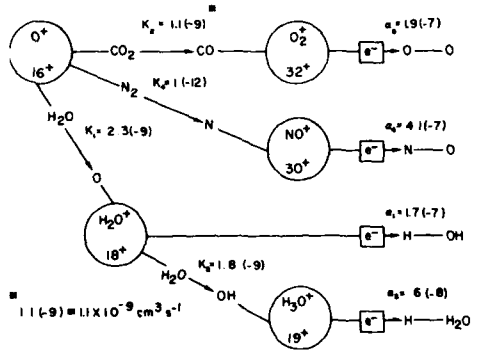


FIGURE 3

Major ion-neutral chemistry for O^+ reacting with H_2O, N_2, and CO_2 (Bernhardt, 1976) produced by NH_4NO_3 - CH_2NO_2 detonation. The reactions $O^+ + H_2O \rightarrow H_2O^+ + O$ followed by $H_2O^+ + e^- \rightarrow H + OH$ represent the primary channel for the loss of ionization in the chemically modified F_2 region.

Near event dynamics. Detonation of the explosive packages occurred at calculated distances of 0.55 and 0.79 km respectively from the 1st and 2nd scientific paylaods. Thirty and sixty milliseconds later, respectively, the BIMS output signals went off scale as if a high pressure was encountered. This arrival time suggests that some material from the event moved past the rocket payloads at velocities in the order of 15 km sec^{-1}. Ion spectra reappeared 1.87 and 4.19 sec after the respective events. Short periods of zero ion current indicate that the near event regions were swept free of ions by material from the events.

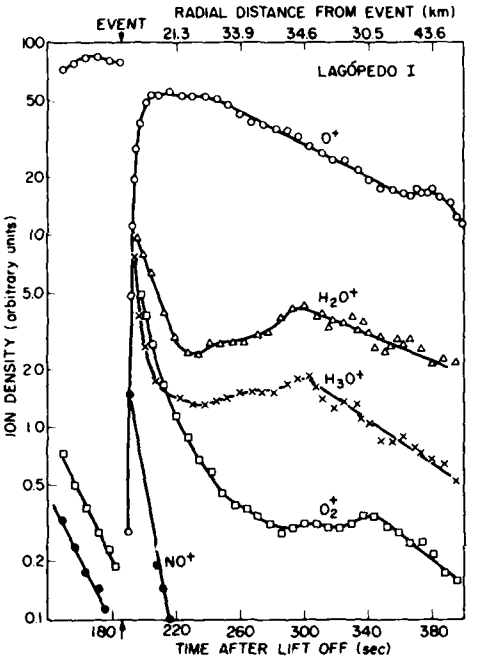


FIGURE 4

Flight I Bennett Ion Mass Spectrometer data for the F_2 region and post-event ionosphere as sampled in time after lift-off and distance from the event. Barking Sands, Hawaii 1950:54 HST, 1 Sept 1977. H_2O^+ data has been corrected for vehicle outgassing. The event occurred at 186 sec after launch at 261 km.

Flight I. Figure 4 shows the pre- and post-event ion composition observations during the first flight (Lagopedo I). The most striking feature in these data is the exponential decrease of O^+ in the 240-360 second time period. (Recall that this period is the near constant radius portion of the flight described above.) This exponential behavior of O^+ indicates that the H_2O and CO_2 densities were basically constant and that no measurable O^+ sources were present. The O_2^+ level in this time period was constant. By using reaction rates shown in Figure 3, the neutral CO_2 density at 33 km radius was determined to be 5 x 10^5 cm^{-3}. The neutral density of H_2O was determined to be 4 x 10^6 cm^{-3} by using the observed O^+ decay time constant of 99 seconds

and the assumptions discussed above. The N_2 injected by the event was less than 10% of the ambient N_2, and it does not contribute in a measurable way to O^+ loss nor to NO^+ enhancement at the 33 km radius. The neutral density values calculated for H_2O and CO_2 at 33 km agree within 30% of event yield estimates (Pongratz, et al., 1977; Bernhardt, 1976).

The dissociation recombination rate constant for H_3O^+ (Fig. 3) determined from these results is a factor of 22 slower than the value of 1.3×10^{-6} $cm^3 sec^{-1}$ at 300 K listed in DNA, (1972). Initial studies of the data showed that the values for K_1, K_2 and α_2(Fig. 3) are consistent with the BIMS observations; however, the relative H_3O^+ density was greater than predicted by the listed α_2 value. By assuming no additional H_3O^+ sources our current understanding of H_3O^+ data for flight I suggests $\alpha_2 = 6 \times 10^{-8} cm^3 sec^{-1}$.

Details of ion composition near the event for flight I are shown in Figure 5. The pre-event level of O^+ was reduced to 0.3% at 4 sec after and 3 km away from the event to produce H_3O^+, H_2O^+, O_2^+, HCO^+ and OH^+. (The secondary OH charge exchange with O^+ is not outlined in Fig. 3.) The formation of HCO^+ was not an expected result, and it suggests a coupling between H_2O and CO_2 which leads to the exothermic reaction

$$H_2O^+ + CO \rightarrow HCO^+ + OH.$$

The reaction rate is not known (Ferguson, 1977).

Flight II. The ion composition results for the Lagopedo II flight are shown in Figure 6. These are raw data and no attempt has been made to study flight II at this writing. These data are shown only for completeness in this initial report and to demonstrate to the reader that there are observational differences between the two Lagopedo experiments.

CONCLUSIONS

Project Lagopedo was the first in situ investigation of an ionospheric region depleted by the injection of chemically reactive gases. The BIMS results showed that the depletion rate of the F_2 ionosphere was enhanced by a factor of 50 to 100 over normal depletion processes. In these tests injected H_2O played the most important role.

The BIMS observations revealed unexpected results as well. The presence of HCO^+ indicates that the reaction between H_2O^+ and CO should be investigated, and the measur[illegible]vels of H_3O^+ suggests that its dis[illegible]ative recombination rate ma[illegible] additional study.

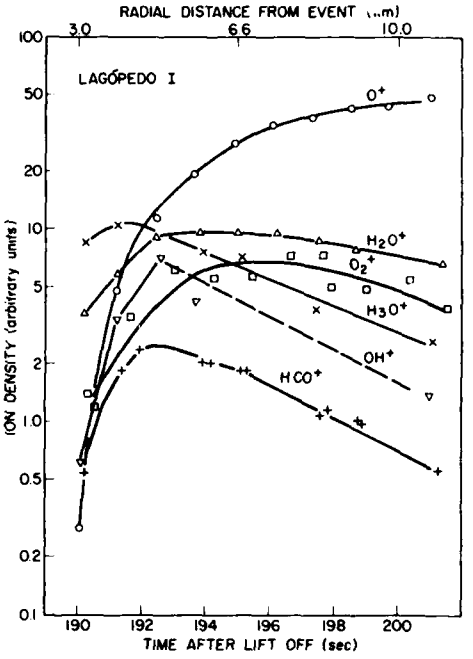


FIGURE 5

Flight I "near event" data showing the ion composition 3 to 10 km from and 4 to 16 seconds after the event.

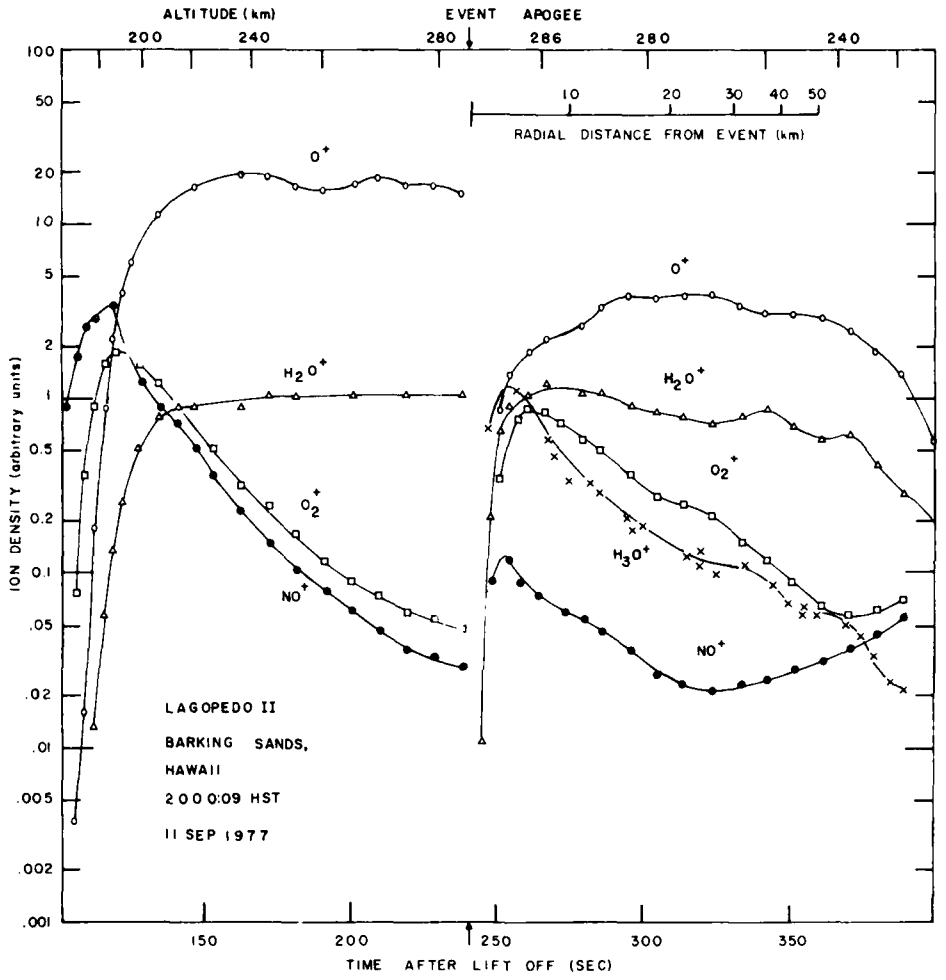


FIGURE 6

Flight II Bennett Ion Mass Spectrometer data for the pre- and post-event ionosphere. The event occurred at 241 seconds after launch at 283 km.

ACKNOWLEDGEMENTS

We wish to thank all participants, Sandia Laboratory, Albuquerque personnel, especially Theodore Krein and Ernest Niper, for the integration and preparation of the Lagopedo rocket payloads, and above all, for the field operations that resulted in two excellent rocket flights. We thank Los Alamos Scientific Laboratory personnel Drs. Morris Pongratz, and Gordon Smith, co-directors, and Dr. Robert Jeffries, scientific advisor for Project Lagopedo.

REFERENCES

Bernhardt, Paul A., The Response of the Ionosphere to the Injection of Chemically Reactive Vapors, Stanford Electronics Laboratories Technical Report No. 17, 1976.

Bernhardt, Paul A., private communication, 1977.

DNA 1948H, Defense Nuclear Agency Reaction Rate Handbook, 2nd. Ed. p. 16-17, March 1972.

Ferguson, E.E., private communication, 1977.

Johnson, Charles Y., "Bennett Radio Frequency Spectrometer", The Encyclopedia of Spectroscopy, Reinhold Publishing Corp., New York, 1960, pp. 587-598.

Pongratz, M. B., G. M. Smith and R. A. Jeffries, private communication, 1977.

Determination of Minimum Scale Sizes in Plasma Cloud Striations

B. E. McDonald, S. L. Ossakow, S. T. Zalesak, and
N. J. Zabusky*

Naval Research Laboratory, Washington, D.C. 20375
*University of Pittsburgh, Pittsburgh, Penn. 15260

ABSTRACT

We propose a fluid model for investigating the possibility that a small scale cut-off is present in plasma cloud striations due to particle diffusion. The model is two dimensional and perpendicular to the magnetic field, and assumes parameters applicable to the F region. The nondimensionalized equations for the model show that the plasma motion is determined by initial plasma distribution and the value of an effective Reynolds number. Numerical simulations carried out for random phase initial conditions and maximum to minimum Pedersen conductivity ratio of 11 to 1 suggest that striation formation stops when the Reynolds number drops below about 400. We calculate a diffusion constant including electron-ion collisions, which can be dominant in high altitude releases. We then scale the results of the numerical simulation and conclude that the fluid model (without kinetic corrections) predicts a minimum scale size in the range 2.4-24 m for a typical choice of ionospheric parameters.

I. INTRODUCTION

Artificial plasma clouds deposited in the ionosphere typically develop visible striations over periods of a few minutes to approximately one hour. Propagation studies carried out simultaneously with plasma cloud experiments have shown that line-of-sight radio communication through a striated region may be subject to attenuations of 10-30 db (Prettie et. al., 1977). Thus local enhancements in the ionospheric plasma density can have a great impact on propagation when small scale structuring occurs. The resolution of the optical data from barium cloud experiments is insufficient to reveal structures smaller than about 100 meters (Davis et. al., 1974). However, recent in situ rocket measurements (Baker et. al., 1977) suggest that structure down to meter sizes may be present. We wish to address theoretically the question of what determines the smallest structures to be found in a striated plasma.

The mechanism by which striations are produced has been fairly well established to be the gradient drift instability (Linson and Workman, 1970; Zabusky et. al., 1973; Scannapieco et. al., 1976). This is an electrostatic fluid process analogous to the Rayleigh-Taylor instability. The results of the model are in need of correction for kinetic effects below scale sizes approximating the ion gyro radius. For a barium ion in the daytime F region, the gyro radius is approximately 7 meters. Before carrying out the kinetic corrections (to be reported elsewhere), one should ascertain whether or not the fluid model tends to generate structures on such a small scale.

II. THE ONE LEVEL, TWO DIMENSIONAL FLUID MODEL

For barium clouds released at altitudes of approximately 200 km or greater, electron and ion collision frequencies are small compared to their gyrofrequencies. As a result (Volk and Haerendel, 1971) the electrical conductivity of the plasma is dominated by the scalar (Pedersen) component,

$$\Sigma \simeq n \frac{ec}{B} \frac{\nu_{in}}{\Omega_i} \qquad (1)$$

where n, e, c, B, ν_{in}, and Ω_i are respectively the ion number density, electronic charge, speed of light, magnetic field strength, ion-neutral collision frequency, and ion gyrofrequency. Assuming all parameters in (1) to be constant except n, Σ obeys a continuity equation,

$$\frac{\partial \Sigma}{\partial t} = - \nabla \cdot (\underline{V} \Sigma) + K \nabla^2 \Sigma, \qquad (2)$$

where $\underline{V}$ is the plasma drift velocity, and K is a diffusion constant used to model the effects of electronic collisions. The value of K will be estimated later. Let us adopt Cartesian coordinates (x,y,z), with z in the direction of the (constant) magnetic field. We assume that all variables depend only on x and y. We express the electric field as

$$\underline{E}\ (x,y) = E_o\ \hat{y} - \nabla\phi\ (x,y) \qquad (3)$$

where E_o is the constant ambient electric field and ϕ is the potential due to the presence of the cloud. It is convenient to use a frame moving with the ambient plasma drift velocity, E_o c/B $\hat{x}$. In this frame, we have a net plasma drift velocity

$$\underline{V} = -\ c/B\ \nabla\phi \times \hat{z} \qquad (4)$$

The equation can be viewed as a solution of the electron momentum equation in which inertia and collision terms have been neglected. The collisional correction has been included in an approximate way in the diffusion term in (2). The set of fluid equations is closed in this simple model by quasineutrality; i.e., the constraint that the electric current be divergenceless:

$$\nabla\cdot\ (\Sigma\nabla\phi) = \underline{E}_o\cdot\nabla\Sigma \qquad (5)$$

The more sophisticated two layer model Scannapieco et. al (1976) allows polarization currents generated by the cloud to flow along the magnetic field direction and close in the lower ionosphere, where cross-field conductivities are high. However, the lower layer becomes unimportant when the cloud's conductivity is high enough to let current loops close in the cloud itself. Some of the barium releases at altitudes of approximately 200 km are known to have integrated conductivities as high as 15 times that of the ionosphere. For such large clouds, the one-level model should be adequate. The one-level model is simple enough to allow a convenient scaling law to emerge. The small number of variables also expedites numerical solution of the equations of motion.

III. SCALING THE EQUATIONS OF MOTION

Equations (2), (4), and (5) can be put into dimensionless form as follows. Let

$$\begin{aligned} \underline{x} &= L_o\,\underline{x}' \\ t &= t_o t' \\ \Sigma &= \Sigma_o \Sigma' \\ \underline{V} &= V_o \underline{V}', \\ \phi &= L_o E_o \phi' \end{aligned} \qquad (6)$$

where L_o is an arbitrary scale, length, and

$$\begin{aligned} V_o &= cE_o/B, \\ t_o &= L_o/V_o \end{aligned} \qquad (7)$$

In (6) quantities with zero subscript are dimensional constants and all primed variables are dimensionless. Upon expressing (2), (4), and (5) in terms of the dimensionless primed variables and dropping primes from x', t', Σ', $\underline{V}'$, and ϕ', we have

$$\frac{\partial\Sigma}{\partial t} = -\nabla\cdot(\underline{V}\Sigma) + K'\nabla^2\Sigma \qquad (8)$$

$$\underline{V} = -\ \nabla\phi \times \hat{z} \qquad (9)$$

$$\nabla\cdot(\Sigma\nabla\phi) = \hat{y}\cdot\nabla\Sigma \qquad (10)$$

$$\text{where } K' = \frac{K}{L_o V_o} \qquad (11)$$

Note that K' is just the inverse of an effective Reynolds number, with the diffusion constant K in place of the usual kinematic viscosity. The fact that all quantities in (8)-(10) are dimensionless means that physical systems of different sizes will evolve in the same way, providing initial conditions and boundary conditions are analogous, and K' has the same value.

Thus the model (8)-(10) can be used to answer the following crucial question: What criterion determines whether or not structures in a striated plasma cloud will develop into structures of smaller scale? It becomes clear from (8) and (11) that as the length scale becomes smaller, diffusion becomes more important. For sufficiently small scales, we expect an equilibrium to be reached between the tendency toward finer structuring and the smoothing out effect of diffusion. Our approach to finding this scale will be to carry out a set of numerical simulations based on (2)-(5) or equivalently (8)-(10) and hopefully to identify an approximate value for K' at which diffusion is just able to halt further structuring of previously formed striations. Knowledge of a critical value of K' combined with estimates for the drift speed and diffusion coefficient allow calculation of a minimum length scale from (11). Drift speeds are fairly well known to be of order 100 m/sec. However, there are no direct measurements of the diffusion constant, so it must be calculated.

IV. CALCULATION OF THE DIFFUSION CONSTANT

Neglecting ion and electron inertia, the momentum equations for electrons and ions are

$$0 = \Omega_i \underline{V}_i \times \hat{z} - \nu_{in}\underline{V}_i + \nu_{ie}(\underline{V}_e - \underline{V}_i) + \frac{e}{m_i}\underline{E} - \frac{kT_i}{m_i}\nabla n/n \qquad (12)$$

$$0 = -\Omega_e \underline{V}_e \times \hat{z} - \nu_{en}\underline{V}_e + \nu_{ei}(\underline{V}_i - \underline{V}_e) - \frac{e}{m_e}\underline{E} - \frac{kT_e}{m_e}\nabla n/n \qquad (13)$$

For the purposes of this derivation we have adopted a coordinate system at rest with the neutral atmosphere. In (12) and (13), the $\underline{V}$'s are fluid velocities, with subscripts i and e referring to ions and electrons. Similarly, ν stands for momentum transfer collision frequency, T for temperature, and m for particle mass. The gyrofrequencies are

$$\Omega_i = \frac{eB}{m_i c} \tag{14}$$

$$\Omega_e = \frac{eB}{m_e c} \tag{15}$$

Note that we have retained the electron-ion collision frequency ν_{ei} in (13). For conditions typical of a 200 km release, ν_{en} is of order 10^2 sec^{-1}, and ν_{ei} is of order 10^3 sec^{-1}. This is because the ions, although less numerous than the neutrals, interact with electrons via a Coulomb cross section. However the ν_{ie} term in (12) is typically much smaller than the ν_{in} term because of the low mass of the electron. Neglecting ν_{ie}, we find from (12) and (13)

$$n\underline{V}_i = M_i^{-1}\left(\frac{e}{m_i} n\underline{E} - \frac{kT_i}{m_i}\nabla n\right) \tag{16}$$

$$n\underline{V}_e = M_e^{-1}\left(-\frac{e}{m_e} n\underline{E} - \frac{kT_e}{m_e}\nabla n\right) + \nu_{ei} M_e^{-1} n\underline{V}_i \tag{17}$$

where

$$M_i^{-1} = \begin{pmatrix} \nu_{in} & \Omega_i \\ -\Omega_i & \nu_{in} \end{pmatrix} / (\Omega_i^2 + \nu_{in}^2) \tag{18}$$

$$M_e^{-1} = \begin{pmatrix} \nu_e & -\Omega_e \\ \Omega_e & \nu_e \end{pmatrix} / (\Omega_e^2 + \nu_e^2), \tag{19}$$

and

$$\nu_e = \nu_{en} + \nu_{ei} \tag{20}$$

Let us invoke quasineutrality in the form

$$\nabla\cdot(n\underline{V}_i) = \nabla\cdot(n\underline{V}_e), \tag{21}$$

and take the divergence of (16) and (17). We assume all parameters except n, $\underline{V}_i$, $\underline{V}_e$, and $\underline{E}$ to be constant. We also assume all vectors have only x and y components. We may then eliminate $\nabla\cdot n\underline{E}$ between (16) and (17), and find, to lowest order in ν_e/Ω_e,

$$-\nabla\cdot(n\underline{V}_i) = -\frac{c}{B}\nabla\cdot(n\underline{E}\times\hat{z}) + K\nabla^2 n, \tag{22}$$

where

$$K = \frac{c}{B}\frac{k(T_e + T_i)/m_e}{\nu_{en}\nu_{in}/\Omega_i}. \tag{23}$$

We use the convenient approximations (Banks and Kockerts, 1973),

$$\nu_{ei} \approx (34 + 4.18 \log_{10}(T_e^3/n_e))\, n_e T_e^{-3/2} \tag{24}$$

$$\nu_{en} \approx 1.8 \times 10^5\, p_n \tag{25}$$

$$\nu_{in} \approx 4 \times 10^{-10}\, n_n, \tag{26}$$

where p_n and n_n are the neutral pressure and number density, T_e is in °K and all other quantities are in cgs units. For typical daytime conditions at 200 km, we have T_e = 950°K, T_i = 700°K, $p_n = 8.64 \times 10^{-4}$ erg cm^{-3}, n_e = 7.0 × 10^5 cm^{-3}, $\Omega_e = 9 \times 10^6$ sec^{-1}, and Ω_i (singly ionized Ba) = 36 s^{-1}. We know that for typical barium releases $10^6 \lesssim n_e \lesssim 10^7$ cm^{-3} (Baker et al., 1977; Zabusky et al., 1973). Thus K is expected to lie between 0.6 and 6 m^2/sec. When the calculation is repeated using data appropriate to 170 km and 150 km, K changes by less than a factor of 2. This is because ν_e is dominated by ionic collisions, and is insensitive to moderate changes in the neutral background.

V. NUMERICAL SIMULATION

In order to estimate a critical value for K', we have solved numerically equations (2)-(5) using diffusion constants K determined so as to give one of four desired K' values. The results to be presented here were obtained using a mesh of 162 x 42 grid points with constant grid spacing in both directions of 310 meters and doubly periodic boundary conditions. Equation (2) was advanced in time using a flux-corrected leapfrog-trapezoid scheme, which is basically second order in time, fourth order in space. Second order centered differences were used in (4) and (5). The elliptic potential equation (5) was solved iteratively using the Chebychev method (Varga, 1962; McDonald, 1977). The initial condition (t=0) was $\Sigma = 1 + 10\, e^{-(x/8km)^2}(1 + \varepsilon(x,y))$, where ε was generated from a k^{-4} power spectrum and random phases. The root mean square value of ε was 0.03. (The dimensionality of Σ may be scaled out of (2) and (5) with no consequence). B is taken to be 0.5 gauss and $V_o = cE_o/B \approx$ 100 m/sec.

We are not concerned here with details of relating an initial condition to a final state achieved after a long period of development. Rather we wish to determine whether or not a given configuration tends toward spontaneous generation of structure smaller than that which is present. Thus we view the time evolution to be presented here as a continuous sequences of initial conditions, hopefully at some point descriptive of structure in a striated cloud. We seek a

value of the Reynold's number just low enough that a long finger-like striation will not evolve into new and smaller scale structures.

Since we hope to identify a critical value of the Reynold's number,

$$R_e \equiv \frac{V_o L_o}{K}$$

$$= 1/K' \qquad (27)$$

we have carried out numerical simulations so as to keep R_e constant throughout a given computer run. Throughout each run presented here, V_o is constant, and L_o is determined self-consistently from the conductivity profile:

$$L_o(t) = \{\Sigma\ (\Sigma(x,y,t)^2/\Sigma(\nabla\Sigma(x,y,t))^2\}^{1/2} \qquad (28)$$

where Σ without the argument (x,y,t) refers to summation over all grid points. This definition of L_o is somewhat arbitrary, but it does have the desirable feature of being sensitive to small scale structure. Since L_o changes in time, K must change if R_e is to be held constant. Thus during a computer run, we evaluate L_o at each timestep and determine K from (27). It may be more conventional to keep K fixed and let R_e vary with L_o, but our approach allows a more direct identification of a critical R_e.

The results of four computer runs ($R_e = \infty$, 800, 400, and 300) are shown at five times (t = 0, 240, 360, 480, and 720 sec) in Figures 1-5. All four cases show the breakup of the initial slab geometry into structure elongated in the x direction. However, only the top two cases $R_e = \infty$ and 800 show a clear tendency toward further structuring. The $R_e = 400$ case appears marginal, and the $R_e = 300$ case is clearly held back by diffusion. Therefore we tentatively propose that the critical Reynold's number for the initial condition presented here ($\Sigma_{max}/\Sigma_{min} \approx 11$) is approximately 400. We intend to carry out several more simulations using finer resolution and different initial conditions to determine if $R_e = 400$ is in fact universal to the short scale limit of the striation process. Using $R_e = 400$, and $V_o = 100$ m/sec. in (27), we find the following range for L_o:

For n: $10^6 - 10^7$ cm^{-3} (29)

and K: 0.6 - 6 m^2/sec.

L_o: 2.4 - 24 m

VI. CONCLUDING REMARKS

The one-level, two dimensional striation model presented here predicts, for the particular set of initial conditions considered, minimum length scales of 2.4-24 meters for large barium clouds near 200 km altitude. These scales are small enough that kinetic effects may be important. The kinetic corrections have not yet been calculated, and will be reported elsewhere. However, the tendency of the fluid mechanism to generate structures this small is in apparent agreement with recent rocket measurements (Baker et. al., 1977) and with the attenuation (probably due to strong scattering or diffraction) observed in propagation experiments (Prettie et. al., 1977). Further work needs to be done to determine the sensitivity of the results to initial conditions. One should also investigate the effects of ion inertia (which may be important at higher altitudes), and electrostatic coupling to the lower ionosphere in the case of small clouds.

ACKNOWLEDGMENT

This work was supported by the Defense Nuclear Agency.

REFERENCES

Baker, K. D., L. C. Howlett, J. C. Ulwick, D. Delorey, and N. Grossbard, Measurements of Electron Density Structure in Barium Clouds - Project STRESS, Utah State Univ. Report, May 5, 1977.

Banks, P. M., and G. Kockarts, Aeronomy (Part A), Academic Press, New York, 1973.

Davis, T. N., G. J. Romick, E. M. Wescott, R. A. Jeffries, D. M. Kerr, and H. M. Peek, Observations of the development of striations in large barium ion clouds, Planet. Space Sci., 22, 67, 1974.

Linson, L. M., and J. B. Workman, Formation of striations in ionospheric plasma clouds, J. Geophys. Res. 75, 3211, 1970.

McDonald, B. E., Explicit Chebychev-iterative solution of nonself-adjoint elliptic equations on a vector computer, NRL Memo Report 3541, June, 1977.

Prettie, C., A. Johnson, J. Marshall, T. Grizinski, and R. Swanson, Project STRESS satellite communication test results, AFAL Technical Report 77-158, July, 1977.

Scannapieco, A. J., S. L. Ossakow, S. R. Goldman, and J. M. Pierre, Plasma cloud late time spectra, J. Geophys. Res. 81, 6037, 1976.

Varga, R. S., Matrix Iterative Analysis, Prentice Hall, Englewood Cliffs, N.J., 1962.

Volk, H. J. and G. Haerendel, Striations in ionospheric ion clouds, J. Geophys. Res. 76, 4541, 1971.

Zabusky, N. J., J. H. Doles, and F. W. Perkins, Deformation and striation of plasma clouds in the ionosphere, 2, Numerical simulation of a nonlinear two-dimensional model, J. Geophys, Res., 78, 711, 1973.

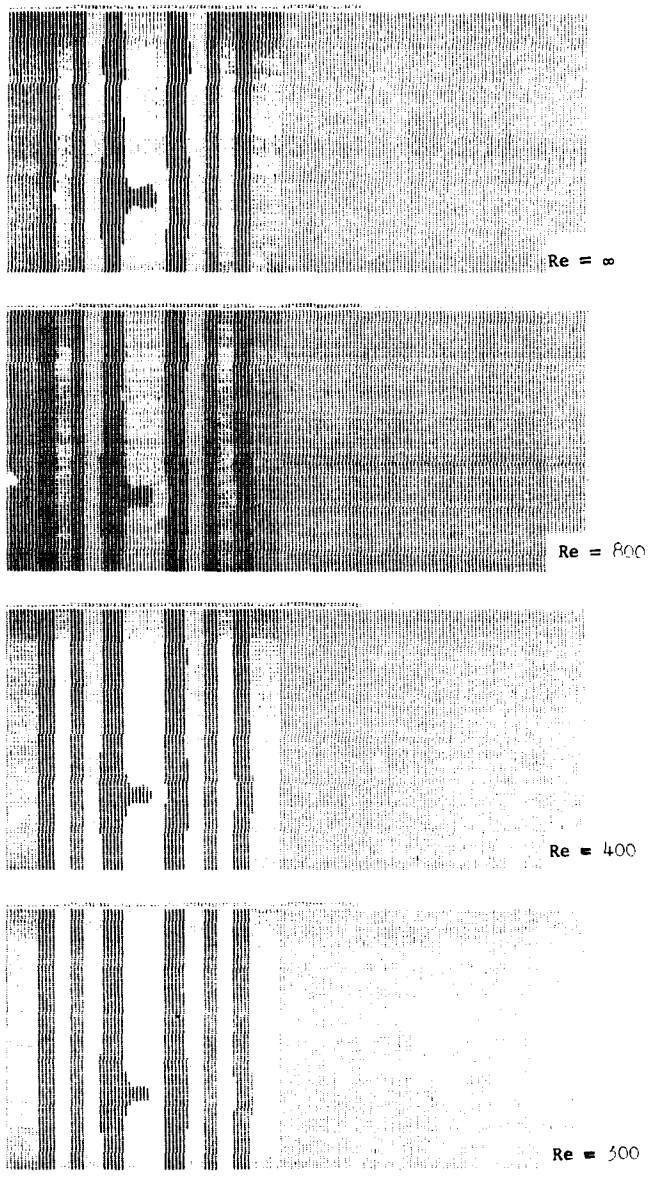


Figure 1. Initial conditions. Contours of constant Σ are boundaries between light and dark areas. A maximum Σ of 11.32 relative to the background occurs in the light vertical band containing the most noticeable bump. The vertical (y) extent of each rectangular box is 12.4 km; the horizontal (x) extent is 49.9 km. The finite difference grid is 42 by 162 points in the y and x directions respectively. Boundaries are doubly periodic. The Reynold's numbers are (top to bottom) $R_e = \infty$, 800, 400, 300.

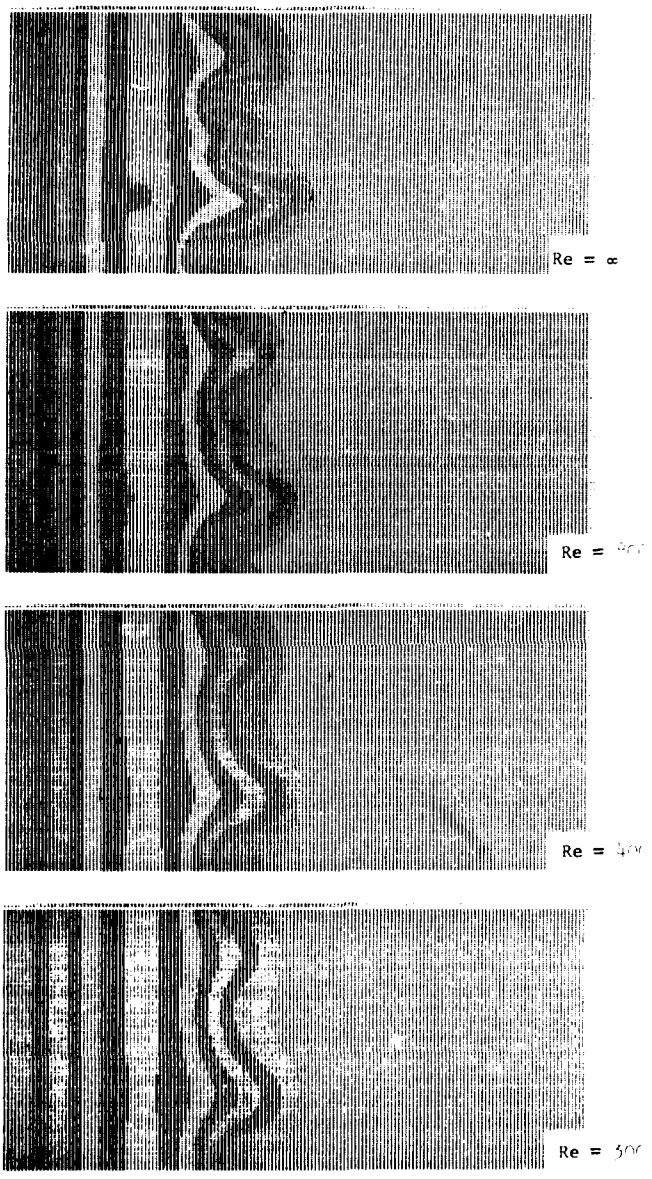


Figure 2. ζ contours at t = 240 seconds.

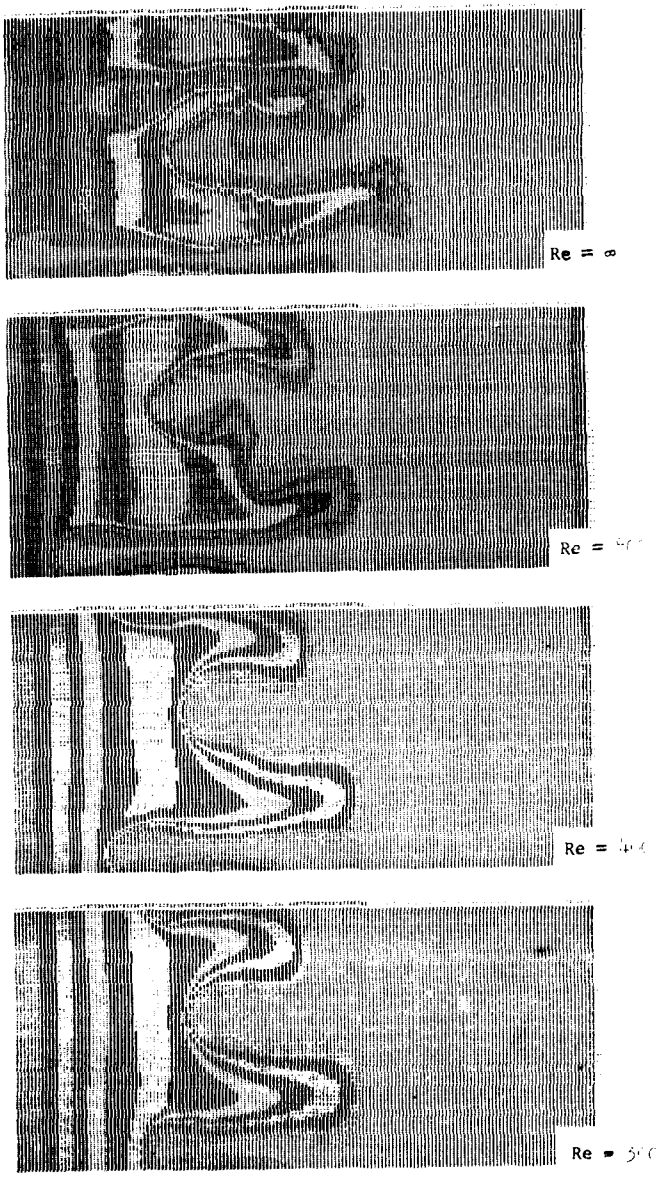


Figure 3. Ξ contours at t = 360 seconds.

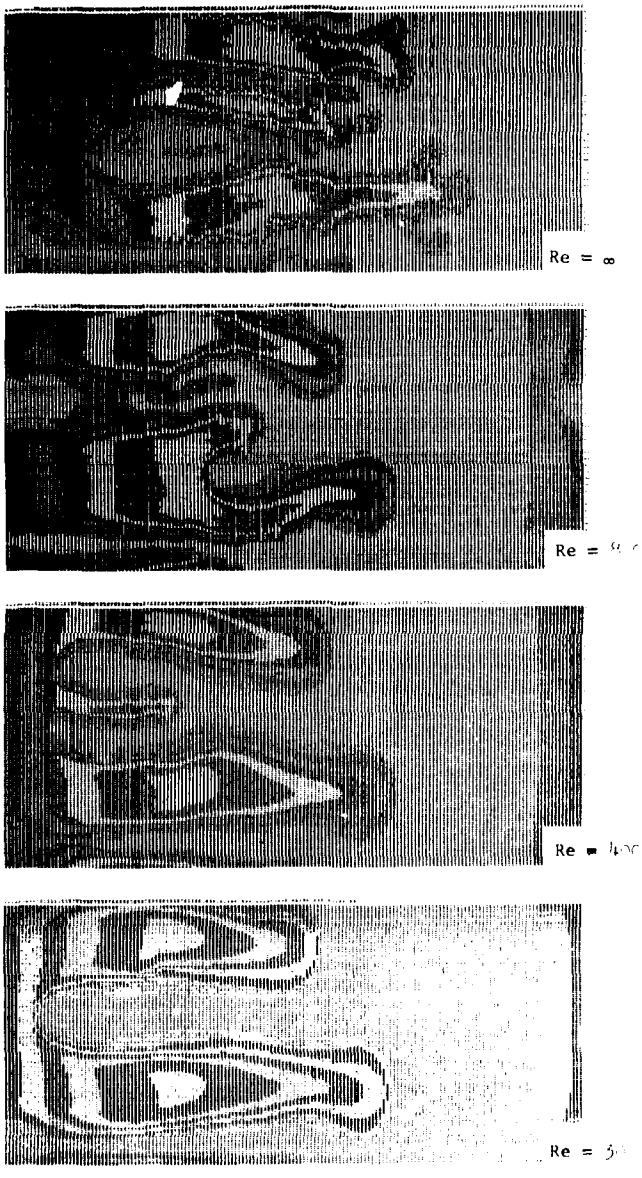


Figure 4. ξ contours at t = 480 seconds.

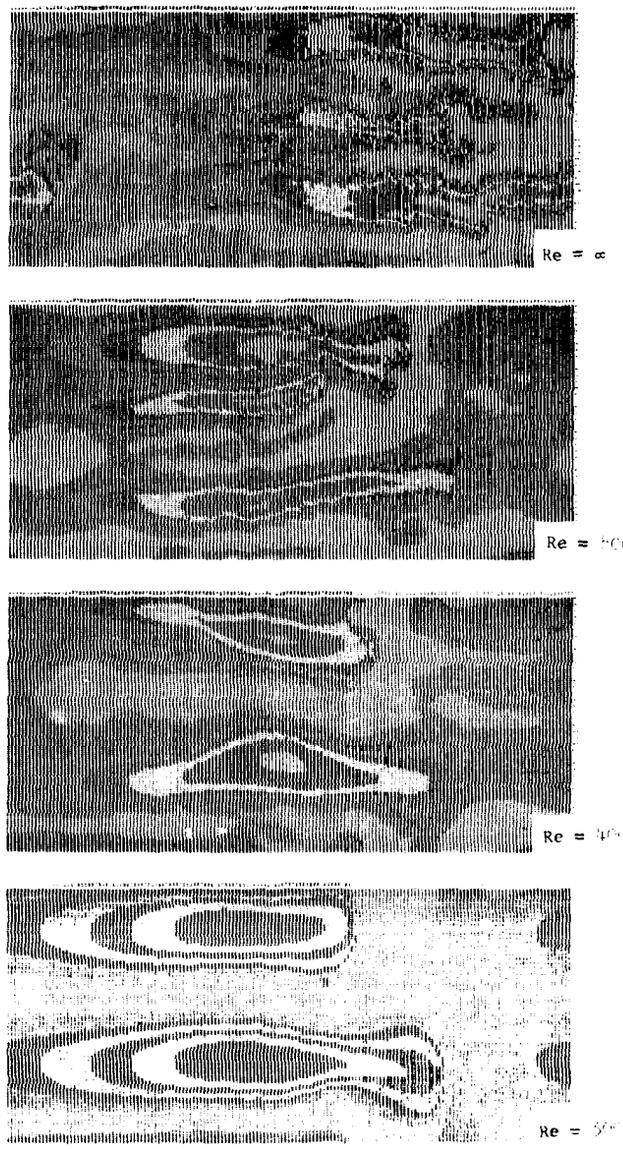


Figure 5. Ξ contours at T = 720 seconds.

EFFECTS OF THE IONOSPHERE ON NAVIGATIONAL AND POSITIONING SYSTEMS
FULL WAVE SOLUTIONS

E. Bahar and B. S. Agrawal
Electrical Engineering Department
University of Nebraska
Lincoln, Nebraska 68588

ABSTRACT

Full wave, numerical and analytical solutions are obtained for LORAN C signals reflected or transmitted through an inhomogeneous anisotropic layer such as the ionosphere. The transfer functions used in these solutions are obtained by using a nonsingular transformation matrix to convert Maxwell's equations (for the transverse electromagnetic field components) into a set of loosely coupled first order differential equations for the forward and backward traveling ordinary and extraordinary wave amplitudes. These transformation matrices which are also suitable for critical coupling regions (characterized by strong reflections or coupling between the ordinary and extraordinary waves) are constructed through the use of generalized characteristic vectors.

Analytical and numerical techniques are used to obtain the field distortions and instantaneous phase anomalies due to the medium. These are used to determine the time of the third zero crossing which is generally regarded as the effective arrival time of the signal.

1. INTRODUCTION

Numerical and analytical expressions are obtained for the transient reflected and transmitted horizontally and vertically polarized waves excited by a vertically or a horizontally polarized LORAN C wave incident upon an inhomogeneous anisotropic layer, such as the ionosphere. To this end it is necessary to determine the transfer functions, (the reflection and transmission coefficients) as a function of frequency and to apply Fourier transform techniques.

Full wave solutions for the (steady-state) transfer functions are used. These full wave solutions are based on a generalized WKB method that employs generalized characteristic vectors to determine the elements of a nonsingular transformation matrix (Bahar 1976). This transformation matrix is used to convert Maxwell's equations for the transverse components of the electric and magnetic fields into a set of loosely coupled first order differential equations for the wave equations. When the four roots of the Booker equation that characterize the inhomogeneous anisotropic media are sufficiently distinct, the solutions for the wave amplitudes are the familiar WKB solutions (Budden 1962). However these WKB solutions fail in critical coupling regions where the characteristic roots tend to merge. For these regions, numerous analytic and numerical techniques employing several special functions have been used to obtain suitable solutions for the fields. The generalized WKB approach used here, does not use these special mathematical functions for the critical coupling regions. Instead, a suitable set of linearly independent wave amplitudes are introduced through the use of generalized characteristic vectors (Bahar 1976).

An approximate analytic expression for the transient response is obtained through a suitable expansion of the transfer function about the carrier frequency. This analytical solution does not account for the singularities of the transfer function. However, since the third zero crossing of the response is regarded as the effective arrival time of the received signal, the analytical solutions can be used to determine the signal delay as well as the time of the third zero crossing. The fast Fourier transform techniques are also used to determine the transient response, the amplitude distortion (due to the singularities of the transfer function) as well as the phase anomaly which determines the effective arrival time of the response.

In Section 5 several illustrative examples are presented for the transfer functions (as functions of frequency) which satisfy the adjoint reciprocity relationships (Kerns 1976). These transfer functions are used to obtain the transient response for LORAN C excitations.

2. STEADY-STATE SOLUTIONS FOR THE WAVE AMPLITUDES

Maxwell's equations for the transverse electromagnetic field components in horizontally stratified anisotropic media, such as the ionosphere, can be expressed in separable form as follows (Budden 1962):

$$e' = \frac{de}{du} = -\frac{1}{ik}\frac{de}{dz} = Te \qquad (2.1)$$

in which

$$e \equiv \begin{pmatrix} E_x \\ -E_y \\ H_x \\ H_y \end{pmatrix}, \quad u-u_o = -ik(z-z_o) \qquad (2.2)$$

$$T = \begin{pmatrix} -\frac{sM_{zx}}{1+M_{zz}} & \frac{sM_{zy}}{1+M_{zz}} & 0 & \frac{c^2+M_{zz}}{1+M_{zz}} \\ 0 & 0 & 1 & 0 \\ \frac{M_{yz}M_{zx}}{1+M_{zz}} - M_{yx} & c^2 + M_{yy} - \frac{M_{yz}M_{zy}}{1+M_{zz}} & 0 & \frac{sM_{yz}}{1+M_{zz}} \\ 1 + M_{xx} - \frac{M_{xz}M_{zx}}{1+M_{zz}} & \frac{M_{xz}M_{zy}}{1+M_{zz}} - M_{xy} & 0 & \frac{-sM_{xz}}{1+M_{zz}} \end{pmatrix} \qquad (2.3)$$

and $k = \omega(\mu_o\varepsilon_o)^{\frac{1}{2}}$ is the free space wave number. The z axis is normal to the horizontally stratified media and the direction of the incident wave normal in free space is

$$\bar{n} = (sa_x + ca_z) \qquad (2.4)$$

where s and c are the sine and cosine of the angle of incidence θ for propagating waves. Thus the fields are independent of y and assuming time harmonic excitations, the factor $\exp(-iksx)\exp(i\omega t)$ is suppressed throughout this work. The transverse electric field components are E_x and E_y while H_x and H_y are the transverse components of the normalized magnetic field $\eta\bar{\mathcal{H}}$ where $\eta = (\mu_o/\varepsilon_o)^{\frac{1}{2}}$ is the free space wave number and $\bar{\mathcal{H}}$ is the magnetic field. The elements of the susceptibility matrix M for the ionosphere are $M_{\alpha\beta}$ $(\alpha,\beta=x,y,z)$ (Budden, 1962).

When the parameters of the susceptibility matrix M (the plasma, collusion and gyrofrequencies as well as the direction of the earth's magnetic field) are slowly varying functions of z and the medium is devoid of critical coupling regions, the matrix transformation

$$e = Sf \qquad (2.5)$$

can be used to convert Maxwell's equations (2.1) into a set of loosely coupled equations for the upward and downward propagating ordinary and extraordinary wave amplitudes $f_i(i=1,2,3,4)$

$$f' = \Delta f + \Gamma f \qquad (2.6)$$

in which

$$\Delta = S^{-1}TS, \Gamma = -S^{-1}S' \qquad (2.7)$$

The matrix transformation $S^{-1}TS$ yields a diagonal matrix, Δ, whose elements are the characteristic values of the matrix T. These characteristic values q_i are solutions to the Booker quartic (Budden 1962).

$$\det(T-q_iI) = |T-q_iI| = 0 \qquad (2.8)$$

in which I is the identity matrix of rank n=4. The columns of the 4 x 4 matrix S are the four characteristic vectors that satisfy

$$TS^i = q_iS^i \qquad (2.9)$$

Since the medium is varying, T, q_i and S^i are functions of z in (2.6). To numerically solve (2.6) for the region

$$-\frac{\Delta z}{2} \le z - z_c \le \frac{\Delta z}{2}$$ it is convenient to evaluate S^i using (2.9) at $z=z_c$ for the entire partition of width dz. Thus for each partition $|z-z_c| \le \frac{\Delta z}{2}$, (2.6) is replaced by

$$f' = S^{-1}TSf \equiv S^{-1}(T_c + T_\delta)Sf$$

$$= (\Delta_c + \gamma)f \equiv Cf \qquad (2.10)$$

in which

$$T_c = T(z_c),\ \Delta_c = \Delta(z_c) \text{ and } \gamma = S^{-1}T_\delta S \qquad (2.11)$$

The elements, γ_{mn}, of the coupling matrix γ vanish at $z=z_c$ since $T_\delta = T-T_c$ vanishes at the center of the partition of width Δz. When the fourth order Runge-Kutta method (Abramowitz and Stegun 1964) is used to solve (2.10), the partition size must be chosen such that $|C_{mn}k\Delta z| << 1$, where C_{mn} are the elements of the matrix C.

For critical coupling regions where the characteristic values q_i tend to merge, the transformation matrix S obtained by using (2.9) becomes near singular and the coupling coefficients γ_{mn} become very large. In these regions, the interaction between the characteristic waves f_i constitute large reflections or strong coupling between the ordinary and extraordinary waves traveling in the same direction. Thus the transformation based on (2.9) is not suitable for critical coupling regions. For these regions it has been shown recently that it is possible to construct a non-singular transformation matrix S by using generalized characteristic vectors A^{pi} of rank p (Bahar 1976). Thus for instance if for the matrix $T=T_o$ the characteristic equations $|T_o-qI| = 0$ has a root q_i of multiplicity m, it has been shown that there exists m linearly independent generalized characteristic vectors A^{ki}, of rank $k \leq \mu \leq m$. The integer μ is called the index of the characteristic value q_i; it is the smallest integer such that the rank of $(T_o-q_iI)^{\mu} \equiv r(T_o-q_iI)^{\mu} = n-m$ (Friedman 1956, Bronson 1969, Brogan 1974). The rank of the system of equations for the problem considered here (2.1) is n=4. The number of linearly independent generalized characteristic vectors of rank k is given by δ_k, where

$$\left.\begin{aligned} \delta_k &= r(T_o-q_iI)^{k-1} - r(T_o-q_iI)^{k}, k=1,2\ldots\mu \\ \text{and} \quad r(T_o-q_iI)^{o} &= r(I) = n \end{aligned}\right\} \quad (2.12)$$

For the characteristic value q_i of multiplicity m, N chains of generalized characteristic vectors A^{qi} of length $\ell_j \leq \mu \leq m$, $\left(\sum_{j=1}^{N} \ell_j = m\right)$ can be found such that

$$(T_o-Iq_i)A^{1i} = 0, (T_o-Iq_i)A^{qi} = A^{q-1i} \quad (2.13a)$$

and

$$(T_o-Iq_i)^{\ell}A^{\ell i} = 0 \quad (2.13b)$$

Corresponding to one such chain of length ℓ_j we can construct ℓ_j columns of the transformation matrix S;

$S^i, S^{i+1} \ldots S^{i+\ell_j-1}$ such that

$$S^{i+p-1} = u \sum_{q=1}^{p} \frac{A^{qi}}{(p-q)!u^q}, \quad p=1,2,\ldots\ell_j \quad (2.14)$$

The ℓ_j generalized characteristic values associated with the chain of generalized characteristic values are

$$\Delta_{i+p-1} = q_i + \frac{p-1}{u}, \quad p=1,2\ldots\ell_j \quad (2.15)$$

In this case even though T_o and A^{qi} (2.13) are independent of z, the transformation matrix S is z dependent. It can be shown that the singular matrix S satisfies the equation (Bahar 1976)

$$S^{-1}(T_oS - S') = \Delta \quad (2.16)$$

in which the elements of the diagonal matrix Δ are given by (2.15). Thus for critical coupling regions, where the characteristic values q_i tend to merge (2.1) can be transformed to the following matrix equation

$$f' = S^{-1}(TS-S')f = S^{-1}\left[(T_o+T_\delta)S-S'\right]f = (\Delta+\gamma)f \equiv Cf \quad (2.17)$$

in which

$$T_\delta = T-T_o \text{ and } \gamma = S^{-1}T_\delta S \quad (2.18)$$

Hence, at a critical coupling point where $T=T_o$, the elements of the coupling matrix γ_{mn} vanish.

Since the transverse components of the electromagnetic fields are continuous for sourceless regions, at the interface $z=z_T$ between two adjacent partitions the continuity condition is

$$e(z_T) = S^{-}f(z_T^{-}) = S^{+}f(z_T^{+}) \quad (2.19)$$

where S^- and S^+ are the transformation matrices for $z=z_T^-$ and $z=z_T^+$ respectively and $z_T^{\pm} = \lim_{\delta \to o} z_t \pm \delta$ ($\delta>0$).

3. NUMERICAL SOLUTIONS FOR THE TRANSFER FUNCTIONS

The transfer functions for an inhomogeneous anisotropic dielectric layer are the horizontally and vertically polarized reflected and transmitted wave amplitudes excited by a horizontally polarized or a vertically polarized incident wave of unit amplitude. Thus for instance, the reflection and transmission coefficients $R_{VH}(\omega)$ and $T_{VH}(\omega)$ are the vertically polarized reflected and transmitted wave amplitudes due to an incident horizontally polarized wave of unit amplitude. For a dielectric layer of width L ($0 < z < L$) there are eight transfer functions $R_{PQ}(\omega)$ and $T_{PQ}(\omega)$ (P,Q=V or H) for waves incident from below the layer ($z < 0$) and eight transfer functions for waves incident from above ($z > L$). They are related through the adjoint reciprocity relationship in electromagnetic theory (Kerns 1976). To obtain these transfer functions it is necessary to numerically solve the first order coupled differential equations (2.10) or (2.17)

$$f' = Cf \quad (3.1)$$

in conjunction with the continuity condition (2.19) at the interface between two adjacent partitions with Δz. For free space above and below the anisotropic dielectric layer the transformation matrix S can be written as (Bahar 1976)

$$S = \begin{pmatrix} 0 & 0 & c & -c \\ 1 & 1 & 0 & 0 \\ c & -c & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \quad (3.2)$$

Thus for an incident horizontally polarized wave of amplitude f_o^{iH} at $z = 0$ the value of the wave amplitudes a $z = 0^-$ and $z = L^+$ are f^{iH} and f^{TH} respectively, where

$$f^{iH} = f_o^{iH} \begin{pmatrix} 1 \\ R_{HH} \\ 0 \\ R_{VH} \end{pmatrix} \quad (3.3a) \qquad f^{TH} = f_o^{iH} \begin{pmatrix} T_{HH} \\ 0 \\ T_{VH} \\ 0 \end{pmatrix} \quad (3.3b)$$

Similarly for an incident vertically polarized wave of amplitude f_o^{iv} at z=0, the values for the wave amplitudes at $z=0^-$ and $z=L^+$ are f^{iv} and f^{TV} respectively, where

$$f^{iv} = f_o^{iv} \begin{pmatrix} 0 \\ R_{HV} \\ 1 \\ R_{VV} \end{pmatrix} \quad \text{and} \quad f^{TV} = f_o^{iv} \begin{pmatrix} T_{HV} \\ 0 \\ T_{VV} \\ 0 \end{pmatrix} \qquad (3.4)$$

Since the values of the wave amplitudes f are unknown initially at either boundary $z = 0^-$ or $z = L^+$, we first solve (3.1) subject to the two separate boundary conditions at $z = L^+$.

$$g^T = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad h^T = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad (3.5)$$

Denoting these solutions as g(z) and h(z) respectively and setting $g(0^-) = g^i$, $h(0^-) = h^i$, the wave amplitudes f(z) can be expressed as a superposition of these two independent solutions. Thus

$$f^H(z) = g(z) + \alpha^H h(z) \qquad (3.6)$$

$$f^V(z) = \alpha^V(z) + h(z) \qquad (3.7)$$

Equating (3.6) for $z = 0^-$ and $z = L^-$ to (3.3a) and (3.3b) respectively we get

$$\alpha^H = -g_3^i/h_3^i \ , \quad f_o^{iH} = g_1^i + \alpha^H h_1^i \qquad (3.8)$$

$$\begin{pmatrix} 1 \\ R_{HH} \\ 0 \\ R_{VH} \end{pmatrix} = \frac{1}{f_o^{iH}} (g^i + {}^H h^i), \quad \begin{pmatrix} T_{HH} \\ 0 \\ T_{VH} \\ 0 \end{pmatrix} = \frac{1}{f_o^{iH}} \begin{pmatrix} 1 \\ 0 \\ \alpha^H \\ 0 \end{pmatrix} \qquad (3.9)$$

Similarly equating (3.7) for $z = 0^-$ and $z = L^-$ to (3.4a) and (3.4b) respectively we get

$$\alpha^v = -h_1^i/g_1^i \ , \quad f_o^{iv} = \alpha^v g_3^i + h_3^i \qquad (3.10)$$

$$\begin{pmatrix} 0 \\ R_{HV} \\ 1 \\ R_{VV} \end{pmatrix} = \frac{1}{f_o^{iv}} (\alpha^v g^i + h^i), \quad \begin{pmatrix} T_{HV} \\ 0 \\ T_{VV} \\ 0 \end{pmatrix} = \frac{1}{f^{iv}} \begin{pmatrix} \alpha^V \\ 0 \\ 1 \\ 1 \end{pmatrix} \qquad (3.11)$$

For low frequencies where the width of the anisotropic dielectric considered L is of the order of several wavelengths, the numerical solutions to (3.1) can be obtained using a fourth order Runge-Kutta method (Abramowitz and Stegun 1964). Typically the number of partitions needed is $20L/\lambda > N > 10L/\lambda$. For high frequencies where $L/\lambda >> 1$ the Runge-Kutta method is not suitable, and it is more convenient to solve for the modified wave amplitudes

$$F_i(z) = f_i(z)\exp(+\int C_{ii}\,dz)$$

4. TRANSIENT RESPONSE FOR LORAN C EXCITATIONS

In this paper both numerical as well as analytical techniques are used to determine the transient response to LORAN C excitations. For convenience the excitation f(t) is expressed in terms of the complex function $f_s(t)$ as follows (Johler and Horowitz 1973)

$$f(t) = \mathrm{Re}\ f_s(t) = \mathrm{Re}\ i \exp(-ct)\sin^2\omega_p t \exp(-i\omega_c t)$$

$$= \exp(-ct)\sin^2\omega_p t \sin\omega_c t = \sum_{j=1}^{3} A_j \exp(-\Gamma_j t)u(t) \qquad (4.1a)$$

in which u(t) is the unit step function and

$$A_1 = i/2 \quad \Gamma_1 = C + i\omega_1 \quad \omega_1 = \omega_c$$

$$A_2 = -i/4 \quad \Gamma_2 = C + i\omega_2 \quad \omega_2 = \omega_c + 2\omega_p$$

$$A_3 = -i/4 \quad \Gamma_3 = C + i\omega_3 \quad \omega_3 = \omega_c - 2\omega \qquad (4.1b)$$

Thus the Fourier Transform of $f_s(t)$ is

$$F(\omega) = \sum_j A_j/(s+\Gamma_j) = \frac{2i\omega_p^2}{[i(\omega+\omega_c) + c][4\omega_p^2 + \{i(\omega+\omega_c) + c\}^2]} \qquad (4.1c)$$

where the symbol (*) denotes the complex conjugate. Thus it can be shown that the full wave expression for the transient response is the real part of the inverse Fourier Transform

The transfer functions $R_{PQ}(\omega)$ and $T_{PQ}(\omega)$ satisfy the relationships

$$R_{PQ}(\omega) = R_{PQ}^*(-\omega) \text{ and } T_{PQ}(\omega) = T_{PQ}^*(-\omega)$$

$$\text{Re } e_s(t) = \text{Re } F^{-1}[F(\omega)R(\omega)] = \text{Re } \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)R(\omega)\exp(i\omega t)d\omega \qquad (4.2)$$

$$\equiv \text{Re}[e_R(t) + ie_I(t)] = e_R(t)$$

in which $R(\omega)$ is any one of the eight transfer functions $R_{PQ}(\omega)$ and $T_{PQ}(\omega)$. To obtain an approximate analytical solution for e(t) we express the transform $R(\omega)$ as follows

$$R(\omega) = |R(\omega)|\exp[i\phi(\omega)] \simeq R(-\omega_c)\exp[-i(\omega+\omega_c)\tau_R] \qquad (4.3a)$$

in which

$$R(-\omega_c) = |R(-\omega_c)|e^{i\phi(-\omega_c)} = R^*(\omega_c) \qquad (4.3b)$$

is the value of R_{PQ} at $\omega = -\omega_c$ and the time delay τ_R is

$$\tau_R = -\frac{d}{d\omega}\phi(\omega)\Big|_{\omega=\pm\omega_c} \qquad (4.4)$$

Thus on substituting (4.2) into (4.1) we get

$$e_s(t) \simeq i\exp[-c(t-\tau_R)]\sin^2\omega_p(t-\tau_R)\exp[-i(\omega_c t-\phi(-\omega_c)]\,u(t-\tau_R) \qquad (4.5)$$

Noting that $\phi(-\omega_c) = -\phi(\omega_c)$ and setting $t' = t-\tau_R$, (4.5) can also be written as

$$e_s(t') = \text{Re}\exp(-ct')\sin^2\omega_p t'\exp\{-i[\omega_c(t'+\tau_R) + \phi(\omega_c)]\}u(t') \qquad (4.6)$$

Thus in the above analytical expression the envelope of the response with respect to the shifted time t' is the same as the envelope of the incident LORAN C pulse, however the instantaneous phase is $\omega_c(t'+\tau_R) + \phi(\omega_c)$. The approximate analytical expression (4.6) for (4.1) only accounts for the residues at the poles of the transform $F(\omega)$. However the expression (4.1) accounts for singularities of both $F(\omega)$ and $R(\omega)$. The singularities of $R(\omega)$ account for the distortion of the transient response which is expressed as follows:

$$\Delta e(t') = \text{Re}\frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)[R(\omega)\exp(i\omega\tau_R) - R^*(\omega_c)\exp(-i\omega_c\tau_R)]\exp(i\omega t)d\omega$$

$$\equiv \Delta e_R(t') + i\Delta e_I(t') \qquad (4.7)$$

Since an analytical expression for $R(\omega)$ is not known for arbitrary anisotropic media both (4.1) and (4.7) can only be evaluated numerically. The fast Fourier transform techniques are found suitable to evaluate these integrals. The instantaneous phase of the transient response (4.1) is given by

$$\phi(t') = \tan^{-1}[e_I(t')/e_R(t')] \qquad (4.8)$$

Thus the phase anomaly of the instantaneous response is

$$\Delta\phi(t') = \phi(t') + \omega_c(t'+\tau_R) + \phi(\omega_c) \qquad (4.9)$$

Provided that $\Delta e(t')$ and $\Delta\phi(t')$ vanish for $t' > \frac{3}{f_c} = t_3$ the shift in the third zero crossing due to the effects of the medium can be predicted accurately. (The third zero crossing is regarded as the effective arrival time of the response.) Thus for instance if $R(\omega)$ is $T_{VH}(\omega)$, the effect of the medium is to introduce a total delay in the third zero crossing equal to

$$\Delta t = \frac{1}{f_c}(N - \phi(\omega_c)/2\pi) \qquad (4.10)$$

in which N is an integer such that $2.5 < (\Delta t-\tau_R)f_c < 3.5$. Similarly if $R(\omega)$ is $R_{HV}(\omega)$ (the reflection coefficient with respect to the reference height z = 0) the expression for Δt (4.10) is the delay in the third zero crossing relative to a signal reflected by a perfect conductor at z = 0.

5. ILLUSTRATIVE EXAMPLES

For the illustrative examples presented in this section, the carrier frequency $f_c = 10^5$ Hz, $f_p = f_c/40$ and $c = 10f_p$. The inhomogeneous anisotropic ionosphere is characterized by the plasma, collision and gyro-frequencies ω_p, ν and ω_H respectively. The inhomogeneous plasma frequency f_p is assumed to be given by

$$X(z) = \left(\frac{f_p(z)}{f}\right)^2 = \frac{X_1}{2}[1-\cos(\pi z/L)],\quad 0<z<L=\lambda_c/2 \qquad (5.1)$$

where

$$X_1 = X(L) = (\omega_p(L)/\omega)^2 = (\lambda/\lambda_p(L))^2 \qquad (5.2)$$

and

$$\lambda_p(L) = 3.5\ 10^3\ \text{m} = c_o/f_p(L),\ \lambda_c = c_o/f_c \qquad (5.3)$$

The collusion frequency ν is given by

$$Z(z) = \nu(z)/\omega = Z_1 \exp(-\gamma z),\ 0 < z < L \quad (5.4)$$

where

$$Z_1 = Z(0) = \frac{\nu(0)}{\omega} = \lambda/\lambda_\nu(0) \quad (5.5)$$

and

$$\lambda_\nu(0) = 2\pi c_o/\nu = 3.10^4 \text{m},\ \gamma = 2.10^{-4}\ \text{m}^{-1} \quad (5.6)$$

The gyro frequency f_H is given by

$$Y(z) = \frac{f_H}{f} = \lambda/\lambda_H \quad (5.7)$$

where

$$\lambda_H = c_o/f_H = 3.10^2 \text{ m} \quad (5.8)$$

The direction of the earth's magnetic field $\bar{a}_H$ is given by

$$\bar{Y} = (0.5\bar{a}_x + \sqrt{0.5}\,\bar{a}_y + 0.5\bar{a}_z)Y = \bar{a}_H Y \quad (5.9)$$

In Figures 1 through 4 the transfer functions $R_{HH}(\omega), R_{VH}(\omega), T_{HH}(\omega)$ and $T_{VV}(\omega)$ are given for waves incident from below the anisotropic layer at an angle $\theta = 45^o$, thus the direction of the wave normal is $\bar{n} = (\bar{a}_x + \bar{a}_y)/\sqrt{2}$. The value of the transfer functions at $\omega = \omega_c/2$, $\omega = \omega_c$ and $\omega = 2\omega_c$ are indicated by symbols $\square$, Δ and * respectively. In these figures $12\Delta f < f < f_{max}$. For the purpose of the fast Fourier transform algorithm used (Brigham, 1974) $f_{max} = 3.2f_c$ and $\Delta f = 2f_{max}/N = f_c/40$, thus $N = 256$.

In Figs. 5 and 6 $e_R(t')$ and the envelope $|e_s(t')|$ together with the envelope of the incident pulse $|f_s(t)|$ are given for the transfer functions $R_{HH}(\omega)$ Fig. 1 and $R_{VH}(\omega)$ (Fig. 2) respectively. In these figures $f_s(t)$ and $e_s(t')$ have been normalized to unity at $t = t_p = \frac{1}{\omega_p}\tan^{-1}(\frac{2\omega_p}{c})$ (peak value). Thus

$$e_{sN}(t') = e_{RN}(t') + i\, e_{IN}(t') = e_s(t')/|e_s(t_p)| \quad (5.10)$$

Note that $|e_{sN}(t')| \approx |f_{sN}(t)|$, however $e_{RN}(t') \neq f_{RN}(t) = \text{Re}\, f_s(t)/|f_s(t_p)|$ thus the third zero crossing of $e_{RN}(t')$ does not coincide with the third zero crossing for $f_{RN}(t)$ which is shown in the figures by the first Δ symbol on the t' axis. The second Δ symbol is shown at $t = t_p$.

In Figures 7 and 8, the phase anomalies $\Delta\phi(t')$ (4.9) are given for the transfer functions $R_{HH}(\omega)$ (Fig. 1) and $R_{VH}(\omega)$ (Fig. 2), respectively. Note that $\Delta\phi(t')$ vanishes near the peak of the pulse.

In addition, the phase anomaly for the inverse fast Fourier transform of $F(\omega)$ is given in these figures. This phase anomaly, which is due to the finite spectrum used in the fast Fourier transform algorithm is negligible except where $|e_{sN}(t)| << 1$.

In Figures 9 and 10 the normalized values of $e_{RN}(t')$ and its envelope $|e_{sN}(t')|$ are given for the transfer functions $R_{HH}(\omega)$ (Fig. 1) and and $R_{VH}(\omega)$ (Fig. 2). These functions are evaluated using the fast Fourier transform (FFT) algorithm as well as the analytic expression (4.5). The difference between the numerical (FFT) results and the analytical results $\Delta e(t')$ (4.7) is negligible for $t > 3/f_o$.

The effect of the medium is to cause the delay τ_R and the shift in the third zero crossing of approximately half the period $1/f_c$. The delay τ_R is not shown in the figures since e_{RN} is plotted as a function of the shifted time scale $t' = t - \tau_R$. For the cases considered in Figs. 9 and 10, τ_R is $6.666\text{x}10^{-6}$sec and $6.660\text{x}10^{-6}$sec respectively. There is excellent agreement between the values for τ_R obtained from (4.4) and the numerical results obtained by using the fast Fourier transform algorithm as indicated in Figs. 5 and 6 where $|e_{sN}(t')| \approx |f_{sN}(t)|$.

6. CONCLUDING REMARKS

The effects of anisotropic inhomogeneous medium (such as the ionosphere) on the reflection and transmission of signals used for navigational and positioning systems have been investigated. A full wave generalized WKB approach, which accounts for the depolarization of the incident wave, is used to obtain the eight transfer functions $R_{PQ}(\omega)$ and $T_{PQ}(\omega)$ (P,Q = H,V) for waves incident from below the ionosphere. Both fast Fourier transform techniques and analytical methods are used to determine the transient response $e_{RN}(t')$, the amplitude and phase anomalies $\Delta e(t')$ and $\Delta\phi(t')$ as functions of the shifted time scale $t' = t - \tau_R$. Since the third zero crossing is generally regarded as the effective time of arrival of the signal, analytical expressions (which are in good agreement with the FFT results) are given for the effective delay of the signal (4.10) in terms of τ_R (4.4) and $\phi(\omega_c)$ (4.3).

REFERENCES

Abramowitz, M. and Stegun, I. A. (1964), Handbook of mathematical functions with formulas, graphs and mathematical tables, Applied Mathematical tables, Applied Mathematical Series 55, National Bureau of Standards, U. S. Government Printing Office, Washington, D.C.

Bahar, E. (1976), Generalized characteristic functions for simultaneous linear differential equations with variable coefficients applied to propagation in inhomogeneous anisotropic media, Canadian Journal of Physics, Vol. 54, No. 3, pp. 301-316.

Brigham, E. Oran (1974), The Fast Fourier Transform, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Brogan W. L. (1974), Modern Control Theory, Quantum Publishers, Inc., New York.

Bronson, R. (1969), Matrix methods--An introduction, Academic Press, New York.

Budden, K. G. (1962), Radio Waves in the Ionosphere, Cambridge University Press, Cambridge.

Friedman, B. (1956), Principles and Techniques of Applied Mathematics, John Wiley and Sons, Inc., London.

Kerns, D. M. (1976), Plane-Wave Scattering Matrix Theory of Antennas and Antenna-Antenna Interactions: Formulation and Applications. NBSIR-75-824 National Technical Information Service, Springfield, Virginia 22151.

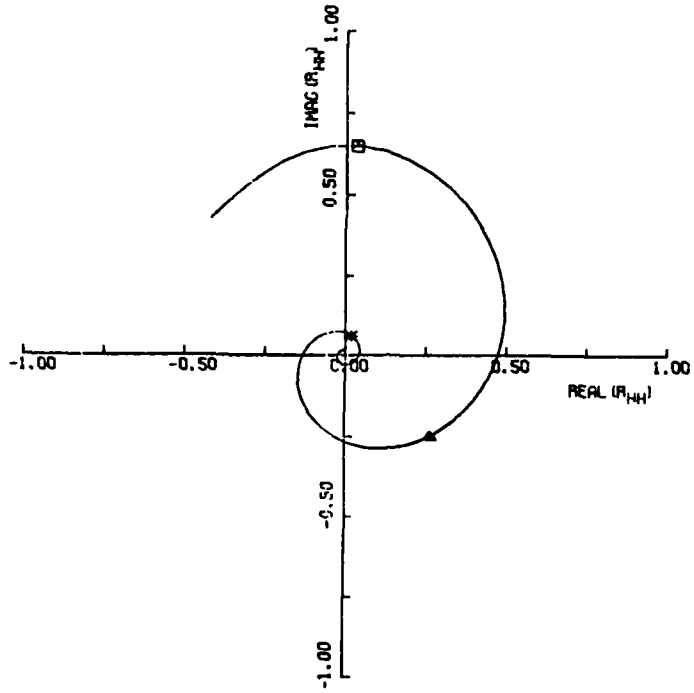


Fig. 1 Reflection coefficient for the horizontally polarized waves, $R_{HH}(\omega)$.

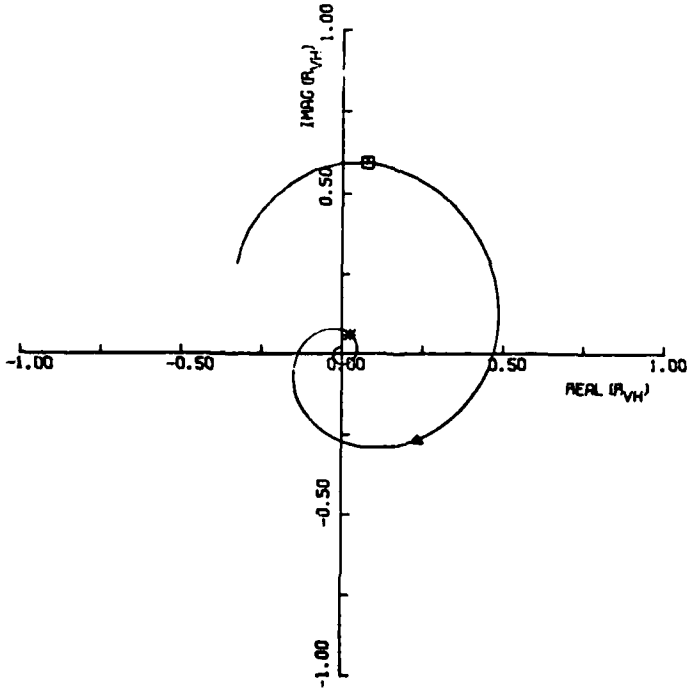


Fig. 2 Reflection coefficient for the depolarized waves (horizontal to vertical), $R_{VH}(\omega)$.

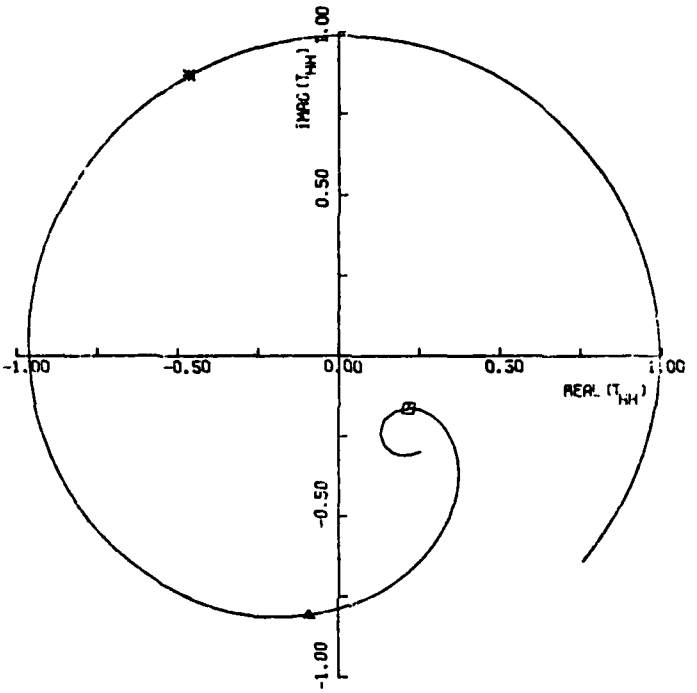


Fig. 3 Transmission coefficient for the horizontally polarized waves, $T_{HH}(\omega)$.

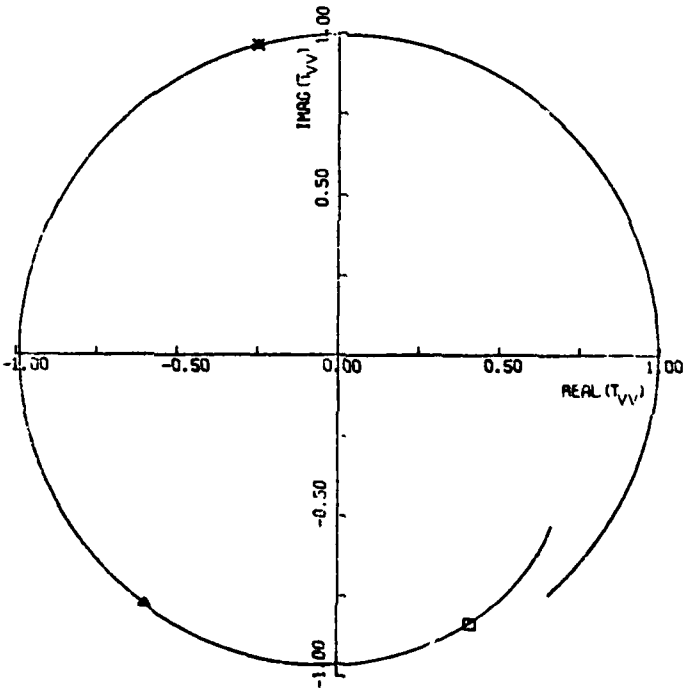


Fig. 4 Transmission coefficient for the vertically polarized waves, $T_{VV}(\omega)$.

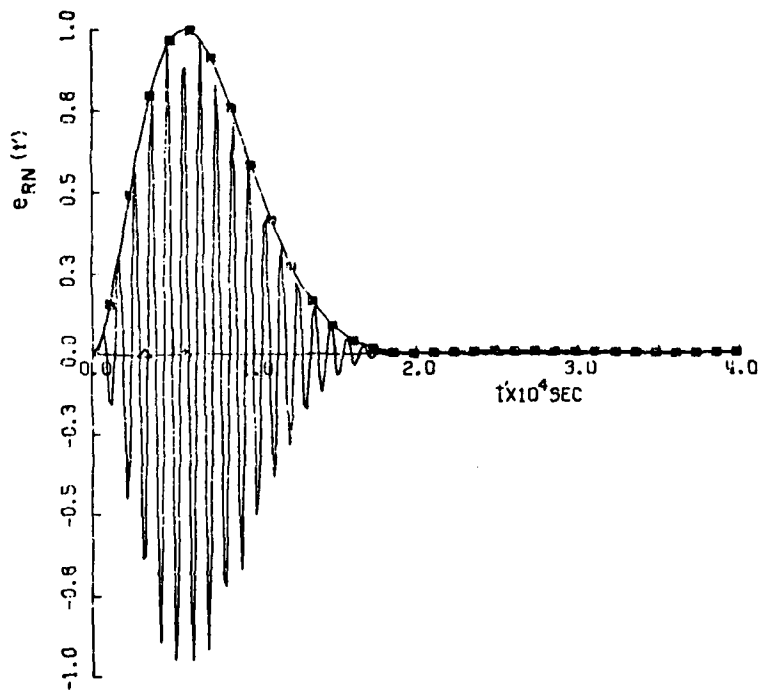


Fig. 5 Normalized transient response $e_{RN}(t')$ and its envelope, $|e_{sN}(t')|$ corresponding to the transfer function, $R_{HH}(\omega)$ in Fig. 1.

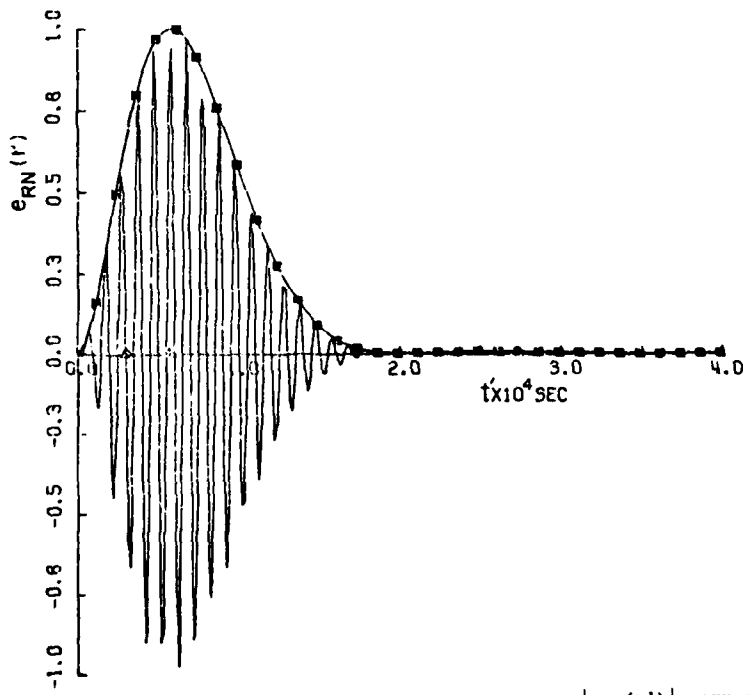


Fig. 6 Normalized transient response, $e_{RN}(t')$ and its envelope, $|e_{sN}(t')|$ corresponding to the transfer function, $R_{VH}(\omega)$ in Fig. 2.

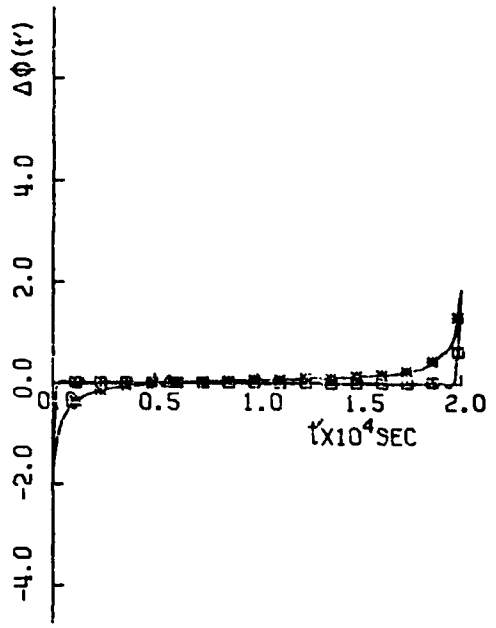


Fig. 7 Phase anomalies, $\Delta\phi(t')$ for the transient response (*) corresponding to the transfer function, $R_{HH}(\omega)$ in Fig. 1 and for the inverse fast Fourier transfer of $F(\omega)$ (□).

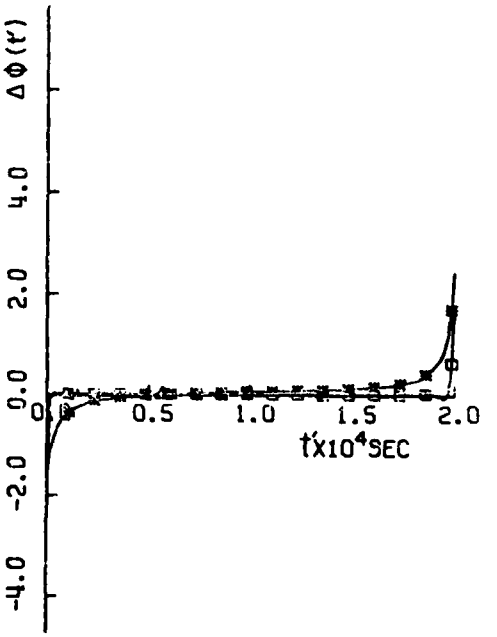


Fig. 8 Phase anomalies, $\Delta\phi(t')$ for the transient response (*) corresponding to the transfer function, $R_{VH}(\omega)$ in Fig. 2 and for the inverse fast Fourier transform of $R(\omega)$ (□).

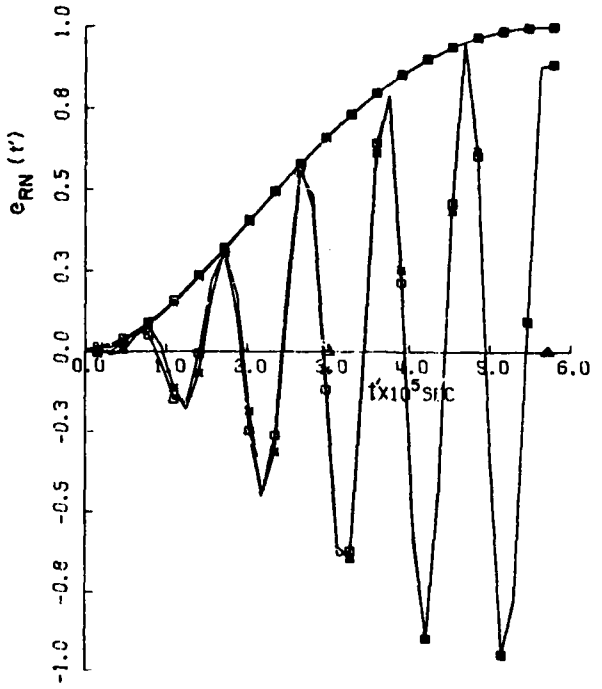


Fig. 9 The normalized transient response, $e_{RN}(t')$ and its envelope, $|e_{sN}(t')|$ corresponding to the transfer function, $R_{HH}(\omega)$ in Fig. 1. (*) Analytic, (□) Fast Fourier transform.

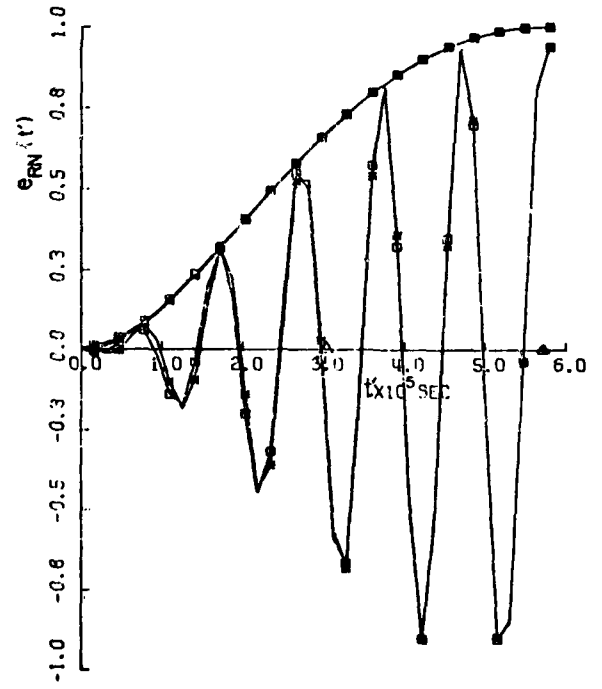


Fig. 10 The normalized transient response, $e_{RN}(t')$ and its envelope , $|e_{sN}(t')|$ corresponding to the transfer function, $R_{VH}(\omega)$ in Fig. 2. (*) Analytic, (□) Fast Fourier transform.

MEAN ARRIVAL TIME AND MEAN PULSE WIDTH OF SIGNALS PROPAGATING THROUGH AN INHOMOGENEOUS IONOSPHERE WITH RANDOM IRREGULARITIES

Y. K. Wong, K. C. Yeh and C. H. Liu

Ionosphere Radio Laboratory, Department of Electrical Engineering
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

ABSTRACT

It has been known for some time that as a dispersive medium the ionosphere will distort any radio signal propagating through it. In particular, the signal will propagate with a group velocity which is smaller than the free space velocity of light and the pulse shape will be degraded by being smeared and stretched. Recent studies of random medium propagation have demonstrated that scattering from random electron density irregularities can give rise to further degradation on signals of the kind just described. The purpose of this paper is to investigate the combined dispersive and scattering effects on radio signals in an inhomogeneous ionosphere.

The ionosphere is modeled as a slab of random irregularities imbedded in an inhomogeneous background. The background electron density is assumed to be vertically distributed with an arbitrary profile, although in actual numerical computations a Chapman profile is used. For almost all space communication applications the wave length is very small when compared with the scale height. In this case an electromagnetic pulse can be described by using the WKB solution. In many applications the precise shape of the pulse is not of great interest. Instead, a hierarchy of coarse scale descriptions will suffice. This suggests the use of temporal moments, for which the first moment is related to the mean arrival time, the second moment is related to the mean pulse width, the third moment is related to the skewness, etc. In solving for these moments the one-position two-frequency mutual coherence function appears naturally. The mutual coherence function in an inhomogeneous ionosphere is derived and solved approximately by the WKB method. The results are applied to propagation of transionospheric signals. Some numerical values will be given. They show that scattering from random irregularities can have non-negligible contribution to the arrival time and can have dominant contribution to pulse lengthening.

INTRODUCTION

As a radio signal propagates through the ionosphere, the signal is being distorted due to dispersion and scattering. The amount of distortion and its dependence on medium parameters have been studied by various authors [Liu and Yeh, 1975; Liu and Wernik, 1975; Liu et al., 1974; Ulaszek, 1974]. Their approach is adapted from the theory of wave propagation in a turbulent atmosphere (for a review see Tatarski, 1971). Central in this approach is the solution of the transport equation for the two-frequency mutual coherence function. At the same time Mark [1974] has suggested a new method of studying the characteristics of stochastic transients and transmission media by using power-moments spectra which characterize the temporal behavior of signals. Temporal moments of the stochastic wave can be derived and have also been studied by different authors [Yeh and Liu, 1977a; Yeh and Liu, 1977b; Yeh and Yang, 1977] under various conditions.

The objective of this paper is to investigate the combined effects of dispersion and random scattering on the mean arrival time and the mean pulse width of a signal propagating through an inhomogeneous ionosphere with random irregularities. These quantities have close relationship with the problems of time delay and pulse distortion. We first investigate the parabolic equation of quasi-optics [Liu et al., 1974] from which we obtain the two-frequency two-position mutual coherence functions generalized for a horizontally stratified inhomogeneous background medium. The forward scatter approximation is used to simplify the problem. The irregularity spectrum discussed here is assumed to follow a power-law as suggested by various experimental results. Equations for the mean arrival time and mean pulse width are then derived. For specific evaluation of the expressions, the background electron density is assumed to have a Chapman's profile. Using typical ionospheric parameters, some numerical results are obtained and shown in a later section.

TWO-FREQUENCY TWO-POSITION MUTUAL COHERENCE FUNCTION

The geometry of the problem of interest is shown in Fig. 1. A uniform plane wave is transmitted at $z=z_S$ and incident normally on top of a slab of random irregularities at $z=0$. Owing to the orientation of the coordinate axes z_S is actually negative. The slab of irregularities has a thickness of z_B. The wave is received, after emerging out of the slab at z_B, by a receiver located at $z=z_R$. The background electron density is not shown in the diagram but it is assumed to have a Chapman's profile with the peak at z_o. Since the medium is the communication channel that is of interest, the effects of the transmitting and receiving antennas will not be considered.

The random electron density irregularities in the ionosphere can be mathematically characterized by taking the dielectric permittivity to be random in time and space. But to simplify the problem, the variation of the permittivity is assumed to be very slow comparing to the time that the signal will take to pass through the medium. Thus, the dielectric permittivity can be written as $\epsilon(\vec{r})$ which is a random function of position only. The permeability is taken to be uniform throughout the region of interest and equal to that in free space. The characteristic of the dielectric permittivity is described as

$$\epsilon(\vec{r}) = \epsilon_0 <\epsilon(\vec{r})>(1+\mu(\vec{r})) \qquad (1)$$

where

$$<\epsilon(\vec{r})> = 1-\omega_p^2/\omega^2 \qquad (2)$$

$$\mu(\vec{r}) = -\frac{\omega_p^2/\omega^2}{1-\omega_p^2/\omega^2} \cdot \frac{\Delta N(\vec{r})}{N(z)} \qquad (3)$$

Here ω_p is the angular electron plasma frequency which is a function of z in our case. ϵ_0 is the vacuum permittivity and N(z) denotes the background electron density which is a function of height with a typical scale height H. The percentage electron density fluctuation $\Delta N(\vec{r})/N(z)$ is assumed to be a homogeneous random field with a typical scale size ℓ_0.

Starting with Maxwell's equations, assuming time dependence $e^{j\omega t}$, we can write the wave equation [Ulaszek, 1974; Yeh and Liu, 1972] for the electric field in the form

$$\nabla^2\vec{E}+k^2<\epsilon(\vec{r})>(1+\mu(\vec{r}))\vec{E} = -\nabla[\vec{E}\cdot\nabla[\ln(1+\mu(\vec{r}))]] \qquad (4)$$

where $k=\omega/c$. If the irregularities are large compared to the signal wavelength, the right hand side of equation (4) can be ignored [Tatarskii, 1971]. This assumption implies that we neglect the depolarization effects of the wave. Therefore equation (4), after decoupling the cartesian components, reduces to the scalar wave equation

$$\nabla^2 E+k^2<\epsilon(\vec{r})>(1+\mu(\vec{r}))E = 0 \qquad (5)$$

Let $u(\vec{r})$ denote the complex amplitude of the scalar field $E(\vec{r})$ by writing

$$E(\vec{r}) = u(\vec{r})\ \exp\{-jk\int_0^z \sqrt{<\epsilon(\zeta)>}\,d\zeta\} \qquad (6)$$

Substituting (6) into (5) gives

$$\nabla^2 u - 2jk\sqrt{<\epsilon>}\frac{\partial u}{\partial z} - jk\frac{\partial\sqrt{<\epsilon>}}{\partial z}u + k^2<\epsilon>\mu\cdot u = 0 \qquad (7)$$

If we assume the field is scattered into a small angular cone centered about z-axis, the forward scatter approximation can be applied [Tatarskii, 1971]. Actually, in deriving (7), the validity of the WKB solution is implicity assumed. $\nabla^2 u$ in (7) can be replaced by the transverse Laplacian $\nabla_T^2 u$ since u is assumed to have a negligible variation in z as compared to one wavelength. Thus the complex amplitude of the time harmonic field component satisfies the so-called parabolic differential equation

$$\nabla_T^2 u - 2jk\sqrt{<\epsilon>}\frac{\partial u}{\partial z} - jk\frac{\partial\sqrt{<\epsilon>}}{\partial z}u + k^2<\epsilon>\mu u = 0 \qquad (8)$$

where $\nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.

In order to study the statistical properties of the signal, such as the temporal moments, rms intensity fluctuation, etc., one needs to first investigate the statistical characteristics of the complex amplitude u. In particular, the two-frequency two-position mutual coherence function Γ will be under invertigation. It is defined as follows

$$\Gamma(z,|\vec{\rho}_1-\vec{\rho}_2|,\omega_1,\omega_2) \equiv <u_1(z,\vec{\rho}_1,\omega_1)u_2^*(z,\vec{\rho}_2,\omega_2)> \qquad (9)$$

where $<\ >$ denotes the ensemble average operator and an asterisk * represents the complex conjugate operation. This function has a unit value for $z\leq 0$ since in this region the slab of random irregularities is absent. On the other hand, though the random irregularity is absent between z_B to the receiver, the function Γ does not equal to unity. This is due to the after effect as the signal passes through the slab of irregularities. This will be investigated later in the session.

The mutual coherence function Γ has been derived by some authors [Liu and Wernik, 1975; Ulaszek, 1974] but only for the homogeneous case. It is a bit more complicated when the background is inhomogeneous. From (9) we know that Γ depends only on the distance between the two points $\vec{\rho}_1=(x_1,y_1)$ and $\vec{\rho}_2=(x_2,y_2)$ instead of their individual position vectors owing to the assumption that the fluctuation of the electron density is a homogeneous random process

and the incident wave is uniform and plane. Let $u_1=u(z,\vec{\rho}_1,\omega_1)$ and $u_2=u(z,\vec{\rho}_2,\omega_2)$ be the signals at angular frequencies ω_1 and ω_2 and at positions $(z,\vec{\rho}_1)$ and $(z,\vec{\rho}_2)$ respectively. From (8) we write the parabolic equations for u_1 and u_2^* as

$$\frac{\partial u_1}{\partial z} + \frac{j}{2k_1\sqrt{<\epsilon_1>}} \nabla_{T_1}^2 u_1 + \left(\frac{jk_1\sqrt{<\epsilon_1>}}{2} u_1 + \frac{1}{4} \frac{\partial \ell n <\epsilon_1>}{\partial z} \right) u_1 = 0 \tag{10}$$

$$\frac{\partial u_2^*}{\partial z} - \frac{j}{2k_2\sqrt{<\epsilon_2>}} \nabla_{T_2}^2 u_2^* + \left(\frac{-jk_2\sqrt{<\epsilon_2>}}{2} u_2 + \frac{1}{4} \frac{\partial \ell n <\epsilon_2>}{\partial z} \right) u_2^* = 0 \tag{11}$$

If we multiply (10) by u_2^* and (11) by u_1 and add them together, it becomes

$$\frac{\partial \gamma}{\partial z} + P\gamma = \frac{-j}{2} L_T \gamma \tag{12}$$

where

$$\gamma \equiv u_1 u_2^* \tag{12a}$$

$$L_T \equiv \frac{\nabla_{T_1}^2}{k_1\sqrt{<\epsilon_1>}} - \frac{\nabla_{T_2}^2}{k_2\sqrt{<\epsilon_2>}} \tag{12b}$$

$$Q \equiv \frac{j}{2} (k_1\sqrt{<\epsilon_1>}u_1 - k_2\sqrt{<\epsilon_2>}u_2) \tag{12c}$$

$$P \equiv Q + \frac{1}{4} \frac{\partial}{\partial z} (\ell n[<\epsilon_1><\epsilon_2>]) \tag{12d}$$

The formal solution for (12), after taking the ensemble average, is given as

$$<\gamma> = \gamma_0 \left[\frac{<\epsilon_1(0)><\epsilon_2(0)>}{<\epsilon_1(z)><\epsilon_2(z)>} \right]^{\frac{1}{4}} \cdot e^{\frac{1}{2}\int_0^z \int_0^z <Q(\zeta')Q(\zeta'')>d\zeta' d\zeta''} - \frac{1}{2} \int_0^z \left[\frac{<\epsilon_1(\zeta)><\epsilon_2(\zeta)>}{<\epsilon_1(z)><\epsilon_2(z)>} \right]^{\frac{1}{4}} \cdot e^{\frac{1}{2}\int_\zeta^z \int_\zeta^z <Q(\zeta')Q(\zeta'')>d\zeta' d\zeta''} \cdot L_T <\gamma> d\zeta \tag{13}$$

where $\gamma_0=\gamma|_{z=0}$. Here we have assumed Q is Guassian with zero mean. Further, we assume that the random medium is "delta-function" correlated along the direction of propagation so that Markov's assumption can be used [Tatarskii, 1971], i.e. using the approximation

$$B_N(\zeta'-\zeta'',0) = A_N(0)\ \delta(\zeta'-\zeta'') \tag{14}$$

where $B_N(\vec{r})$ is the correlation function of $\Delta N/N$ which has a spectrum $\Phi_N(\vec{\kappa})$ such that

$$A_N(\vec{\rho}) = \int_{-\infty}^{\infty} B_N(\vec{\rho},z)dz = 2\pi \iint_{-\infty}^{\infty} \Phi_N(\vec{\kappa}_T,0) e^{-j\vec{\kappa}_T\cdot\vec{\rho}} d^2\kappa_T \tag{15}$$

$$\Phi_N(\vec{\kappa}_T,0) = \left(\frac{1}{2\pi}\right)^3 \iint_{-\infty}^{\infty} A_N(\vec{\rho})\ e^{j\vec{\kappa}_T\cdot\vec{\rho}}\ d^2\rho \tag{16}$$

and $\vec{\kappa}=(\vec{\kappa}_T,\kappa_z)$. Thus, differentiating (13) with respect to z and letting $\Gamma=<\gamma>$ and assuming the wave is incident normally on the slab, we obtain

$$\frac{\partial \Gamma}{\partial z} + \frac{1}{2} [k_1^2-k_p^2)^{-\frac{1}{2}} - (k_2^2-k_p^2)^{-\frac{1}{2}}] \nabla_T \Gamma + \{\frac{1}{4} \frac{\partial}{\partial z} [\ell n(1-k_p^2 k_1^{-2})(1-k_p^2 k_2^{-2})] + \frac{k_p^4}{8} [\frac{k_1^2+k_2^2-2k_p^2}{(k_1^2-k_p^2)(k_2^2-k_p^2)} A_N(0) - 2\ [(k_1^2-k_p^2)(k_2^2-k_p^2)]^{-\frac{1}{2}} A_N(\rho)]\} = 0 \tag{17}$$

where $\rho=|\vec{\rho}_1-\vec{\rho}_2|$ and $k_p=\omega_p(z)/c$. To solve for Γ in (17), let us put Γ in the following form

$$\Gamma \equiv \exp[-\int_0^z \frac{k_p^4}{8} \cdot \frac{(k_1^2+k_2^2-2k_p^2)}{(k_1^2-k_p^2)(k_2^2-k_p^2)} \cdot A_N(0)\ d\zeta + \frac{1}{4} \ell n[(1-k_p^2k_1^{-2})(1-k_p^2k_2^{-2})] + \phi(\rho,z)] \tag{18}$$

When we put Γ given in (18) back to (17), it will give a different equation in terms of ϕ which looks like the following

$$\frac{\partial \phi}{\partial z} + \frac{1}{2} [k_1^2-k_p^2)^{-\frac{1}{2}} - (k_2^2-k_p^2)^{-\frac{1}{2}}] [\nabla_T^2\phi + (\nabla_T\phi)^2] = \frac{k_p^4}{4} [(k_1^2-k_p^2)(k_2^2-k_p^2)]^{-\frac{1}{2}} \cdot A_N(\rho) \tag{19}$$

For weak random irregularities, the non-linear term $(\nabla_T\phi)^2$ in (19) may be ignored. The resulting approximate solution is known as the Rytov's solution. With this assumption, (19) becomes

$$\frac{\partial\phi}{\partial z} + \frac{j}{2}\left[(k_1^2-k_p^2)^{-\frac{1}{2}} - (k_2^2-k_p^2)^{-\frac{1}{2}}\right]\nabla_T^2\phi = \frac{k_p^4}{4}\left[(k_1^2-k_p^2)(k_2^2-k_p^2)\right]^{-\frac{1}{2}} A_N(\rho) \tag{20}$$

To solve (20), we introduce a new two-dimensional Fourier transform

$$\psi(\vec{\kappa}_T,z) = \left(\frac{1}{2\pi}\right)^2 \iint_{-\infty}^{\infty} \phi(\vec{\rho},z)\, e^{j\vec{\kappa}_T\cdot\vec{\rho}}\, d^2\rho \tag{21a}$$

$$\phi(\vec{\rho},z) = \iint_{-\infty}^{\infty} \psi(\vec{\kappa}_T,z)\, e^{-j\vec{\kappa}_T\cdot\vec{\rho}}\, d^2\kappa_T \tag{21b}$$

Using (21b) and (15), (20) becomes

$$\frac{\partial\psi}{\partial z} - j\frac{\kappa_T^2}{2}\left[(k_1^2-k_p^2)^{-\frac{1}{2}} - (k_2^2-k_p^2)^{-\frac{1}{2}}\right]\psi = \frac{\pi k_p^4}{2}\left[(k_1^2-k_p^2)(k_2^2-k_p^2)\right]^{-\frac{1}{2}}\phi_N(\vec{\kappa}_T,0) \tag{22}$$

Let us assume ψ has a form

$$\psi = W\cdot\exp\{j\frac{\kappa_T^2}{2}\int_0^z [(k_1^2-k_p^2)^{-\frac{1}{2}} - (k_2^2-k_p^2)^{-\frac{1}{2}}]\}\, d\zeta \tag{23}$$

From (22) and (23), W and ψ can be obtained and thus ϕ which is found to be

$$\phi = \frac{\pi}{2}\int_0^z k_p^4\left[(k_1^2-k_p^2)(k_2^2-k_p^2)\right]^{-\frac{1}{2}} d\zeta \cdot \iint_{-\infty}^{\infty}\phi_N(\vec{\kappa}_T,0)\exp(-j\vec{\kappa}_T\cdot\vec{\rho}) \cdot \exp\{j\frac{\kappa_T^2}{2}\int_\zeta^z [(k_1^2-k_p^2)^{-\frac{1}{2}} - (k_2^2-k_p^2)^{-\frac{1}{2}}]d\zeta'\}d^2\kappa_T \tag{24}$$

Putting ϕ given in (24) back to (18) with the initial condition $\Gamma(z=0) = 1$, we obtain finally,

$$\Gamma(z,\rho,\omega_1,\omega_2) = \left\{\frac{\langle\epsilon_1(0)\rangle\langle\epsilon_2(0)\rangle}{\langle\epsilon_1(z)\rangle\langle\epsilon_2(z)\rangle}\right\}^{\frac{1}{4}} \cdot \exp\{\frac{-A_N(0)}{8}\int_0^z \frac{k_p^4(k_1^2-k_2^2-2k_p^2)}{(k_1^2-k_p^2)(k_2^2-k_p^2)}\, d\zeta\} \cdot e^{\phi} \tag{25}$$

This is the two-frequency two-position mutual coherence function for a horizontally stratified background ionosphere under the combined WKB and Rytov approximations. This solution is valid for an observer inside the slab of random irregularities. If the observer is outside the slab as shown in Fig. 1, the solution (25) evaluated at the base of the slab ($z=z_B$) will serve as the initial condition for diffraction of waves below it. Using the same approach as shown and putting $A_N(\rho)=0$ in (20) for the absence of irregularities, Γ is obtained. It has the exact form as in (25) except the limits in ϕ are changed and is given as follows

$$\phi = \frac{\pi}{2}\int_0^{z_B} d\zeta\left[(k_1^2-k_p^2)(k_2^2-k_p^2)\right]^{-\frac{1}{2}} k_p^4 \cdot \iint_{-\infty}^{\infty}\phi_N(\vec{\kappa}_T,0)\exp(-j\vec{\kappa}_T\cdot\vec{\rho}) \cdot\exp\{j\frac{\kappa_T^2}{2}\int_\zeta^{z_R}[(k_1^2-k_p^2)^{-\frac{1}{2}} - (k_2^2-k_p^2)^{-\frac{1}{2}}]d\zeta'\}d^2\kappa_T \tag{26}$$

where $z_R \geq z_B$. The first two factors given in (25) are also changed. Instead of the variable z, it is replaced by z_B. The first factor is approximately equal to unity under high frequency approximation and it is assumed in our discussion.

It is useful to expand the mutual coherence function in a series

$$\Gamma = \Gamma_0 + \Gamma_1\eta + \Gamma_2\eta^2 + \ldots\ldots \tag{27}$$

where $\eta=(k_1-k_2)/k_1$ and all the $\Gamma_i(i=0,1,2,\ldots.)$ are functions of k_1 only. The first three terms are obtained using the result in (25) and (26) and are given as follows

$$\Gamma_0 = 1 \tag{28a}$$

$$\Gamma_1 = j\frac{\nabla_T^2 A_N(0)}{8k_1^3}\int_0^{z_B} k_p^4\cdot(z_R-\zeta)\, d\zeta \tag{28b}$$

$$\Gamma_2 = \frac{k_1^2}{2}\{\frac{-1\nabla_T^2 A_N(0)}{8k_1^4}\int_0^{z_B} k_p^4\cdot(z_R-\zeta)d\zeta\}^2 + \frac{k_1^2}{2}\{\frac{-A_N(0)}{4k_1^4}\int_0^{z_B} k_p^4\, d\zeta + \frac{\nabla_T^2\nabla_T^2 A_N(0)}{16k_1^6}\int_0^{z_B} k_p^4\cdot(z_R-\zeta)^2 d\zeta + \frac{j\nabla_T^2 A_N(0)}{2k_1^5}\int_0^{z_B} k_p^4\cdot(z_R-\zeta)d\zeta\} \tag{28c}$$

The first term in (28c) will be neglected for its value is very small comparing to other terms in the expression under high frequency approximation. ρ is put equal to zero because the receiver is assumed to be a point.

IRREGULARITY SPECTRUM

It is shown in the previous section that the propagation of an EM-wave in a random medium is closely related to the correlation, or the spectrum, of the random field $\Delta N/N$. In the ionosphere, through many measurements, they all seem to suggest the irregularity spectrum has the spatial wave number dependence κ^{-p}, commonly referred to as a power-law spectrum. But there are a few difficulties for a 3-dimensional power spectrum of the form κ^{-p}. Firstly, for p>2, its associated correlation will not exist. Secondly, for any finite value of p some spectral moments will always fail to exist. Fortunately, Shkarosky [1968] introduced an inner scale r_o along with the outer scale ℓ_o, introduced by Tatarskii [1971] to remove the difficulties by writing the spectrum as

$$\Phi_N(\kappa) = \frac{[\kappa_o r_o]^{(p-3)/2} \cdot r_o^3}{(2\pi)^{3/2} \; K_{(p-3)/2}(\kappa_o r_o)} \cdot \frac{K_{p/2}(r_o\sqrt{\kappa^2+r_o^2})\sigma_N^2}{(r_o\sqrt{\kappa^2+\kappa_o^2})^{p/2}} \tag{29}$$

where $\kappa_o=1/\ell_o$ and K is a Hankel function of imaginery argument. σ_N^2 is the variance of the fluctuation function $\Delta N/N$. It has an advantage in using (29) for all the related correlation functions can be evaluated analytically.

The 2-dimensional correlation $A_N(\rho)$ defined by (15) for an isotropic random field can be expressed in the form [Yeh and Liu, 1977a]

$$A_N(\rho) = A_o + A_2\rho^2 + A_4\rho^4 + \ldots. \tag{30}$$

Thus, using (15) and (29), the coefficients in (30) are obtained and are given as follows

$$A_o = \sqrt{2\pi r_o/\kappa_o} \cdot \sigma_N^2 \; K_{(p-2)}(\kappa_o r_o)/K_{(p-3)/2}(\kappa_o r_o) \tag{31a}$$

$$A_2 = -\sqrt{\pi\kappa_o/2r_o} \; \sigma_N^2 \; K_{(p-4)/2}(\kappa_o r_o)/K_{(p-3)/2}(\kappa_o r_o) \tag{31b}$$

$$A_4 = \sqrt{\pi\kappa_o^3/2^5 r_o^3} \; \sigma_N^2 \; K_{(p-6)/2}(\kappa_o r_o)/K_{(p-3)/2}(\kappa_o r_o) \tag{31c}$$

MEAN ARRIVAL TIME OF SIGNALS

Let us assume the signal be a real pulse P(z,t) which is a superposition of plane waves, i.e.

$$P(z,t) = \int_{-\infty}^{\infty} f(\omega)u(z,\omega) \exp\{j\omega(t-\int_{z_S}^{z}\sqrt{<\varepsilon>}/c \; d\zeta)\}d\omega \tag{32}$$

where $f(\omega)$ is the amplitude spectrum and u is the complex amplitude with boundary condition $u(z_S,\omega)=1$ and satisfies approximately the parabolic differential equation (11). The pulse enters the slab normally at z=0 and propagates through the random medium. It is received at a point $z=z_R$ after emerging out of the slab at $z=z_B$. For a real pulse, $f(\omega)$ and $u(z,\omega)$ must be even in ω for an isotropic medium which is assumed here. The pulse is assumed to have a narrow bandwidth, i.e. if $\omega=\omega_c+\Omega$ where ω_c is the carrier angular frequency, then $\Omega<<\omega_c$. Hence, the pulse can then be expressed as

$$P(z,t) = A(z,t)\exp\{j(\omega_c t - k_c \int_{z_S}^{z}\sqrt{<\varepsilon_c>} \; d\zeta)\}+c.c. \tag{33}$$

where c.c. denotes the complex conjugate of the first term and the subscript c represents the corresponding parameters at the carrier frequency. The complex envelope A(z,t) is given as

$$A(z,t) = \int_{-\infty}^{\infty} d\Omega \; g(\Omega) \; V(z,\Omega) \cdot \exp\{j(\Omega t-(k\int_{z_S}^{z}\sqrt{<\varepsilon>}d\zeta - k_c\int_{z_S}^{z}\sqrt{<\varepsilon_c>}d\zeta))\} \tag{34}$$

where $g(\Omega)=f(\omega_c+\Omega)$ and $V(z,\Omega)=u(z,\omega_c+\Omega)$. Since we have assumed that the bandwidth of the pulse is small, thus the function $g(\Omega)$ is sharply peaked at $\Omega=0$. Therefore the limits of the integral in (34) can be extended from $-\infty$ to ∞ as shown.

Now, let us define the n^{th} temporal moment by the equation

$$M^{(n)}(z) \equiv \int_{-\infty}^{\infty} <A^*(z,t) \; t^n \; A(z,t)>dt \tag{35}$$

where n=0,1,2,.... This definition has been used in computations of quantum mechanical packets [Baird, 1972], in studies of wave dispersion [Anderson and Askne, 1974] and in calculating the spatial fluctuations of a light beam [Kon et al., 1974]. The ensemble average operator is present in (35) because A is random. When the expression (34) for A is substituted in (35) one can tell that the mutual coherence function will appear naturally.

We note from (35) that the zeroth temporal moment $M^{(0)}(z)$ is just the energy in the pulse.

Under the forward scatter approximation it can be shown that $M^{(0)}(z)$ equals to constant, i.e. the total energy in the pulse is conserved. In this case it is convenient to let $M^{(0)}(z)=1$. For the time reference, it is convenient to set $M^{(1)}(z_S)=0$ which means that a pulse is emitted at the sending end z_S when t=0. In general, $M^{(1)}(z_S)$ is not equal to zero but only if the impressed signal has a real symmetric envelope, i.e. $A(z_S,t)= A(z_S,-t)=A^*(z_S,t)$. This implies that $g(\Omega)$ must be real and even in Ω. For simplification, such a symmetric envelope is assumed. For general pulses, $M^{(1)}(z_S)$ must be subtracted out in order to establish the time origin.

Having all these assumptions, the mean arrival time is given by

$$t_a(z_R,z_S) = M^{(1)}(z_R) = \int_{-\infty}^{\infty} \langle A^*(z_R,t)\ t\ A\ (z_R,t)\rangle dt \qquad (36)$$

Putting A(z,t) given by (34) into (36) and using the results from last section, we obtain

$$t_a(z_R,z_S) = \frac{(z_R-z_S)}{c} + \Delta t + t_1 + t_2 \qquad (37)$$

where

$$\Delta t = \frac{1}{2ck_c^2}\int_{z_S}^{z_R} k_p^2 \cdot \left(1 + \frac{3}{4}\cdot\frac{k_p^2}{k_c^2}\right)d\zeta \qquad (37a)$$

$$t_1 = \frac{3}{4ck_c^2}\cdot\frac{\overline{\Omega^2}}{\omega_c^2}\cdot\int_{z_S}^{z_R} k_p^2 \cdot \left(2+5\frac{k_p^2}{k_c^2}\right) d\zeta \qquad (37b)$$

$$t_2 = \frac{-A_2}{2ck_c^4}\int_0^{z_B} k_p^4 \cdot (z_R-\zeta)d\zeta \qquad (37c)$$

$\overline{\Omega^2}$ is called the mean square bandwidth of the signal and is defined as

$$\overline{\Omega^2} \equiv 2\pi\int_{-\infty}^{\infty} g(\Omega)\ g^*(\Omega)\ \Omega^2\ d\Omega \qquad (38)$$

If the receiver is inside the random medium, (37) is modified to have different limits. The contributions to the mean arrival time in (37) are contained in four terms. The first term represents the time required for the signal to propagate a distance z_R-z_S in free space. The second term Δt is a correction factor to the first term due to the dispersive characteristics of the medium. The third term t_1 is a higher order correction to the first two terms and is caused by the finite bandwidth in the signal. This t_1 is proportional to $\overline{\Omega^2}$ and the dispersive characteristics of the medium and can be called a term of higher order dispersion. For a non-dispersive medium, t_1 vanishes. The fourth term t_2 accounts for the random scattering and diffraction in the random irregularity slab. The numerical value of A_2 in (37c) is negative and all the integrals in (37a,b,c) have positive values. Thus all these terms tend to increase the mean time of arrival of the signal.

MEAN SQUARE PULSE WIDTH

The mean square pulse width is defined as

$$\tau^2 \equiv \frac{M^{(2)}(z_R)}{M^{(0)}(z_R)} - \left[\frac{M^{(1)}(z_R)}{M^{(0)}(z_R)}\right]^2 \qquad (39)$$

where $M^{(2)}(z)= \int_{-\infty}^{\infty} \langle A^*(z,t)\ t^2\ A(z,t)\rangle dt$ and both $M^{(1)}(z_R)$ and $M^{(0)}(z_R)$ have obtained previously. The final expression for τ^2 is given by

$$\tau^2 = \tau_0^2 + \tau_1^2 + \tau_2^2 + \tau_3^2 \qquad (40)$$

where

$$\tau_0^2 = M^{(2)}(z_S) \qquad (40a)$$

$$\tau_1^2 = \frac{A_0}{4_c^2k_c^4}\left(1+10\frac{\overline{\Omega^2}}{\omega_c^2}\right)\int_0^{z_B} k_p^4\, d\zeta \qquad (40b)$$

$$\tau_2^2 = \frac{4A_4}{c^2k_c^6}\left(1+21\frac{\overline{\Omega^2}}{\omega_c^2}\right)\int_0^{z_B} k_p^4(z_R-\zeta)^2 d\zeta \qquad (40c)$$

$$\tau_3^2 = \frac{1}{c^2k_c^4}\cdot\frac{\overline{\Omega^2}}{\omega_c^2}\left[\int_{z_S}^{z_R} k_p^2 d\zeta\right]^2 \qquad (40d)$$

In (40), τ_0^2 is actually the mean square pulse width of the signal at the sending end, i.e. the original pulse width of signal. τ_3^2 is a correction term which is contributed by the dispersive characteristics of the medium only. τ_1^2 and τ_2^2 are the terms that come about because of scattering from the random irregularities. There is no A_2-term in (40) because it is negligible compared to all other terms under high frequency approximation. If we assume the signal has a Guassian frequency spectrum where

$$g(\Omega) = (2\sigma\pi^{3/2})^{-\frac{1}{2}} \exp(-\Omega^2/2\sigma^2) \qquad (41)$$

then $\tau_0^2=(4\overline{\Omega^2})^{-1}$. Out of these four terms in (40), with typical ionospheric parameters, τ_0^2 and τ_2^2 are the dominant terms. At low frequency, τ_2^2 is the largest.As the frequency increases, these two terms will have comparable values until at a much higher frequency, τ_0^2 dominants. Some numerical computations are shown in the next section.

NUMERICAL RESULTS

In this section, we are going to compute some numerical values of the mean arrival time and the mean pulse width for some typical ionospheric conditions. The background electron density is assumed to have a Chapman's profile. With the geometry shown in Fig. 1 the z-axis is pointing downward instead of the usual upward direction, the square of the plasma wave number for a Chapman layer is expressed as

$$k_p^2(\zeta) = \frac{N_o q^2}{c^2 m\epsilon_o} \cdot \exp\left[\frac{1}{2}\left(1+\frac{(\zeta-z_o)}{H}\right) - \exp\left(\frac{\zeta-z_o}{H}\right)\right] \tag{42}$$

where N_o is the maximum electron density; q--the charge of an electron; m--the mass of an electron; c--the velocity of light; H--the scaled height and z_o--the location of the peak of the electron density. The integrals in (37) and (40) can be obtained by making use of numerical analysis and a digital computer. They can also be worked out analytically.

Let us consider some typical ionospneric parameters where $N_o=10^{12}$ m^{-3}, H=80 km, p=4, $\kappa_o=10^{-4}$ m^{-1}, $r_o=10^{-1}$ m and $\sigma_N=0.2$. The peak of the inhomogeneous electron density is assumed to be at $z_o=0$. The distance between the satellite and receiver is usually 35,000 km and the top of the irregularity slab is about 500 km from ground having a thickness of about 100 km. Assuming the signal has a carrier frequency of 250 MHz and having $\sqrt{\overline{\Omega^2}}=2\pi\times10$ MHz, the mean arrival time is evaluated according to (37) which gives

$$t_a(z_R,z_S) = 0.117 + 7.095\times10^{-7} + 3.411\times10^{-9} + 4.278\times10^{-9} \text{ s.} \tag{43}$$

The time taken for the signal to travel that distance in free space is 0.117s. Out of the rest three correction terms, $\Delta t=7.095\times10^{-7}$ s. which is proportional to the total electron content is the dominant term. For the given values that are used in here, $t_2=4.278\times10^{-9}$ s. which is due to random scattering and diffraction in the random irregularities slab has a higher value than that correction factor $t_1=3.411\times10^{-9}$ s. For different set of parameter values, t_1 can be greater than t_2 where they are given in (37b) and (37c) respectively. These three correction terms are plotted in log-log scale in Fig. 2 with carrier frequency ranging from 10^2 to 10^4 MHz. They formed three straight lines and their values decrease as the carrier frequency increases. This is true because as the frequency increases, the factor $(1-\omega_p^2/\omega^2)^{1/2}$ is closer to unity. Thus, the signal seems to propagate in a medium more nearly approaching free space. But for frequencies between 100 and 200 MHz, the scattering can increase the time delay by 10 to 100 ns.

Using the same set of parameter values, the mean square pulse width is obtained as follows

$$\tau^2 = 6.333\times10^{-17} + 2.744\times10^{-16} + 2.518\times10^{-11} + 1.287\times10^{-14} \text{ s}^2. \tag{44}$$

For frequency below 1000 MHz, τ_2^2 given by (40c) dominates the mean pulse width of the signal. When frequency is above 4000 MHz, $\tau_0^2=6.333\times10^{-17}$ s^2 dominates over all the others, i.e. the pulse is not being distorted much. It is found that the frequency range in which the pulse is not distorted is very sensitive to the mean square bandwidth $\overline{\Omega^2}$. If instead of $\sqrt{\overline{\Omega^2}}=2\pi\times10$ MHz, we use $\sqrt{\overline{\Omega^2}}=2\pi\times1$ MHz, τ_0^2 will become 6.333×10^{-15} s^2 and the undistorted signal frequency range will be 1500 MHz and above. Furthermore, τ_2^2 given in (40c) is very sensitive to the choice of the inner scale r_o. In Fig. 3, the mean pulse width is plotted in log-log scale with three different values of r_o and remaining other parameters same as those used in (44). For $r_o=0.1$ m, it is still a question as to the validity of the forward scatter approximation in the parabolic equation method when the inner scale is so small. But even when $r_o=10$ m, at 100 MHz, the pulse width is lengthened by more than two orders of magnitude.

CONCLUSION

In this paper, the expressions for the mean arrival time (37) and the mean square pulse width about the mean (40) for an EM pulse passing through a random irregularities slab with an inhomogeneous background electron density are derived. The model that is used is more realistic in characterizing the ionosphere. For example, the random irregularity spectrum has a power-law profile instead of a Gaussian profile used by Yeh and Yang [1977] and the background electron density is allowed to be inhomogeneous instead of homogeneous. For some typical ionospheric parameters, the greatest contribution to the correction in time delay is Δt which is proportional to the total electron content in the medium and is given in (37a). Nevertheless, the random scattering factor brings an addition of 10 to 100 ns to the time delay. Actually, the importance of the thickness of the irregularities slab is also being investigated. It only affects the values t_2 given in (37c) and it is shown in Fig. 4, using the same parameters as in (43). When $z_B>150$ km, t_2 does remain the same. It is reasonable because we have assumed a Chapman profile for the electron density. For some other sets of ionospheric parameters, the correction factor t_1 and t_2 can be larger and this shows that the results due to random scattering and higher order dispersion have important implications for accurate satellite based navigational and communication systems.

The pulse lengthening effect depends on the carrier frequency being used. With a

higher carrier frequency, the signal is less distorted. It is also sensitive to the inner scale value that is being assumed. Here, it should be cautioned that if the inner scale is too small, the forward scatter approximation may not be valid. Further, the mean square bandwidth of the pulse also has a great effect on the range of frequency where the signal is less distorted.

ACKNOWLEDGEMENT

This work was supported partly by the National Science Foundation under Grant ATM 75-21755 and partly by the U.S. Army Research Office under the Grant DAAG 29-76-G-0286.

REFERENCES

Anderson, D. G. and J. I. H. Askne, "Wave packets in strongly dispersive media," Proc. IEEE, Vol. 62, USA, 1974, pp. 1518-1523.

Baird, L. C., "Moments of a wave packet," Am. J. Phys., Vol. 40, USA, 1972, pp. 327-329.

Kon, A. I., V. L. Mironov and V. V. Nosov, "Fluctuations of the centers of gravity of light beams in a turbulent atmosphere," Radiofizika, Vol. 17, USSR, 1974, pp. 1501-1511.

Liu, C. H. and A. W. Wernik, "A Characterization of transionospheric fading communication channel," IEEE Trans. Commun., USA, July 1975, pp. 773-776.

Liu, C. H. and K. C. Yeh, "Frequency and spatial correlation functions in a fading communication channel through the ionosphere," Radio Sci., Vol. 10, 1975, pp. 1055-1061.

Liu, C. H., A. W. Wernik and K. C. Yeh, "Propagation of pulse trains through a random medium," IEEE Trans. Ant. Prop., USA, July 1974, pp. 624-627.

Mark, W. D., "Characterization of stochastic transients and transmission media: The method of power-moments spectra," J. of Sound and Vibration, Vol. 22, No. 3, USA, 1972, pp. 249-295.

Shkarofsky, I. P., "Generalized turbulence space-communication and wave-number spectrum-function pairs," Can. J. Phys., Vol. 46, Canada, 1968, pp. 2133-2153.

Tatarskii, V. I., "The effects of the turbulent atmosphere on wave propagation," U.S. Department of Commerce, National Technical Information Service, Springfield, Va., USA, 1971.

Ulaszek, S. J., Jr., "A theoretical study of the transionospheric fading communication channel," IRL Tech. Rep. #54, University of Illinois, Urbana-Champaign, Il., USA, 1974.

Yeh, K. C. and C. H. Liu, "Theory of ionospheric waves," New York, USA, Academic Press, 1972.

Yeh, K. C. and C. H. Liu, "An investigation of temporal moments of stochastic wave," Radio Sci., Vol. 12, USA, 1977a, pp.671-680.

Yeh, K. C. and C. H. Liu, "Pulse delay and pulse distortion by random scattering in the ionosphere," presented at the AGARD EPP Symposium on Aspects of Electromagnetic Scattering in Radio Communications, Boston, USA, 3-7 Oct., 1977b.

Yeh, K. C. and C. C. Yang, "Mean arrival time and mean pulsewidth of signals propagating through a dispersive and random medium," IEEE Trans. Ant. Propagat., Vol. AP-25, No. 5, USA, 1977, pp. 710-713.

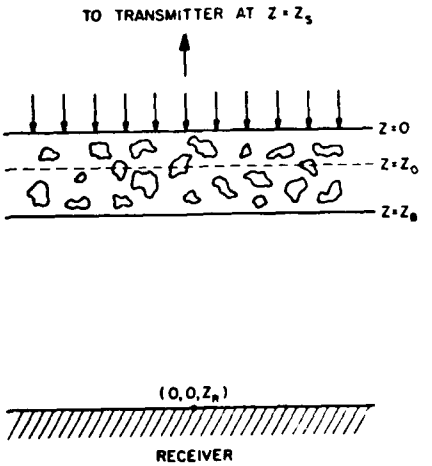


Fig. 1 The geometry of the problem. The plane wave signal is transmitted at z_S, is incident normally on top a slab containing random irregularities at z=0, emerges from the slab at $z=z_B$, and is received at z_R. The background ionosphere is assumed to be inhomogeneous with a peak density at $z=z_o$.

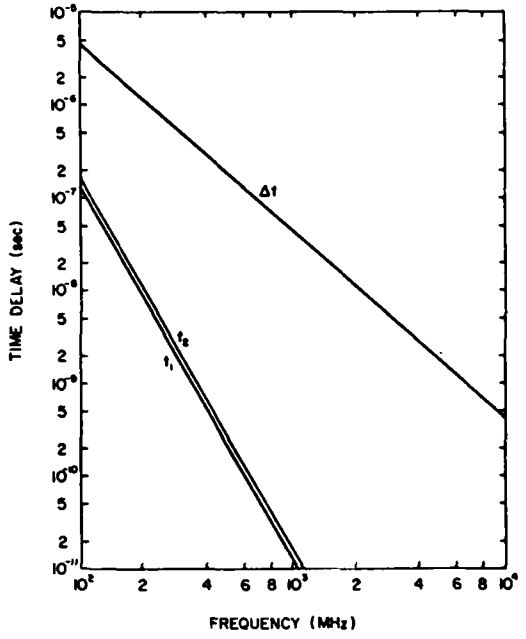


Fig. 2. Excess time delay as a function of frequency for parameters used in (43).

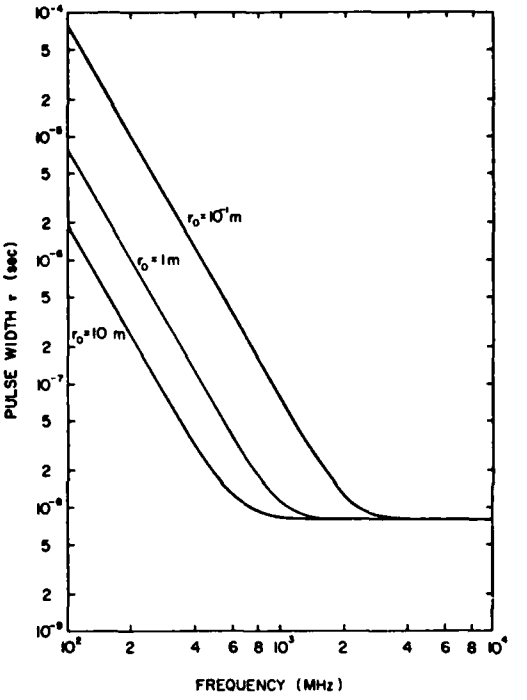


Fig. 3. Mean pulse width as a function of frequency for parameters used in (44). Three values of r_o are used and plotted here.

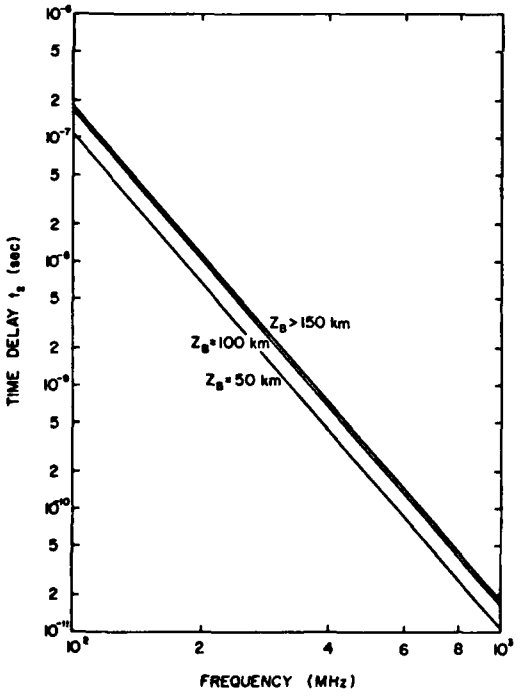


Fig. 4. The time delay t_2 as a function of frequency for parameters used in (43). Three different thicknesses of slab z_B are used and plotted here.

THE VARIABILITY OF IONOSPHERIC TIME DELAY

J. M. Johanson
Emmanuel College
Boston, MA 02115

M. J. Buonsanto
Boston University
Boston, MA 02115

and

J. A. Klobuchar
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

INTRODUCTION

Modern satellite navigation and satellite detection radar systems can be limited in accuracy due to the time delay caused by the ionosphere through which their RF signals must pass. The ionospheric time delay is directly proportional to the number of electrons encountered along the path from satellite to user and is inversely proportional to the square of the system operating frequency. Some systems, for example the NAVSTAR/Global Positioning System, GPS, are able to take advantage of the fact that the ionosphere is a dispersive medium, by measuring the ionospheric time delay at two widely spaced frequencies and thus are able to directly correct for this effect. Other precision ranging systems do not have the two frequency capability, however, and must instead rely upon models of ionospheric time delay to make corrections to increase system accuracy. Such models generally do a good job of predicting the monthly average behavior but, unless they are updated with a nearby measurement within a few hours of their use, do little to predict the day to day changes which occur.

The purpose of this paper is to outline the errors which result when only monthly average time delay values, without updating, are available for corrections to precise ranging systems. With these results systems design engineers will be able to determine how much improvement can be expected by near-real-time updating or by using a direct two frequency time delay measurement as compared with monthly average time delay models used without updating.

THE DATA BASE

Since the total number of electrons along the path from satellite to system user is directly proportional to the ionospheric time delay it is this parameter which was studied from a number of stations to determine the variability of time delay. This total number of electrons, called TEC, was determined from Faraday rotation measurements of linearly polarized VHF radio waves transmitted from geostationary satellites. Since the TEC is generally greatest during the mid-afternoon period, and the corresponding ionospheric time delays will be greatest during that period also, the deivations from average behavior are particualrly important during this part of the day. Further, since these systems must operate during all conditions of solar and geomagnetic activity we made no attempt here to separate magnetically quiet periods from disturbed times. Data for solar maximum years of 1968 and 1969 were reduced separately from the near solar minimum periods.

The Total Electron Content parameter is capable of describing many features of ionospheric behavior. The variability of this parameter may be easily described in terms of the departure from the monthly mean behavior by constructing the standard deviation,

σ. The percentage standard deviation may then be obtained via:

$$\sigma\ (\%) = 100\ \sigma/\bar{x}$$

where $\bar{x}$ is the monthly mean value. From the large quantity of data available, the diurnal, seasonal, solar cycle, and geographic variations of the percentage standard deviations of monthly mean TEC have been determined.

Total Electron Content data from eleven northern mid-latitude monitoring stations whose world-wide distribution is shown in Figure 1 was used in this study of ionospheric time delay. The TEC data base contains solar maximum data from six stations including Edmonton, Alberta, Canada; Aberystwyth, Wales; Hamilton, Massachusetts; Stanford, California; Honolulu, Hawaii; and Hong Kong. Solar minimum data was collected from the following stations: Narssarssuaq, Greenland; Goose Bay, Labrador; Hamilton, Massachusetts; Kennedy Space Flight Center, Florida; Athens, Greece; and Osan, Korea. Many people were responsible for the data used in this paper and they are each noted in the acknowledgements section.

DIURNAL VARIATION

Typical monthly mean diurnal TEC curves are shown in Figure 2 where data from Hamilton, Massachusetts is plotted for the solar maximum month of March 1969 and for the solar minimum month of March 1975 along with their standard deviations. It is clear in Figure 2 that as the TEC decreases, so does the absolute value of the variability with the result that changes in percentage standard deviation are small. Figure 2 also points out the fact that when operating at a frequency of 1.2GHz, one TEC unit corresponds to approximately one nanosecond of time delay. It should be noted that while Figure 2 represents a typical monthly mean diurnal TEC curve for a mid-latitude station, a near-equatorial station may have monthly mean TEC values which are 2 or 3 times greater than those shown in Figure 2. The corresponding magnitude of σ will also be greater for near-equatorial stations.

Figure 3 shows diurnal curves of σ (%) for TEC from six stations during the solar maximum year 1969, except for Hong Kong, for which 1968 data was used. Separate curves for winter (November - February), summer (May - August), and equinox (March, April, September, October) are given. Figure 4 shows similar curves for data from six stations during the period April 1974 through March 1975, a period not far from solar minimum. Figures 3 and 4 indicate lowest σ (%) values during the midday period where the curves are fairly flat, ranging from 15 to 25 percent. One must be careful when comparing Figures 3 and 4 for solar cycle effects because data from only one station, namely Hamilton, Massachusetts, was included in both figures. It may be concluded, however, that no obvious solar cycle dependence of σ (%) is evident. There is also no evidence of a geographical dependence except for the winter nighttime period under solar minimum conditions where a strong latitudinal gradient can be seen.

The data in the two figures seem to fall into one of two categories, the first being those periods exhibiting regular day-to-day variability of approximately 25 percent. The second includes those periods where the extremely high variability can be related to auroral and magnetic storm effects. The Narssarssuaq and Goose Bay nighttime data fall into this second group. At Narssarssuaq, the satellite to station raypath traverses the auroral oval during nighttime and at Goose Bay, the raypath intersects the auroral oval during magnetic storms. We unfortunately have no explanation for the large variability in the Honolulu data, observed especially during the nighttime period in the equinox and winter months.

MONTHLY VARIATIONS

An alternative way of examining the seasonal dependence of σ (%) for the two sets of stations is given in Figure 5. Here the average daytime (12 to 16 hours local time) TEC variability is plotted as a function of month for both the solar maximum and solar minimum periods. Again, we see no evidence of a solar cycle dependence. There is, however, a tendency for the curves to maximize in the equinoctial months. This may be due to the fairly rapid seasonal change between summer and winter diurnal patterns which occurs during those months.

DISTRIBUTION OF PERCENTAGE DIFFERENCES FROM MONTHLY MEAN TEC

Histograms giving the frequency distribution of percentage differences from the monthly mean daytime values of TEC are given in Figures 6 and 7.

Daytime TEC values from Hamilton, June 1971 are plotted in Figure 6 along with a Gaussian curve computed from the standard deviation of the frequency distribution and normalized to the area under the histogram. One can see that the normal curve gives an excellent fit to the data.

Figure 7 shows percentage differences during the daytime period for Hamilton, December 1971. This month includes a massive magnetic storm and may be considered as an example of worst case errors. The December 1971 data was chosen because it includes the greatest absolute deviations from monthly average values ever observed in ten years of observations made at Hamilton. In Figure 7, it can be seen that the normal curve (solid line), calculated in the same way as in Figure 6, is not representative of the data. If we eliminate the three hourly values of December 17, 1971 due to the storm, the standard deviation of the remaining data is greatly reduced, and the resulting normal fit (dashed line) is a better indication of the shape of the residuals.

In the months that we examined, the distribution of errors resulting from use of the monthly mean model were found to be approximately Gaussian. The December 1971 distribution was included only as a worst case example and even here it can be seen that by neglecting 3 hourly values, or 2 percent of the daytime values, the fit is also nearly Gaussian.

CONCLUSIONS

We have presented typical errors that a user of a single frequency system can expect when using monthly mean models, without updating, to correct for ionospheric time delays. If one number were to be chosen as the overall standard deviation of monthly mean time delay, one can assume approximately a 25 percent variability, although slightly higher values can be expected during the nighttime period when absolute time delay values are generally much lower.

There is no evidence of a solar cycle dependence of the percentage standard deviation and only a slight geographic dependence as one approaches the auroral zone. The daytime variability exhibits a peak during the equinox and is much more well behaved than the nighttime variability.

We have shown that the distribution of time delays about the monthly mean is nearly Gaussian for the cases examined.

ACKNOWLEDGEMENTS

Several experimenters were responsible for the data collection from various stations. The authors wish to express their appreciation for use of the data by acknowledging the following: TEC data from Stanford and Edmonton was provided by A.V. da Rosa; from Aberystwyth by L. Kersley; from Hong Kong by G.O. Walker; from Hawaii by H.T. Roelofs; from Narssarssuaq by I. Mikkelsen; and from Athens by D. Matsoukas.

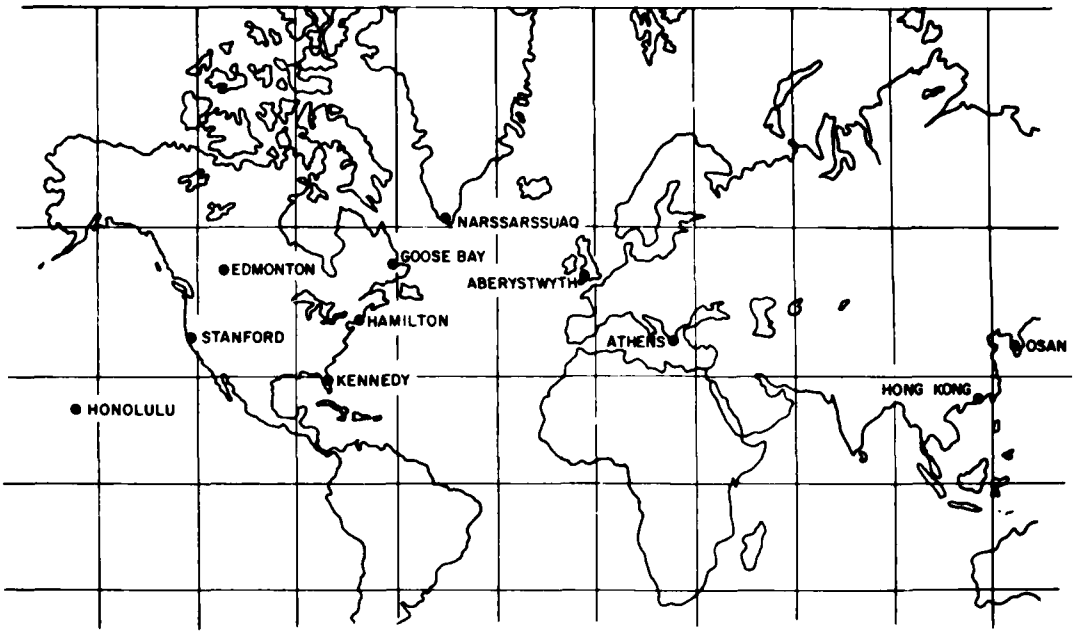


Fig. 1. Location of Ionospheric Monitoring Stations.

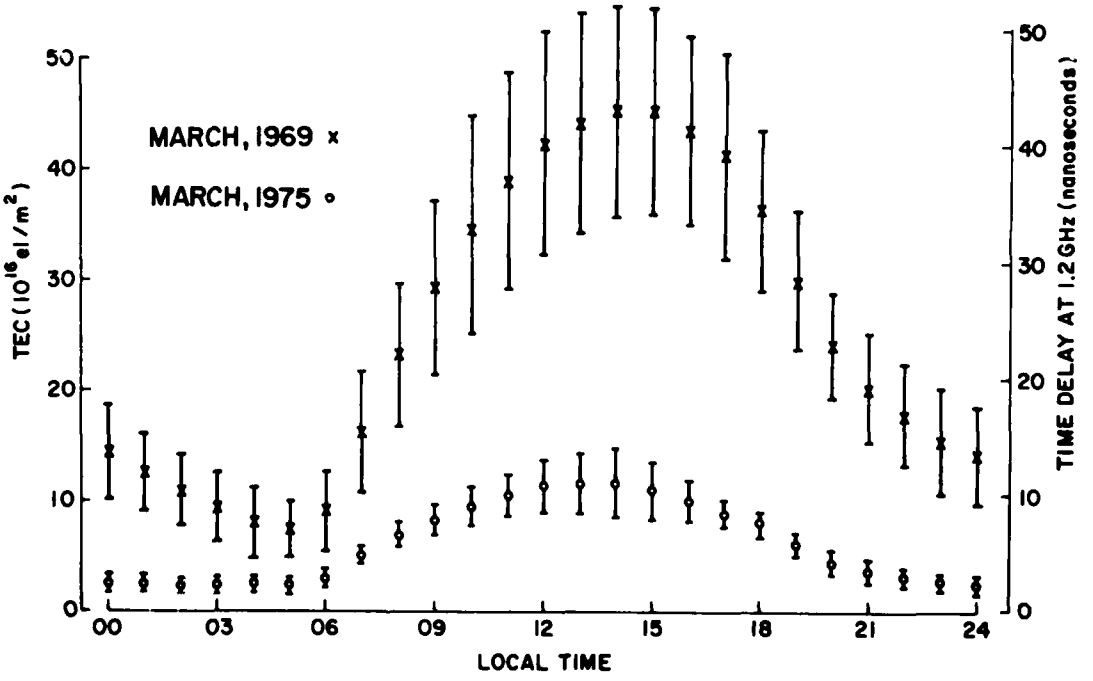
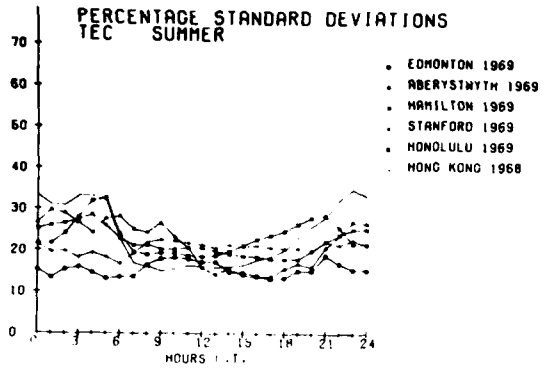
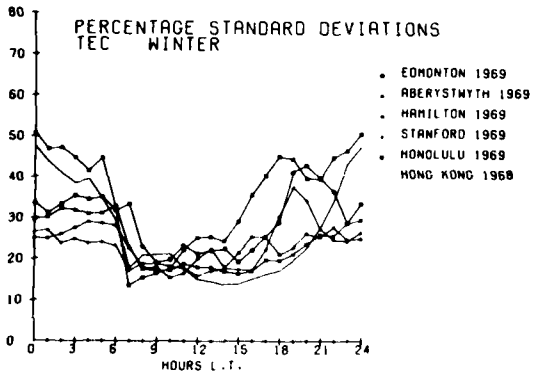


Fig. 2. Typical diurnal monthly mean curves and their standard deviations during solar maximum and minimum conditions observed at Hamilton, Massachusetts.





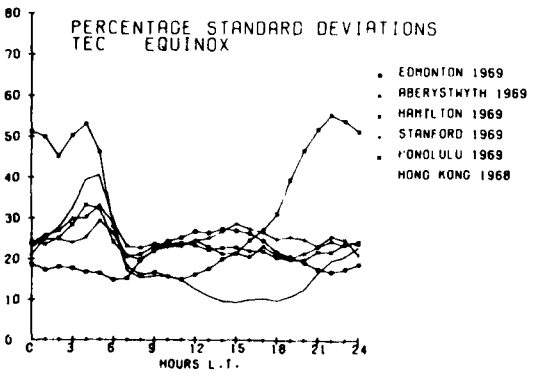
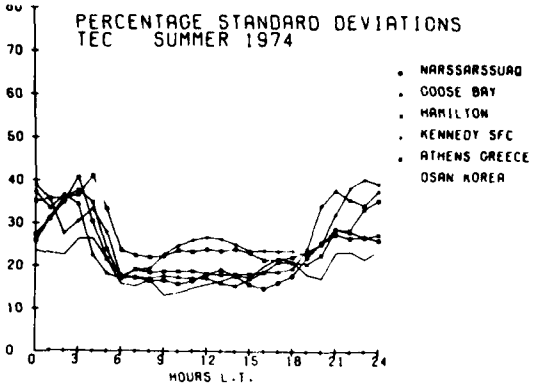
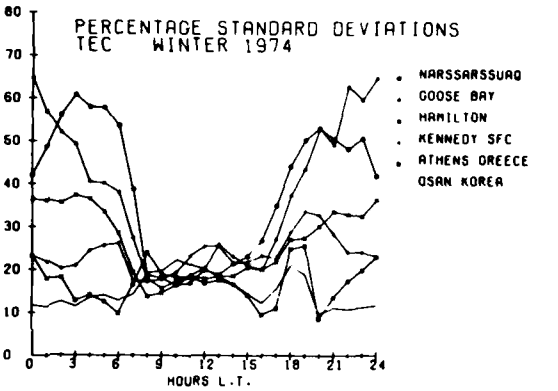


Fig. 3. Diurnal percentage standard deviation TEC curves for six stations during solar maximum.





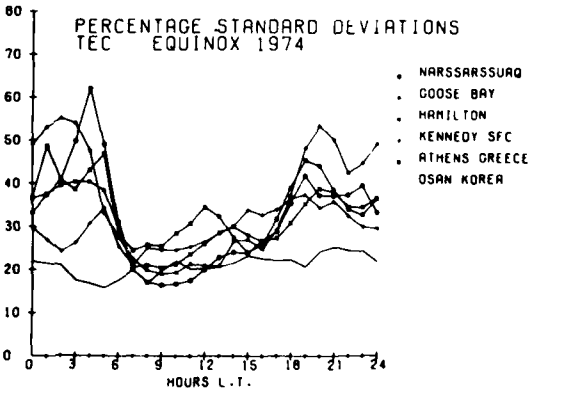


Fig. 4. Diurnal percentage standard deviation TEC curves for six stations during solar minimum.

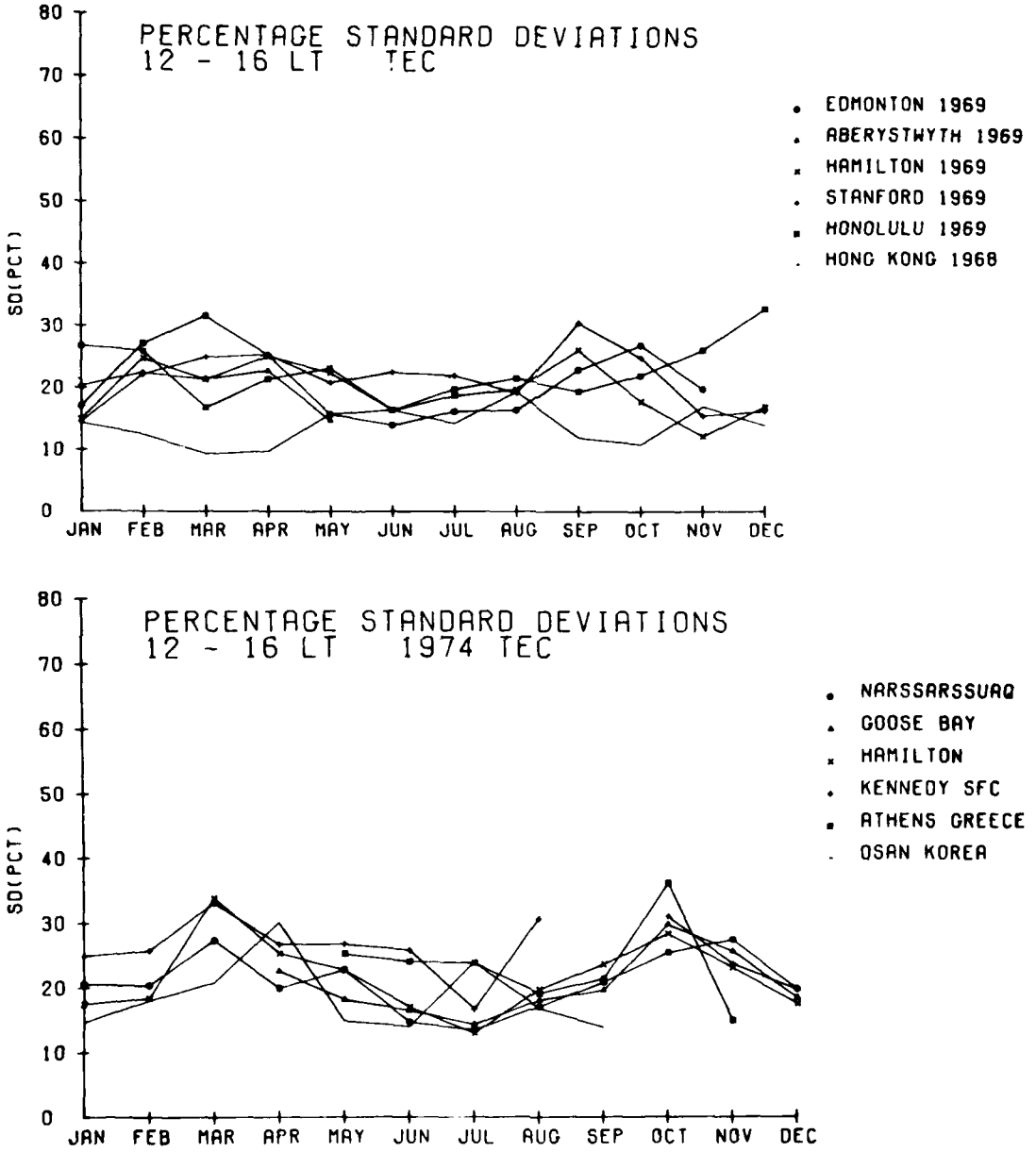


Fig. 5. Daytime percentage standard deviation vs. month for solar maximum and solar minimum conditions.

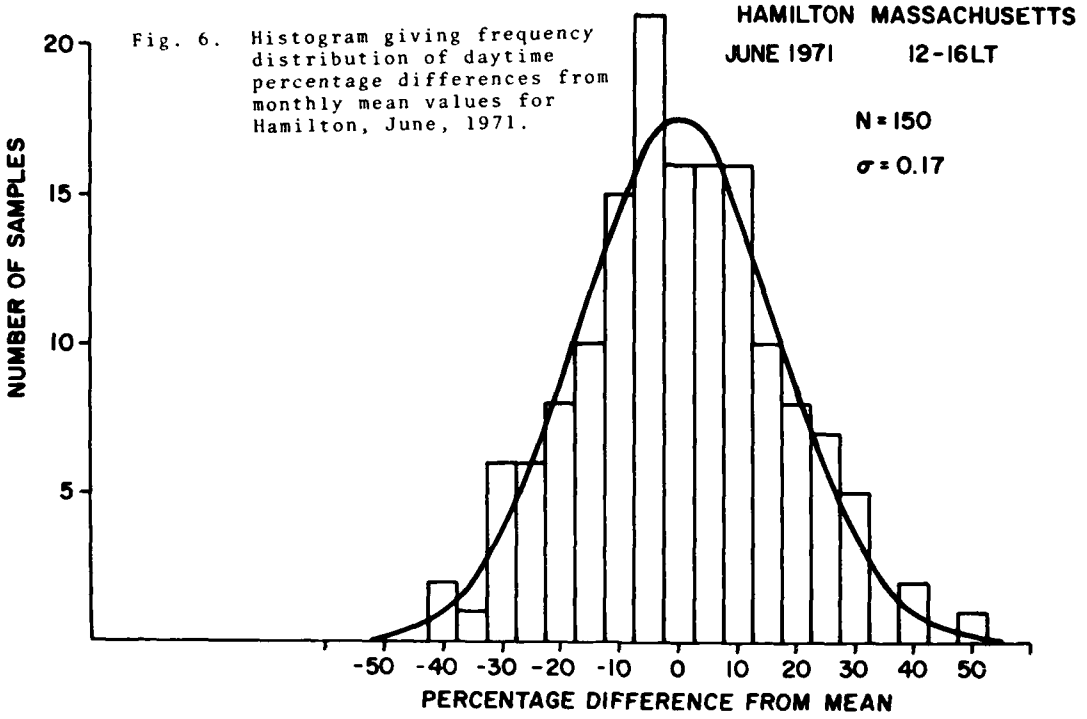


Fig. 6. Histogram giving frequency distribution of daytime percentage differences from monthly mean values for Hamilton, June, 1971.

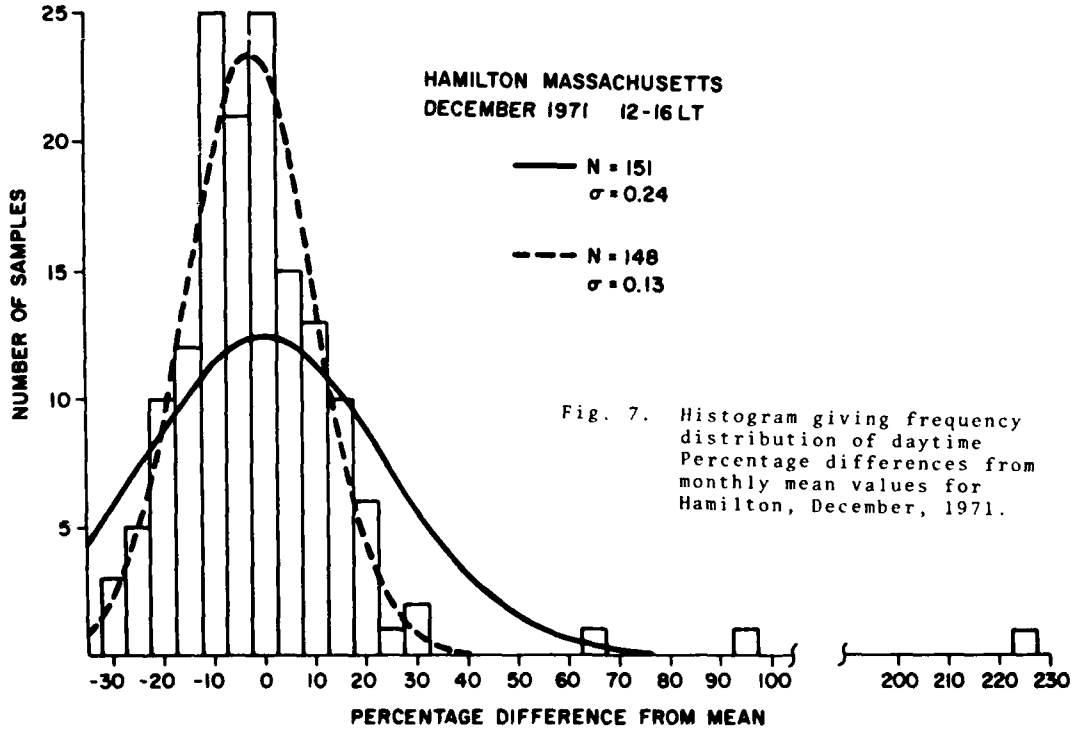


Fig. 7. Histogram giving frequency distribution of daytime Percentage differences from monthly mean values for Hamilton, December, 1971.

THE CONTRIBUTION OF THE PLASMASPHERE TO TOTAL TIME DELAY

J.A. Klobuchar
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731

M.J. Buonsanto and M.J. Mendillo
Boston University
Boston, MA 02215

J.M. Johanson
Emmanuel College
Boston, MA 02115

INTRODUCTION

The Total Columnar Electron Content (TEC) of the earth's ionosphere produces group delay of modulation of radio frequency waves which traverse the ionosphere. This group delay can produce significant range errors in advanced systems which require accuracies of a few feet in measured range. The NAVSTAR-Global Positioning System[1] (GPS) is an example of a precise ranging system currently under development by the Department of Defense which will undergo significant ionospheric group delay. Fortunately, the GPS will have two, widely spaced frequencies available from which the group delay of the ionosphere can be directly measured and subtracted from the apparent range to yield the true range from satellite to user. Some users may not choose the sophistication of the two frequency version of the system, however, and may instead choose to use an analytic model representation of the ionospheric time delay. To construct and to test such models and to enable advanced ranging system designers to obtain estimates of the magnitude of the ionospheric group delay error, use has been made of the available Faraday rotation data which is a measure of the TEC to a height of approximately 2,500 kilometers. Since advanced satellite ranging systems will be at orbital heights considerably greater than 2,500 kilometers, 20,000 kilometers in the case of the GPS, it is of interest to determine the additional contribution to ionospheric time delay not measureable from the available Faraday TEC data.

AVAILABLE PLASMASPHERIC ELECTRON CONTENT DATA

There are approximately 1000 station-months of TEC data obtained from the Faraday effect available from representative stations around the world, covering both solar maximum and minimum conditions. Until the launch of the geostationary satellite ATS-6 in mid-1974, there were no simultaneous Faraday and group delay measurements from which the additional contribution to TEC not measureable from the available Faraday TEC data could be determined. The ionospheric beacon transmitter carried on the ATS-6 geostationary satellite, Davies, et al[2], was designed to enable the measurement of simultaneous Faraday rotation and true group delay from its multi-frequency beacon transmissions. This is the first satellite to carry a beacon transmitter specifically designed for simultaneous Faraday and group delay measurements.

The difference between the group delay measurement of slant TEC, minus the Faraday electron content, defined here as a measure of the electrons in the region up to approximately 2,500 kilometers in vertical height, is called the plasmaspheric electron content, N_p. The currently available published monthly mean values of N_p are relatively few and consist of approximately 11 months of data from Boulder, Colorado, Fritz[3], a few months from Ft. Monmouth New Jersey, Soicher[4], one month each from Natal, Brazil and from Kiruna, Sweden, Soicher[5], and nine months of N_p monthly median values from

Aberystwyth, Wales, Hajeb-Hosseinieh[6]. In addition, Soicher[7], Webb and Lanzerotti[8], Poletti-Liuzzi[9], et al and Hajeb-Hosseinieh[6] have published studies of the behavior of N_p during a few magnetic storms, and Klobuchar and Johanson[10] presented a comparison of the monthly average N_p results at Boulder with those from Hamilton, Mass. Since all the available N_p data is from the period later than May 1974 it is representative of solar minimum conditions only.

Kersley[11] has recently pointed out that plasmaspheric electron content measurements taken from stations at nearly the same latitude, but differing in longitude, while viewing the ATS-6 satellite at the same satellite location, actually probe nearly the same plasmaspheric region even though the stations may have relatively large longitudinal differences. This is simply because the ray paths from the ground stations all converge at the satellite at 36,000 kilometers and in the plasmasphere they are also not significantly different. Klobuchar and Johanson indeed showed that the difference in monthly behavior between data sets taken at Hamilton, Mass. and Boulder, Col. was small. In view of this we shall take the Hamilton, Mass. data as being typical of the US region and the data from Aberystwyth, Wales as typical of the European sector.

EXPERIMENTAL RESULTS

With a substantial amount of data available only from Wales from the European sector and with the Hamilton, Mass. data representative of the US data it is of interest to compare the absolute values of the monthly average N_p values versus local time. Data from Hamilton are from July 1974 through May 1975, taken when ATS-6 was at 94°W longitude, while data from Aberystwyth are from November 1975 through July 1976 during which time ATS-6 was located at 35°E longitude. The scale on the left in Figure 1 is in TEC units of $m^{-2} \times 10^{16}$, while the right hand scale is in nanoseconds of group delay at a system operational frequency of 1.6 GHz, the single frequency mode GPS system operational frequency. Also plotted in Figure 1 is the one month's available N_p data from Sao Paulo, Brazil for May 1975 and the one month's available N_p data from Kiruna, Sweden for October 1975.

It is immediately obvious from Figure 1 that the Wales data is significantly higher than the values from the US. Kersley and Klobuchar[12] have attributed the difference between the N_p values in the American and European sectors as being due to the lower geomagnetic latitudes reached by the Aberystwyth ray path to ATS-6 as opposed to the ray from Hamilton, Mass. to ATS-6. At the lower geomagnetic latitudes the scale height at the base of the plasmasphere is predominantly controlled by H^+ ions, with consequent greater scale heights, as opposed to the higher magnetic latitudes where more O^+ ions yield smaller scale heights. The Wales ray path reaches a minimum L shell of 1.7 as opposed to 2.2 for the Hamilton ray path.

While the differences in the diurnal behavior of the two sets of data shown in Figure I can be explained by their different minimum geomagnetic latitudes, for the systems engineer it is of more interest to point out that the day-to-night variation for both sets of data is less than the difference between the two data sets. The absolute value of the time delay is small in terms of present system accuracy requirements. In view of the small absolute values of time delay no attempts were made to fit diurnally changing functions to the monthly average data from the station. Simple constant values of 2 nanoseconds for the Wales data and 1 nanosecond for the US representative data are probably sufficient to account for the monthly average values of plasmaspheric electron content.

PLASMASPHERIC GROUP DELAY DURING MAGNETIC STORMS

Major magnetic storms produce electrodynamical effects which cause the emptying of the plasmasphere into the ionosphere and/or the peeling off of the plasmasphere away from the near-earth region. The time for refilling of the plasmasphere has been studied by use of whistler measurements by Park[13], and recently for the N_p data by Soicher[7], Poletti-Liuzzi[9], et al, Webb and Lanzerotti[8], and by Hajeb-Hosseinieh[6]. The time for refilling of the plasmasphere as viewed from the mid-latitudes is on the order of several days. These results from a few representative magnetic storms illustrate that from 3 to 8 days are required for N_p to return to its approximate average value.

CONCLUSIONS

With the limited plasmaspheric

data available thus far the additional contribution to ionospheric group delay not previously available from Faraday rotation measurements of TEC is of the order of 1 to 2 nanoseconds at 1.6 GHz. This value is certainly small compared to the present system operational requirements, but it is an additional amount which can be added to existing models of time delay. The difference between the European and US data is significant and perhaps represents the two limiting cases of N_p from the mid-latitudes as the US longitudes have the highest magnetic latitudes compared to geographic latitudes, while the European magnetic latitudes are lower than their geographic latitudes. Insufficient data are presently available from higher latitudes or from near equatorial latitudes to determine whether either of these regions will have values of N_p significantly different from those in the mid-latitudes. The one month of data available from Kiruna, Sweden compares well with the Aberystwyth data. The Sao Paulo data is at a much lower latitude than any of the other stations, yet it is consistent with the other data sets.

During magnetic storms N_p drops to very small values and takes from 3 to 8 days to recover to the average value. If information on occurrence of magnetic storms is available the model value of N_p can be modified accordingly. Of course, during magnetic storms much greater changes occur in the TEC of the ionosphere and the N_p depletions will be generally insignificant compared to the greater changes from average conditions which occur in the ionosphere. Thus, any TEC model must be capable of representing the ionospheric changes from average conditions, before the much smaller N_p values need be of concern.

All available N_p data are for solar minimum conditions and are from only two representative mid-latitude regions of the world, with small exception. It is expected that the electron content of the plasmasphere will increase with increasing solar activity as does the electron content of the ionosphere. Measurements of N_p using signals from ATS-6 are continuing.

REFERENCES

1. Lassiter, E.M., and B. Parkinson, "The Status of the Development of the NAVSTAR-Global Positioning System", Navigation, Volume XXV, No. 97, pp 13-26, January 1977.

2. Davies, K., R.B. Fritz, R.N. Grubb, and J.E. Jones, "Some Early Results from the ATS-6 Radio Beacon Experiment", Radio Science, Vol. 10, No. 8 and 9, pp 185-799, 1975.

3. Fritz, R.N., "ATS-6 Radio Beacon Electron Content Measurements at Boulder, July 1974-May 1975", Report UAG-58, World Data Center A for Solar-Terrestrial Physics, Boulder, Colorado, September 1976.

4. Soicher, H. "Diurnal, Day-to-Day, and Seasonal Variability of N_f, N_t, and N_p at Fort Monmouth, New Jersey", Cospar Satellite Beacon Symposium, Boston, Ma., Boston University, Ed by M. Mendillo, June 1976.

5. Soicher, H., "Comparative Ionospheric and Plasmaspheric Electron Contents from the Three World Regions", Nature, Vol. 264, pp 46-48, Nov. 4, 1976.

6. Hajeb-Hosseinieh, H., "Ionospheric Studies Using Geostationary Satellites", Thesis submitted to the University of Wales, Aberystwyth, Wales, June 1977.

7. Soicher, H., "Response of Electrons in Ionosphere and Plasmasphere to Mangetic Storms", Nature, Vol. 259, pp 33-35, Jan. 1 & 8, 1976.

8. Webb, D.C. and L.J. Lanzerotti, "Temporal Variations in Total Equatorial Plasmasphere Content and their Relationships to the Ring Current Intensity and the Plasmapause", presented at Spring meeting of Americal Geophysical Union, Washington, D.C. June, 1977.

9. Poletti-Liuzzi, D.A., K.C. Yeh, and C.H. Liu, "Radio Beacon Studies of the Plasmasphere", Journ. of Geophysical Res., Vol. 82, No. 7, pp 1106-1114, March, 1977.

10. Klobuchar, J.A. and J.M. Johanson, "A Comparison of Average Plasmaspheric Electron Content at Two Mid-Latitude Stations", presented at Spring meeting of American Geophysical Union, Washington, D.C., June 1977.

11. Kersley, L., private communication.

12. Kersley, L. and J.A. Klobuchar, "Comparison of Protonospheric Electron Content Measurements from the American and European Sectors", in preparation.

13. Park, C.G., "Some Features of Plasma Distribution in the Plasmasphere Deduced from Antarctic Whistlers", Journ. of Geophys. Res., Vol. 79, No. 1, pp 169-173, January 1, 1974.

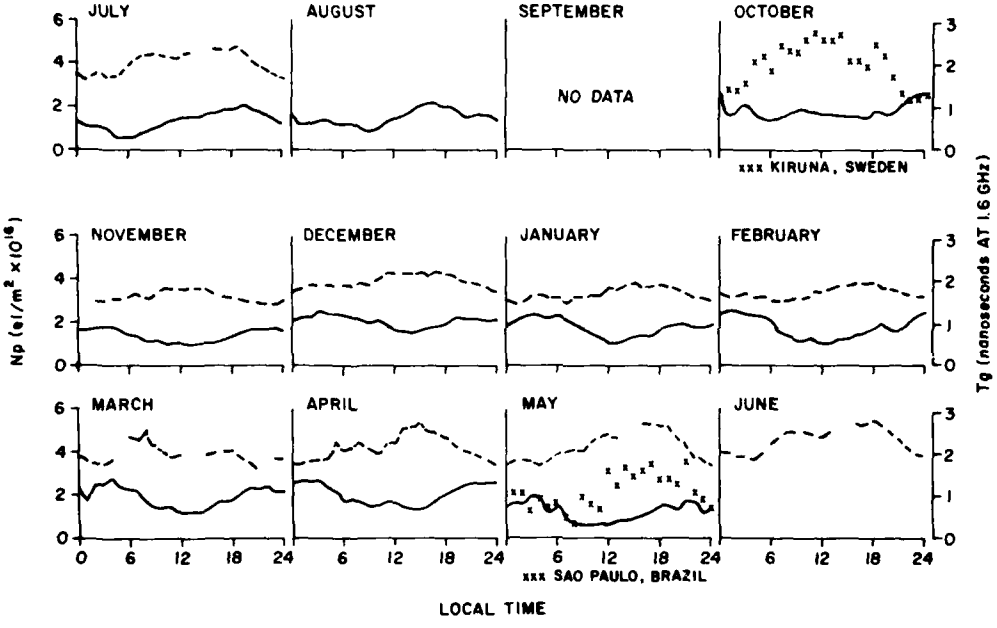


Fig. 1. Monthly average plasmaspheric electron content versus local time for Aberystwyth, Wales (dashed line) and for Hamilton, Mass. (solid line). Also plotted are values from Kiruna, Sweden for October 1975 and from Sao Paulo, Brazil for May 1975. *The right hand axis is the equivalent time delay in* nanoseconds at 1.6 GHz.

TEMPORAL VARIABILITY OF IONOSPHERIC REFRACTION CORRECTION

D. E. Donatelli
Regis College Research Center
Weston, Massachusetts

R. S. Allen
Air Force Geophysical Laboratory
Hanscom AFB, Mass. 01731

INTRODUCTION

Precision radar and navigation systems now require corrections which consider current ionospheric conditions over their field of view. Techniques are being developed using measurements from calibration satellites to adapt the numerical maps which provide monthly median corrections. These basic maps are derived from a world-wide climatology of ionospheric parameters; their use alone reduces the residual error in range or time delay measurements to 20 to 25 percent of the median correction in daytime and to 30 to 35 percent at night, reflecting the day-to-day variability of the ionosphere about its monthly median values. To reduce this residual, several techniques have been proposed which would scale the maps up or down by a constant percentage, change gradients, and reproduce large local features. This presentation examines the potential effectiveness of such techniques.

Since the refraction corrections are directly proportional to the electron content along the slant path through the ionosphere, measurements of total electron content from Hamilton, Mass. and Goose Bay, Labrador, 30.7°N, 70.7°W, and 47.5°N, 62.2°W geographic, respectively, can be used as calibration measurements for updating median estimates of refraction correction. The data were obtained from measurements of the Faraday rotation of the VHF beacon on the ATS-3 geostationary satellite and were reduced to equivalent vertical electron content (TEC). The years 1970 and 1976 at Hamilton, Mass., taken as representative of solar maximum and minimum, respectively, were used to determine solar cycle and seasonal variations. The data for 1972 from Hamilton, Mass. and Goose Bay, Labrador were used for an initial study of spatial variation in the effectiveness of the updating technique.

EFFECTIVENESS OF AN UPDATING TECHNIQUE

The potential for reduction of residual error while using an updating technique is demonstrated in Figure 1 for March, 1972, with data from Hamilton, Mass. The heavy solid line represents the expected residual error, δR_m, when only the predicted median was used for correction. The light lines depict the time-decay in residual error when a calibration measurement was used. At each calibration time a scaling factor was determined from the ratio of the calibration measurement and the predicted median. This factor was then used to scale the predicted median of the subsequent 12 hours, in 15-minute intervals. There was a 0-error at the time of calibration, shown at every second hour, and these curves show the increase in error with time, as well as the length of time required for the error to approach δR_m.

The dashed lines represent the diurnal variation in the residual rms error using a scaling factor determined 30-minutes previously, δR_{30m}, and 1-hour previously, δR_{1h}. After 30 minutes the maximum error increased to about 3 meters and after 1 hour it increased to 5 meters. These residual errors are significantly below the expected residual error of 18 meters for this time of day. This 70-80 percent reduction occurs when refractive effects are greatest and the need for improved accuracy is most critical. Note that a scaling factor should not be used to project an update across either the sunrise or sunset terminator as this may increase error, particularly near sunrise. Unless calibration measurements can be made at these times in 15-30 minute intervals it would be preferable to simply use the climatological prediction.

This applies at any time in which steep gradients in TEC are expected to occur, such as during magnetic disturbances. A "worst case" example of this is noted in Figure 2. On March 8, 1970, a severe magnetic storm occurred, and the effects observed in the TEC data at Hamilton, Mass. were the largest for the total 9 years examined from 1968 through 1976, both for the extremes and rapid changes in TEC.

The ordinate of the upper scale is range correction in meters at 425 Mhz. The predicted median range correction for March, 1970, ΔR, was the range correction the radar would have used on this particular day if no updating procedure existed. The actual range correction that should have been made, ΔR_d, shows that a correction of nearly 195 meters was required at about 1500 hours on that day, more than twice the maximum of 95 meters for ΔR. If an updating technique, such as the one previously described had been used, the range correction made using calibration measurements after 30 minutes would have been ΔR_{30m}, and after 3 hours, ΔR_{3h}.

The ordinate of the lower scale is the absolute error, in meters, that would have been experienced with the corrections presented for this day. The error between: the observations on that day, ΔR_d, and the monthly median, ΔR, is δR_m; between ΔR_d and the 30 minute update, ΔR_{30m}, is δR_{30m}; and between ΔR_d and the 3 hour update, ΔR_{3h}, is δR_{3h}. Using the 30 minute update resulted in a maximum error of less than 50 meters compared to the maximum of nearly 110 meters if only the median had been used. On the other hand, the 3 hour update produced errors on the order of δR_m or greater during this disturbance. The salient feature of this result occurred near 1800 hours when the 3-hour delay caused a large error during the time when a negative gradient was observed. It follows that whenever steep and variable gradients are likely to occur, calibration measurements should be made in intervals no greater than 30 minutes, not only to achieve significant accuracy, but also to avoid introducing errors.

The effectiveness of an updating technique varies seasonally and over solar cycle as well as diurnally. The principal features of these variations are summarized in Figure 3 for the years 1970 and 1976 representing solar maximum and minimum, respectively, at Hamilton, Mass. Included are the maximum values each month for: the predicted median range correction, ΔR; the residual error when the median is used for correction, δRm, which represents the day-to-day variability of the ionosphere; the reduction in residual error when calibration measurements are used in updating techniques each 15 minutes, δR_{15m}; 30 minutes, δR_{30m}; 1 hour, δR_{1h}; and 3 hours, δR_{3h}.

Comparison of ΔR and δR_m showed the 75 percent reduction that can be achieved by simply using a median correction for ionospheric effects. This is possible for both solar maximum and solar minimum conditions, and at all seasons. A further reduction, through use of local measurements in an updating procedure, showed less consistent results. At solar maximum, in daytime, even an update that is 3 hours old will provide a significant reduction in residual error. But at solar minimum, a measurement much more than 1 hour old will provide no significantly greater accuracy than using only the expected median. Whereas the monthly median, ΔR, and the day-to-day variability, δR_m, are a factor of 4 greater at solar maximum than at solar minimum, the residual error using an updating procedure, as shown by δR_{15m}, δR_{30m}, δR_{1h}, and δR_{3h}, varies by only a factor of 2 over the solar cycle. Thus it appears that at solar maximum there are longer period variations that are readily corrected with the updating procedure. The shorter period fluctuations which are superimposed on these long-term variations exist at both solar maximum and solar minimum and impose a limit on the effectiveness of updating techniques.

Seasonally, summer is the time of least day-to-day variability and least residual error after an updating procedure. The greatest residual error appears at the equinoxes and can be considerably reduced with frequent updating, as shown by δR_{15m} and δR_{30m}. It is in winter that updating is least effective. At solar maximum, the residual error can be reduced to a maximum of 25 meters with 15-minute updating, and to 40 meters with 30-minute updating. The maxima occur in winter and at the equinoxes. This becomes even more apparent at solar minimum, where δR_{15m} and δR_{30m} show a definite maximum in winter and minimum in summer.

The same updating procedure was used with data from 1972 for Hamilton, Mass. and Goose Bay, Labrador. The results from both stations are compared in Figure 4, using the same parameters as in Figure 3. It is seen that at both stations the curves for each parameter are comparable; the fact that Goose Bay is a higher latitude station does not appear to be significant. The general results are similar to those described for 1970, at Hamilton in Figure 3, which may be attributed to the fact that the year 1972 was one of increased solar activity. Therefore it is not surprising that the results resemble solar maximum conditions rather than a median which is more representative of the descending phase of the solar cycle.

CONCLUSION

This study has shown that reduction in residual error due to the day-to-day variability of the ionosphere is possible through use of an updating technique coupling local

measurements with climatological predictions of median ionospheric behavior. Reductions of 70 to 80 percent are possible in daytime, particularly during periods of increased solar activity. It has been shown that while the residual error varied by a factor of 4 between solar maximum and minimum, use of the updating procedure reduced this variation to a factor of 2. Therefore, this technique was more effective when the residual error in range correction was greatest.

It has also been shown that caution is required in choosing the length of time a particular scaling factor is used. For an average day, the constraint on this time is set by the sunrise-sunset terminators. A scaling factor may be used for any length of time within the day or nighttime period in which it is determined, but it must not be projected across the terminator.

The large, variable gradients of the sunrise and sunset periods can be equated with those occurring in periods of severe magnetic disturbance. The interpretation of these gradients provided by a median prediction can be improved upon by an updating procedure using calibration intervals of less than 30 minutes. Longer intervals are likely to introduce errors of larger magnitude than the variation from the predicted median.

The temporal decay in effective updating is dependent on the period of the TEC variations. Long period fluctuations, which are readily corrected, are a consistent feature at any site. These include day or nighttime periods that are enhanced or depressed with respect to the median behavior. These periods must be longer than the time interval between calibrative measurements for updating to be effective. Shorter period fluctuations are generally present and the percentage amplitude of these variations is the minimum percentage residual error achievable.

By considering these results with the following factors: range correction is inversely proportional to the radio frequency squared and directly proportional to the total electron content (TEC) along the wave path; TEC, to a first approximation, varies linearly with 12-month running mean, sunspot number and solar flux at 2800 Mhz, it is possible to estimate performance for any system, at any location, at any time.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. J. Klobuchar of the Air Force Geophysics Laboratory for use of the TEC data from Hamilton, Mass. and Goose Bay, Labrador. This work was performed under Air Force Contract F19628-76-C-0255.

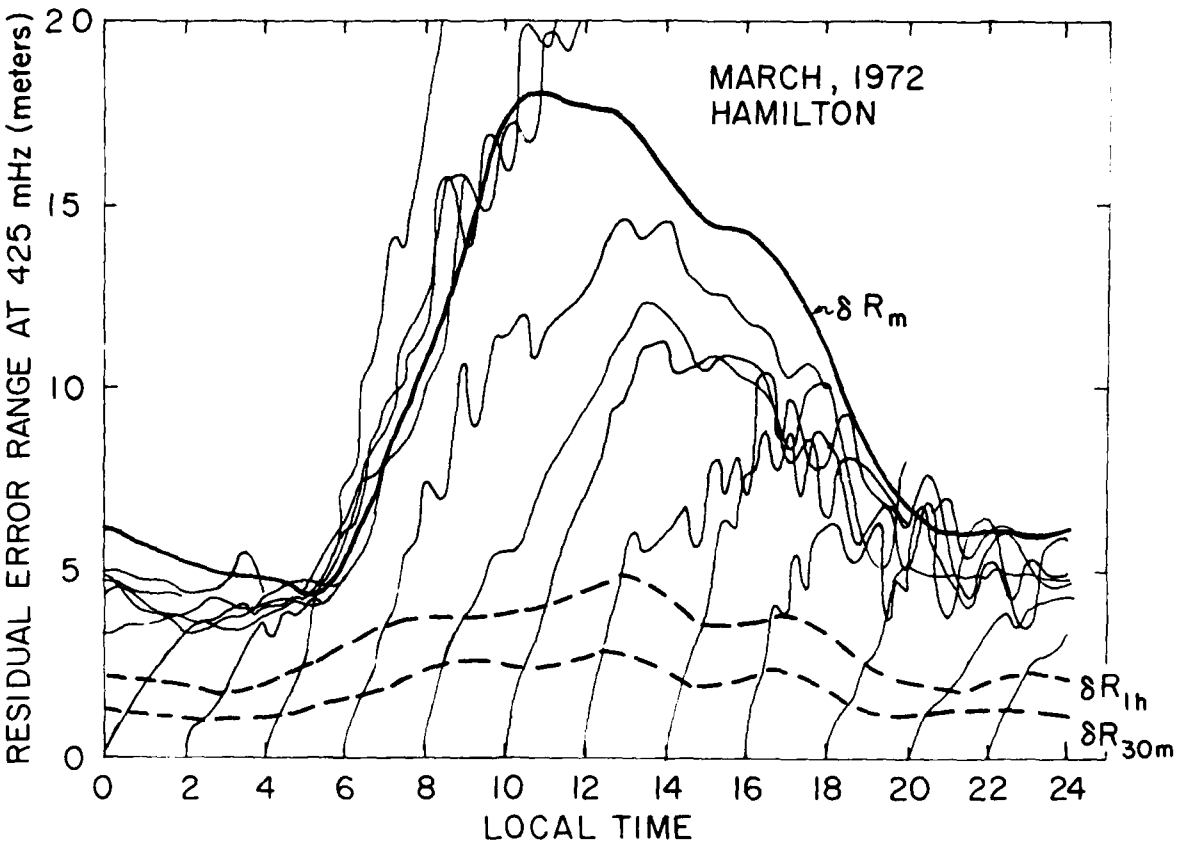


Figure 1. The residual error in range correction using an updating technique ranging from 0-residual error at the time of calibration to the residual error using only the median correction, δR_m, at a time depending primarily on the interval between calibration and subsequent terminator.

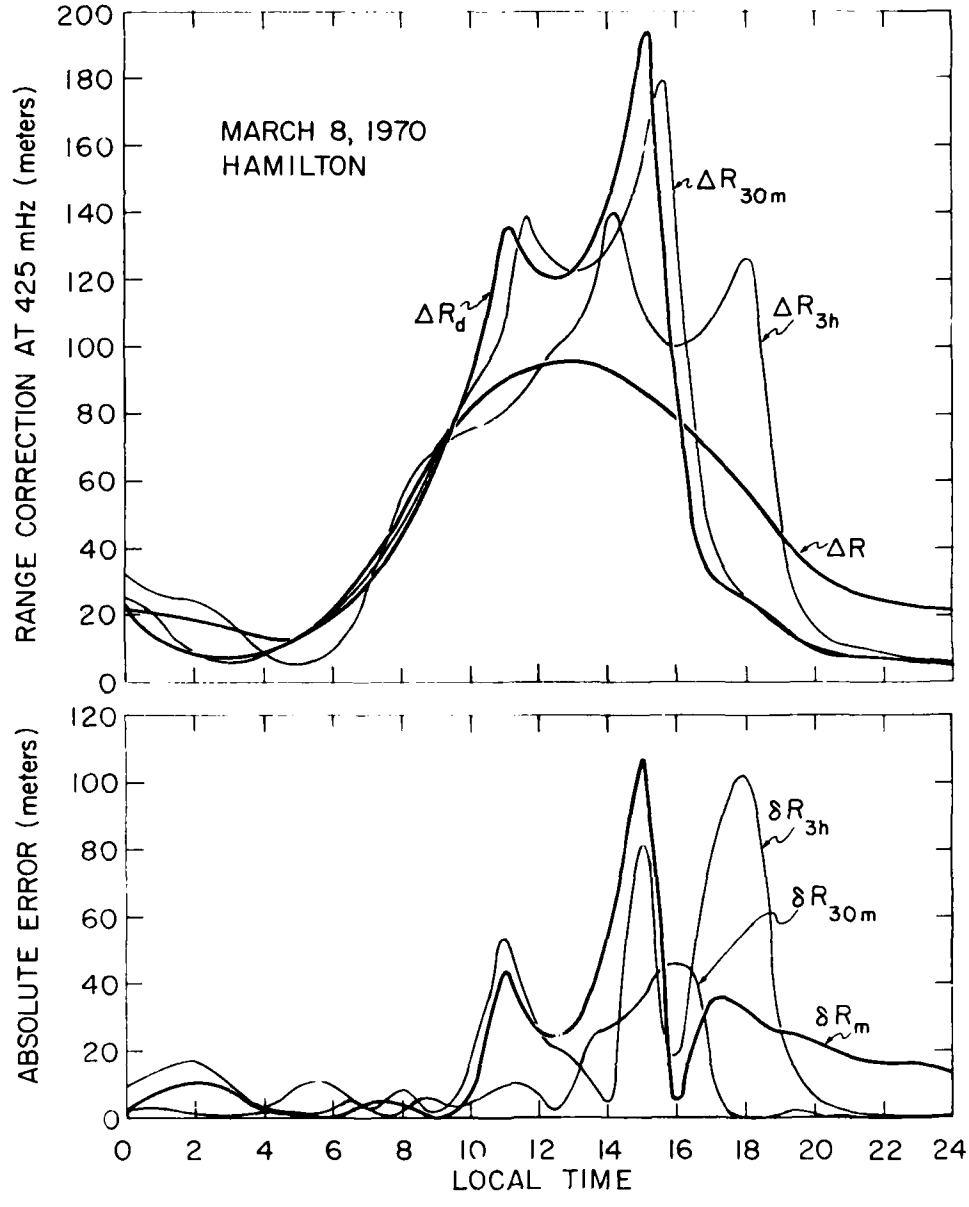


Figure 2. The effectiveness of an updating procedure during a severe magnetic disturbance, comparing the actual required range correction, ΔR_d, with the predicted median, ΔR, and the 30 minute, $\Delta R_{30\,m}$, and 3 hour, $\Delta R_{3\,h}$, updated predictions. The differences in the range correction curves of the upper scale are presented as absolute error on the lower scale.

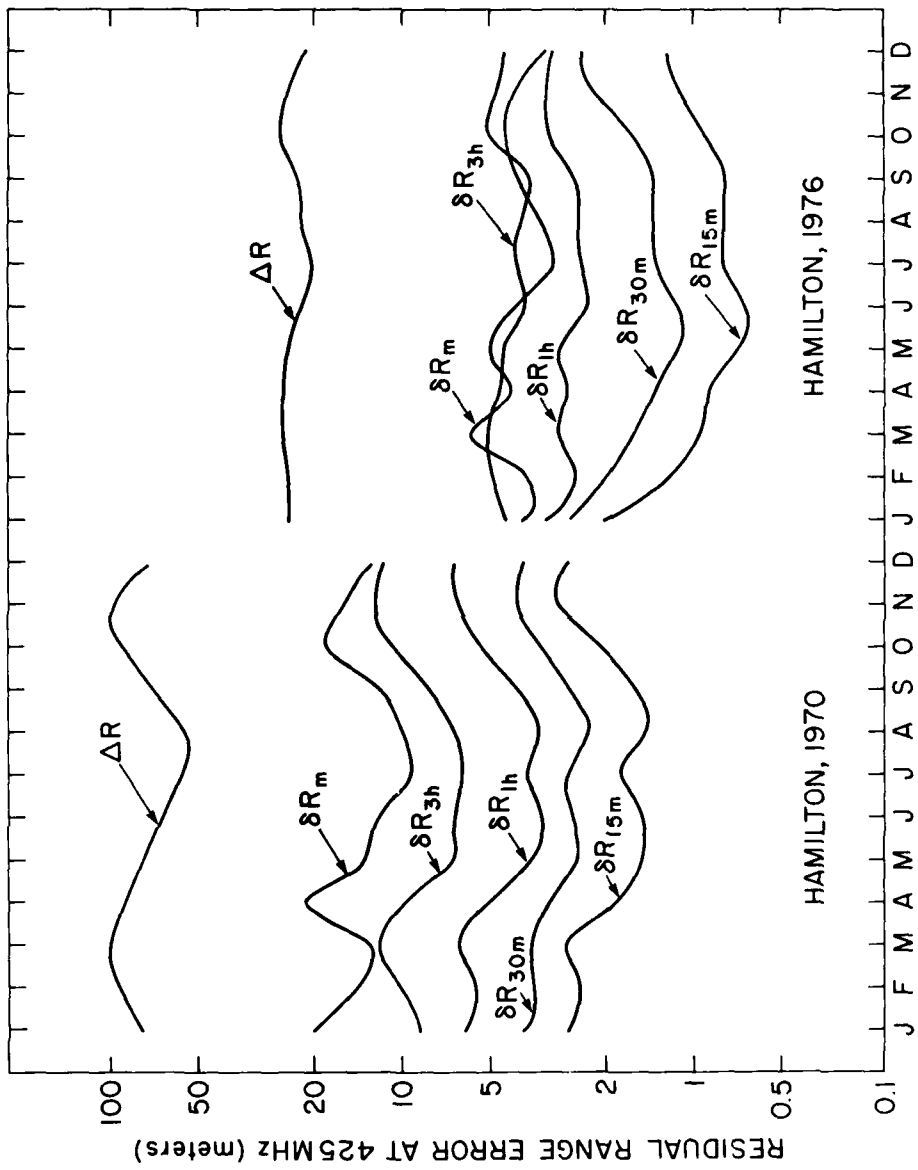


Figure 3. The reduction in maximum monthly residual error in range correction at solar maximum, 1970, and solar minimum, 1976, using updated predictions 15 minutes, 30 minutes, 1 hour and 3 hours after calibration. These are compared to the maximum residual error, δR_m, after the non-updated median correction, ΔR, is applied.

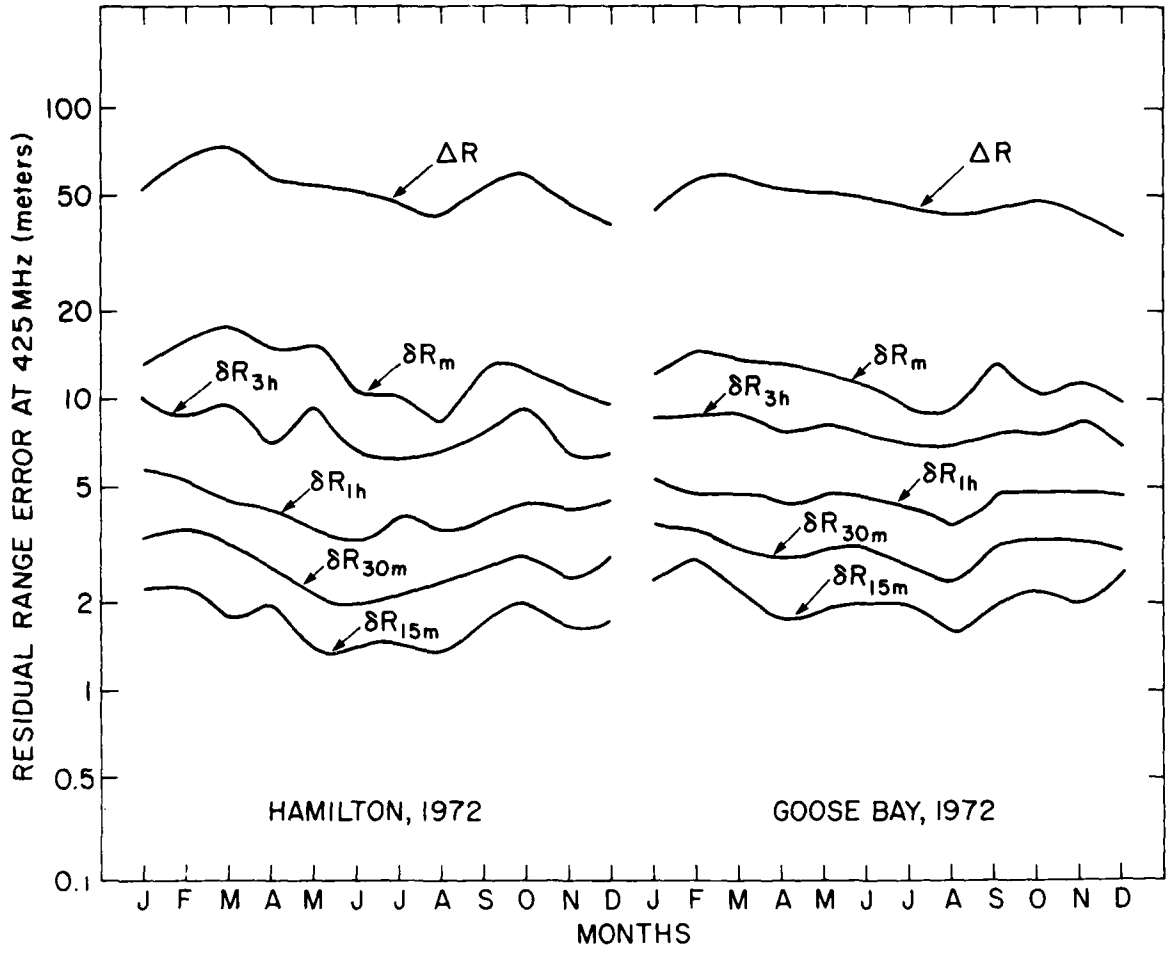


Figure 4. The reduction in maximum monthly residual error in range correction at two different stations, Hamilton, Mass., and Goose Bay, Labrador, for 1972 using the same parameters as in Figure 3.

IONOSPHERIC RANGE-RATE EFFECTS IN SATELLITE-TO-SATELLITE TRACKING

Rodney B. Bent, Judy R. Lipofsky, Sigrid K. Llewellyn
Atlantic Science Corporation
and
Paul E. Schmid-NASA Goddard Space Flight Center

Abstract

Investigations of ionospheric range and range-rate corrections in satellite-to-satellite tracking have pointed out major problems and defined the magnitude of errors that have to be considered for communications between satellites and related experiments.

The two major causes of the sudden and sometimes large variations apparent in the ionospheric range and range-rate corrections along the satellite arc are the geometric effects of the raypath and the curved ionosphere, and the localized perturbations in electron density along the satellite-to-satellite line of sight. The raypath may follow along the region of maximum electron density for a considerable distance, it may pierce the ionosphere at two separated locations, pass through it once, or only partially, or remain above it, allowing for a great range in the magnitude of the effect. Significant gradients in the electron density, as occur in the equatorial anomaly, result in very large perturbations.

Ionospheric tracking errors are greatly effected by the following parameters, whose influence was thoroughly investigated: satellite height, height of maximum electron density, localized ionospheric perturbations and gradients, and electron density variations with diurnal, day-to-day, seasonal and solar cycle patterns. Changes in these parameters can produce large variations in the magnitude of errors as well as time shifts of the peak errors along the satellite arc.

The results point to the need of using a sophisticated modeling approach incorporating daily solar data, and where possible actual ionospheric measurements as update information, as a simple median model cannot possibly account for the complex interaction of the many variables. The findings provide a basis from which the residual errors can be estimated after ionospheric modeling has been incorporated in the reduction.

Simulations were performed for satellites at various heights: Apollo, Geos, and Nimbus tracked by ATS-6; and in two different geometric configurations: coplanar and perpendicular orbits.

INTRODUCTION

The need for this study was indicated by the results of a NASA/GSFC experiment[1] to detect gravity anomalies using the Apollo-Soyuz/ATS-6 satellite configuration. This experiment was intended to improve the knowledge of the earth's gravitational field in localized areas by measuring local gravity variations by detecting small velocity changes of the low orbiting spacecraft. Two of the gravity anomalies "detected" in this experiment are shown in Figure 1b. The corresponding Apollo-Soyuz orbits are shown in Figure 1a. The Indian Ocean anomaly as well as an anomalous region in the Himalayas are indicated by the relatively large variations in range-rate residuals.

Earlier analysis[2] of satellite-to-satellite ionospheric range-rate corrections for the ATS-6 to Apollo-Soyuz geometrical configuration indicated values that were of significant magnitude compared to the gravitational effects. Such results are illustrated in Figures 1b and c and imply that most of the suggested gravitational anomaly may, in fact, be due to ionospheric effects. The geometric configurations of the high (ATS-6) and the low

(Apollo-Soyuz) satellite are shown in Fig. 1d. The low satellite is shown at three different positions in its orbit which, in this case, is below the height of maximum electron density. In Figure 1e it can be seen that ionospheric range corrections (ΔR) reach a peak at point B and are greater at C than at A. This is because the line of sight between the high and low satellite passes through a greater portion of the ionosphere at B than at A or C. Range-rate corrections ($\Delta \dot{R}$) will pan through zero at B.

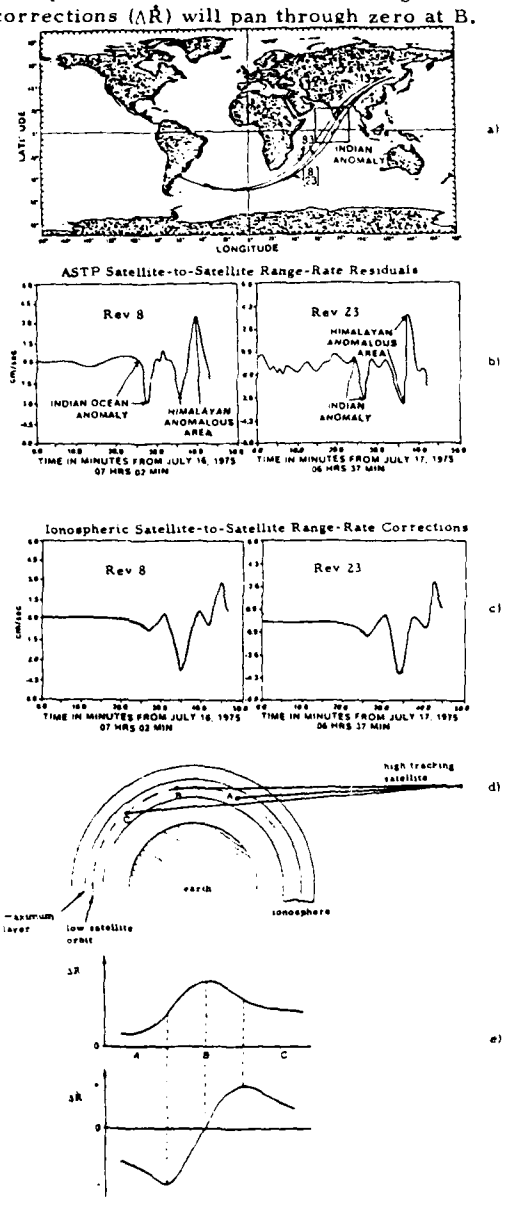


Figure 1. Comparison of gravitational effects and ionospheric effects on Apollo-Soyuz orbits.

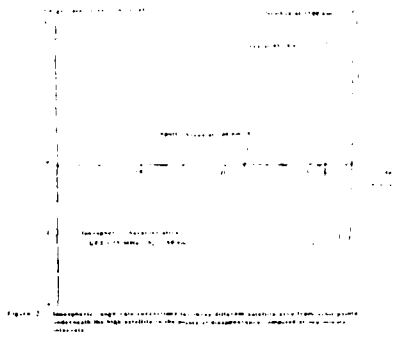
Ionospheric effects, therefore, lead to errors in the gravitational data unless they can be adequately removed by modeling. The purpose of this study is to investigate these effects and understand their origin and characteristics so that a better approach to gravitational experiments can be organized. In order to understand the basic ionospheric range-rate errors, a simple uniform ionospheric model was investigated with certain orbits; the simple model was then varied in height, and perturbations were added in discreet places. Seasonal and solar cycle variations were examined as well as actual orbits of varying inclination and height.

EFFECTS OF A SIMPLE UNIFORM IONOSPHERE ON COPLANAR SATELLITE ORBITS

In satellite-to-satellite tracking the effects of ionospheric refraction are difficult to predict because of complex orbit configurations with changing ionospheric conditions. Evaluating the effects of a simple uniform ionosphere on coplanar satellite orbits at various altitudes clarified the size and duration of refraction errors incurred purely by the geometry of the slanted raypath through the curved ionosphere. Modification of the ionospheric characteristics of the maximum electron density and the corresponding height illustrated the variations that refraction errors will undergo for changing conditions. The height of the ionosphere often changes by $\pm$ 50 km per day, and differences of 5 to 10 MHz in critical frequency can easily be caused by localized perturbations, diurnal variations, day-to-day changes, by seasonal effects and solar cycle patterns. To further investigate the effects of local perturbations, steep ionospheric gradients were injected into the otherwise uniform ionosphere, as they are observed in the equatorial anomaly region, the sunrise region and during irregular occurring ionospheric disturbances.

In the following analysis, the high satellite was considered stationary at a height of about 36,000 km, the low satellite height is varied, and the orbits are coplanar. A tracking frequency of 1 GHz is used and one-way corrections are computed. Each arc was started where the high and low satellites were aligned with the earth's center, and terminated when the low satellite disappeared from the view of the high satellite. The simple model used was based on the Bent Ionospheric Model[3]. This is an emperical worldwide model capable of accurately estimating

the electron density profile and the associated delay and directional changes of a wave due to refraction. The model computes an electron density versus height profile with five exponential topside layers, a bi-parabolic bottomside layer, and an interconnecting parabola. The simplified model used in this study does not include any horizontal gradients as it was held constant world-wide and in time. Different conditions were simulated by varying the critical frequency, f_0F2, and the height of maximum electron density, h_m.



Effects of Satellite Height

The ionospheric range-rate corrections due to a geometric effect along the satellite arc are illustrated in Figure 2 for different altitude satellites. The low satellite passes over the horizon and becomes invisible to the high satellite. The very large range-rate corrections immediately before the invisible region are caused by the changing geometric conditions of the satellite raypath through the curved ionosphere as it disappears (see Figure 1). The higher the low satellite orbit is, the larger are the range-rate corrections caused by this effect. The reason for this is that for a low satellite at a relatively high altitude the tracking raypath remains above most of the ionosphere over large segments of its orbit and penetrates significant electron density layers only when the satellite is near the point of disappearance. The higher the satellite is, the faster it passes from the region where the satellite raypath pierces the ionosphere to the point of invisibility. This faster rate of change produces larger amplitudes in the range-rate corrections. The lower the satellite orbit is, the shorter is its period of visibility from the high tracking satellite, and the longer is the time period over which ionospheric corrections are orders of magnitude larger than in ground-to-satellite tracking. Considering coplanar orbits, for Apollo-Soyuz at 200 km the last 13 minutes of its arc show such large corrections, for Geos at 850 km the last 7.5 minutes, and for Nimbus at 1100 km the last 5.5 minutes. However, if low and high satellites are moving at right angles to each other and are close to the point of disappearance, then such large effects can be observed for extended periods of time.

Range-rate correction curves as a function of time along the arc have steeper slopes and larger positive and negative peaks the higher the satellite altitude is. At h_m = 250 km, f_0F2 = 15 MHz and a tracking frequency of 1.0 GHz the maximum corrections are 1.5 m/sec, 3.5 m/sec, 4.3 m/sec for Apollo, Geos and Nimbus respectively.

Effects of Ionospheric Height

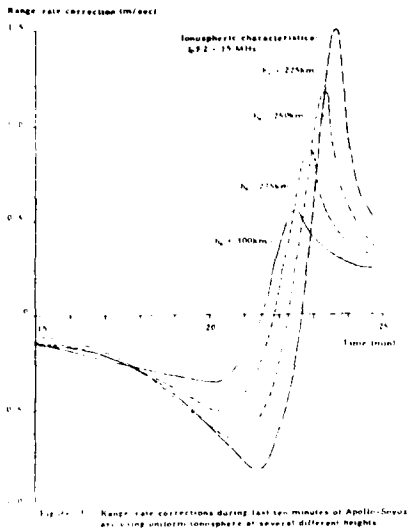
Changes in the height of the ionosphere which are typically $\pm$ 25 km, produce very significant effects on the range-rate corrections, as shown in Figure 3. Time shifts of the peak values along the arc are experienced; for a 50 km shift in h_m they are of the order of 0.9 minutes for Apollo and 0.35 minutes for Geos and Nimbus. Such time changes are significant when attempting to derive localized gravitational effects. In addition, for the Apollo spacecraft which always remains below h_m, large changes in the maximum correction values are apparent. Increases of about 0.3 m/sec in the peak value occur for each 25 km drop in h_m at f_0F2=15MHz. Such changes in peak values are, however, not apparent from the higher satellite data from Geos and Nimbus.

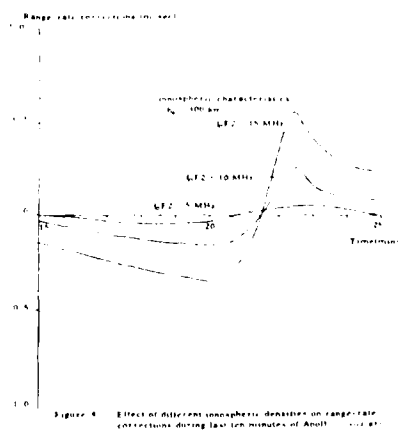
Effects of Electron Density

Changes in electron density effect the amplitude of the range-rate corrections such that the peak correction values for each satellite are approximately in the same ratios as the squares of the f_0F2's. Time shifts of the peak corrections along the satellite arc are not associated with changes in f_0F2. See Figure 4.

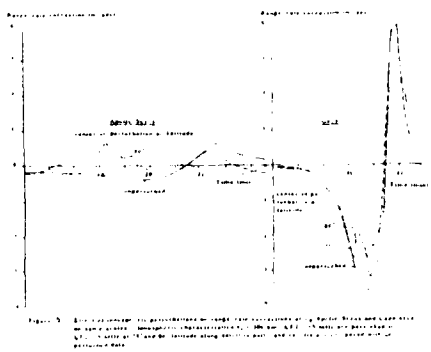
Effects of Ionospheric Perturbations

Ionospheric perturbations with associated steep gradients in electron density can greatly modify the smooth range-rate correction curves as predictable from geometric

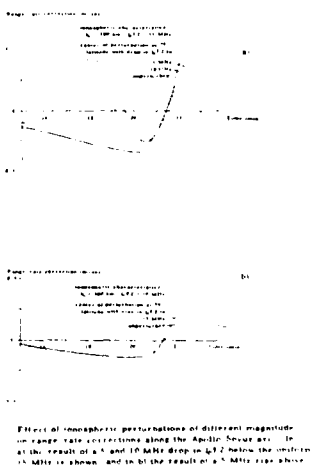




effects in satellite-to-satellite tracking. The peak values may experience drastic increases or decreases in amplitude as well as shifts in time, and additional maxima and minima can occur along the arc. Such effects were simulated by injecting into the previously uniform ionosphere variations in electron density, which have maximum effect at the perturbation center and gradually decrease with increasing distance from the center. An exponential weighting function was used, $w = e^{-(d/D)^2}$, with D = 400 NM; d is the distance from the center of perturbation. As shown in Figure 5, the effects of the perturbations on the Geos corrections are of larger amplitude, but the effects on the Apollo arc are of greater relative severity persisting over a longer period of time.



For different size disturbances under otherwise fixed conditions the range-rate corrections show similar patterns. However, the greater the amplitude of the electron density perturbation and the steeper the slopes of the gradients, the more severe are the deviations from the smooth unperturbed correction curve. See Figure 6.



Shifting the location of the perturbation center to different latitudes close to the Apollo-Soyuz arc manifested itself mainly in a large time shift, of the order of minutes, of the additional single or double peak extremum. The earlier the perturbation center appears along the arc, the earlier does the extra peak occur.

SPACIAL AND TEMPORAL EFFECTS

In this section the effects of actual variations and gradients in the ionosphere on satellite-to-satellite range-rate corrections are investigated. These gradients occur along the arc of a low satellite being tracked by a high satellite. For this purpose the

3-dimensional world-wide Bent ionospheric prediction model was used to show the effects of steep spacial gradients as well as of short and long term temporal gradients, such as diurnal variations, day-to-day deviations from the monthly mean, seasonal and solar cycle variations.

Simulated Orbits

To capture the ionospheric variations, the orbit configuration was chosen quite different from the coplanar orbits used for the simple model investigations. The satellite orbits were simulated perpendicular to each other, the stationary high satellite in an equatorial orbit, and the low satellites at various heights in polar orbits both prograde and retrograde. The angular separation between the satellites was chosen in the proximity of the disappearance points, and under such circumstances significant range-rate corrections exist along most of the orbit. Tracking frequencies of 1GHz were used in the simulations. The range-rate corrections are presented as the one-way values.

In Figure 7 one of the orbit configurations is shown. The low satellite in a prograde orbit disappears from the sight of the high satellite for a portion of its orbit and becomes visible again as it crosses the equator. Three such cases were considered for low altitude satellites at 200, 300 and 850 km and the high tracking satellite at about 36,000 km. Figure 7 shows the 300 km orbit superimposed on a world map of the ionospheric characteristic of f_xF2. Where the f_xF2 equal value lines are concentrated, large ionospheric gradients exist, and it is indicated where the ionospheric pierce point curve crosses these gradients. The intersection of the satellite-to-satellite line of sight with the region of maximum electron density establishes the ionospheric pierce point. When the low satellite is at an altitude above the height of maximum density, as is the case for Geos and Nimbus, two distinct pierce points may occur.

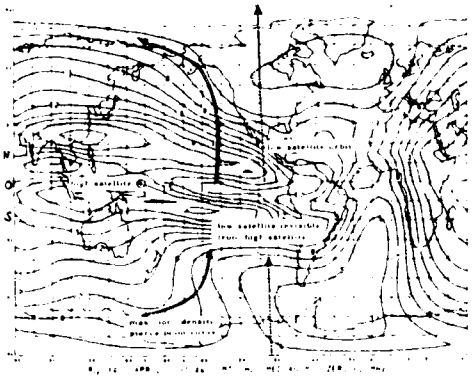


Figure: Ground tracks of high satellite position, low satellite orbit and maximum electron density pierce point curve superimposed on world map (right). Low satellite altitude equals 300 km

Spacial Gradients

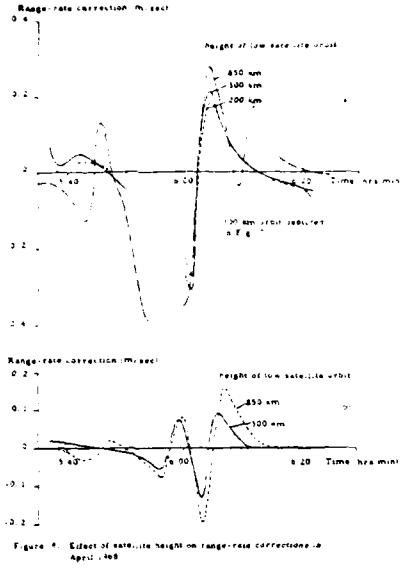
To examine the effect of the low satellite altitude on the range-rate corrections along the arc, the geometric conditions were fixed, such as the high satellite position and the ionospheric pierce point location location at the equator crossing and with that also the satellite line of sight, and only the height of the orbit was varied. The range-rate corrections for the orbit depicted in Figure 7 and for four additional orbits are correlated in Fig. 8. The low satellite is in a polar orbit and crosses the equator at 6 UT, April 1, 1968. The orbits corresponding to Figure 8a are prograde, in 8b they are retrograde.

In Figure 8a there is an invisible portion for all satellite orbits, but at 6 UT they become visible again as they cross the equator. The 200, 300 and 850 km satellites all lie along the same line of sight from the high satellite. The very large deviations to both sides of the invisible region are caused by the changing geometric conditions of the satellite raypath through the curved ionosphere as it disappears, which is illustrated in Figure 1.

Superimposed on these geometric effects in Figure 8 a are the ionospheric gradient effects, because the ionospheric density and height along the line of sight changes as the low satellite travels along its orbit.

The geometric effects are minimized in Figure 8 b by decreasing the high to low satellite separation angle by about 3 degrees at the equator crossing and maintaining visibility throughout the half-orbits. The difference between the range-rate corrections along the 300 and 850 km orbits are now mainly due to the gradients in ionospheric density and height. Again the higher altitude orbit displays the larger amplitude refractive effects, but they are only about half as large as those caused by the geometric conditions in Fig. 8a.

To better understand how the variations in the refraction corrections along the satellite path are caused by the ionospheric conditions, two low satellite orbits, the ionospheric densities depicted in the f_xF2 world map, the ionospheric heights and the satellite



angular separation shown in Figure 9 were examined in detail and were correlated to the range-rate corrections also shown in Fig. 9.

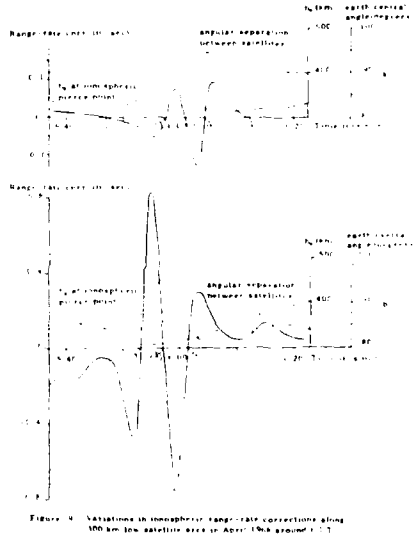
In the orbit corresponding to Figure 9a the low satellite, at 300 km, passes east of the high satellite and remains visible throughout the simulated half-orbit. A retrograde orbit was used to provide better geometric conditions, keeping the satellite distance fairly constant over the equatorial anomaly region, assisted by the slight longitude shift due to the earth rotation.

Figure 9b corresponds to another retrograde orbit, this time crossing the equatorial anomaly west of the high satellite over some very dense regions.

From the curve giving the angular separation between the satellites it is apparent that while this separation varies somewhat at the beginning and end of the arc, it is fairly constant over the central region of the equatorial anomaly where most of the large range-rate deviations occur. Hence, in the region of interest the corrections are mostly due to ionospheric gradients, because the geometric configuration remains nearly fixed.

The marked zero crossings of the range-rate corrections (Figure 9) correspond very clearly to the maxima or minima of f_xF2 shown on the world map (Figure 7). This is to be expected since extrema in f_xF2 cause extrema in electron density and range corrections, hence zero crossings in range-rate correction. The range-rate corrections give a very clear picture of the gradients encountered as the satellite connecting beam sweeps the equatorial anomaly.

Not just the electron density, but also the height at the maximum undergoes steep gradients in the anomaly region. In Figure 9 it is shown that the height of the pierce point has a definite maximum at the magnetic equator around which the equatorial anomaly is centered, and its values vary between 314 and 456 km over the arc. The smaller variations in the range-rate corrections toward the ends of the arc in Figure 9b might be influenced by the ionospheric height changes.



It should be noted that different vertical scales are employed in Figure 9. Maximum absolute amplitudes in b) are about seven times as large as in a). The major reason for this lies in the satellite configurations. The satellite with range-rate results in a) remains barely above the horizon when crossing the equator, 97° away from the high satellite, and the line of sight pierces the ionosphere at a slanted angle. The satellite with results in b), however, is closer to the high satellite separated by 81° at the equator. The 300 km satellite used in both these examples is below the height of maximum density which varies in b) between 350 and 456 km in the equatorial region, and the line of sight between the

satellites pierces the ionosphere at extremely slanted angles, remaining in the vicinity of the maximum electron density over very long distances, hence the very large refraction effects. In addition, slightly steeper gradients exist in ionospheric height and density for b) than exist for a) which also contributes to the larger range-rate corrections in b). The maximum pierce point of f_0F2 in b) was 14.2 MHz, in a) 12.9 MHz.

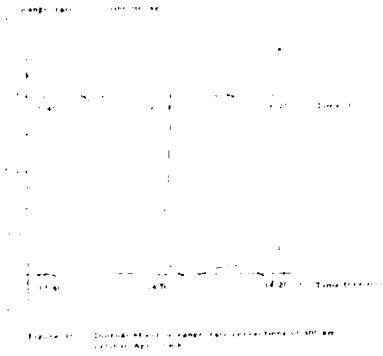
In using the April 1968 data, the spacial gradient investigations were performed for average solar maximum conditions; the smoothed sunspot number was 107. It should be recognized, however, that during a more severe solar cycle the range-rate corrections could be significantly larger than the 0.8 m/sec in Figure 9. During the maximum cycle recorded in 1958, the smoothed sunspot number went as high as 201 in March.

Temporal Gradients

The effects of short and long term ionospheric temporal gradients on satellite-to-satellite range-rate corrections was investigated by comparing the corrections along the orbits, which were obtained for ionospheric situations where all conditions, except the one tested parameter, were held constant.

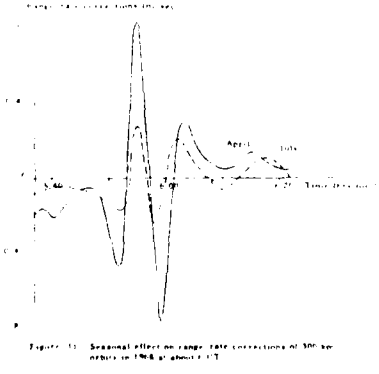
Diurnal Effect

The diurnal effect on range-rate corrections was examined by computing corrections for satellites crossing the equator at 6 hours and at 14 hours universal time to provide information with high and low gradients, and the results show large differences in amplitude. In Figure 10a the ionospheric pierce point occurs at longitude 222° or 21 hours local time. At this time very rapid changes in electron density are encountered as the satellite connecting ray sweeps across the equatorial anomaly while following the low satellite orbit. The steep gradients are apparent along the pierce point curve in Fig. 7, and the resulting range-rate corrections are large. In Figure 10b the satellite crosses the equator at 14 hours UT. Here the ionospheric pierce point occurs at 5 hours local time, when the ionospheric gradients along the pierce point curve and hence, the corresponding range-rate corrections are about as small as they can get for such an orbit configuration.



Seasonal Effects

Range-rate corrections were compared for April 1968 and July 1968 to examine the effect of seasonal variations on range-rate corrections. The solar activity did not influence the result significantly, because the smoothed sunspot numbers were very similar, 107.2 in April and 105.2 in July. Polar orbits at 300 km and 850 km altitude were examined, both having the equator crossing fixed at 6 hours UT. Range-rate corrections corresponding to the 300 km orbit are presented in Figure 11. In each case, maximum ionospheric effects occur in April and minimum effects in July. For other months the effects would be expected to be of intermediate amplitude. Significant changes in peak amplitudes are apparent, as well as time shifts of the peak correction values.



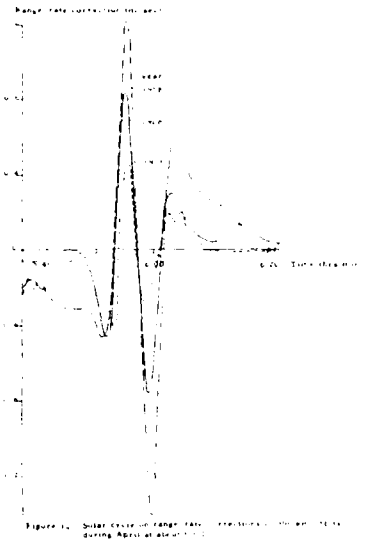
Solar Cycle Effects

The effect of the solar cycle on range-rate corrections was also considered. Three years were chosen to represent the extremes of solar activity, and in all cases the month of April was selected to eliminate seasonal

effects.

Year	Smoothed Sunspot Number
1958	196.8
1968	107.2
1975	18.6

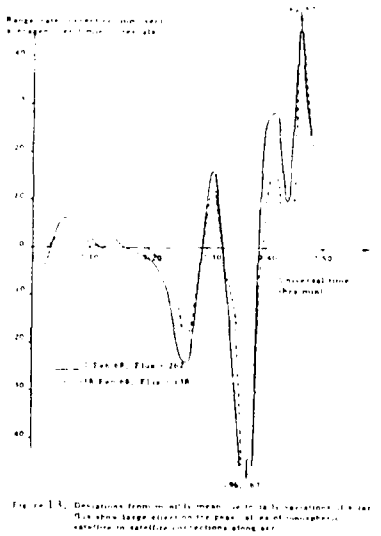
The 1958 solar cycle was the all-time maximum recorded. During 1968 and 1975, the maximum and minimum activity occurred for the last solar cycle. Figure 12 shows the results for a 300 km orbit. The 1958 corrections are in all cases significantly larger than the 1968 results by factors of 1.5 to 2.5. The 1968 results in turn exceed the 1975 results by similar amounts. Shifts in the time of the peak range-rate correction occur for several peaks but without a simple pattern.



Temporal Gradient Tests with Apollo-Soyuz Orbits

The effect of day-to-day variations in the ionospheric density and height on the range-rate corrections along the satellite link were investigated using an orbit of the Apollo-Soyuz data. Revolution 8 on 16 July 1975, with the tracking frequency of 2.16 GHz was used, and the ionospheric conditions of interest were simulated.

To examine the influence of the deviations in f_0F2 from the monthly mean due to daily variations of solar flux, the month of February 1968 was chosen for the simulations; 1968 was the year of average maximum solar activity. The monthly mean of the solar flux was 173, and corrections were computed for February 1st when solar flux was at a monthly maximum 262, and for February 18th when the flux was at the minimum of 138. As shown in Figure 13, the increased solar flux causes a sharp increase in peak ionospheric range-rate corrections by a factor of 1.1 to 1.8, while the same pattern is maintained for the variations along the arc. This points to the necessity to adjust the monthly mean ionospheric correction model for daily variations as is done in the Bent model.



Apollo-Soyuz Orbit

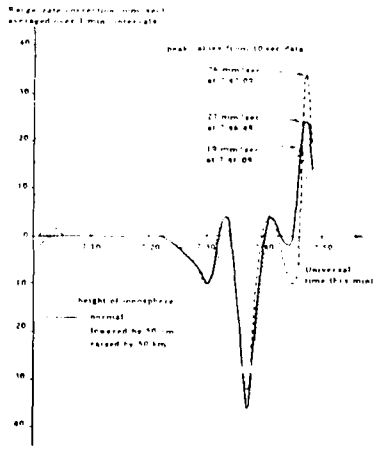


Figure 14

Date: 16 July 1975

Apollo-Soyuz Orbit

The height of maximum electron density was computed using the ionospheric model, and then it was raised and lowered by 50 km during the simulations. Such variations represent the possible day-to-day changes in the ionospheric height, and result in changes in the amplitude of the range-rate corrections as well as in a time shift of the peaks, see Figure 14. Lowering the ionosphere by 50 km causes an increase in the amplitude by a factor of 1.4 and a shift of the peak in increasing time along the arc. Raising the ionosphere by 50 km causes a decrease in amplitude and a 40 second shift of the peak in decreasing time along the arc.

Acknowledgements

The majority of the work described herein has been performed under contract NAS5-23899 from NASA/Goddard Space Flight Center under the direction of Mr. Paul Schmid.

References

1. F. O. Vonbun, W. D. Kahn, J. W. Bryan, P. E. Schmid, W. T. Wells, T. D. Conrad, "Gravity Anomaly Detection-Apollo-Soyuz", NASA/GSFC document X-920-75-308, December, 1975.

2. R. B. Bent, S.K. Llewellyn, J.R. Lipofsky, "Apollo-Soyuz/ATS-6 Ionospheric Refraction Corrections", unpublished, NASA contract NAS5-20744, Mod 4, July, 1976

3. S. K. Llewellyn, R. B. Bent, "Documentation and Description of the Bent Ionospheric Model" prepared for Space and Missile Systems Organization, Department of the Air Force, July, 1973.

PRELIMINARY EVALUATION OF A NEWLY DEVELOPED SATELLITE-TO-SATELLITE IONOSPHERIC REFRACTION CORRECTION MODEL

John A. Behuncik
NASA-GSFC
Greenbelt, Maryland

INTRODUCTION

With the decision of NASA to implement in the near future the Tracking Data Relay Satellite System (TDRSS), one of many questions raised by the user in general and the orbit computation engineer in particular is the following. What is the effect upon the tracking signal and consequently upon the spacecraft orbital state of the ionospheric traverse of the signal on satellite-to-satellite legs of the total signal path. Heretofore, all tracking measurements and data reception were obtained from ground based stations. Refraction corrections were computed for and applied to the uplink on downlink legs for tropospheric and ionospheric effects. With the advent of spaceborne trackers, TDRSS spacecraft, the number of path links is doubled and on the satellite-to-satellite legs of the total signal path the signal can encounter the ionosphere in many complex geometric forms. Because the ionosphere extends into space hundreds of times as far as the troposphere any data editing scheme including the ionosphere would eliminate the use of all or a prohibitive portion of, tracking data for orbit determination computations for most of the scientific spacecraft tracked by NASA. A data editing scheme for data points whose satellite-to-satellite legs of the tracking configuration intersect or traverse the troposphere, usually assumed to extend to 30 kilometers, would eliminate perhaps ten to fifteen percent of all tracking data for the same satellites, a more acceptable situation. A task was, therefore, set to develop the means of computing accurate corrections for ionospheric traverse of the signal for the satellite-to-satellite legs of the total path length resulting from such a tracking configuration. A model was developed and implemented into a large orbit determination computer program. It was applied to real tracking data taken during the satellite-to-satellite experiment employing the ATS-6 as the tracker spacecraft, GEOS-3 as the target spacecraft and Rosman, North Carolina as the ground transmitter as a means of model evaluation. The method and results of the evaluation of the newly developed model are the subject of this paper.

DISCUSSION

Satellite-to-Satellite Refraction Model

The satellite-to-satellite refraction model is essentially a generalization of a previously developed and verified model that produced tropospheric and ionospheric refraction corrections for measurements obtained at ground based stations. The base model is global in nature, capable of predicting the ionospheric state parameters at any geographic location at any point in time. It is an empirical model developed by Bent, et al., on contract to NASA and is fully described in Reference 1. In addition to predicting ionospheric state parameters the necessary model parameters are provided and the required refraction corrections are produced. As a first order approximation to horizontal gradients the ionospheric state and model parameters are generated at the point of maximum electron density-ray path intersection through an iterative process. The base model was fully evaluated by comparing the model results with those obtained by ray tracing through the same ionospheres used by the model. The results of that study are documented in detail in Reference 2 but the model results generally agreed with ray trace values to one percent at elevation angles greater than 20 degrees and less than 20 percent below. It was also demonstrated in Reference 1 that the predicted ionospheric state was consistently in excellent agreement with actual measured values.

A task was let to develop, using the base model as a starting point, a model capable of computing refraction corrections for the satellite-to-satellite leg of the satellite-to-satellite tracking configuration. The base model was selected as the starting point because of its global application, its demonstrated ionospheric state prediction capability, and its proven accuracy of corrections vis-a-vis ray trace results. A detailed description of the satellite-to-satellite refraction correction model developed under contract to the author is described in Reference 3 and will not be described in detail herein. A brief overview of the model is, however, now presented.

The intersection of the target-tracker line with the ionospheric, assumed to be a geocentric sphere of radius of 9871.2 kilometers for model purposes, is determined. The point of closest approach of the ray, or its extension, to the earth is also found. Knowing the ionospheric path length and the height of the point of closest approach of the ray there is computed the number and locations of required ionospheric states. With the model parameters, the electron density is computed at the specified points along the ionospheric line segment as well as at the reference point, taken as the point of maximum electron density. A deviation function which defines the electron density at each point relative to the electron density of the reference ionsophere for the height of the point under consideration is formed as the integrand. The resulting line integral is next evaluated by means of Gaussian Quadrature technique to obtain the total electron content for the ionospheric path segment of the satellite-to-satellite leg of the total signal path common to this mode of tracking. The ionospheric range correction for this measurement is then obtained from the following equation.

$$\Delta R = 40.3 N_T / f^2$$

where N_T is the total electron content for the satellite-to-satellite leg of the four legged path, and f is the operating frequency in Hertz.

As the model was so developed to handle all ray path-ionosphere geometric intersections the ground tracking configuration became just one specific case. Consequently the resultant model is capable of computing refraction corrections for ground-borne and space-borne tracking. In this paper, however, only the satellite-to-satellite range and range rate corrections are considered. Furthermore, only the accuracy of the model-produced corrections as determined by comparison with ray traces results is addressed in this paper.

Ray Trace Program

A ray trace program was written and developed using the method and approach presented in Reference 4. It was capable of ray tracing from the ground to a space point. It employed double precision arithmetic and the numerical results of Reference 4 were accurately reproduced. The equations relating ray path bending, group delay and local elevation angle to local refractivity and initial elevation angle were analyzed and modified where necessary to make the program capable of ray tracing from space point to space point. The new ray trace program was verified in the following manner. The previously verified ground-space point ray trace program was used to trace a ray from the ground to the top of the ionosphere with an initial elevation angle of zero degrees. A typical daytime ionosphere with height of maximum electron density of 300 kilometers and a maximum electron density of 1×10^{12} electrons per cubic meter was assumed. The operating frequency was chosen as 2100 MegaHertz. The final local elevation angle from this tracing was used, with a change in sign, as the initial elevation angle to trace the ray with the newly developed space point-to-spacepoint program from the top of the ionosphere to the ground with the same ionospheric state and frequency. The range corrections agreed to 0.1 millimeters or 0.001 percent (the correction is 7.3964 meters). This difference is consistent with the numerical differences between procedures in the two programs.

RESULTS

An experiment was conducted to prove the feasibility of satellite-to-satellite tracking. It consisted of a ground station tracking the GEOS-3 spacecraft through the ATS-6 satellite. Consequently, actual satellite-to-satellite measurements were available for model evaluation purposes. The test data span selected was 3-4 May 1975. During that two day period there were four distinct passes of satellite-to-satellite tracking extending from almost 10 hours U. T. on 3 May to approximately 9 hours U. T. on 4 May. The passes and their start and end times are presented in Table 1. During this time interval Rosman, North Carolina was the ground tracking station.

The procedure was to process the test data in a large orbit determination program in which was implemented the satellite-to-satellite ionospheric refraction correction model. Obtained from orbit determination program for each measurement were the ionospheric refraction corrections for range and range rate for the satellite-to-satellite leg as well as the model profile parameters for each of the profiles used in the computation of the corrections and time. The ray trace program had been modified to accept all the profiles obtained from the orbit determination program for a measurement and, when

Table 1

Evaluation Data Spans

Span Number	Start Time H:M:S	End Time H:M:S	Number of Observations
1	10:02:00	11:04:00	70
2	21:05:00	22:07:50	70
3	22:44:00	23:48:00	72
4	08:34:30	90:12:30	76
Total Number of Observations			288

(Data Spans Occur on May 3 and 4 1975)

computing an electron density for a particular point, to do so using four point interpolation in density values obtained from four profiles for the height of the point in question. There were obtained in this manner ray trace values of ionospheric refraction range corrections which were used as the standard for comparison with the model produced values. Presented below are model-ray trace comparisons and the results of the evaluation.

Presented in Figure 1 are the model and ray trace range corrections for data span one. Also presented in the figure are the differences between the two, the residual error that remain in the measurement due to this effect. The range corrections are very large on target acquisition and just prior to target occultation by the earth. In these areas the corrections reach as high as 135 and 791 centimeters respectively with the model underestimating the corrections at these points by 6 and 29 centimeters, or 4 and 3.5 percent low. As the target crosses the earth beneath the tracker satellite the model produces corrections that average near 25 centimeters or approximately 6 percent larger than ray trace values. Presented in Figure 2 are the model and ray trace range rate corrections for data span one. The range rate corrections reach a maximum of -5 millimeters per second just after target acquisition and, just prior to occultation, oscillates from -87 to +94 millimeters per second, with the model overestimating by 106 percent for the first point, by 0.1 percent for the second and underestimating by 3 percent for the third point.

Presented in Figure 3 are the model and ray trace range corrections for data span two. Just after acquisition the corrections reach a maximum of 423 centimeters while just prior to occultation a value of 957 centimeters is obtained with the model underestimating the corrections by 20 and 37 centimeters or 4.5 and 3.7 percent respectively. In Figure 4 are the model and ray trace range rate corrections for this data span. Just after acquisition the range rate corrections reach a maximum of -14 millimeters per second while just prior to occultation they reach maximum values of -110 and +97 millimeters per second, with the model underestimating the correction by 3 percent at the first point, 3.3 percent at the second point, and 3.8 percent at the third point.

Presented in Figure 5 are the model and ray trace range corrections for data span three. Just after acquisition the corrections reach a maximum of 408 centimeters and just prior to occultation reach a maximum of 814 centimeters, with the model underestimating the corrections by 4.3 percent at both points. Over the middle of the data span the corrections average about 30 centimeters, overestimating the corrections by almost 6 percent. In Figure 6 are presented the range rate corrections for this data span. Just after acquisition the corrections reach a maximum value of -17 millimeters per second while just prior to occultation they reach maximum values of -37 and +106 millimeters per second, with the model underestimating the correction at the first point by 2.7 percent, at the second point by 4.9 percent and at the third point by 3.9 percent. Over the remainder of the pass the corrections average about 0.2 millimeters per second, an overestimation of almost 15 percent.

Presented in Figure 7 are the model and ray trace range corrections for data span four. Since data span four comprises just the latter half of a normal pass the corrections reach just one maximum of 987 centimeters with the model underestimating the correction by 3.7 percent. Over the first part of the span the corrections average about 24 centimeters underestimating the correction by almost 2 percent. In Figure 8 are presented the range rate corrections for this data span. Over the first part of the span the corrections are near zero and just prior to occultation reach maximum values of -100 and +94 millimeters per second. At these points the model underestimates the corrections by 2.9 and 2.3 percent respectively.

CONCLUSIONS

From this preliminary evaluation it is concluded that the model does quite well in computing ionospheric refraction range and range rate corrections for the satellite-to-satellite leg of the satellite-to-satellite tracking configuration. Generally, in the regions that produce large refraction effects the model range corrections differ from ray trace values by an average of minus four percent while the model range rate corrections are less than ray trace values by an average three percent. In regions of minimum refraction effect, where the tracker-target ray path does not penetrate the ionosphere to the regions of maximum ionization, the model overcorrects by almost 3 and 15 percent respectively for range and range rate. Furthermore, there are points where the model-ray trace differences reach into the high tens of percent. These extrema and the higher relative error across the center of each data span can be corrected by computing more ionospheric states per measurement path length.

In conclusion it appears that the model is more than adequate to be used in routine orbit determination and with slight modification can be used in areas where accuracy requirements are more stringent.

ACKNOWLEDGMENTS

The author acknowledges the diligent efforts of Mr. Raymond Luczak, of Computer Sciences Corporation, who produced the data contained herein and modified the ray trace program used to provide the

ray trace corrections. The author thanks Mr. John Cook, of Goddard Space Flight Center, who prepared the figures contained herein.

REFERENCES

1. Schmid, P. E., R. B. Bent, S. K. Llewellyn, G. Nesterczuk, and S. Rangaswamy, NASA-GSFC Ionospheric Corrections to Satellite Tracking Data, X-591-73-281, Dec. 1973.

2. Nesterczuk, G. and J. K. Kozelsky, Evaluation, Modification and Improvement of the Ionospheric Refraction Algorithm in the GTDS - Final Report for Contract No. NAS 5-22302 - Part 4, August 1976.

3. Nesterczuk, G. and J. K. Kozelsky, Ionospheric Refraction Corrections in the GTDS for Satellite-to-Satellite Tracking Data - Final Report for Contract No. NAS 5-22423, July 1976.

4. Bean, B. R. and G. D. Thayer, CRPL Exponential Reference Atmosphere NBS Monograph No. 4, October 1959.

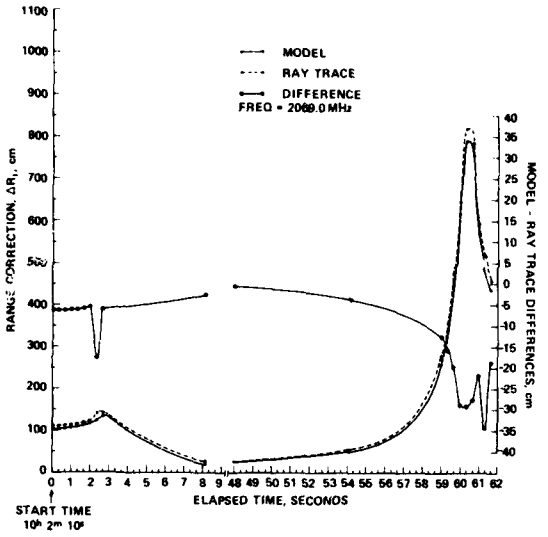


Fig. 1. Range corrections for data span one.

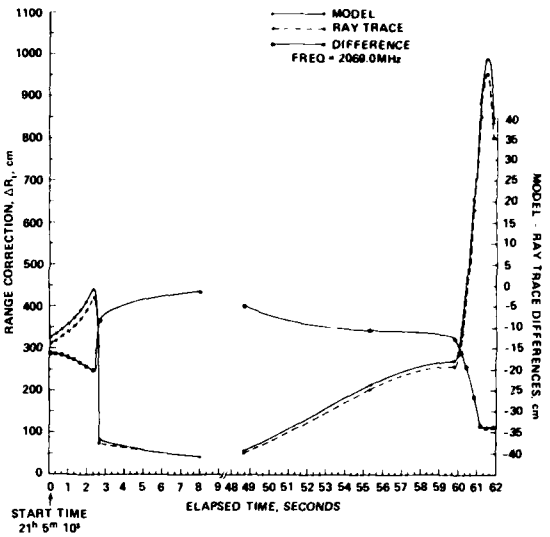


Fig. 3. Range corrections for data span two.

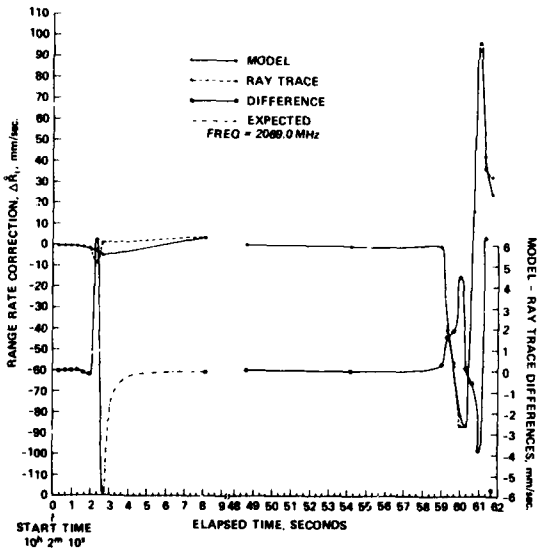


Fig. 2. Range rate corrections for data span one.

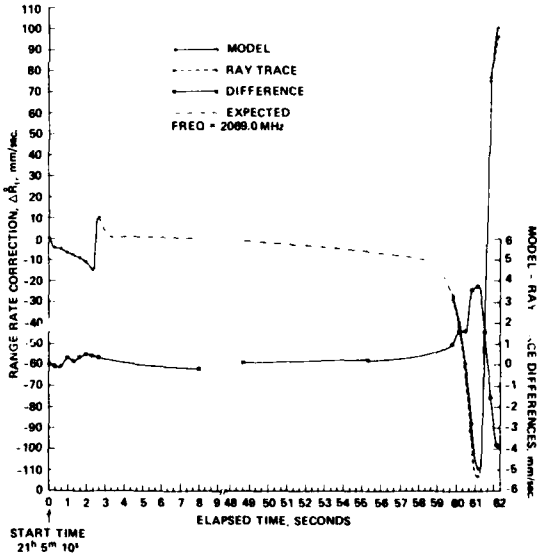


Fig. 4. Range rate corrections for data span two.

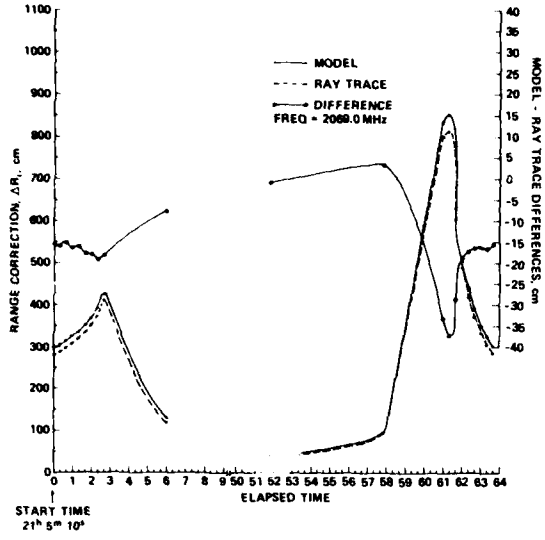


Fig. 5. Range corrections for data span three.

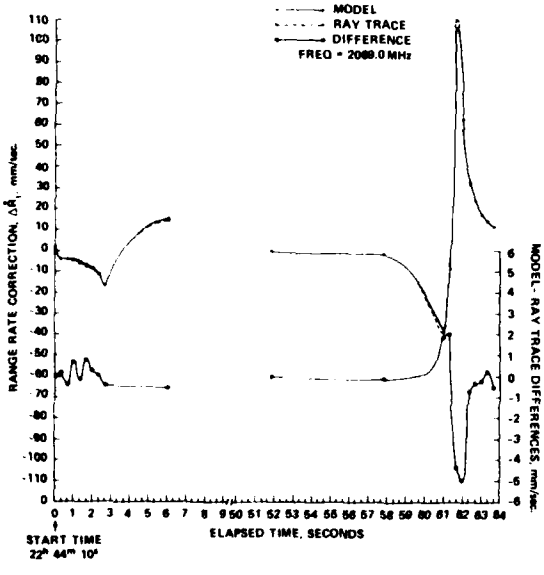


Fig. 6. Range rate corrections for data span three.

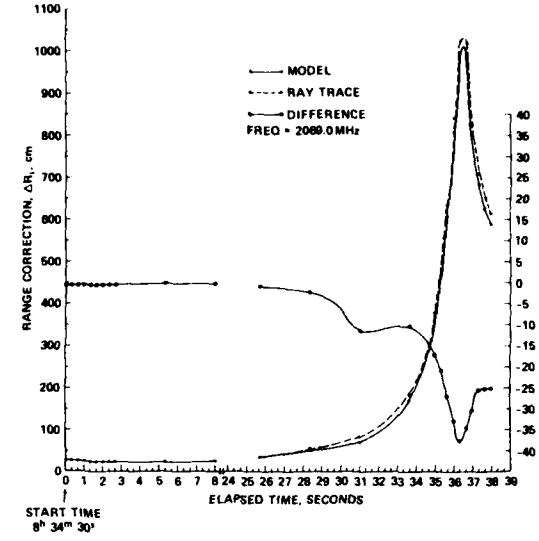


Fig. 7. Range corrections for data span four.

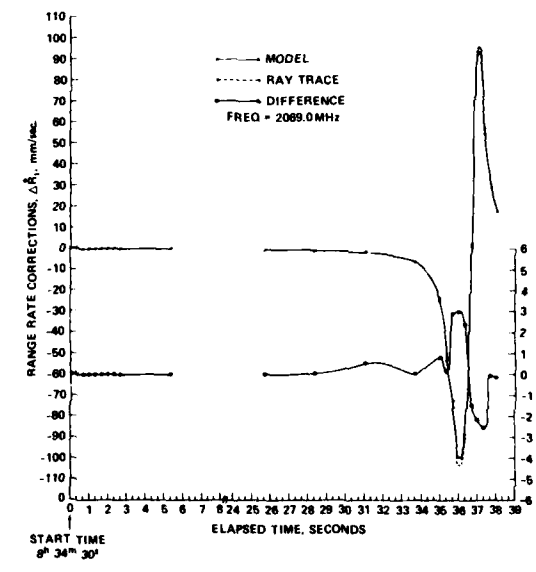


Fig. 8. Range rate corrections for data span four.

ADAPTIVE CORRECTION OF THE EFFECT OF THE IONOSPHERE ON RANGE DETERMINATION BY TERRESTRIAL RADARS

A.H. Katz and M.D. Grossi
Raytheon Company, Wayland, Massachusetts 01778

R.S. Allen
Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts 01731

D.E. Donatelli
Regis College, Weston, Massachusetts 02193

ABSTRACT

An effective correction of the range errors induced by the ionosphere in terrestrial radar systems (the error can be reduced down to 10 ft at UHF) is based on the joint use of an ionospheric model and of a few discrete samples of the dispersive ionospheric medium taken by the radar itself (when dual-frequency ranging is available) or by other means, such as reception at the radar site of dual-frequency emissions from navigation satellites. The scheme makes use of a model of the monthly median ionosphere constructed from existing worldwide climatology (and capable of providing by itself a residual range error not larger than 35 ft) and updates this model with real-time dual-frequency measurements (accurate in themselves to about 4 ft) of the columnar electron content. This is done for directions and at times for which targets of opportunity (such as satellites and target vehicles embedded in the ionosphere) become available. A practical implementation of this real-time adaptive scheme, prepared by the authors, has confirmed that a hybrid hardware-software approach leads to an efficient utilization of radar resources. An experimental evaluation has recently been performed of this novel procedure by using both the data from a precision L-band radar and from ground-based observations of the emissions of the NNSS navigation satellite. This experimental verification has sho shown that the method is both effective in terms of reduction of the ionospheric range errors and advantageous inasmuch as it minimizes the utilization of the radar resources for this scope.

1. INTRODUCTION AND SUMMARY

Precision radar systems require corrections for time or range errors caused by the ionosphere, when that medium is crossed in all or in part by the propagation path. To first order such errors are directly proportional to the integrated electron content along the path to the target. In real time these radar errors may be substantially reduced by predictions of the expected propagation effects using a model of the monthly median ionosphere constructed from existing worldwide climatologies. Such median corrections are shown to have a residual day-to-day r.m.s. variability about the median on the order of 20-25 percent of the median value.

For instance, the expected worst median case (daytime, equinox, sunspot maximum) for range errors experienced by an L-band radar at mid-latitudes is shown in Table 1.

By use of such a model, a goal of 35 ft for the 1-sigma residual range error can be achieved at L-band even during expected worst median conditions. In precision radar systems more stringent requirements exist and the ionospheric induced error must be typically reduced to a few feet.

Table 1
Worst Case Monthly Median Uncorrected Range Errors (ft) at L-Band and Their Variability (1 Sigma)

Target Altitude (km)	Radar Elevation 0° Range Error (ft)	0° Sigma	20° Range Error (ft)	20° Sigma	90° Range Error (ft)	90° Sigma
1000	147	30	102	20	47	9
400	107	20	74	16	34	7
300	57	11	40	8	18	4

To satisfy such requirements, a real-time adaptive scheme has been developed consisting of a model of the median ionosphere for the particular radar coverage region (derived from more complicated models of worldwide climatology) and of a set of algorithms to provide error corrections for targets at specified coordinates, which together attempt only to deliver an r.m.s. error on the order of 25% of the expected median correction at that location. This estimate is then specified by a dual-frequency dispersive probing of the ionosphere in the target area by the radar itself, using satellites of opportunity or even target vehicles embedded in the ionosphere.

The dual-frequency measurement, with accuracy on the order of 4 feet, is used to derive a normalization factor for the median ionospheric model in the surveillance volume, dependent on the radar range, elevation and altitude of the radar target.

The interval between the model's adaptive updatings by the dual-frequency probing depends on the space and time characteristics of the ionosphere.

At times of low sunspot number, a 10-ft range error specification can be met at mid-latitude by using a spherically symmetric monthly median ionospheric model, stored in the radar processor memory and adapted with dual-frequency measurements made at half hourly intervals. This correction is effective even in the presence of a strong Traveling Ionospheric Disturbance (TID).

At times of high sunspot number, the radar algorithms must consider realistic horizontal gradients modeled by the median ionosphere. The adaptive dual-frequency updating is still sufficient even at these times to meet the 10-ft range error.

During sporadic disturbed periods, such as when impulsive TIDs generated in auroral substorms pass over narrow regions of the radar coverage volume, the space-time cell over which the 10-ft specification can be met by a single normalization will shrink in both space and time, therefore more frequent (in time and space) dual-frequency probing may become necessary.

2. MODEL'S SOFTWARE MECHANIZATION

2.1 General

The range correction model consists of three 33-element vectors. Let's assume, as an example, a radar azimuth span of 130°, an elevation angle span from 0.0° to 80°, and a height span from 0 to 1000 km. Ionospheric range corrections (ΔR) are required at each location in the volume. The value of ΔR is generated by multiplication of the three 33-element matrices (i for height, j for elevation angle, and k for azimuth). ΔR_{ijk} is generated by the following equation:

$$\Delta R_{ijk} = (H_i \; E_j \; A_k) \; C_0(U)$$

where

H_i is a 33-element vector which represents the variation in height referenced to 1000 km at a constant elevation angle (0.0°) and constant azimuth;

E_j is a 33-element vector which represents the variation in elevation angle referenced to 0.0° at a constant azimuth and constant height (1000 km);

A_k is a 33-element vector which represents the variation in azimuth referenced to boresight at a constant height (1000 km) and constant elevation angle (0.0°);

C_0 is a normalization constant which makes the largest value in the H_i vector (which occurs at 1000 km) equal to 1;

U is a factor which updates the model either from a real-time pulse pair measurement from a satellite within the radar coverage or from an ionospheric update service such as the Air Weather Service (AWS) Prediction Service.

Figure 1 provides an example of the median range error predicted as a function of height and elevation angle for an azimuth located at the array boresight. The 33-element elevation and height vectors are determined from this data.

2.2 Height Vector

At radar boresight and 0.0° elevation angle, the range corrections at the radar frequency are determined in 33 steps in height. If we adopt a function I defined as

$$I = K\left[(1 + h/R)^2 - 1\right]$$

it can be shown that

$$I = K\left[(r/R)^2 + 2(r/R)\sin E\right]$$

where h = height of target above earth in kilometers, R is the radius of the earth in kilometers, K is the scaling constant such that I = 33 at h = 1000 km, r is the target range in kilometers, and E is the elevation angle. As r and sin E are normally available in a real-time system, it is relatively straightforward to calculate I. Thus, storing the ionospheric corrections in I space reduces computational requirements. Table 2 shows the value of i, I and the corresponding height for 33 steps along a 0.0° elevation angle. The constant K = 97.46. Note that at h = 0 km, I = 0 and the ionospheric correction is zero. If in the calculation of I we have I = 5.4, the value of H at i = 5.4 is determined by linear interpolation between the value of H at i = 5 and i = 6. The constant C_0 is equal to $1/H_{33}$ and all values of H are divided by H_{33} before entry into the H_i vector.

2.3 Elevation Vector

The elevation vector E are arrayed in (sin E) space where sin E = v cos E_0 + w sin E_0 (E_0 is the tilt of the array from zenith). The 33 elements of the vector are generated in equal increments of sin space from 0.0 (j = 1) to 80 (j = 33) degrees. Table 3 lists the element number "j", the elevation and sin E for the 33 elements. If the calculated sin E from the radar data falls between vector elements, linear interpolation between vector elements will be used.

2.4 Azimuth Vector

The azimuth vector A_k [Kahrilas, 1976] is arrayed in (Tan ΔAz) space where:

$$\text{Tan } \Delta Az = \frac{u}{w \cos E_0 - v \sin E_0}$$

and

$$\Delta Az = Az_0 - Az \quad (Az_0 = \text{radar boresight})$$

ΔAz varies from 0 to 65° off boresight. The use of equal steps in tan ΔAz will lead to a large first step in azimuth from boresight. That is, if a 33-element vector is used the first step in ΔAz space is 7.6° (i.e., 0 to 7.6°). Thus, near the radar boresight, where the greatest accuracy is required, the fewest azimuthal vector elements will be located. However, by using the function $\tan(\Delta Az)/(1 + 0.2 \tan^2(\Delta Az))$, the first step is 4° and further the function provides a fairly uniform spacing in azimuth space. As such, the 33-element azimuth vector is generated in equal steps of $\tan(\Delta Az)/(1 + 0.2 \tan^2(\Delta Az))$ at 0.0° elevation angle (j = 1) and h = 1000 km (i = 33). Table 4 lists the 33 elements of the azimuthal vector including the element number "k", the true azimuth, the difference between true and boresight azimuth, and $(\tan(\Delta Az)/(1 + 0.2 \tan^2(\Delta Az))$ from the

Table 2
Height Vector H_i

i	I	h (km)	H_i
	0	0	
1	1	32.6	
2	2	65.0	
3	3	97.3	Generated By Model
.	.	.	
.	.	.	
31	31	943.3	
32	32	971.7	
33	33	1000.0	

Table 3
Elevation Vector E_j

(j) Vector Element	E	sin E	E_j
1	0.00	0.0000	
2	1.76	0.03078	
3	3.53	0.06156	
.	.	.	Generated By Model
.	.	.	
.	.	.	
31	67.43	0.9234	
32	72.59	0.9542	
33	80.05	0.9849	

Table 4
Azimuth Vector A_k

K	ΔAz (°)	$\frac{\tan(\Delta Az)}{(1 + 0.2\tan^2(\Delta Az))}$	A_k	
1	-65.0	-1.117		
2	-57.2	-1.047		
3	-52.8	-0.977	Generated By Model	
.	.	.		
.	.	.		
.	.	.		
31	52.8	+0.977		
32	57.2	+1.047		
33	65.0	+1.116		

radar data falls between vector elements, linear interpolation between vector elements will be used.

2.5 Update Procedure

In our study we have considered two update procedures, one adaptive and one non-adaptive. In both cases a single value multiplicatively updates the values of ΔR_{ijk} for all i, j and k. The first update procedure is based on the use of a pulse pair (dual frequency) measurement of the ionosphere by using a satellite of opportunity as the target. This update procedure must be repeated every half-hour. The second procedure is a non-adaptive update approach based on data provided by an ionospheric update service such as AWS. If the dual-frequency measurement is not available or otherwise unacceptable for whatever reason, the AWS update is used instead. In addition, the AWS update is used to ensure that the pulse pair (dual frequency) update does not generate unreasonable corrections

The update algorithm for the pulse pair (Upp) is

$$Upp = \frac{\Delta b}{\Delta R_{ijk}}$$

where

$$\Delta b = \frac{(R_L - R_H)\, f_H^2\, f_L^2}{f_{OP}^2 (f_H^2 - f_L^2)}$$

R_L is the apparent range to satellite at f_L

R_H is the apparent range to satellite at f_H

f_{OP} is the operating frequency of the radar

Δb is the ionospheric correction factor obtained directly from the pulse pair measurements and ΔR_{ijk} is the prediction of the model at that location.

The Air Weather Service (AWS) provides an update U_{AWS} on a three-hour basis. This number is used in the same manner as the number generated from the dual-frequency measurement except that it

requires an operator-initiated action. The update is defined as the ratio U_{AWS} where

$$U_{AWS} = \frac{\text{AWS Predicted TEC}}{\text{Median TEC}}$$

where TEC is the vertical Total Electron Content. The update U_{AWS} is provided for the "nominal ionospheric characteristic location" for the radar coverage.

Finally, a brief comment on when the update occurs and the time sequence involved in using the vectors from one hour to the next. A given set of vectors will be used from 30 minutes before the hour until 29 minutes, 59 seconds after the hour (e.g. the 00Z deck is used from 2330:00 until 0029:59). If an event occurs during the last few minutes of the half-hour and will continue into the next hour, the new table for the next hour should not be entered until the event has ended.

In addition to updating the model for the one-hour time period which encompasses the pulse-pair measurement, a pulse-pair measurement can also be used to update the model for succeeding hours. However, three hours after the pulse-pair update was obtained, the median model should be used with the AWS update unless a new pulse-pair measurement is made. That is, each pulse-pair update should be considered valid for at most three hours. The update ratio that is obtained during a specific hour can be used to update the set of vectors which represent the following two hours. It must be remembered, however, that the update factor or ratio once calculated does not change even though a new set of vectors representing the next hour is used.

2.6 Software Structure for the Ionospheric Range Correction Procedure

For each hour of the day (24 hours per day) ten computer cards will form the set of inputs required by the model. The 24 sets of ten computer cards will be valid for each day of the month. Each new month will require a new set of cards. Figure 2 shows a listing of the 10 computer cards for April 1977 at 0000Z. The first card is a header card which specifies the month, year, hour (in universal time) and the constant C_0 in feet. The first card also identifies the source of the ionospheric range correction, the elevation coverage sector, and the azimuthal coverage sector. The last 7 columns of each card (columns 74-80) contain the year, month, hour and card number, with the header card Card No. 0. The next three cards contain the height vector H_1. The location of H_i for i = 1 to 33 is noted on Figure 2. In like manner, cards 5 and 7 contain the elevation vector E_j in sin space and cards 8, 9 and 10 contain the azimuth vector A_k in tan space. Figure 3 is a listing of the computer cards which provide the ionospheric range error model for the month of April 1977.

3. VERIFICATION PROCEDURE

3.1 General

Range correction, to the first order, is directly proportional to the integrated electron content along the slant ray path and for an L-band radar tracking a target at 5 degrees elevation angle at a height of 1000 km, the proportionally constant is 7.1 feet for each TEC unit of 10^{16} el/m^2. A recent study of the temporal variability of TEC at a station very similar to the target area of the L-band radar referred to in this paper [Dulong, 1977] can be used to estimate the effectiveness of the updating techniques across the overall radar's operational time.

The average day-to-day variability of the residual range correction after using the monthly median as a predictor will be 20% to 25% of the monthly median at night. The absolute error is about 2 to 3 times greater in daytime than in nighttime.

An evaluation of the effectiveness of the correction method illustrated in this paper was performed using nine years of continuous TEC measurements at Hamilton, Mass., a station very similar to the target area of the radar under evaluation, in geographic and geomagnetic coordinates. This has shown that

one calibration update performed according to our method in the target area can reduce the residual error by 60% when the update is one hour old and by 30% (in daytime) even when the update is three hours old.

3.2 Effect of Local Features

When the highest precision is required, then large local enhancements such as latitudinal gradients over the whole field of view, medium scale traveling disturbances and persistent localized features such as the mid-latitude trough are of concern. To assess their effects, a study is underway using field observations taken with the Navy Navigation Satellite System satellites [Leitinger and Hartmann, 1976]. This can be illustrated by Figure 4. The procedure consists of first determining the equivalent vertical range corrections for the ionosphere along a NNSS track as for 1120 local time, 12 December 1975. This information, plotted versus latitude, is then compared to the prediction of the median expected for the month. Thus, scaling factors are obtained as a function of latitude. These scaling factors are then used to update the median prediction over the whole field of view for the next pass, in this case at 1240 local time. In this way a sequential set of NNSS tracks can be used to simulate first the update calibration and then the actual radar measurement.

As seen in Figure 4, the gradients of the uncorrected median, derived from a smooth climatology of the ionosphere, do not match the actual gradients observed at 1240, resulting in large absolute residual errors in some regions. The adapted prediction, even though it is an hour and twenty minutes old and from a somewhat different longitude region, still significantly reduces the error. Note however that some of the medium scale features of the 1120 ionosphere are either displaced or changed in the 1240 ionosphere, which suggests that there would be the need of some filtering of the spatial information in the update process.

The effect of the mid-latitude trough is sampled in Figu. (after Leitinger and Hartmann, 1976). This series of sequential NNSS passes shows that the ionosphere was depressed below the expected medians during the predawn hours and that a deep trough formed near 49 N geographic latitude ($L \approx 4$) at the minimum. The NNSS pass near midnight local time (0005 LT) indicates the start of this feature, which was quite deep four hours later. With such a trough near the horizon of our radar, the residual error with respect to the predicted median would be increased by a factor of about three to account for the slant thickness of the ionosphere.

An on-going study [Allen et al, 1977] suggests that general enhancements of the ionosphere, changes in latitudinal gradients and large local features, may be tracked by local measurements, such as the dual-frequency update, and that they may be extended over the field of view. The density of the updating grid in space and time will be determined by the stringency of the metric requirements of the radar.

3.3 L-Band Radar Data

The proposed approach has been evaluated by comparing L-band radar observations with predicted (and updated) ionospheric range errors. Figure 6 shows the comparison between median predictions and radar observations of ionospheric range error (the radar observations of the ionosphere were obtained using differential time delay from dual-frequency transmission from the radar). In this case the median predicted and radar observations (at the normalization point) were equivalent and no update was required. The difference between the prediction and the observation was in all cases less than 5 ft.

The second example (Figure 7) shows that the median predictions were underpredicting the required range correction. However, updating the median prediction substantially reduces the discrepancy between predicted and measured data.

4. CONCLUSIONS

The effectiveness of the adaptive correction of the deteriorating influence of the ionosphere on the range accuracy

of terrestrial radars [Allen et al, 1976; DuLong et al, 1977; Katz et al, 1977] has been confirmed by experimental findings.

By updating climatological models of the expected ionospheric errors with real-time dual-frequency measurements of the medium, it was shown that L-band radar range errors could be kept at all times below 10 ft with a spatially and temporally rare application of the updating process. This statement holds true even taking into account the presence of localized gradients and features, of difficult inclusion in climatological models. To accommodate then exceptional ionospheric circumstances or more stringent radar range accuracy requirements, our method must simply be applied with a higher spatial and temporal density.

REFERENCES

Allen, R.S., D.D. DuLong, M.D. Grossi, A.H. Katz, Ionospheric Range Error Correction in Precision Radar Systems by Adaptive Probing of the Propagation Medium, AGARD Conf. Proc. #209 on Propagation Limitations of Navigation and Positioning Systems, Istanbul, Turkey, 1976.

Allen, R.S., D.E. Donatelli, G.K. Hartmann, R. Leitinger, Adaptive Mapping of Mid-Latitude Ionosphere, AFGL-TR-77-0176, August 1977.

DuLong, D.D., Reduction of the Uncertainties of Range Correction, AFGL-TR-77-0125, June 1977.

Kahrilas, P.J., Electronic Scanning Radar Systems (ESRS) Design Handbook, Chapter 7, Artech House, Inc., 1976.

Katz, A.H., M.D. Grossi, T.I.S. Boak, Practical Implementation of Adaptive Ionospheric Range Error Correction for High Accuracy Radars, URSI Meeting, San Francisco, Cal., June 1977.

Leitinger, R. and G.K. Hartmann, Time and Latitude Dependence of Ionospheric Electron Content from the Combination of NNSS and ATS-6 Data, COSPAR Symposium Proceedings, Geophysical Use of Satellite Beacon Observation, Boston University, June 1976.

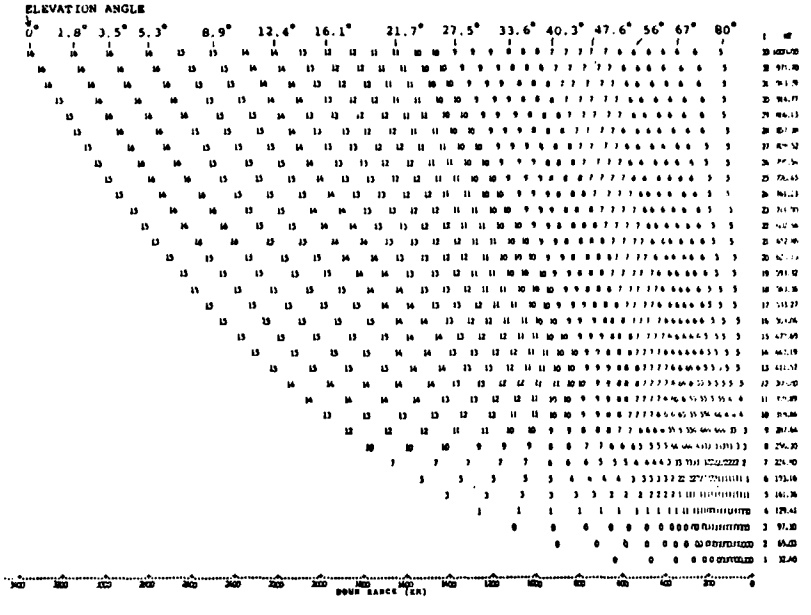


Figure 1 Plot of Range Error (feet) at L-Band Vs Height and Down Range for Selected Elevation Angles for 0000Z, April 1977

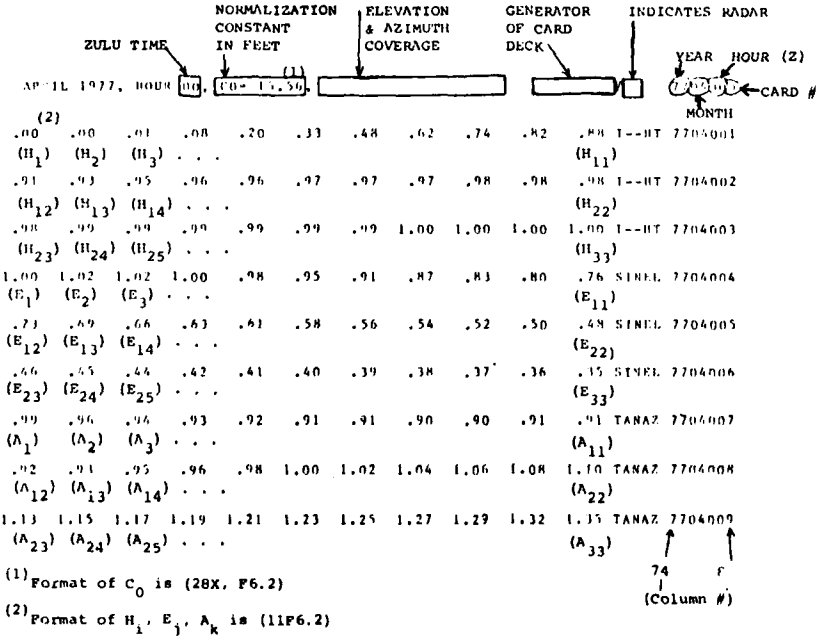


Figure 2 Listing of Computer Cards for One Hour in April 1977

```
  APRIL 1977, HOUR 18, CO=  4.03,                                                7704180
.00   .00   .00   .00   .01   .02   .03   .05   .08   .22   .40 I--HT 7704181
.55   .67   .74   .79   .83   .85   .87   .89   .90   .91   .92 I--HT 7704182
.93   .94   .94   .95   .96   .97   .97   .98   .99   .99  1.00 I--HT 7704183
1.00  1.06  1.15  1.26  1.39  1.51  1.67  1.72  1.69  1.62  1.55 SINEL 7704184
1.49  1.43  1.37  1.32  1.27  1.23  1.19  1.14  1.11  1.07  1.04 SINEL 7704185
1.00   .97   .95   .92   .89   .87   .85   .83   .81   .79   .77 SINEL 7704186
2.28  2.13  2.04  2.00  2.07  2.09  2.08  2.05  2.01  1.98  1.70 TANAZ 7704187
1.38  1.21  1.10  1.04  1.01  1.00  1.00  1.01  1.02  1.04  1.07 TANAZ 7704188
1.09  1.13  1.16  1.20  1.24  1.28  1.32  1.37  1.41  1.45  1.51 TANAZ 7704189
  APRIL 1977, HOUR 19, CO=  8.75,                                                7704190
.00   .00   .00   .06   .15   .24   .33   .42   .53   .63   .71 I--HT 7704191
.79   .84   .88   .90   .92   .93   .94   .95   .95   .96   .96 I--HT 7704192
.97   .97   .97   .98   .98   .98   .99   .99   .99  1.00  1.00 I--HT 7704193
1.00  1.04  1.04  1.03  1.01   .99   .96   .93   .90   .87   .83 SINEL 7704194
.80   .77   .74   .72   .69   .67   .64   .62   .60   .58   .56 SINEL 7704195
.55   .53   .51   .50   .49   .47   .46   .45   .44   .43   .42 SINEL 7704196
1.20  1.14  1.11  1.08  1.06  1.03  1.02  1.00  1.00  1.03  1.04 TANAZ 7704197
1.05  1.04  1.04  1.03  1.02  1.00   .97   .94   .90   .87   .85 TANAZ 7704198
.84   .85   .86   .87   .90   .92   .96  1.00  1.05  1.11  1.20 TANAZ 7704199
  APRIL 1977, HOUR 20, CO=  9.75,                                                7704200
.00   .00   .01   .09   .21   .36   .50   .63   .72   .79   .84 I--HT 77
.87   .90   .91   .93   .93   .94   .95   .95   .96   .96   .97 I--
.97   .97   .98   .98   .98   .99   .99   .99   .99  1.00  1.00                  08
1.00  1.01  1.01  1.00   .99   .97   .95   .92   .89   .86                       7704209
.80   .77   .74   .71   .69   .66   .64   .62   .60                             7704210
.54   .53   .51   .50   .48   .47   .46   .45   .43                       .88 I--HT 7704211
1.19  1.14  1.11  1.08  1.06  1.04  1.02  1.01                              .97 I--HT 7704212
.96   .95   .95   .96   .98  1.00  1.02  1.03                        .00  1.00 I--HT 7704213
1.10  1.12  1.15  1.17  1.19  1.22  1.24  1.26                        .90   .87 SINEL 7704214
                                                                      .59   .57 SINEL 7704215
  APRIL 1977, HOUR 21, CO= 10.51,                                     .43   .42 SINEL 7704216
.00   .00   .01   .10   .24   .42   .58                             1.03  1.02  1.01 TANAZ 7704217
.91   .92   .93   .94   .95   .95                                   1.12  1.15  1.18 TANAZ 7704218
.97   .98   .98   .98   .99   .99                                   1.48  1.52  1.59 TANAZ 7704219
1.00  1.03  1.05  1.06  1.05                                                      7704220
.83   .80   .77                                                                  7704221
.55   .53                                .40   .56   .70   .80   .86   .90 I--HT 7704221
1.20  1.15                             .96   .96   .97   .97   .97   .98   .98 I--HT 7704222
1.01                                  .99   .99   .99   .99   .99  1.00  1.00  1.00 I--HT 7704223
1.23                                 1.05  1.04  1.01   .98   .95   .91   .87   .83 SINEL 7704224
                                .73   .70   .67   .65   .62   .60   .57   .55   .53 SINEL 7704225
                          .50   .48   .47   .45   .44   .43   .42   .41   .40   .39 SINEL 7704226
1.13  1.09  1.07  1.05  1.04  1.03  1.02  1.01  1.00   .99   .99 TANAZ 7704227
.98   .98   .97   .96   .97  1.00  1.04  1.08  1.11  1.14  1.18 TANAZ 7704228
1.21  1.24  1.28  1.31  1.34  1.37  1.40  1.43  1.46  1.50  1.56 TANAZ 7704229
  APRIL 1977, HOUR 23, CO= 13.95,                                                7704230
.00   .00   .01   .09   .21   .36   .52   .66   .77   .85   .89 I--HT 7704231
.92   .94   .95   .96   .96   .97   .97   .97   .98   .98   .98 I--HT 7704232
.98   .99   .99   .99   .99   .99   .99  1.00  1.00  1.00  1.00 I--HT 7704233
1.00  1.03  1.04  1.03  1.01   .98   .94   .90   .87   .83   .79 SINEL 7704234
.75   .72   .69   .66   .63   .61   .58   .56   .54   .52   .50 SINEL 7704235
.49   .47   .45   .44   .43   .41   .40   .39   .38   .37   .36 SINEL 7704236
1.04   .99   .98   .97   .96   .95   .94   .94   .93   .93   .92 TANAZ 7704237
.92   .92   .93   .95   .97  1.00  1.03  1.05  1.08  1.11  1.13 TANAZ 7704238
1.16  1.18  1.21  1.23  1.25  1.27  1.30  1.32  1.34  1.37  1.43 TANAZ 7704239
```

Figure 3 Example of Listing of Computer Deck for Ionospheric Correction (April 1977)

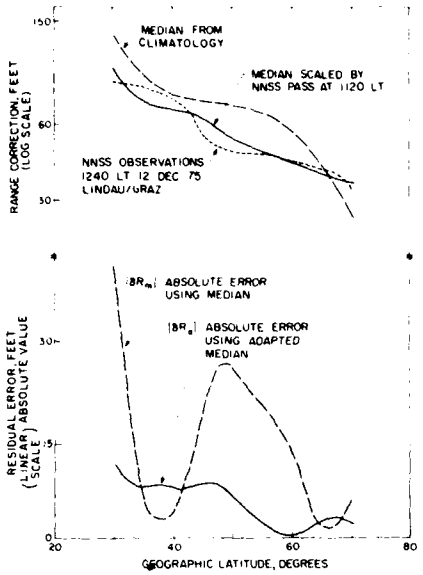


Figure 4 Adaptive Modeling Using Data from NNSS Measurements. 1st pass at 1120 LT is used to scale predicted latitudinal variation at 1120 LT and then the scaled values are predicted forward to 1240 LT.

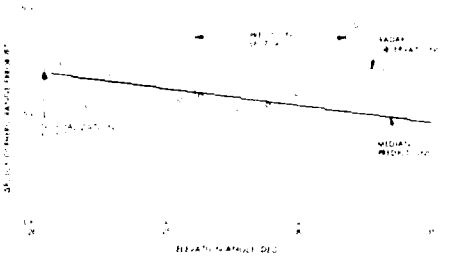


Figure 6 Comparison of L-Band Radar Observations with Predicted Ionospheric Range Errors

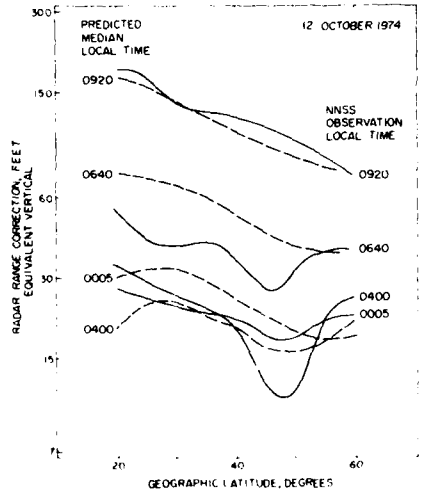


Figure 5 Sample of the Effect of the Trough in Electron Density, Seen Near 49°N Latitude on 12 Oct. 1974 on the Expected Range Correction for an L-band Radar

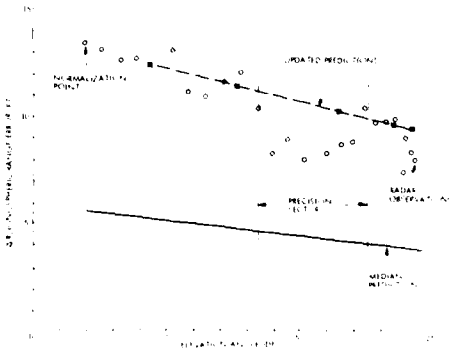


Figure 7 Comparison of L-Band Radar Observations with Predicted and Adaptively Updated Ionospheric Range Errors

TRANSIONOSPHERIC PROPAGATION EFFECTS ON COHERENT INTEGRATION BY GROUND-BASED RADARS

R. Wm. Bush
HMED, General Electric Company
Syracuse, New York 13221

ABSTRACT

Coherent processing of multiple radar returns is utilized in a wide variety of applications to extract critical signal information and to enhance the performance capabilities of a given radar system. The quality of information and degree of enhancement improves in proportion to the length of coherent observation. Phase perturbations induced by the ionosphere from spatial and temporal variations in electron density are shown to ultimately limit the useful coherence intervals to times from fractions of a minute at VHF to several minutes at L-band frequencies.

The degradation in coherent processing is a function of both the particular processor being considered and the nature of the phase perturbations encountered. In recent years, the coherent processor has generally taken the form of a digital Fast Fourier Transform (FFT) to map the time (or spatial) domain into the frequency domain. The performance of an FFT processor is evaluated parametrically for typical phase perturbations associated with various observed ionospheric states. The observed ionospheric effects treated include: the slow but very large phase changes which result from diurnal variations in total electron content (TEC); the sinusoidal, more rapid phase changes which result from traveling ionospheric disturbances (TID's); and the random phase fluctuations which result from signal scintillation in the presence of small-scale irregularities.

I. INTRODUCTION

Advancements in equipment technology are making coherent processing more and more attractive as an efficient technique to exploit the basic capability of a modern radar. As a result there is a large community interested in utilizing the technique for very diverse applications. Examples include radar imaging of satellites from complex receive signal returns for Space Object Identification (SOI); enhancement of angular resolution through synthetic aperture, vehicle-borne radars for terrain mapping; and coherent integration of multiple pulses to extend the per-pulse detection and tracking ranges of ground-based radars out to and beyond synchronous satellite altitudes. The performance achieved for each of these techniques is ultimately limited by phase perturbations induced by the external environment.

Numerous investigators have observed a wide range of effects on RF signals propagated through the ionosphere. The phenomena which account for the observed effects are dependent upon both spatial and temporal variables, some of which are geographic location, season, local time, solar flux, and magnetic activity. The observed phenomena not only differ in form but also vary greatly in level of activity and frequency of occurrence. A brief review of the observed phenomena is presented to form a basis for analyzing the effects of ionospheric-induced phase perturbations upon coherent integration.

II. SURVEY OF IONOSPHERIC EFFECTS

The ionosphere may be viewed as a non-homogeneous cloud of free electrons (and other particles). Both spatial and temporal variations in the distribution and number of electrons will, in general, result in an associated variation in the signal phase of a satellite transmission (or radar echo return). Slow variations in phase delay are a result of gradual changes in electron content and the presence of large-scale irregularities, such as occur during diurnal variations in total electron content (TEC) and traveling ionospheric disturbances (TID's). Random fluctuations occur during signal scintillations caused by small-scale irregularities.

Diurnal Variations in Electron Content

The total electron content in a local region of the ionosphere has been shown[1] to vary in a near-cyclic manner with a period of a day, as illustrated in Figure 1. The data represents one month of con-

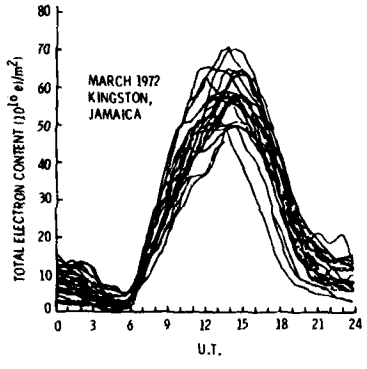


Figure 1. Variations in Total Electron Content from Kingston, Jamaica, West Indies, March 1972 (After Klobuchan, 1973)

tinuous recordings of TEC superimposed over a 24-hour period. The minimum TEC occurs during early morning and increases more than tenfold to a peak value at early evening. During the early morning and early evening, the TEC varies in an extremely non-linear fashion, which induces non-linear phase perturbations that tend to spread the signal spectrum of RF transmissions propagated through the medium. These variations are of potential concern and are treated in Section IV.

If we consider nearly vertically directed signal propagation paths, the two-way phase shift ϕ is given by

$$\phi_{TEC} = 16.9 \frac{TEC}{F} \qquad (1)$$

where TEC is the total electron content in units of 10^{12} electrons/cm^2 and F is the frequency in GHz (10^9 Hz). If the path of propagation is at an angle of ψ with respect to the zenith, then the path length through the ionosphere increases proportional to sec ψ.

Since the diurnal variations in TEC are periodic, they can be approximated by a harmonic series:

$$\phi_{TEC}(t) = 16.9 \frac{TEC_{MAX}}{F} \sum_{K=0}^{n} A_K \cos\left(\frac{2\pi K}{P} t + \theta_K\right) \qquad (2)$$

where

TEC_{MAX} is the maximum TEC observed over the day
A_K, θ_K are the Fourier component amplitude and phase
P is the period of a day

Traveling Ionospheric Disturbances (TID's)

Traveling ionospheric disturbances (TID's), are large-scale compressions and rarefactions of electron-density in the F-region of the ionosphere possessing a normal component of velocity relative to the path of propagation. They have been observed since the early 1950's, through the use of a variety of techniques, including ionosondes, spaced doppler sounders, and incoherent-scatter radars. The presence of TID's has been shown[2] to induce quasi-sinusoidal phase perturbations, as illustrated in Figure 2.

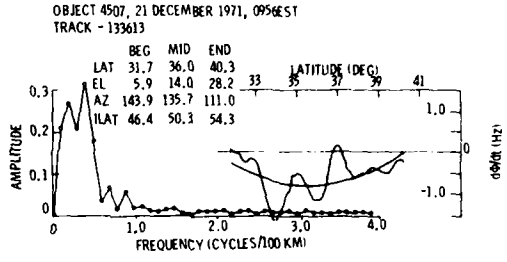


Figure 2. Differential-Doppler Residuals and Their Power Spectra in the Presence of a Long-Period TID (After Evans, 1973)

Observations compiled by Davies and Jones[3] and Hunsucker and Tvetten[4] indicated that the describing parameters for TID's vary over a rather large range of values. The typical wavelengths for observed TID's vary from 150 km (at the 10 percentile) to 1000 km (at the 90 percentile). The observed velocities range from 95 m/second (at the 10 percentile) to 200 m/second (at the 90 percentile). The reported periods varied from 16 minutes (at the 10 percentile) to 120 minutes (at the 90 percentile), with a median value of approximately 28 minutes.

The frequency of occurrence and characteristics of the TID have been shown by Evans[5] to depend upon time of day, season, and geographic latitude. The most frequent and most severe TID's occur during the daylight hours. The degree of modulation is typically 1% to 2%.

The magnitude and rate of phase change caused by TID's are a function of the ambient TEC, the percent modulation, the location in time along the TID wave, and the operating frequency

$$\phi_{TID}(t) = 16.9 \frac{M * TEC}{F} \cos\left(\frac{2\pi t}{T}\right); \quad T = \frac{L}{v} \qquad (3)$$

where

TEC is the ambient total electron content (10^{12} electrons/cm^2)
M is the fractional modulation
T is the period of the TID (seconds)
L is the spatial wavelength (km)
v is the horizontal velocity component (km/second)
F is the operating frequency (GHz)

Signal Scintillations

Scintillations in the amplitude and phase of an RF signal propagated through the ionosphere are caused by refraction and diffraction of the signal as it passes through small-scale irregularities which possess a normal component of velocity relative to the path of propagation. For the case of satellite transmissions (or radar echo returns), the relative velocity is generally a combination of the motion of the satellite and the ionospheric plasma drift. Amplitude and phase scintillation have been frequently[6] observed at VHF, as illustrated in Figure 3.

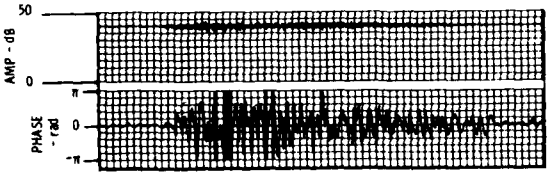


Figure 3. Observed VHF Amplitude and Phase Scintillation (After Fremouw, 1976)

The characteristics of signal scintillation have been described statistically in terms of intensity and phase scintillation indices, S_4 and σ_ϕ. Both short-term and long-term statistics have been compiled.[6] The short-term statistics include quantitative values for S_4 and σ_ϕ. The long-term statistics include frequencies of occurrence for several different archetypes of activity.

III. PROCESSOR CHARACTERIZATION

Analytic Description

Coherent integration of a sequence of N complex signal samples is achieved by systematically combing the samples to align the signal vectors in phase. A Fast Fourier Transform (FFT) algorithm is commonly used to coherently process a sequence of complex samples with linear time-varying phase according to the Discrete Fourier Transform

$$S_n = \frac{1}{N} \sum_{k=0}^{N-1} S_k \, e^{-j\left(\frac{2\pi}{N}\right) nk} \qquad (4)$$

The evaluation of ionospheric phase perturbation which follows assumes that an FFT processor, or its equivalent, has been implemented to perform coherent integration. Therefore, any non-linear (i.e., non-uniform, time-varying) phase change will in general degrade the coherent integration. For moderate phase variations, the performance of the FFT can be easily evaluated by shifting the signal spectrum down to baseband and applying Equation 4 for n=0.

We begin by examining the response of the FFT to systematic phase perturbations. A unit amplitude signal which exhibits a sinusoidal phase perturbation such as induced by the presence of a TID of amplitude $\Delta\theta$ and frequency ω_T can be expressed as

$$S(t) = e^{j\Delta\theta \sin(\omega_T t)} \qquad (5)$$

Substituting Equation 5 into Equation 4 and replacing the summation by an integral for large values of N, it follows that:

$$S_o = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} e^{j\Delta\theta \sin(\omega_T t)} \, dt \qquad (6)$$

If the coherent integration is taken over the full period of the sinusoid, then Equation 6 can be recognized as the zero-order Bessel function of the first kind:

$$J_o(\Delta\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\Delta\theta \sin(x)} \, dx \qquad (7)$$

The coherent integration loss is simply computed from a tabulation of $J_o(\Delta\theta)$. The family of sideband components for the familiar FM spectrum are given by the higher-order Bessel functions. For large phase excursion (modulation indices $\Delta\theta > \pi$) the energy of the sidebands can exceed the carrier and must therefore be evaluated.

If the integration is taken over a portion of the sinusoid, the non-linear phase perturbations can be expressed to a first approximation in terms of a quadratic phase trend.

$$s(t) = e^{j\frac{a}{2}t^2} \qquad (8)$$

Substitution of Equation 8 into Equation 4 and replacing the summation by an integral yields the normalized complex Fresnel integral

$$z(x) = \frac{1}{T} \int_{o}^{x} e^{j\frac{\pi}{2}t^2} \, dt \qquad (9)$$

The coherent integration loss is simply computed from a tabulation of the real Fresnel integrals C(X) and S(X). If we evaluate Equation 4 for all values of n, we can obtain the spectral components of the familiar linear frequency modulation waveform. Again for large phase excursions, the energy of the sidebands can exceed the carrier and must therefore be evaluated.

The analysis of the random phase fluctuations, associated with scintillation, is based upon a conservative assumption that the phase errors are in-

dependent sample to sample. Since we are considering a random process, the quantity of interest is the expected value. Without loss in generality, we can again consider the output of the zero doppler cell. Two distributions are considered, the normal or Gaussian pdf, which is characteristic of a noise-like environment, and the uniform pdf which is characteristic of a dispersive distribution.

The calculations are greatly simplified if one recognizes that the quadrature components cancel, and Equation 4 reduces to simply

$$S_o = \frac{1}{N} \sum_{k=0}^{N-1} R_e(s_k) = \frac{1}{N} \sum_{k=0}^{N-1} \cos \phi \qquad (10)$$

The expected value of Equation 10 can be expressed in terms of the probability distribution $f_\phi(\phi)$

$$< S_o > = < \cos \phi > = \int \cos \phi \, f_\phi(\phi) \, d\phi \qquad (11)$$

For Gaussian distributed phase perturbations with zero mean and variance σ_ϕ^2, the loss is easily calculated as

$$L(dB) = 4.343 \ \sigma_\phi^2 \qquad (12)$$

For uniformly distributed phase perturbations with zero mean and standard deviation σ_ϕ, the loss is calculated as

$$L(dB) = 20 \log \left[\frac{\sin(\sqrt{3}\,\sigma_\phi)}{\sqrt{3}\,\sigma_\phi} \right] \qquad (13)$$

The analytic equations developed have been validated for the assumed models using a computer simulation which emulated the radar FFT processor. The results are consistent for moderate integration losses in which the peak signal output occurs in the zero doppler cell. Large phase perturbations become difficult to evaluate analytically for two reasons: first, the maximum signal output does not occur in the zero doppler cell, and hence the calculations become somewhat more involved; second, spectral fold-over occurring in the FFT process further complicates the calculations.

Computer Simulation

To assess the sensitivity of coherent integration to the variety of parameters which characterize the observed ionospheric effects, a computer simulation was implemented to emulate the characteristics of the radar. Specifically, the various models developed to mathematically describe the phase perturbations were exercised using parameter values representative of those experimentally observed. A digital FFT algorithm was used to coherently integrate phase-perturbed sequences of time samples.

The coherent integration loss resulting from sinusoidal and quadratic deterministic phase excursions, and from uniformly and Gaussian distributed, random, independent phase perturbations are illustrated in Figure 4 as a function of the phase excursion. The phase excursion represents the peak phase error over the observation interval for deterministic variations and three times the standard deviation of the phase perturbations for the random variation. The losses shown for sinusoidal phase variations are those associated with an integration time interval equal to a half-period centered about *the peak of the phase excursion*. As illustrated, the coherent integration loss increases rapidly (nearly quadratically as a function of phase) for phase excursions beyond 90^o.

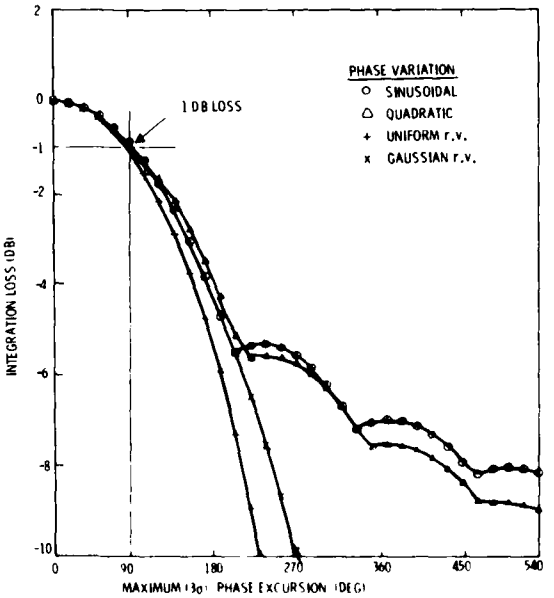


Figure 4. Coherent Integration Loss for Deterministic and Random Phase Fluctuations

The degradation in signal-to-noise ratio from quadratic time varying losses is shown on Figure 5.

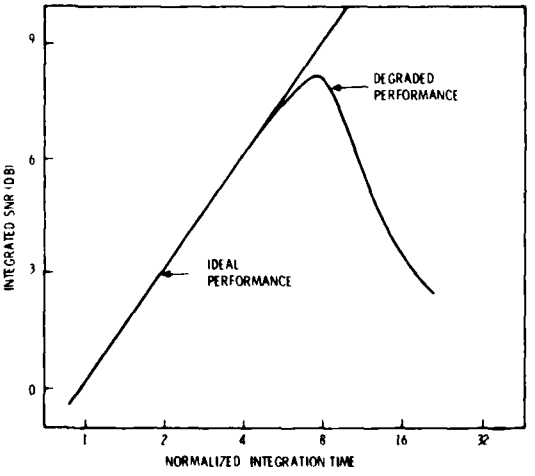


Figure 5. Degradation in Integrated Signal-to-Noise Ratio from Time-Varying Phase Errors

The significant point being illustrated is that once the maximum phase error reaches a value of approximately 90^{o}, the signal-to-noise ratio ceases to improve and actually decreases as time progresses. This is intuitively satisfying, since the signal vectors tend to cancel as the phase difference exceeds 90^{o}.

IV. PARAMETRIC EVALUATION

A set of curves are presented which quantitatively describe the expected degradation in coherent integration for typical and extreme conditions within the ionosphere. The curves were generated using a software FFT algorithm operating on simulated data. The presentation of the data allows an estimate of the useful length of coherent integration as a function of the parameters which characterize the particular ionospheric effects. Each of the phenomena described in Section II are treated separately. In practice, perturbations resulting from each ionospheric effect are all present to some degree.

Limitations from Diurnal Variations in TEC

The mathematical model presented in Section II, together with reported(5) diurnal variations in TEC, was used to generate phase perturbations of simulated radar signals as a function of time of day. As shown in Figure 6, several thousand radians of phase change occur at UHF throughout the period of a day. The most non-linear phase variations are noted to occur in the early morning hours of the day.

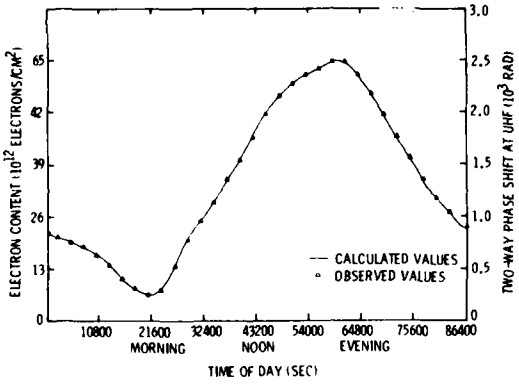


Figure 6. Typical Diurnal Variation in Electron Content and Corresponding Phase Shift at UHF (Reference 5)

To assess the sensitivity of the loss to both the operating frequency and length of integration time, the losses were computed for the worst part of the day (during sunrise). Figure 7 presents the coherent integration loss experience at VHF, UHF, and L-band as a function of the length of coherent integration time. The data indicates that for VHF, the coherent integration time is limited to a few hundred seconds. At UHF, useful coherent times are available up to eight minutes. At L-band, there is no significant limitation to coherent integration for the time intervals considered.

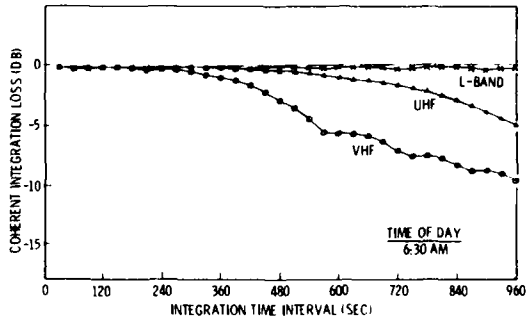


Figure 7. Coherent Integration Loss from Diurnal Variations in Electron Content

Limitations from TID's

The phase perturbations induced by the presence of a TID are a function of percent modulation and ambient TEC, the period of the TID, the time of integration within the period of the TID, and the frequency of the signal being propagated. A set of four graphs are presented which parametrically describe the expected coherent integration loss and useful coherent integration time for the VHF, UHF, and L-band frequencies of interest. The values used for the parameters cover the range of extreme and typical values observed. (2, 3, 4)

The coherent integration loss as a function of coherent integration time at the maximum point of a TID is presented on Figure 8. The parameters, for the TID shown, represent a short period TID occurring during late afternoon where the TEC is maximum. These parameters represent near-worst-case conditions. The coherent integration time is limited to approximately 30 seconds at VHF, 60 seconds at UHF, and 120 seconds at L-band.

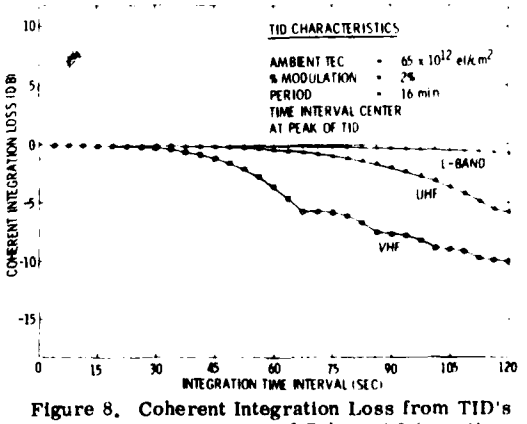


Figure 8. Coherent Integration Loss from TID's as a Function of Coherent Integration Time Interval

The coherent integration loss as a function of the midpoint of the integration time interval is shown on Figure 9. The previous illustrations

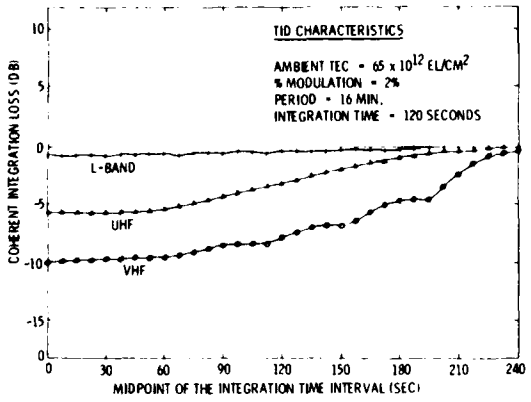


Figure 9. Coherent Integration Loss from TID's as a Function of the Midpoint of the Integration Time Interval

assumed that the integration time interval was centered about the peak of the TID, which produces the greatest non-linear phase excursions. At UHF and L-band, the coherent integration loss decreases to zero when the integration time interval is centered about the point of inflection of the TID waveform. The change in integration loss is seen to vary gradually as a function to the midpoint position.

The coherent integration loss as a function of the period of the TID's is illustrated in Figure 10. A coherent integration time of 120 seconds was assumed. At UHF, the loss is significant for TID periods less than 28 minutes, which corresponds to approximately 50% of the observed periods. At VHF, significant losses are present for TID periods up to 50 minutes, or for nearly 75% of the periods observed. The magnitude of 2% modulation for an ambient TEC of 65 x 10^{12} el/cm^2 represents typical daily extremes. The TEC and, consequently, the magnitude are significantly lower over most parts of the day.

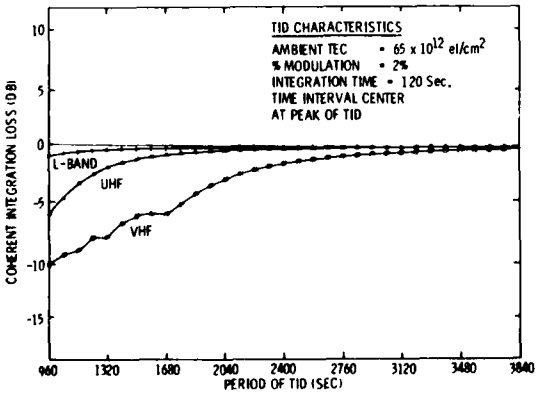


Figure 10. Coherent Integration Loss from TID's as a Function of the Period of the TID

The coherent integration loss as a function of the amplitude of the TID is presented in Figure 11. The amplitude of the TID is characterized by the ambient TEC and fractional modulation M. Change in amplitude, M * TEC, can be interpreted as either changes in ambient TEC or fractional modulation or both. For example, the midportion of the figure represents the expected coherent integration loss for a midday TID with a 1% modulation. The left-hand portion of the plot represents the expected coherent integration loss for 2% modulation during the night-time hours of the day in which the TEC is generally an order of magnitude below the peak.

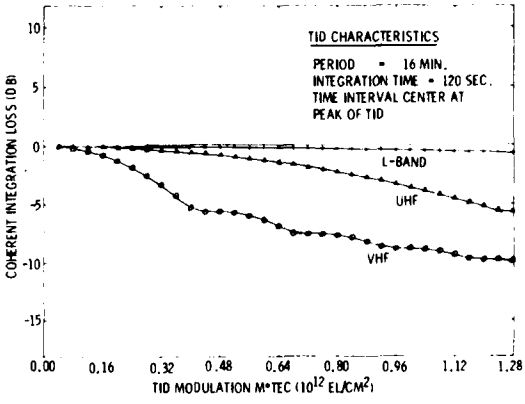


Figure 11. Coherent Integration Loss from TID's as a Function of the Magnitude of the TID

Limitations from Scintillation

An estimate of the expected losses for coherent integration in the presence of phase scintillations is made based upon the observations analyzed by Fremouw[6] and the assumptions that the phase fluctuations are independent and Gaussian distributed.

The loss has been shown in Section III to vary proportionally with the square of the standard deviation σ_ϕ. The standard deviation has been observed to vary inversely with frequency. Thus, for a given degree of scintillation, the loss varies as the inverse of the square of the frequency.

The minimum, average, and maximum coherent integration loss at VHF, UHF, and L-band associated with the reported[6] minimum, average, and maximum phase scintillation index σ_ϕ are presented in Figure 12 for six archetypes of activity observed. The maximum losses computed for VHF are significant even during the quiet periods observed. The average losses computed for UHF are insignificant for all activities except those that are very active.

A cumulative distribution of the frequency of occurrence for the various archetypes of observed[6] activity level is presented on Figure 13 for 250 observations collected at Ancon, Peru, Stanford, California, and Poker Flat, Alaska. Over 95% of the scintillations observed were less

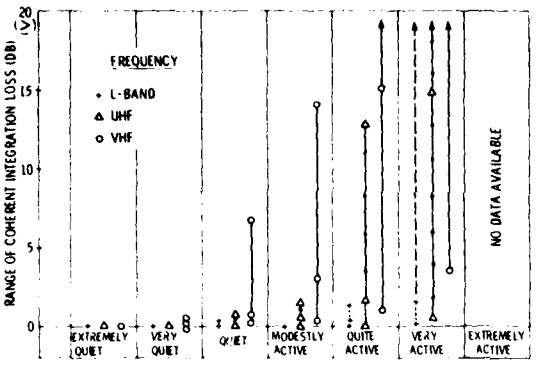


Figure 12. Range of Coherent Integration Loss for Various Activity Levels of Scintillation

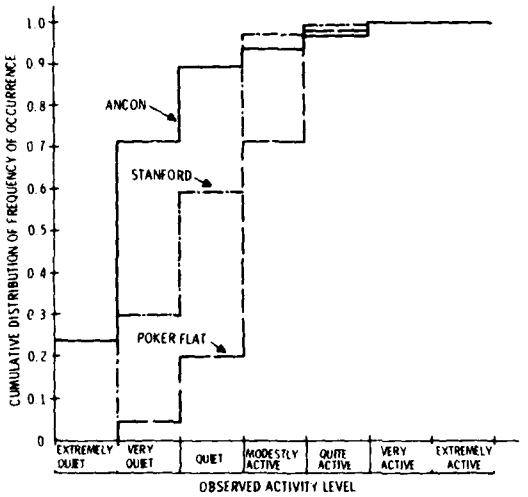


Figure 13. Cumulative Distribution of Frequency of Occurrence for Various Activity Levels of Scintillation

than very active for all latitudes. The corresponding useful coherence time represents 20 seconds for observations of the DNA Wideband satellite at an altitude of 1,000 km. The useful coherence times for satellites at geostationary altitudes increase as a consequence of the slower velocity between the irregularities and line-of-sight. The scaling factor is related to ratio of the satellite velocity and ionospheric plasma drift velocity taking into account the orientation of the irregularities. Recent comparisons of low-altitude and high-altitude satellite observations indicate a scaling factor on the order of ten.

The coherent integration losses calculated from the observed signal scintillations represent conservative estimates which apply to the given observation interval of 20 seconds. If the phase perturbations were completely uncorrelated, and were derived from a stationary process, then the RMS phase error and the associated loss would be invariant with the observation interval. For correlated phase perturbations, however, it is not possible to quantitatively interpolate or extrapolate the results to other time intervals using the phase variance as the only statistical parameter.

V. CONCLUSION

The evaluation of the effect of various types of induced ionospheric phase perturbations on coherent integration indicates that the useful coherence times can be expected to vary from tens of seconds to several hundred seconds. The actual limitation depends upon the operating frequency and state of the ionosphere. Coherent integration losses which are moderate at UHF become intolerable at VHF and insignificant at L-band. The limitations imposed by diurnal variations in TEC and from the presence of TID's are typically modest at UHF. In principle, their effect can be minimized through appropriate operational procedures (such as time diversity) and through processing techniques (such as phase compensation) since they are systematic in nature. The random phase perturbations associated with signal scintillation result in processing losses, which can become excessive for the time interval considered, but on an infrequent basis.

The results of the study provide a useful design criterion for evaluating a coherent radar system. Specifically, for moderate phase perturbations, the coherent integration loss (in dB) is shown to vary nearly quadratically as a function of the maximum systematic or 3-sigma random phase excursion. The significance of this characteristic is that the integration loss gradually increases to a value of approximately 1 dB for a maximum (or 3 sigma) phase excursion of $\pi/2$ radians. Beyond this point the integration loss, associated with time-varying phase perturbations, exceeds the integration gain. The net result is that the signal-to-noise ratio of the processed radar returns not only ceases to improve but actually decreases as more samples are collected.

VI. CONTINUED INVESTIGATIONS

A preliminary evaluation of the effects of induced ionospheric phase perturbations on coherent integration by ground-based radars has been presented. Further evaluations are under current investigation in a joint study with SRI International. Parameters derived from empirical data are being evaluated with an analytic model to determine the effects of correlated random phase perturbations in the presence of large systematic phase trends. The results of the data reduction and analysis of several hundred hours of coherent transmissions from the ATS-6 satellite are presented at this symposium by Rino, et al. [7]

ACKNOWLEDGMENTS

The work was sponsored by the Electronic Systems Division, Air Force Systems Command under Contract F19628-76-C-0251.

REFERENCES

1. Klobuchar, J.A., Total Electron Content Studies of the Ionosphere, Air Force Cambridge Research Laboratories, Report No. AFCRL-TR-73-0098, February 1973.

2. Evans, J.V., Millstone Hill Radar Propagation Study: Scientific Results - Part II, Lincoln Laboratory, Technical Report No. 509, November 1973.

3. Davies, K., and Jones, J.E., "Three Dimensional Observations of Traveling Ionospheric Disturbances", Journal of Atmospheric and Terrestrial Physics, Vol. 33, pp. 39-46, 1971.

4. Hunsucker, R.D., and Tvetten, L.H., "Large Traveling - Ionospheric - Disturbances Observed at Midlatitudes Utilizing the High Resolution H.F. Backscatter Technique", Journal of Atmospheric and Terrestrial Physics, Vol. 29, pp. 909-916, 1967.

5. Lincoln Laboratory, Semiannual Technical Summary, ARPA/STO Program, June 1974.

6. Fremouw, E.J., Wideband Satellite Operations, Stanford Research Institute, Bimonthly Progress Reports, Contract DNA001-75-C-01, 1976.

7. Rino, C.L., Dawson, C.H., Livingston, R.C., and Petriceks, J., "The Ionospheric Limitations to Coherent Integration In Transionospheric Radars," Proceedings of the Symposium on the Effect of the Ionosphere on Space and Terrestrial Systems, January 1978.

ADAPTIVE MAPPING OF IONOSPHERIC FEATURES

R. Leitinger
Institut fur Meteorologie und Geophysik
Universitat Graz, Graz, Austria

R. S. Allen
Air Force Geophysics Laboratory
Hanscom AFB, Mass. 01731

D. E. Donatelli
Regis College Research Center
Weston, Mass.

G. K. Hartmann
Max-Plank-Institut fuer Aeronomie
D-3411 Katlenburg-Lindau 3, FGR

ABSTRACT

Precision navigation and radar systems use local measurements to adapt predictions of refractive effects within their field of view. The intent is to reduce the day to day variability of the refraction correction by tracking local features. To simulate a particular adaptive technique, data from a sequence of passes of the Navy Navigation Satellite System satellites are used to represent both the local measurements of refractive effects and the subsequent system range measurements. This is augmented by measurements of total electron content taken in the same area using VHF beacons on nearly synchronous satellites. Local features such as the nighttime trough and the plasmapause enhancement of the mid latitude ionosphere are tracked by sequential measurements and may be mapped in space and local time across the entire system field of view during quiet daytime geophysical periods. During quiet daytime periods systematic changes in the ionosphere are usefully predicted over a few hours. When the ionosphere is disturbed, very large features can develop in local regions that should only be introduced into the adapted maps if the grid of observations in space and time can resolve their behavior.

INTRODUCTION

The problem being addressed here is generic to modern navigation and radar systems. As their requirements for precision time and range measurements become more stringent, methods must be developed to remove the effects of refraction within the ionosphere. Previously Millman (1974) has shown that, from the first order correction for change in apparent range (ΔR) which is directly proportional to the slant electron content, a correction can be derived for the first order change in elevation angle. Leitinger and Hartman (1977) have shown that for both the range error that affects ranging systems (for navigation purposes, for determination of satellite orbits, etc.) and for the group delay (for dispersion correction of systems), the corrections based on the electron content integrated along the slant path is second order in fineness.

The principle correction, ΔR, may be modeled from existing climatologies of ionospheric parameters (Lewellyn and Bent, 1973), and provided to the system as maps covering the entire field of view.

Since the basic climatology of ionospheric

parameters is dependent only on smoothed solar conditions, these maps can be prepared several months in advance, and when used, will typically remove 50 to 75 percent of the monthly r.m.s. refractive error. For the many systems which require more precise correction, these maps may be adapted by pertinent local measurements to reflect the current solar-terrestrial conditions and major local ionospheric features (DuLong and Allen, 1976). Many local measurements have been proposed: vertical incidence ionosonde data, total electron content derived from beacons of near synchronous satellites, in situ particle measurements from low altitude satellites, etc. The residual refractive error after using any adaptation is dependent on the information within the local measurement and the density of measurements in space and time. These and similar points are discussed in companion presentations (Katz et al, DuLong et al, this Symposium). This working paper will examine some of the promise and problems found when making adapted predictions using slant electron content derived from observations of the Navy Navigation Satellite System (NNSS) satellites.

ADAPTIVE MODELING

To illustrate how refractive effects can be corrected in an adaptive manner, consider a hypothetical radar, figure 1, with a field of view encompassing the ionosphere over Europe. Observations of two radio beacons aboard the NNSS satellites were reduced to provide equivalent vertical electron content for latitude strips representing the satellite tracks within $\pm 8^{o}$ longtitude of the central stations, Graz, Austria and Lindau, FGR. These NNSS measurements were normalized by referencing to simultaneous electron content data reduced from beacons on the nearly geostationary satellite ATS-6, as observed from Lindau. By converting the NNSS data to radar range correction, the results can be used to adapt maps of median range correction and the reduction in refractive error can be examined over the field of view.

In this presentation we want to ask some basic questions about the usefulness of such an adaptive procedure. Can it predict:

1. enhanced ionospheric conditions?
2. changes in latitudinal gradients?
3. local features?

such that the residual error after adapting for these items is significantly better than the error resulting from use of the monthly median maps alone? We must also ask the question:

4. are the results model dependent?

In simulating a procedure for adaptive modeling, data from the first in a sequence of NNSS passes may be used to normalize the prediction of median range correction to conform to the actual conditions at that time. The scaling used along the latitudinal intervals for this pass would be used to adapt the median along the path at the time of the next NNSS pass. The first pass may be considered a representation of a local ionospheric measurement and the subsequent passes may represent system measurements, such as a radar system measuring the range to a target. The error, after adaptation, δR_a, can be compared to the error using just the median, δR_m, to determine the effectiveness of the adaptive modeling procedure. This may be repeated, using each pass to adapt the model and comparing the predictions of the adapted model at the time of subsequent NNSS passes with the actual observations.

The procedure is illustated in Figure 2 for a pass of NNSS satellite No.5 starting about 11:18 local standard time, 4 Dec 1975 at subionospheric point, 18E Longtitude, 61.5^{o}N Latitude and ending at 20.4^{o}E Lon., 41.5^{o}N Lat. The continous phase variation of the two NNSS radio beacons were reduced to equivalent vertical electron content at 0.5 degree of latitude increments from 41.5^{o}N to 61.5^{o}N. These were converted by multiplying with the constant 2.23 per 10^{16} el/m^2 to give vertical range correction in meters for a 425 Mhz radar. The prediction of the monthly median refraction correction is from a standard Air Force Global Weather Central program using the Bent formulation based on the ITS-78 coefficients. Near the point of closest approach to the ground station, Graz, Austria, at the center of the hypothetical radar coverage area, the ionospheric correction was about 23 percent lower than the predicted median. At the next recorded pass, 1403 L.S.T., the ionospheric correction was still below the predicted median. The technique for making an adapted prediction can now be considered in arbitrary increments.

1. Assume that the monthly median prediction can be scaled by reference to a single point, say the point corresponding to closest approach to the radar, which might have the greatest reliability. This is equivalent to normalizing to a single value derived from a local vertical incidence ionosponde, a measurement along the path to a near-synchronous satellite, etc. As can be seen from the relation between the line labeled "Adapted GWC" and the actual observations along the NNSS track, this may be model dependent.

2. Assume that the diurnal behavior can be defined by the model in a well behaved region, say mid latitudes (47^{o}N in this case) and that the unique daily latitude gradient persists from pass to pass. The latitude gradient could have been derived from in-situ measurements of topside penetration frequencies, from a network of vertical incidence sounders, etc. As shown by the line labeled "Adapted gradient" in figure 2, the gradient at 1403, 2 hrs. 40 minutes later was remarkably the same as that found earlier at 1118. This behavior persisted for the subsequent recorded passes at 1458 and 1610 LST. Note that our linear gradients have minimized the r.m.s. percentage error. Note, also, that there appears to be merit in considering any measurement in real time which could tie the previous

latitude gradient to the current condition to improve on the behavior predicted by the median model. For example, any simultaneous measurement of total electron content along the path could tie the two gradients together.

3. Assume that almost all of the information along the previous pass may be used to predict the current pass. Again this information could have come from sources other than NNSS measurements, such as in-situ data from the DMSP satellites normalized with a local measurement. As a test procedure, observations from the first of a pair of passes were compared to the GWC prediction along the satellite path and the resulting scaling factors were used to scale the GWC prediction at each corresponding point along the subsequent pass. The results of this procedure, labeled "Update", for the 4 December daytime passes contained in figure 3, show that it is ionospherically dependent. During the daytime, measurements consistently show local features of a characteristic ±5 percent amplitude and 3-degree latitude scale size that are effervescent from hour to hour.

A large local feature, which may often be tracked, is the trough at the edge of the auroral zone. At 0241 LST, 2 December 1975, the position of the minimum, determined from the NNSS pass (figure 4), was near 59°N Geographic. In less than two hours, at 0427 LST, the trough had moved southward by about 3 degrees. The minimum in the general trough region and enhancement in the plasmapause region of the standard prediction program was up to a factor of 2 higher than the measured values. The linear gradient, normalized to the expected diurnal behavior at 47°N, greatly reduced the residual error, even though it smoothed both the trough and the enhanced region under the plasmapause.

By using joint observations from Graz, Lindau and Uppsala, Sweden, these features may often be tracked over several days. In the late afternoon through early morning hours in the data for December 11-13, a quiet geomagnetic period, there are well-developed, persistent features, as illustrated by the sequence of figure 5. Near 2000 LST in the region just above 60 degrees latitude (L value between 4 and 6) there is a well developed trough in the ionosphere, a sharp latitude gradient toward the equator and a persistent enhancement close to the region associated with the plasmapause. These can be discerned as smoothed features in the GWC predictions, figure 4, but here they may be tracked throughout the late afternoon and nighttime hours and from day to day.

CONCLUSION

Actual observations made with the Navy Navigation Satellite System (NNSS) satellites have been used to test techniques which adapt maps predicting ionospheric refractive effects over the field of radar or navigation systems. In these techniques each adapted map is then used to predict the local behavior of the refraction correction and the result compared with ensuing NNSS passes.

Systematic changes which represent the daily response of the local ionosphere to solar-terrestrial conditions, such as the general enhancement seen over Europe on 4 December, 1975, are usefully predicted by this technique. A necessary restriction is that a prediction should not be carried across either sunlit terminator as this could increase error.

During quiet geomagnetic periods, those local features with large geophysical scale size and a consistent identification in climatology, such as the nighttime trough and the plasmapause enhancement, seen persistently in the 10-13 December TEC observations from Europe, may be profitably tracked by local measurements. During the daytime some local features may be as large in amplitude but so transitory in either space or time that they are not observed on subsequent passes. In this case it is better to smooth these unknown features prior to adapting the median maps.

ACKNOWLEDGEMENT

The authors would like to thank Dr. A. Hedberg, Ionospheric Observatory, Uppsala, Sweden, for NNSS observations at Uppsala.

REFERENCES

DuLong, D.D. and Allen, R. S. (1976) "Specification of Navigation and Radar Errors caused by the Ionosphere," COSPAR Symposium Proceedings, The Geophysical Use of Satellite Beacon Observations, Boston University.

Leitinger, R. and Hartmann, G.K., (1977) "Time and Latitude dependence of transionospheric propagation errors," Kleinheubacher Bericte, Bd 20, ed FTZ Darmstadt, FRG.

Llwewllyn, S.K. and Bent, R.B. (1973) "Documentation and Description of the Bent Ionospheric Model," AFCRL-TR-73-0657.

Millman, G.H. and Reinsmith, G.M., (1974) "An analysis of the Incoherent Scatter-Faraday Rotation Technique for Ionospheric Propagation Error Correction," General Electric CO., TIS R74EMH2.

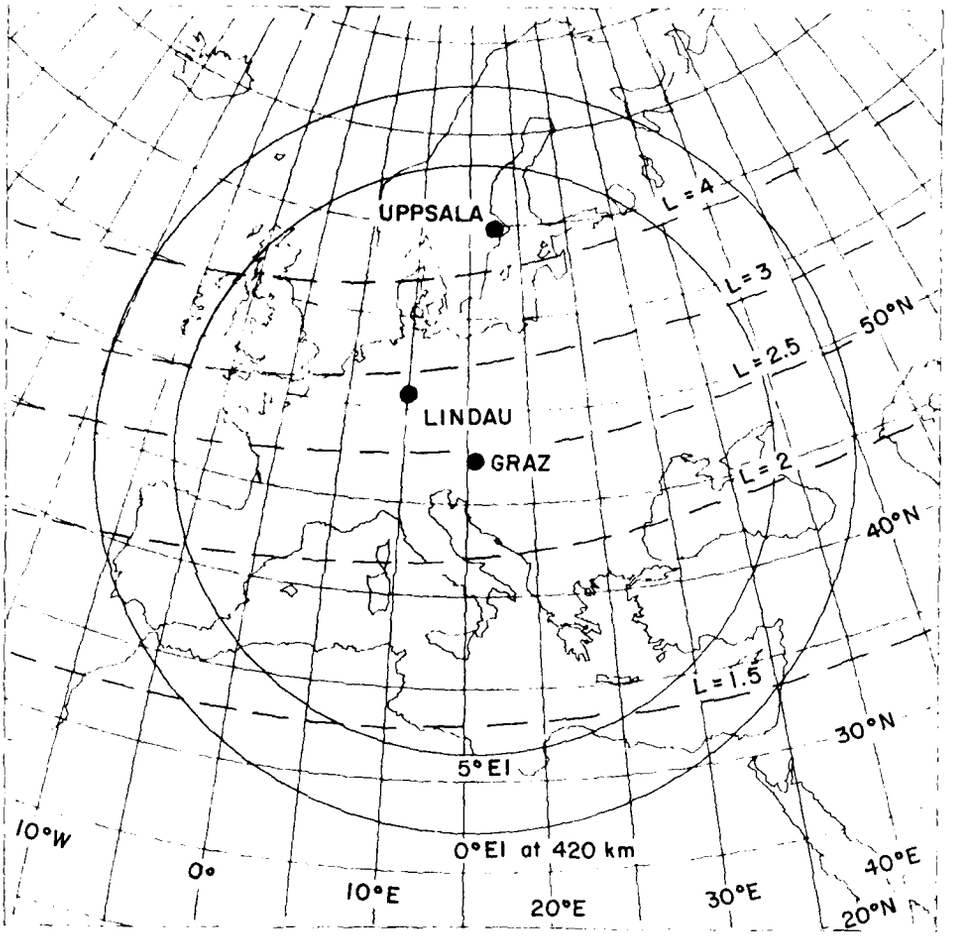


Figure 1. Coverage diagram of a hypothetical radar. Simultaneous measurements of total electron content using NNSS satellites are combined from three stations to represent radar measurements. Coverage extends into the auroral region of the ionosphere.

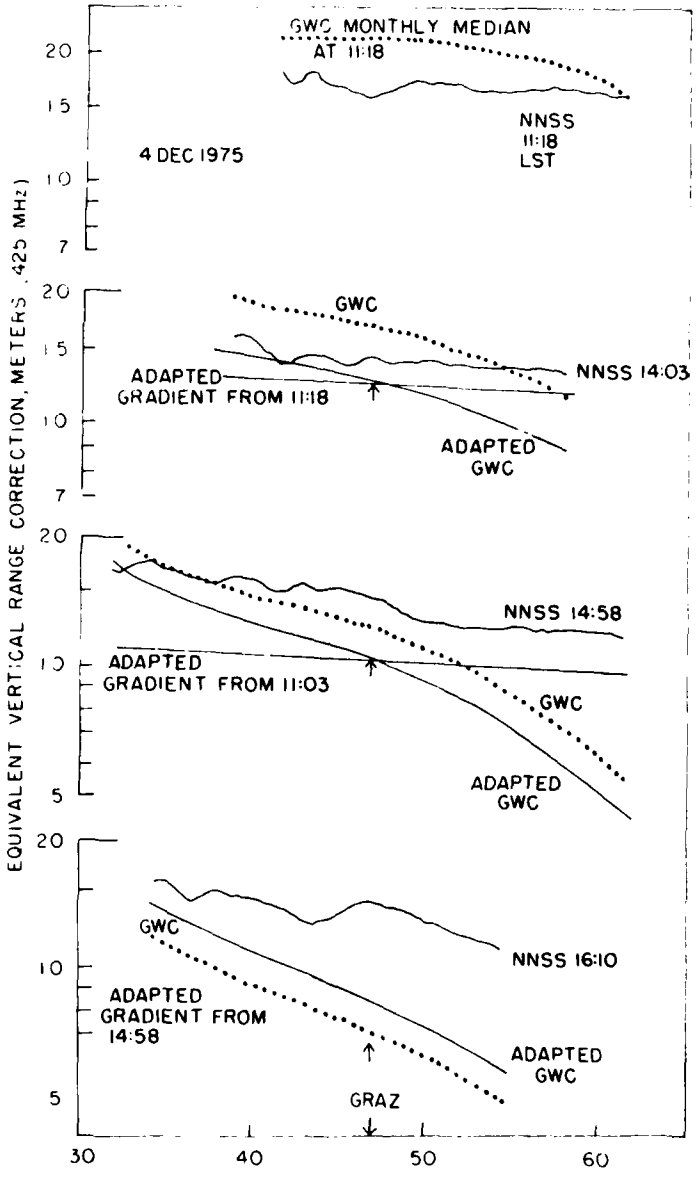


Figure 2. Illustration of adaptive modeling of radar range correction along a series of NNSS tracks. Each pass is predicted by the basic monthly median (GWC); by the GWC prediction scaled by a single observation from the previous pass (labeled Adapted GWC); and by the previously observed latitude gradient predicted forward in tiem at the point of closest approach (labeled Adapted Gradient). Note that general enhancement and simple gradients are usefully predicted.

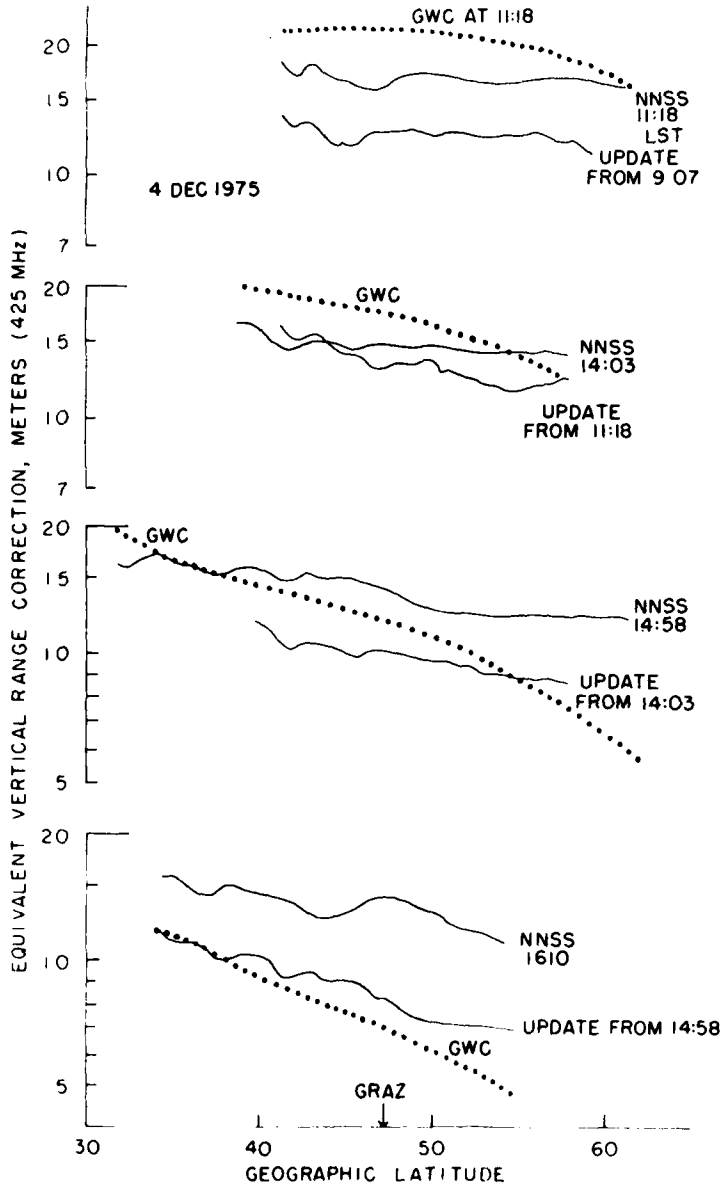


Figure 3. Illustration of adaptive modeling where each point of a previous pass is used to modify the latitude behaviour of the GWC prediction. Local features of 5 to 10 percent amplitude do not reproduce from pass to pass when the inter-pass period exceeds an hour.

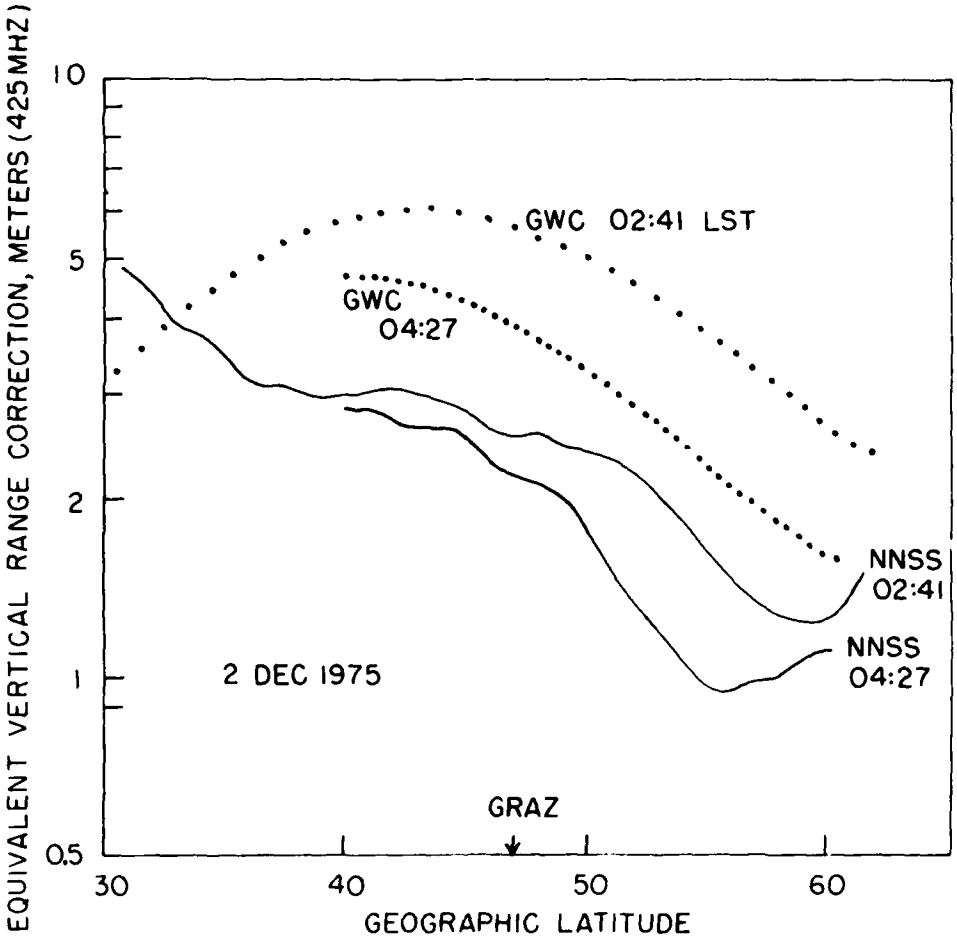


Figure 4. Sample taken 2 Dec 1975 which shows the ability to specify the location and movement of the mid-latitude trough and the plasmapause gradient.

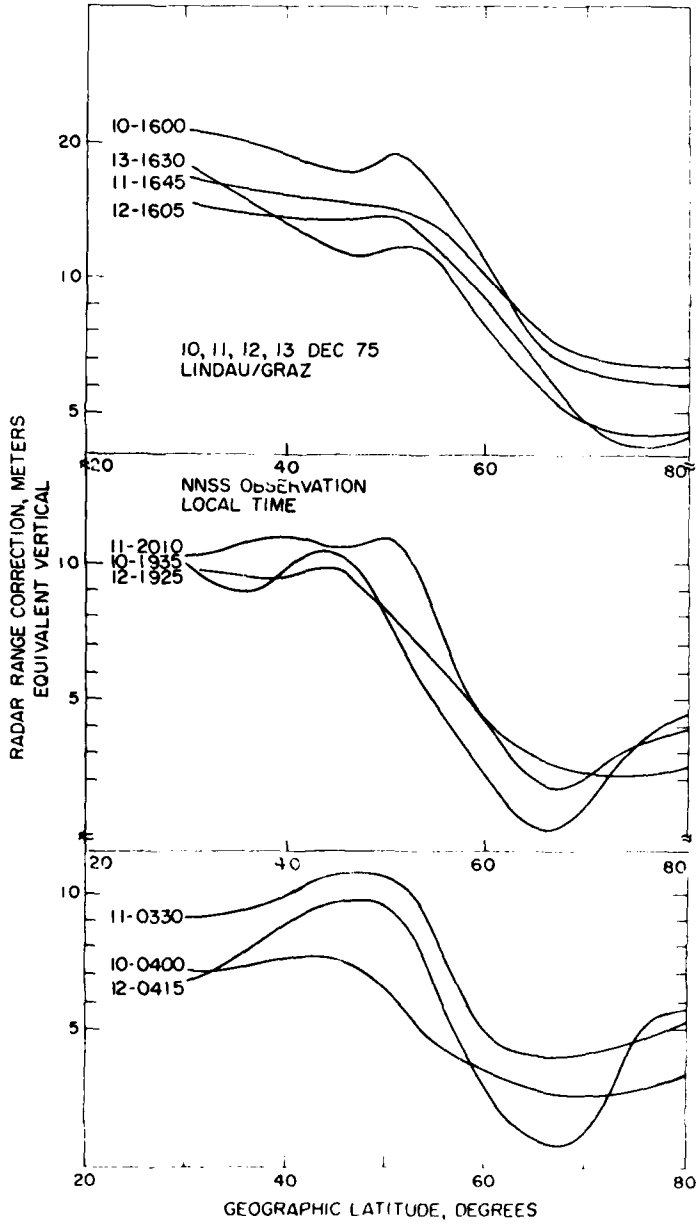


Figure 5. Consecutive days illustrating tracking of the trough, the plasmapause gradient and the plasmapause enhancement from hour to hour and from day to day. Adapted predicitons scaled from such observations could reduce the error of range correction by several meters.

EFFECTS OF IONOSPHERIC IRREGULARITIES ON SATELLITE DOPPLER SHIFT

P. L. Dyson

NAS/NRC Senior Resident Research Associate
NASA/Goddard Space Flight Center
Laboratory for Planetary Atmospheres
Greenbelt, Maryland 20771

J. A. Bennett

Monash University
Department of Electrical Engineering
Clayton, Australia 3168

ABSTRACT

The purpose of this paper is to discuss the effects of fairly large scale ionospheric irregularities on observations of Doppler shift of radio signals emitted from satellites. In this context "fairly large scale" means irregularities which have a scale size $\geq$ 1 km and which, at the frequencies of concern, do not give rise to multiple propagation paths between the satellite and the observer. For frequencies in the VHF band and above the refraction caused by the ionosphere may be regarded as a first order perturbation of the free space ray. A consequence is that the Doppler shift of a satellite signal passing through an ionosphere containing irregularities can be represented as the sum of contributions due to free space, the smooth background ionosphere and the irregularities. Thus it is possible to evaluate separately the contribution due to the irregularities, i.e. to determine the "signatures" which irregularities produce on Doppler records. Mathematically the approach involves the determination of the second variation of the phase path. The theoretical analysis shows that fluctuations in satellite Doppler records may be due to (a) non-stationarity of the ionosphere, (b) irregularities lying below the satellite height and near the straight line path between the satellite and observer, (c) irregularities at the satellite or (d) some combination of (a), (b) and (c).

The effects produced by gravity wave and field-aligned irregularity models have been calculated. The Doppler shift produced by an irregularity is quite sensitive to the orientation of the irregularity with respect to the line of sight from the observor to the satellite. In fact the absence of fluctuations in the Doppler shift does not mean the absence of irregularities.

In the case of gravity wave effects (T.I.D.'s) it is possible to determine from the Doppler signature, the wave length and orientation of the phase fronts of the irregularity.

INTRODUCTION

The purpose of this paper is to discuss the effects of fairly large scale ionospheric irregularities on observations of Doppler shifts of radio signals emitted from satellites. In the context of this paper "fairly large scale" means irregularities which have a scale size $\gtrsim$ 1 km and which, at the frequencies of concern, do not give rise to multiple propagation paths between the satellite and observer. There are a number of different types of irregularity which are relevant. These include the effects of gravity waves, field-aligned ducts and the larger scale irregularities associated with spread F which are also aligned along the earth's magnetic field. Irregularities can occur at a variety of heights. In a given situation the irregularities may be confined to a narrow height region or they may be present over an extended height range which includes the satellite height. Therefore, there is a wide variety of situations which could be considered. We will only consider two main types of irregularity here. The first, a gravity wave, which produces its major effect on the ionosphere near the peak of the ionosphere. The second type is a columnar field-aligned irregularity which could be indicative of either spread F irregularities or of the M.F. ducts observed by topside sounder satellites.

For the fairly large scale irregularities to be considered here, refraction rather than diffraction predominates so that the effects of these irregularities on a probing radio wave may be understood in terms of ray theory. The effect of the irregularities is to distort the rays and the consequent ionospheric contribution to the Doppler shift will be a perturbation from the value it would have for just the smooth background ionization distribution. Such perturbations to an otherwise smooth background Doppler frequency time curve are apparently quite common (e.g. De Medonça, 1962; Al'pert, 1964; Al'pert et al., 1971; Tucker 1967). If the explanation based on refraction is accepted, then the observation of these Doppler fluctuations indicates that such irregularities are present. However it is not clear at what height the irregularities occur. For example, de Mendonça (1962) explained such a fluctuation as due to an irregularity assumed to be located near the F_2 peak whereas Al'pert and co-workers (Al'pert 1964, Al'pert et al. 1965, Al'pert and Sinelinkov 1969, Al'pert et al. 1971, Tsedilina and Kharybina, 1964) and others have assumed such fluctuations are due to irregularities at or immediately below the satellite trajectory.

In this paper we obtain an expression for the Doppler contribution due to the irregularity. The expression shows that when interpreting experimental records, it is not possible, in general, to remove ambiguities such as those described above unless additional measurements are made. For anisotropic irregularities, the Doppler perturbation is found to be very sensitive to the orientation of the irregularity with respect to the unperturbed ray. Indeed the absence of fluctuations cannot be taken as a positive indication that irregularities are absent. Similar conclusions were reached earlier by de Mendonça (1962) using less detailed arguments.

THEORY

A general variational theory useful in the study of ray propagation (Bennett 1969; 1973) allows us to obtain quite easily an expression for the contribution to the Doppler shift due to irregularities. The basis of the method is that the phase path may be expanded in a multiple Taylor-McLaurin series and the coefficients evaluated assuming that for all combinations of values of the variables the ray continues to exist. It is thus essentially an imbedding technique. The calculations are greatly simplified because the ray satisfies a variational equation (Fermat's Principle). Specifically we adopt a quasi-stationary approach to time variations and assume that we can regard the refraction caused by the background ionosphere to be a first order perturbation of the free space ray. We may then write

$$P \sim P_o + (\delta_m P + \delta_r P) + 1/2\ (\delta_m^2 P + 2\delta_m \delta_r P + \delta_r^2 P) \qquad (1)$$

where P is the phase path. P_o is the value of the phase path in the absence of the refracting ionosphere. $\delta_m P$, the first m-variation of P represents the first order (linear) contribution due to the smooth background ionosphere, and $\delta_r P$ represents the corresponding contribution due to the irregularity. It is assumed that all quantities are functions of time. (For a discussion of the limitations of the quasi-stationary approach to the calculation of Doppler shifts, see Bennett (1976). In the present context this approximation is very good. At VHF frequencies and above the ionospheric refraction is small and treating it as a first order perturbation is also a very good approximation).

Making use of the Doppler shift formula,

$$\Delta f = -\frac{f}{c}\frac{dP}{dt} \qquad (2)$$

we therefore find that

$$\Delta f \simeq -\frac{f}{c}\left(\frac{dP_o}{dt} + \frac{d}{dt}\delta_m P + \frac{d}{dt}\delta_r P + \ldots.\right) \qquad (3)$$

In (3) the Doppler shift is represented as the sum of the free space contribution and the first order contributions of the smooth ionosphere and the irregularity. The second and third terms of (3) (which are essentially of the same form) may conveniently be evaluated using the general formula for the second variation of P (Bennett, 1969; 1973). Since we are interested primarily in irregularities we will generally only consider the third term.

Suppose we have a co-ordinate system as shown in Figure 1, where a fixed observer is at A and a satellite is at B at height H above the horizontal plane and moving with velocity $\underline{V} = (V_1, V_2, V_3)$. For convenience the 1-2 co-ordinate plane has been taken to be the plane containing the satellite and observor. Assuming that the frequencies of interest are sufficiently high so that the effect of the earth's magnetic field can be neglected, the third term in equation (3) can be evaluated to give

$$\frac{d}{dt}\delta_r P = \left\{\frac{H}{C}\left\langle\frac{\delta\mu_r}{\partial t}\right\rangle + (V_2 - \frac{S}{C}V_1)\left(\langle G_{2r}\rangle/C + \langle\mu_r\rangle S\right) + V_3\langle G_{3r}\rangle/C + V_1\mu_r|_B/C\right\}\delta r \qquad (4)$$

μ is the refractive index and subscripts of μ indicate the derivative with respect to the subscript.

$C = \cos X_o$ and $S = \sin X_o$ where X_o is the zenith angle of the satellite.

The symbols $\langle\ \rangle$ denote the average along the straight line AB and

$$\langle G_{\alpha r}\rangle\delta r = \frac{1}{H}\int_o^H h\mu_{rq^\alpha}\,\delta r'h \qquad \alpha = 2,3 \qquad (5)$$

and $q = (q^1, q^2, q^3)$ is the position vector of a point in space.

We may choose r so that

$$\mu_r \delta r = -K\, \delta_r N/2f^2 \qquad (6)$$

where N is the electron density

f is the frequency

and $K = e^2/4\pi^2 m\epsilon_o$ where e and m are the electronic charge and mass respectively.

δr itself is just a convenient label in performing the expansion. For example we might take $\delta r = \delta_r N_{max}$. Corresponding results apply also to $\frac{d}{dt}\,\delta_m P$.

The second and third terms of (3) are both essentially of the same form as that obtained by Al'pert (1964) although each refers to only one part of the ionospheric electron distribution whereas Al'pert's result refers to the total distribution. For some purposes it is more convenient to express the results in a form like that obtained by Pavlenko (1966). Rotating the axes in Figure 1 so that C=1 and S=0 (4) becomes

$$\frac{d}{dt}\,\delta_r P = \left\{ L < \frac{\partial \mu_r}{\partial t} > + V_{\perp c} < G_{\perp cr} > + V_{\perp N} \cdot \right.$$

$$\left. < G_{\perp Nr} > \; + V_{\parallel}\, \mu_{r|B} \right\} \delta r \qquad (7)$$

where L is the length of the straight line AB and the subscripts $\parallel$, $\perp c$, and $\perp N$ indicate components parallel to AB, orthogonal to AB but in the plane AB, and orthogonal to both AB and the plane AB, respectively.

Examination of either (4) or (7) shows that ionospheric irregularities may produce fluctuations in a satellite Doppler record in several ways, viz., (a) non-stationarity of the irregularities (first term in (7)), (b) irregularities lying near the straight line AB and below the satellite height, (second and third terms in (7)), (c) irregularities at the satellite height which change the local refractive index (fourth term in (7)); or (d) some combination of (a), (b) and (c).

Making use of (6), we find the contribution (a) is of the form

$$\Delta f = \frac{KL}{2fc} < \frac{1}{\mu}\, \frac{\partial}{\partial t}\, (\delta_r N) >$$

Doppler fluctuations of only a few Hertz are observed with radio waves of a few megahertz reflected by the ionosphere (Davies and Baker 1965). The presence of the 1/f and $1/\mu$ factors shows Δf falls of with f more rapidly than 1/f. Thus unless rapid rates of change exist in the topside ionosphere the contribution of (a) will be negligible.

If situation (b) exists then the Doppler perturbation is due to the electron density gradients perpendicular to AB. Gradients at all heights can make significant contributions but, of course, certain irregularities will produce the largest perpendicular gradients at particular heights (e.g. gravity waves) and it will be in these regions that the largest contributions to the variation in Doppler shift will be made.

It is not always convenient to use a reference frame in which one axis is coincident with the line of sight of the satellite, and so we return to (4) in our discussion. Inspection of (4) shows that case (b) gives rise to terms involving the horizontal gradients (the $< G_{2r} >$ and $< G_{3r} >$ terms) and a term related to the change in total electron content (the $< \ldots_r >$ term). It is most likely that it is the gradient terms rather than the total electron content term which give the predominant contribution to the variation in Doppler shift. To see this consider an irregularity with characteristics dimensions ℓ_1, ℓ_2, ℓ_3 in the co-ordinate directions and located at a height h. An order of magnitude calculation shows that

$$\frac{d}{dt}\,\delta_r P \simeq \left\{ \frac{1}{c}\,\frac{\ell_1}{T} + (V_2 - \frac{S}{C}\,V_1)\,(\frac{\ell_1}{H}\,\frac{h}{\ell_2} + \frac{\ell_1}{H}) + \right.$$

$$\left. V_3\,\frac{\ell_1}{H}\,\frac{h}{\ell_3} \right\} \left\{ \frac{-K}{2f^2}\; \delta_r N_{max} \right\} \qquad (8)$$

where T is the time scale of the irregularity and $\delta_r N_{max}$ is the maximum deviation from the background concentration. (We assume that the irregularity does not extend to the satellite height). The coefficients representing the contribution due to horizontal gradients are respectively $\frac{h}{\ell_2}$ and $\frac{h}{\ell_3}$ times that representing the contribution due to the change in integrated electron content . These factors are usually large numbers so that the gradient terms will predominate. It is a consequence of the limited scale of the irregularities that these gradient terms are even more important here than are the corresponding terms which influence the determination of the smooth contributions $<\delta_m N>$ and $\delta_m N|B$ (Ross et al. 1968).

For case (c), (7) shows that the effect depends upon the change in electron density occurring at the satellite and the component of satellite velocity along the line of sight. Hence for a satellite moving horizontally, there will be no contribution to the variation in Doppler shift when the satellite is overhead. Thus it is most likely that irregularity gradients below and up to the satellite are the sole cause of variations in Doppler shift when the satellite is near the zenith. Clearly there is most ambiguity in determining the source of Doppler shift variations when the satellite is at low elevation angles.

It has already been mentioned that irregularities may be present at the satellite height but produce a zero or negligible change in the Doppler shift. Similar situations can also arise for case (b). Consider the gradient terms in equation (7). If an irregularity exists below the satellite but does not produce any gradients perpendicular to the line of sight it will not produce a variation in the Doppler shift. Such a situation will be approximated when the line of sight propagation from the satellite is perpendicular to the wave fronts of a gravity wave. Alternatively, since it is the average gradients which are important, these may be near zero if there are increases and decreases in electron density along the line of sight which produce both positive and negative gradients. For a gravity wave, this situation will occur approximately when the variations in electron density produced along the line of sight to the satellite have an equal number of positive and negative deviations. While the examples given here indicate that the variation in Doppler shift produced by irregularities can go to zero in specific instances, it is likely that irregularities will produce a negligible Doppler shift under a variety of conditions and this point is illustrated in the next section.

Finally, it may be noted from (7) that if the vehicle velocity is along the line of sight, then, provided time changes of the irregularities are small, only the irregularities at the vehicle height produce a Doppler shift. This would only be applicable to satellites in particular instances when a satellite is in a very eccentric orbit. However, it would be often applicable to rocket vehicles.

DOPPLER FLUCTUATIONS PRODUCED BY IRREGULARITIES

In this section representative Doppler fluctuations due to different fairly large scale irregularities are calculated.

Internal Gravity Waves

We adopt the model used by George and Stephenson (1969) and write

$$N_r \delta r = W\,(N_m \delta m)$$

where $N_m \delta m$ is the smooth background electron density distribution and

$$w = \exp\left\{-(q^1 - h_o)/H_s\right\}^2 \cos\left\{2\pi\left[\frac{q^1}{\Lambda_1} - \frac{q^2}{\Lambda_2}\right]\right\} \delta r \quad (9)$$

where h_o is the height of the maximum percentage wave amplitude, H_s is the scale height of the wave and Λ_1 and Λ_2 are the vertical and horizontal wavelengths of the wave. Now

$$\mu_m \delta m = -K \delta_m N / 2f^2$$

and

$$\mu_r \delta r = K \delta_r N / 2f^2$$

where K is as defined before.

In Figure 2, the background Doppler shift $\delta_m(\Delta f)$ and the contribution due to the irregularity, $\delta_r(\Delta f)$ are plotted for a satellite moving horizontally at an altitude of 1000 km. (Curvature of the earth and ionosphere have been neglected). The background ionosphere was a Chapman layer of critical frequency 6.5 MHz, a scale height of 62 km, and peak height of 300 km. The parameters of the wave were H_s = 100 km, h_o = 250 km, $\Lambda_1 = \Lambda_2$ = 100 km. The contribution $\delta_m^o(\Delta f)$ is an antisymmetric function of X_A but $\delta_r^m(\Delta f)$ is negligible for X_A negative. It can be seen that $\delta_r(\Delta f)$ has the greatest magnitude when the zero order ray is at sin X = 0.74, i.e., when it is almost aligned with irregularity phase fronts (sin X = 0.707).

It is also apparent that for a wide range of zenith angles (including all negative values) the irregularity causes a negligible change in the Doppler shift. Thus, in the experimental situation, there is no reason to suppose an irregularity was concentrated at a particular place in the ionosphere simply because fluctuations are only observed when the ray passes through that region. This is understandable in the light of the discussions of the previous section.

Figure 3 shows the satellite trajectory and the sloping phase fronts of the gravity wave irregularity. The observer is at A. Also displayed are contours of constant electron density which show the effect of the irregularity. Along the phase fronts marked "+" the electron density is increased, along the alternate fronts it is decreased. Equidistanct between these fronts the electron density has its maximum gradient. The gradient is essentially normal to the phase front and alternates in sign at successive maxima and minima.

If AB coincides with a line of maximum or minimum ΔN there is no effect on the Doppler shift because the gradients perpendicular to the line of sight are zero. Referring to Figure 2, the Doppler shift is zero near sin X = 0.70, indicating that this is approximately the situation for the model used. If, however, AB is parallel with the phase fronts and lies on a line along which ΔN is zero, the gradients are the largest and the effect on the Doppler will be greatest.

If the satellite is at B_3 or B_4 the lines of sight intersect the height at which the irregularity is a maximum (near 300 km) at points where the gradient of electron density is a maximum. Moreover, in each case the line of sight is nearly aligned with the phase fronts so that the contributions from heights greater and less than 300 km will combine constructively. We would therefore expect that the Doppler shift would be greatest when the satellite is

near these points. Because the gradients have different signs for the unperturbed rays to B_3 and B_4, one ray will correspond to a maximum in a plot of Doppler shift versus satellite position and the other will correspond to a minimum. From (3) and (7) it is apparent that positive and negative average gradients will produce negative and positive Doppler shifts respectively. Thus B_3 corresponds to a maximum in the Doppler shift and B_4 to a minimum. It can be seen that the perturbations to the Doppler shift will alternate in sign at successive points B_1, B_2, ... At the intermediate positions where the gradient is zero the Doppler perturbation is zero. As the line of sight makes larger and larger angles with the phase fronts, further maxima and minima can be expected but they will be diminished in amplitude partly because the magnitude of the gradient perpendicular to the line of sight will be smaller and also because there is a tendency for the contributions of positive and negative gradients to cancel.

Returning now to Figure 2, we see that the Doppler signature is asymmetric for the model irregularity. If we plot the amplitude of the maxima and minima as a function of sin X, then in the light of our discussion of figure 3, the alignment of the irregularity will be at an angle somewhere between the two largest peaks. By drawing the envelope to the points in Figure 2, the alignment can be estimated to occur where sin X = 0.73, i.e., X ≈ 46°53'. Now the wave vector of the gravity wave is perpendicular to this direction and of the angles of adjacent maxima, minima and zeros are determined from the Doppler signature, the wavelength of the irregularity can be estimated using the geometry shown in figure 4. Using the largest maximum and minimum a value of 74 km is found. If the two zeroes at sin X = 0.64 and 0.76 are used, a value of 75 km obtains. From the model, the wavelength is $100/\sqrt{2}$ or 71 km approximately. In the calculations a height of 300 km was used for h_{max}. The result is proportional to h_{max} so that altering h_{max} by 50 km would alter λ by ± 12.5 km. Thus simply by assuming that the irregularity Doppler signature is caused by a gravity wave having its major effect near the F_2 peak it is possible to make a reasonable determination of the orientation of the phase fronts and the wavelength of the irregularity.

Columnar Field-Aligned Irregularities

In order to gain some insight consider the case where the irregularity cross-section, in the plane given by the unperturbed ray, is a two dimensional Gaussian function, ie.

$$N_r\delta r = \delta r \exp\left\{-\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)\right\} \qquad (10)$$

This is a useful model because it does not only represent the case when the plane given by the satellite movement intersects orthogonally an irregularity of elliptical section. It also represents the situation in which the intersection is oblique, provided one of the major axes of the column cross-section lies in the plane. (In particular it applies to oblique intersection with an irregularity of circular section.) The intersection of a plane and an elliptical cylinder is not an ellipse unless the latter condition is satisfied. Since the assumption of ellipticity is a matter of convenience only, the results may be taken as an indication of the effect of columnar irregularities and the dependence of the affect on irregularity - path geometry.

Assuming the irregularity is of negligible size in comparison to the satellite - observer distance, the expression may be evaluated explicitly. The result is expressed in terms of the distance X from the maximum of the irregularity to the unperturbed ray, Figure 5. Normalizing to a=1 (i.e. X in (12) represents X/a) we have

$$\frac{H}{h_m}\,\frac{c}{V_\perp c}\,\frac{f\Delta f}{k\delta r} = \frac{\sqrt{\pi}\,\epsilon^{1/2}X}{(\epsilon\mathcal{S}^2+\mathcal{C}^2/\epsilon)^{3/2}} \exp\left\{\frac{-X^2\epsilon}{\epsilon\mathcal{S}^2+\mathcal{C}^2/\epsilon}\right\} \qquad (12)$$

where ϵ is the eccentricity of the ellipse of constant $N_r\delta r$ and $\mathcal{C} = \cos\theta$ and $\mathcal{S} = \sin\theta$ where θ is the angle between the unperturbed ray and the major axis of the ellipse. H is the satellite height and h_m the height at which the maximum of the irregularity is intersected by the unperturbed ray.

The expression (12) is plotted in Figure 6 for several values of ϵ and θ. Again it is clear how sensitive the Doppler shift is to the geometry.

CONCLUSION

For the relatively large irregularities of small amplitude it is possible to calculate the effect of the irregularity on the Doppler shift frequency received from a satellite by using a variational approach to geometrical optics. "Relatively large" usually means of a scale of several kilometers or more. The type of satellite Doppler fluctuations to be expected have been calculated for representative irregularities. Such fluctuations in satellite Doppler are frequently observed and they are an indication that irregularities are present. Unfortunately, because of the sensitivity of the effect to the irregularity line of sight geometry, the absence of fluctuations does not mean irregularities are absent.

In the case of gravity wave effects it is possible to estimate the orientation of the phase fronts and the wavelength.

ACKNOWLEDGEMENTS

This work has been supported in part by the Australian Radio Research Board and the Australian Research Grants Committee. One of us (P.L.D.) is grateful to La Trobe University for leave.

REFERENCES

Al'pert Ja. L., 1964, Geomag. and Aeronomy, IV, 505.

Al'pert Ja. L., Vitshas L. N. and Sinelnikov V. M., 1965, Geomag. and Aeronomy V, 502.

Al'pert Ja. L. and Sinelnikov V. M., 1960, Geomag. and Aeronomy IX, 331.

Al'pert Ja. L., Sinkelnikov V. M., Vitshas L. N. and Krajushkina, 1971, J. Atmos. Terr. Phys. 33, 1779.

Bennett J. A., 1969, Radio Sci. 4, 667.

Bennett J. A., 1973, Radio Sci., 8, 737.

Bennett J. A., 1976, Radio Sci. 11, 621.

Davies K. and Baker D. M., 1965, Radio Sci. 1, 545.

De Medonça F., 1962, J. Geophys. Res. 67, 2315.

Georges T. M. and Stephenson J. J., 1969, Radio Sci. 4, 679.

Pavlenko Yu. G., 1966, Geomag. and Aeronomy VI, 609.

Ross W. J., Garriott F., De Mendonca F. and Da Rosa A. V., 1968, J. Geophys. Res., 73, 1102.

Tsedinlina Ye. Ye. and Kharybina A. A., 1964, Geomag. and Aeronomy IV, 399.

Tucker A. J., 1967, DRL Report No. 551, University of Texas, Austin.

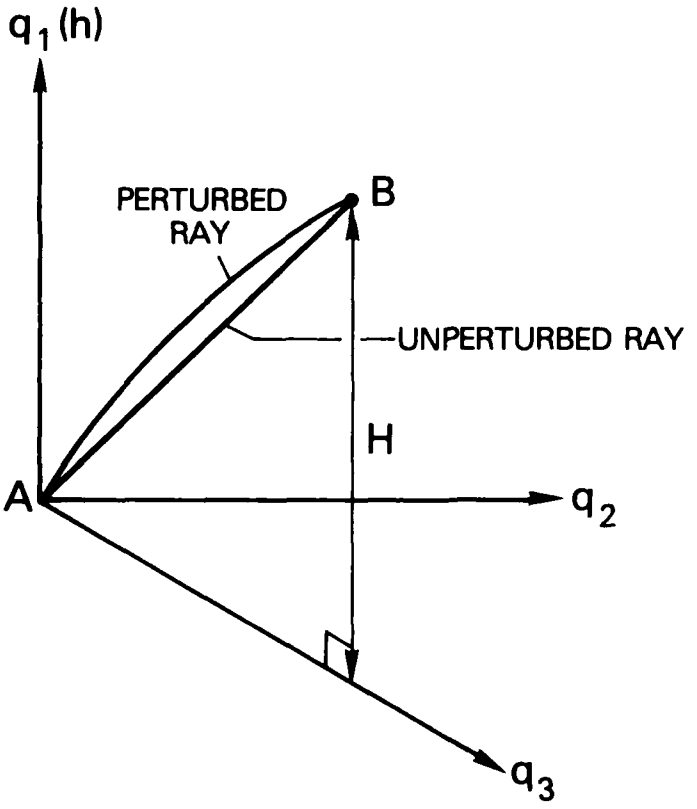


Figure 1. Co-ordinate system for observer at A and satellite at B.

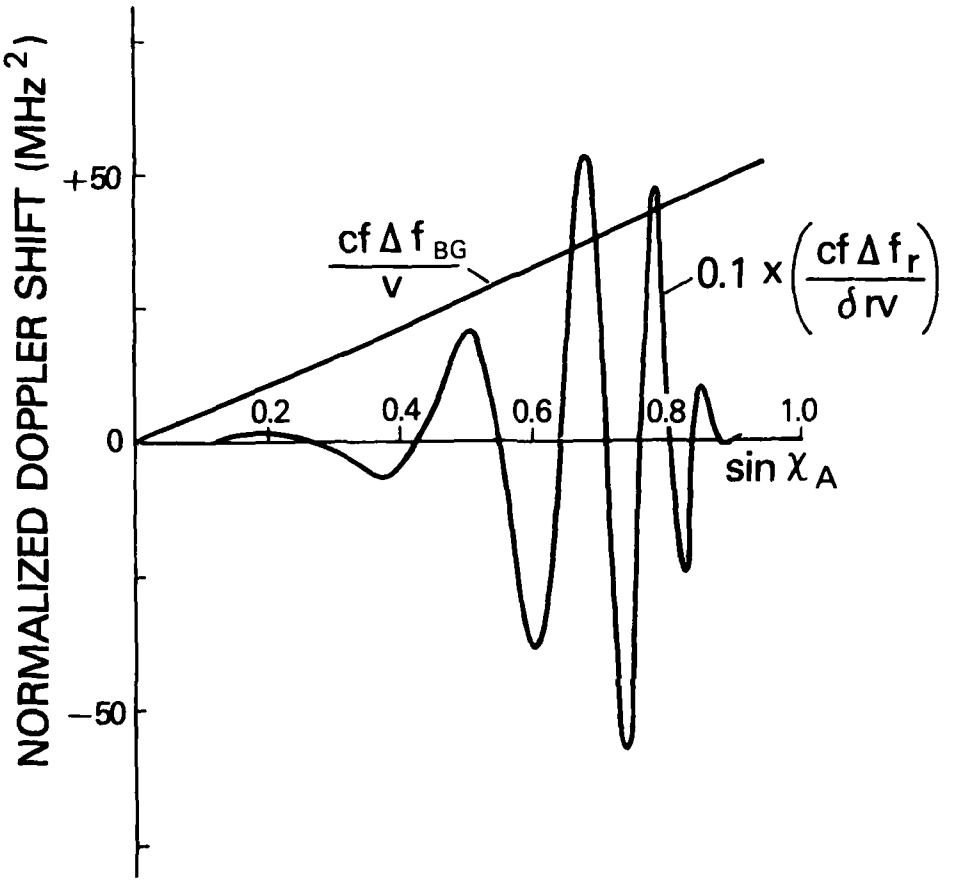


Figure 2. The Doppler frequency shift due to a simulated gravity wave and the Doppler contribution of the smooth background ionosphere. For this model $\Lambda_1 = \Lambda_2 = 100$ km, $H_s = 100$ km and $h_o = 250$ km. The calculations were performed for a frequency of 50 MHz but are normalized so that the curves are substantially independent of frequency. (At 50 MHz, with a satellite velocity of 7 km sec^{-1}, the maximum background Doppler contributions is about 2 Hz.)

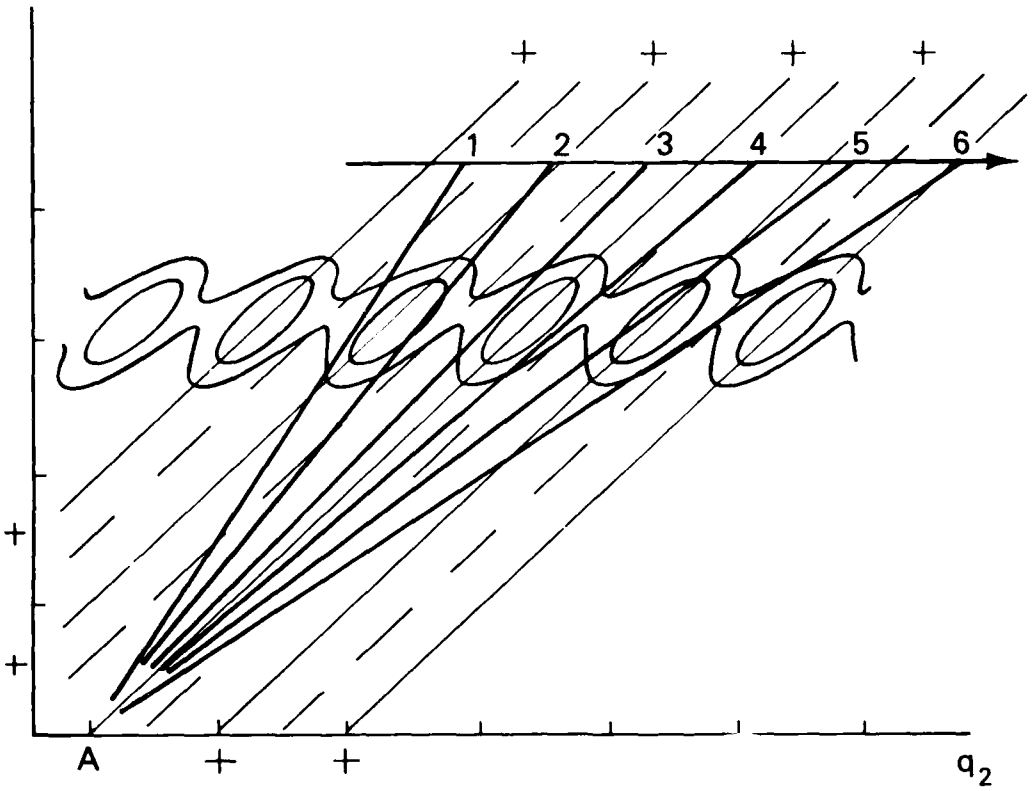


Figure 3. Schematic diagram showing phase fronts of gravity wave, unperturbed ray paths for several satellite positions and electron density contours near F_2 peak.

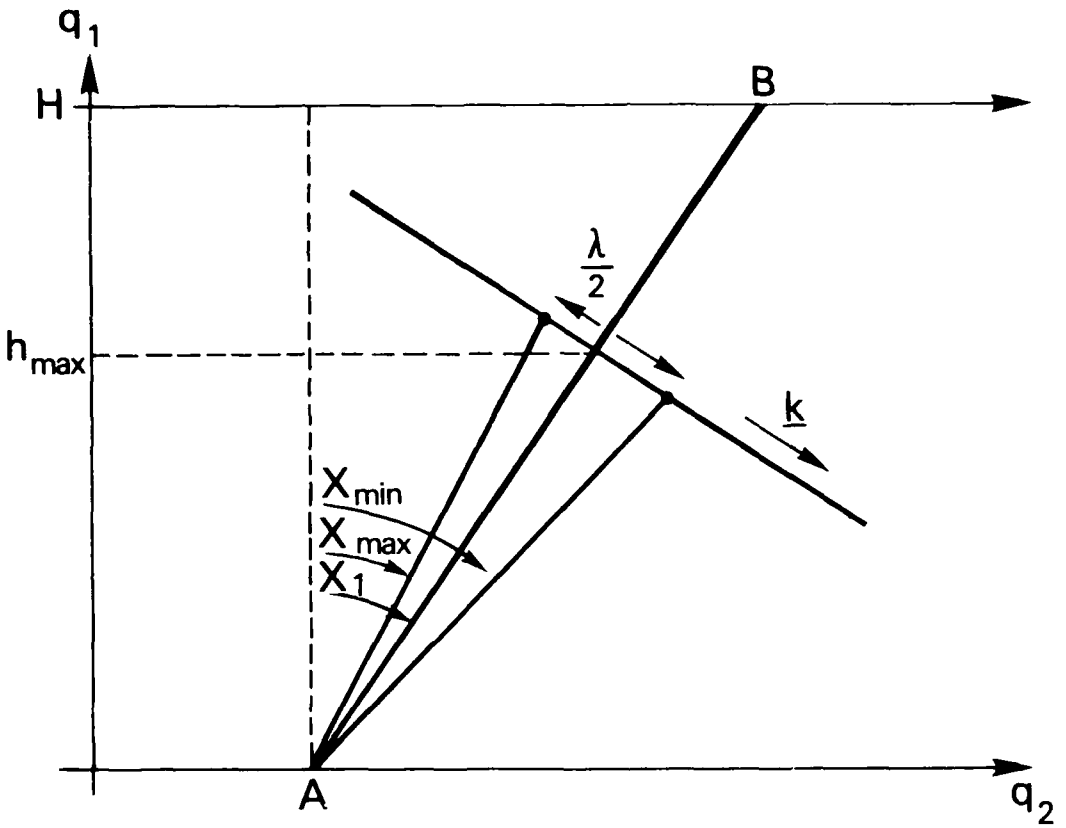


Figure 4. Simple geometry showing relationship of gravity wave wave vector to line of sight when satellite is aligned with phase fronts. X_{max} and X_{min} are angles at which largest positive and negative Doppler shifts are observed.

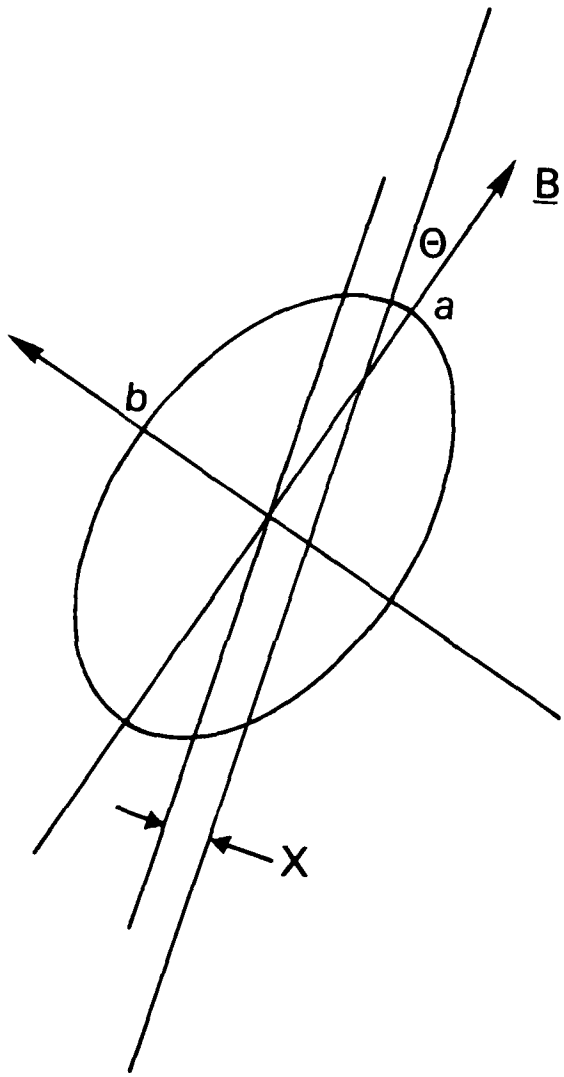


Figure 5. Model for columnar irregularity.

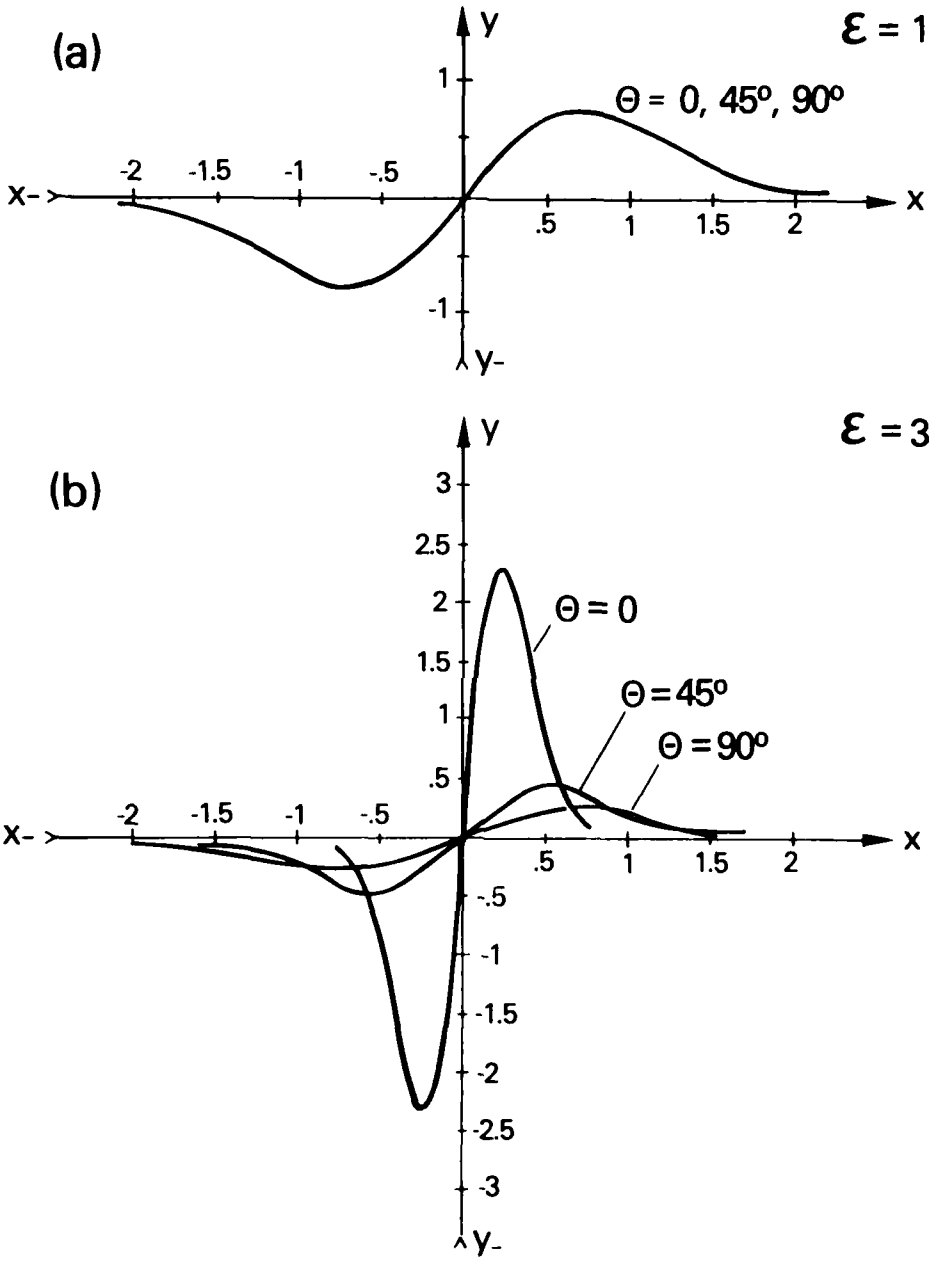


Figure 6. Doppler shifts produced by columnar irregularities.

A Proposed Experiment to Measure the Refractive and Coherence Properties of GHz Radio Wave Propagation in Turbulent Ionospheres Using Global Positioning Satellites

Y. T. Chiu and F. A. Morse
Space Sciences Laboratory
The Ivan A. Getting Laboratories
The Aerospace Corporation
El Segundo, Calif.

ABSTRACT

In this paper we wish to point out that the signal structure and the frequency coverage (UHF and L-band) of the satellites of the NAVSTAR Global Positioning System (GPS) afford them to be excellent scientific instruments for the study of the strong ionospheric turbulence. The GPS satellites carry highly accurate clocks which allow absolute wave phase measurements to be made. The availability of the absolute wave phase not only eliminates the difficulties of phase comparisons with a high frequency reference, which may itself be scintillated in strong turbulence conditions, but also allows, for the first time, the statistical refractive properties of turbulent ionospheres to be measured to accuracies limited only by tropospheric conditions. In the multiple scatter regime, the refractive properties of wave propagation (ionospheric range delay) is no longer linearly related to density fluctuations since ray path fluctuations also contribute; therefore, the advantage of having an accurate time standard in the GPS affords it to be the unique instrument for studying multiple scatter in turbulent ionospheres.

I. INTRODUCTION

It has long been recognized that natural and man-made turbulence in the ionosphere impose significant limitations upon satellite communication, navigation and remote-sensing systems, as well as upon radio and far-infrared astronomical observations. Ironically, the physical characterization of such an important geophysical condition is as yet sketchy. In particular, the physical relationship between ionospheric spread-F conditions detected by incoherent-backscatter radar (Woodman and La Hoz, 1976), ionospheric density irregularities detected by rocket-borne instruments (Morse et al., 1977; Kelley et al., 1976) and scintillation of GHz radio wave signals (Taur, 1976) has only recently come to the forefront of ionospheric research. These observational advances are in stride with recent advances in the theory of electromagnetic wave multiple scattering in random media (Gurvich and Tartarskii, 1975; Liu and Yeh, 1975; Ishimaru and Hong, 1975) on the one hand, and with recent consideration of the process of ionospheric spread-F irregularity generation (e.g. Hudson and Kennel, 1975) on the other. In any case, a major component of the scientific information to be obtained, whether by observational or theoretical means, is the spectral distribution of the ionospheric turbulence, although, because of the primitive state of theories, the observations of the irregularity spectrum also serve as stringent tests upon both wave propagation and irregularity generation theories.

The direct measurement of ionospheric irregularities by either radar observations or vehicle-borne instruments are frequently limited by the instrument response or telemetry rate available which does not allow a complete resolution of the irregularity spectrum. Therefore, the major share of information upon ionospheric turbulence remains to be obtained in the near future from wave propagation studies. It is only very recently that VHF and UHF propagation studies specifically designed for observation of ionospheric turbulence are made possible with the WIDEBAND satellite (Fremouw et al., 1977). The WIDEBAND satellite experiment consists of ground-based observation of signal propagation characteristics over a band of CW frequencies ranging from VHF to S-band transmitted by satellite-borne beacons. The WIDEBAND satellite experiment has yielded a wealth of information concerning the relatively weak scintillation phenomenon obtained in VHF and UHF at mid-latitudes, and has provided data on some unexpected features of the strong equatorial scintillation. Phase measurements relative to S-band propagation have been made. The deep-layered equatorial irregularities and the associated multiple scatterings at times

cause complete saturation of scintillation in the VHF and UHF channels and cause phase scintillations even at S-band. However, since WIDEBAND observations do not include ranging, absolute phase measurements cannot be made.

In this paper we wish to point out that the signal structure and the frequency coverage (UHF and L-band) of the satellites of the NAVSTAR Global Positioning System (GPS) afford them to be excellent scientific instruments for the study of the strong ionospheric turbulence. Furthermore, the GPS satellites carry precise time standards which allow absolute wave phase measurements to be made. The availability of the absolute wave phase not only eliminates the difficulties of phase comparisons with a high frequency reference, which may itself be scintillated (Taur, 1976), but also allows, for the first time, the refractive properties of turbulent ionospheres to be measured to accuracies limited only by tropospheric conditions. This new information will provide insight into the theory of multiple scattering in random media. Perhaps best of all, these scientific efforts, both observational and theoretical, can be carried out at minimal cost because we can utilize existing instrumentation already in orbit.

II. RATIONALE OF THE PROPOSED EXPERIMENT

The objectives of the experiment are: (a) to make absolute phase measurements in signal transmissions through turbulent ionospheres, and (b) to analyze amplitude and phase coherence of the signal in these trans-ionospheric propagations. Hereafter we shall refer to (a) as measurements of the refractive properties of transmission in turbulent ionospheres; whereas (b) will be referred to as measurements of coherence properties.

The measurement and analysis of coherence properties in signal transmission through turbulent ionospheres have been treated by numerous authors (e.g. Fremouw et al., 1977; Zirm and Goodman, 1977), and we shall not consider these topics here.

The refractive properties of the turbulent ionosphere can be measured because the GPS system yields very precise measurements of time, and therefore range, at the frequencies used. If a user at vector location $\vec{x}$ acquires a GPS satellite at vector location $\vec{s}$, then the user measured "pseudo-range" in frequency channel i (measured signal delay in channel i multiplied by the speed of light), defined as m_i, is given by

$$m_i = d + n_i + \int_{\vec{x}}^{\vec{s}} \mu_i \, d\ell_i \quad , \qquad (1)$$

where d is the range delay due to the known bias between satellite and user clocks, and n_i is the range delay due to receiver noise and tropospheric refraction. The ionospheric range integral in (1) involves the path integral of the index of refraction μ_i over the channel ray path ℓ_i connecting $\vec{x}$ and $\vec{s}$. A schematic illustration of the geometrical situation is shown in Fig. 1.

For GPS satellites in general, the pseudo-range m_i and the clock bias delay d are precisely measured, while the less significant receiver and tropospheric noises n_i can be easily calibrated. Therefore, the ionospheric range integral is precisely measured in all frequency channels.

To study the refractive properties of turbulent ionospheres, it is important to determine the ionospheric range delay Δ_i, given by

$$\Delta_i = \int_{\vec{s}}^{\vec{x}} \mu_i \, d\ell_i - |\vec{s} - \vec{x}| \quad . \qquad (2)$$

Δ_i is the difference between the range integral and the straight-line distance between $\vec{s}$ and $\vec{x}$. In the absence of ionospheric irregularities and for propagation at frequencies ω_i, much higher than the plasma frequency $\omega_p = \sqrt{(4\pi e^2 N_e/m_e)}$, the index of refraction at channel i is given by

$$\mu_i = (1 - \omega_p^2 / \omega_i^2)^{\frac{1}{2}} \quad , \qquad (3)$$

and so,

$$\Delta_i \simeq - \frac{2\pi e^2}{m_e} \cdot \frac{1}{\omega_i^2} \int_{\vec{s}}^{\vec{x}} N_e \, d\sigma \quad , \qquad (4)$$

where N_e is the local ionospheric electron density along the line of sight element $d\sigma$ between the points $\vec{s}$ and $\vec{x}$. Equation (4) is the usual approximate relation between ionospheric path delay, wave frequency and total electron content along the line of sight, in the absence of ionospheric turbulence, and is valid when $\omega_i >> \omega_p$.

In the presence of ionospheric turbulence in general, Δ_i, defined in (2), acquires a random component $\delta\Delta_i$, although how $\delta\Delta_i$ is related to the propagation conditions requires careful analysis into two regimes. Consider the case in which (3) and the approximation $\omega_i >> \omega_p$ can still be taken as valid. In the single scatter regime, in which the ionospheric turbulence is either weak or thin-layered, it is a good approximation to assume the wave to have scattered only once in its passage between $\vec{s}$ and $\vec{x}$. Since the scattering angle is small and only one scattering has taken place, the ray path ℓ_i can again be approximated by the line of sight σ. Therefore, in the single scatter regime, (4) remains valid and the random component

$\delta\Delta_i$ is linearly related to the random component of the electron density δN_e. In the multiple scatter regime, in which the ionospheric turbulence is either strong, thick-layered and/or multiple-layered such as revealed in equatorial spread-F conditions, the ray path ℓ_i can no longer be approximated by the line of sight σ, i. e.

$$\Delta_i \simeq -\frac{2\pi e^2}{m_e}\cdot\frac{1}{\omega_i^2}\int_{\vec{s}}^{\vec{x}} N_e \, d\ell_i \quad , \qquad (5)$$

where the ray path element itself has a random component. Therefore, in the multiple scatter regime, $\delta\Delta_i$ is non-linearly related to δN_e in a manner which depends on the random component of the ray path. Therefore, in order to understand the refractive properties of turbulent ionospheres in the multiple scatter regime, it is necessary to measure and theoretically interpret the statistical properties of Δ_i which depends not only upon the electron density fluctuations, but also upon the ray path fluctuations. In particular, we note from (5) that the random fluctuations of N_e and of ℓ_i contribute, through their coherence function $\langle N_e \, d\ell_i \rangle$, to the average value $\langle \Delta_i \rangle$.

The foregoing discussion may be restated with more rigor, but with much more complexity, in terms of the formal scattered wave function phase at $\vec{x}$ referenced to the wave function phase at $\vec{s}$ in channel i. We have chosen to discuss the problem in the language of geometrical optics primarily for the sake of emphasizing the physical concepts involved. However, it is also crucial to point out the distinction between the absolute wave phase fluctuation $\delta\Delta_i$, which is wave phase fluctuation at $\vec{x}$ relative to wave phase at $\vec{s}$, and the relative wave phase fluctuations referenced to another channel both measured at $\vec{x}$. Without a satellite time standard, only the relative wave phase fluctuations can be measured. Further, comparison between $\langle \Delta_i \rangle$ measured in the presence of ionospheric turbulence with that measured in the absence of turbulence will reveal the largest-scaled nonrandom properties of the two types of ionospheres, thus allowing for interpretation or confirmation of the "ionospheric hole" theories of equatorial spread-F (Woodman and La Hoz, 1976; Hudson and Kennel, 1975).

The measurement of the statistical properties of the ionospheric range delay Δ_i, given by (2), requires very precise position measurements and imposes very strict limits on the error budget. In order to discuss these requirements, it will be convenient to rewrite Δ_i in terms of observational quantities. From (1) and (2), we obtain

$$\Delta_i = m_i - d - n_i - |\vec{s} - \vec{x}| \quad . \qquad (6)$$

Because of the accurate clocks available, there is virtually no error in the measurement of the pseudo-range m_i and of the user-satellite clock bias d. The error budget assigned to m_i and d are of order of 1 meter each (Parkinson and Lassiter, 1976). The tropospheric noise error n_i is about 2 meters; however, this error can be reduced by calibration with respect to a transmitter located at a distance of approximately one tropospheric scale height ($\sim$ 6 km) from the receiver. Therefore, the major source of error in Δ_i derives from the error in determining $\vec{s}$ and $\vec{x}$. The location $\vec{x}$ of the observation site can be accurately determined together with the GPS orbit determination program by ranging during quiet ionospheric conditions; hence, the precision to which we measure Δ_i is mainly determined by satellite ephemerides errors, which are $\sim$ 2 meters cross-track and 6-20 meters in-track. Since the value of $\langle \Delta_i \rangle$ is of the order of 10 meters, geometric considerations must be used to reduce the projection of in-track ephemerides errors onto the line of sight, $(\vec{s} - \vec{x})$. For this reason, we require that the data suitable for absolute phase analysis must be taken when the satellite velocity vector $\vec{v}$ is within $\sim \pm 10^{\circ}$ of perpendicular to the line of sight $(\vec{s} - \vec{x})$. It should be noted that corrections for the satellite Doppler is also minimized by this observational geometry. Further, points of zero Doppler are easily identified by the sign change of the Doppler shift. These are the satellite turning points at the extreme north and south latitudes, and the points of closest approach to the observer. Discussion of the geometrical features of the observation is given in the next section.

To measure the second order statistical moment of Δ_i, it is only necessary to determine $\delta\Delta \equiv \langle \Delta_i \Delta_j \rangle - \langle \Delta_i \rangle \langle \Delta_j \rangle$. From (6), we see that $\delta\Delta$ is identical to $\delta I \equiv \langle I_i I_j \rangle - \langle I_i \rangle \langle I_j \rangle$, where

$$I_i = m_i - d - n_i \qquad (7)$$

Since I_i is independent of the uncertainties associated with $\vec{s}$ and $\vec{x}$ measurements, the statistical fluctuations of Δ_i can be determined with great accuracy. In this regard, let us reiterate that the fluctuation of Δ_i is a range fluctuation, i. e. fluctuation of phase at $\vec{x}$ relative to the absolute phase of the wave at the moment of its transmission from the satellite. This range fluctuation, or "absolute" phase fluctuation, is not the same as the fluctuations of wave phase measured at $\vec{x}$ relative to the phase of another channel measured simultaneously at $\vec{x}$. The latter measurement is what is usually obtained in coherence experiments such as the WIDEBAND observations. As we have pointed out, these fluctuations are the same only if we assume that the microscopic ray paths of the two channels are the same and that the two rays are scattered by the same irregularities. In the multiple scatter regime, such an assumption needs to be demonstrated by experiments such as the one described here.

III. NAVSTAR GLOBAL POSITIONING SYSTEM AND OBSERVATIONAL GEOMETRY

The NAVSTAR-GPS is a satellite-based navigation system which will provide extremely accurate position and timing information to properly equipped users anywhere on or near the earth (Parkinson and Lassiter, 1976). The system concept evolved from Air Force and Navy studies initiated in the mid-1960's. Presently, it is a Joint Service Program in which the first satellites designed to prove the feasibility of the system are provided by the Naval Research Laboratory. The total systems responsibility has been assigned to the Air Force with technical support provided by The Aerospace Corporation. The system eventually calls for the deployment of 24 navigation satellites in circular, twelve-hour orbits to provide global navigation coverage. In June 1977, the first fully operational satellite of the system, Navigation Technology Satellite II (NTS-2) was launched. We are proposing to use the transmissions of NTS-2 and subsequent GPS satellites as a probe of the turbulent ionosphere. Since the scientific measurements are concerned with the refractive and coherent properties of signal propagation in turbulent ionospheres, no hindrance or infringement upon the operational aspects of the system will be imposed. Further, the analysis of the purely physical aspects of the signal will have no connection with the operational use and test of the signal.

The passive satellite ranging concept simply requires that the user measures the time delay of a received signal relative to the satellite time when the transmission was made. While ionospheric propagation time delays, clock bias and other time delay factors have to be taken into account for accurate ranging measurements, the unique and basic element in the GPS satellite and user systems is that highly accurate time standards are available for both. The navigational time standards for the GPS system use cesium beam oscillators whose frequency is stable to 2 parts in 10^{13}. No other satellite system carries time standards with such precision. The basic concept of our proposed experiment is to exploit this unique property of the GPS system to definitively measure and theoretically analyze the refractive and coherence properties of wave propagation in severely turbulent ionospheres. The proposed experiment and the associated theoretical analysis will yield information about the refractive properties of turbulent ionospheres which would not be obtainable by any other means short of launching a new satellite experiment with the special capability of precise absolute phase determination similar to the GPS system.

A number of Navigation Development Satellites (NDS) of the GPS will be launched in 1977-1978 to form a test constellation with the operational NTS-2 satellite which was launched into near circular orbit at an approximate altitude of 20,000 km and at an inclination of 63°. The satellite period is 12 siderial hours with a repetitive ground track suitable for operational ranging tests over the continental United States. Figure 2 illustrates the NTS-2 satellite ground track and the observational horizons for the Panama site. From the satellite orbital geometry, it is seen that a complete coverage of the geomagnetic dependence of the equatorial scintillation effects can be achieved with observations at two sites, Panama and Ancon, Peru, although observations made at any one of the two sites are equally valuable. In order to reduce errors in absolute phase measurements, the satellite velocity vector at the time of observation must be nearly perpendicular to the ray path. Thus, at Panama, the segment of data suitable for absolute phase measurements is to be obtained at the southern turning points of the satellite. Similarly, at Ancon, the absolute phase measurements are to be made at or near the northward or southward transits of the geomagnetic equator. The ray path and ionospheric irregularity geometries at the two sites are also very different. These are illustrated schematically in Figure 3. It is seen that, because of the known location and altitude of the spread-F irregularities, the Panamanian observational ray path is necessarily of low elevation angle ($\sim 5^{\circ}$) while that of Ancon is of high elevation angle ($\sim 40^{\circ}$). Further, the Panamanian ray path will be nearly in the same plane as the magnetic field lines while that of Ancon will be nearly perpendicular to the magnetic field lines. Therefore, observations at the two sites complement one another. However, if one is to choose only one of the two sites, the Panamanian site is preferred because the usable observation time at the satellite southern turning point lasts nearly 3 hours, whereas the transit time of the spread-F zone near the geomagnetic equator is only about 5 minutes.

The equipment relevant to the proposed experimental program are two navigational subsystems - the sidetone ranging (STR) system which is integrated with the orbit determination and tracking system (ODATS), and the pseudorandom noise (PRN) pulse ranging system which is the main navigation signal. For our proposed experiment, we shall measure the signal characteristics of both subsystems in order to obtain coverage over UHF and GHz frequencies. The STR-ODATS system transmits ranging tones at approximately 335 MHz and 1580 MHz, each consisting of a carrier, a single side band reference tone, and ten ranging tones spanning up to 6.4 MHz from the reference tone. Thus, the STR-ODATS system allows us to measure fine-scaled coherence properties over very narrow bandwidths. To consider coherence properties over wider frequency bandwidths, we make use of the main navigation subsystem which transmits at 1575.42 MHz and at 1227.6 MHz. Since the pulsed navigation signal

employs a PRN waveform of binary bits at a rate of 10 Megabits/sec, the bandwidths at the two frequencies are approximately 20 MHz. Therefore, the frequency structure of the GPS signals are suitable for the study of propagation in turbulent ionospheres in the GHz range. It should also be noted that the UHF frequencies of the STR-ODATS subsystem overlap the UHF frequencies of the WIDEBAND experiment so that coordination with WIDEBAND observations will be possible.

IV. SUMMARY

We point out that the NAVSTAR Global Positioning System can be viewed as a unique scientific tool to probe the properties of turbulent ionospheres. In particular, the availability of extremely accurate time standards in both the user and satellite segments of the system allows measurements of refractive and coherence properties of GHz wave transmissions through turbulent ionospheres to be made with unprecedented accuracy and at minimal cost.

We propose a campaign at the Panama observation site to make observations of GPS transmissions which would yield the first absolute measurement of ionospheric refraction through quiescent and turbulent ionospheres. The major share of the analytical work will be to exploit the data base obtained to achieve a basic theoretical understanding of wave propagation in the multiple scatter regime. Upon completion of this segment of the experiment, and if the data anlysis indicates the value of a changing geometry, we expect to make further measurements at a magnetic equatorial site such as Ancon, Peru.

REFERENCES

Fremouw, E. J., M. D. Cousins, G. K. Durfey and R. C. Livingston, WIDEBAND satellite observations, progress reports 1976-1977, Stanford Research Institute, Menlo Park, CA.

Gurvich, A. S. and V. I. Tartarskii, Coherence and intensity fluctuations of light in the turbulent atmosphere, Radio Sci., 10, 3, 1975.

Hudson, M. K. and C. F. Kennel, Linear theory of equatorial spread-F, J. Geophys. Res., 80, 4581, 1975.

Ishimaru, A. and S. T. Hong, Multiple scattering effects on coherent bandwidth and pulse distortion of a wave propagating in a random distribution of particles, Radio Sci., 10, 637, 1975.

Kelley, M. C., G. Haerendel, H. Kappler, A. Valenzuela, B. B. Balsley, D. A. Carter, W. L. Ecklund, C. W. Carlson, B. Hansler and R. Torbert, Evidence for a Raleigh-Taylor type instability and upwelling of depleted density regions during equatorial spread-F, Geophys. Res. Lett., 3, 448, 1976.

Liu, C. H. and K. C. Yeh, Frequency and spatial correlation functions in a fading communication channel through the ionosphere, Radio Sci., 10, 1055, 1975.

Morse, F. A., B. D. Edgar, H. C. Koons, C. J. Rice, W. J. Heikkila, J. H. Hoffman, B. A. Tinsley, J. D. Winningham, A. B. Christensen, R. F. Woodman, J. Pomalaza and N. R. Teixeira, EQUION, an equatorial ionospheric irregularity experiment, J. Geophys. Res., 82, 578, 1977.

Parkinson, B. W. and E. M. Lassiter, Ionospheric effects in NAVSTAR-GPS, North Atlantic Treaty Organization Conference Proceedings, 1976.

Taur, R. R., Simultaneous 1.5 GHz and 4 GHz ionospheric scintillation measurements, Radio Sci., 11, 1029, 1976.

Woodman, R. F. and C. La Hoz, Radar observations of F region equatorial irregularities, J. Geophys. Res., 81, 5447, 1976.

Zirm, R. R. and J. M. Goodman, Propagation effects observed in connection with NTS-1 observations near the magnetic equator, North Atlantic Treaty Organization Conference Proceedings, 1976.

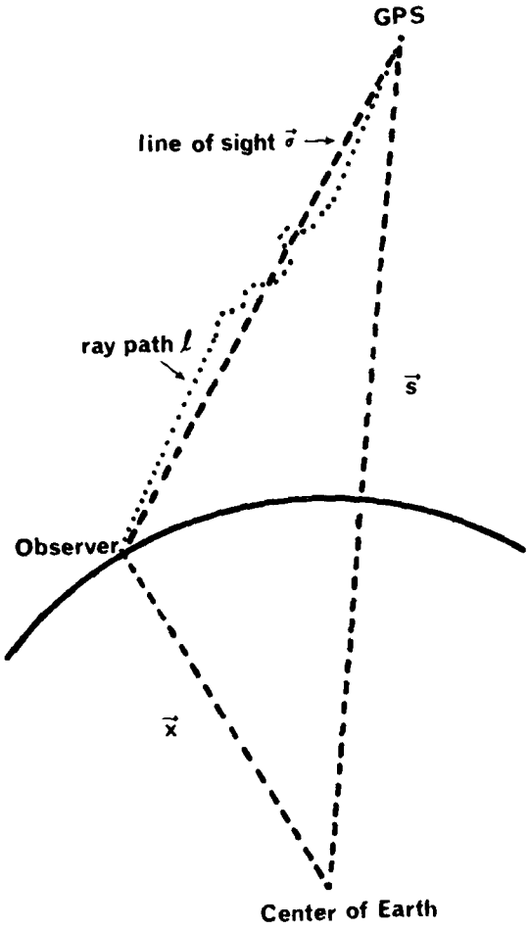

GPS
line of sight $\vec{\sigma}$
ray path l
$\vec{s}$
Observer
$\vec{x}$
Center of Earth

Figure 1

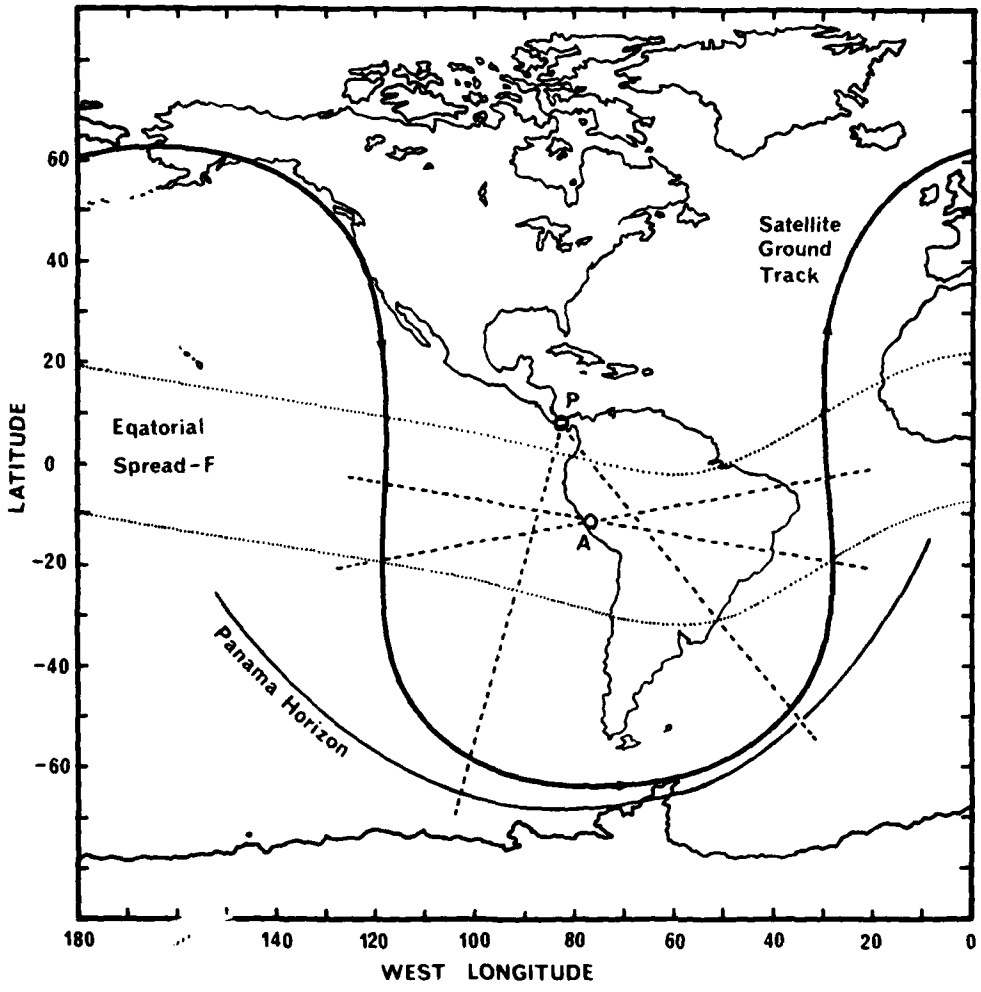

Satellite
Ground
Track
Eqatorial
Spread-F
P
A
Panama Horizon
LATITUDE
60
40
20
0
-20
-40
-60
180
140
120
100
80
60
40
20
0
WEST LONGITUDE

Figure 2

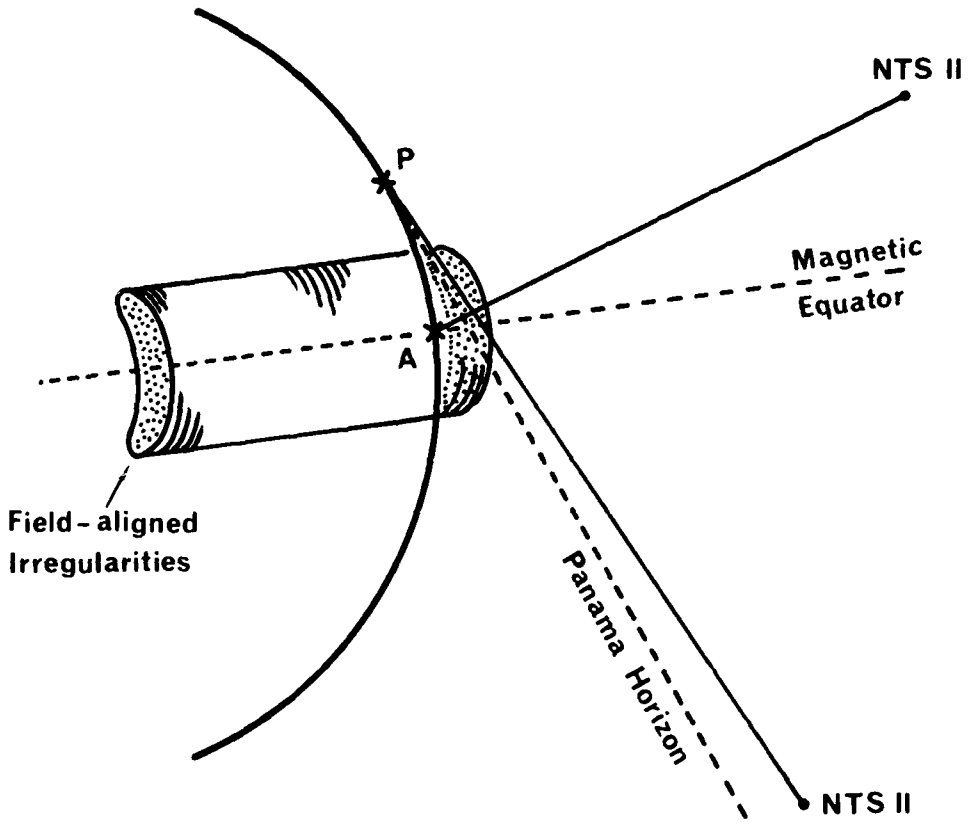

NTS II
P
Magnetic
Equator
A
Field - aligned
Irregularities
Panama Horizon
NTS II

Figure 3

A REVIEW OF THE SERVICES PROVIDED BY THE SPACE ENVIRONMENT SERVICES CENTER

Gary R. Heckman
Space Environment Laboratory
Boulder, Colorado

I. INTRODUCTION

The Space Environment Services Center (SESC), located in Boulder, Colorado is a joint operation of the U.S. Air Force Air Weather Service and the Space Environment Laboratory of the National Oceanic and Atmospheric Administration. The Air Weather Service is responsible for providing operational space environment support to the Department of Defense through its center at Global Weather Central (GWC) at Offutt A.F.B., while SESC provides a similar service for the civilian scientific and technical community. The SESC is concerned primarily with providing solar and geophysical information, leaving responsibility for ionospheric products to the GWC. The flow of space environment services for operational purposes are shown in Figure 1.

Though SESC does not make ionospheric forecasts specifically, it provides a range of real time services in direct support of systems that operate in, near, or through the ionosphere. Those services are described in the following sections.

II. SOLAR GEOPHYSICAL INDICES

A common tendency among operational users of space environment support is the use of various indices to simplify communication about the state of solar or geomagnetic disturbances. Though some of the indices in current use have been used for decades, most of the recent advances in improved services have been in the introduction of new indices or the refinement of existing ones.

Sunspot Number

Sunspot numbers are only indirectly related to ionospheric systems, but a number of long term (months and years) prediction models of HF radio propagation conditions and satellite orbit decay rely on sunspot numbers as an approximation for extreme ultra violet (EUV) radiation levels. The fortuitously good correlation between EUV fluxes and sunspot numbers, when smoothed over months and years, and the observed quasi-periodicity of the sunspot numbers, enables useful predictions for model input to be made. Because there are still no operational satellites making solar EUV flux measurements, a number of system operators also rely on the day-to-day sunspot number as an approximation to the short-term EUV flux level despite the lower correlation of the daily values (Donnelly, 1977). When SOLRAD 11 data are calibrated, EUV fluxes will be available in real time, but the satellite must be operated for a number of years for the atmospheric models themselves to be calibrated sufficiently to make best use of the EUV fluxes.

Solar Flare Indices

The traditional method of classifying solar flares relies on a measure of the flare area in an optically visible line, generally Hydrogen-alpha (6563 Å), and on the brightness of the flare in that line. Beginning in 1969, personnel of the Space Environment Services Center (Baker, 1970) began classifying solar flares according to their emission in the 1 to 8 Å soft x-ray band as measured on the NRL SOLRAD satellites. The SESC now uses the 1 to 8 Å emission as the basic flare classification parameter. The x-ray classification system is shown in Figure 2.

Geomagnetic Indices

Though the usefulness of alternative indices of geomagnetic activity such as the time rate of change of the field at a particular location (Andersen, et. al., 1974) or rapid field pulsations (Jacobs, 1970) have been described for some specific geophysical applications, the most widely used indices in real time continue to be the 24 hour A and three hour K-indices measured at single stations or approximated from a sample of real time stations for planetary values.

The SESC is currently working on implementing a system of classifying substorms for near real time use that would include the east and west boundaries of substorms, along with their time and some measure of their intensity (see Pytte, et. al., 1976).

Proton Events

The primary definition of a "proton event" in the NOAA/Air Force system is a flux of 10 protons $(cm^2\text{-sec-sterad})^{-1}$ with energy greater than 10 MeV. For polar cap absorption, the definition of an event is a daytime absorption of 2dB or 0.5dB nighttime absorption at 30 Mhz.

III. SESC SERVICES

Figure 3 provides a summary of the services provided by SESC. These services are described in the same order below.

Long Range Predictions

Using the method of McNish and Lincoln (1948), the National Geophysical and Solar-Terrestrial Data Center of NOAA predicts monthly smoothed sunspot numbers. The McNish and Lincoln method relies on a comparison of annual mean sunspot numbers from the cycle being predicted to the annual means of previous cycles. The method is subject to considerable variability early in the cycle when the sample for comparison is small. The SESC provides and uses real time observations of sunspot numbers to enable a more accurate prediction over the period of a few months. Predictions of the general level of solar activity and of geomagnetic activity (low, moderate, high) are made seven and twenty-seven days in advance.

Three Day Forecasts

SESC makes forecasts for the next three days of the levels of geomagnetic activity (Fredericksburg A-indices), solar radio flux at 2700 MHz, solar proton events, class X and class M flares. The predictions for proton events and for flares are cast as the probability of the occurrence of each particular type of event for each universal (Greenwich) day. The accuracy of the class M flare forecasts made for a 24 hour period one day in advance are shown in Figure 4.

Alerts of Events

Given the current capability of predicting solar-geomagnetic disturbances, about 40 percent of SESC customers rely on rapid alerts of events once the event is detected. Categories and thresholds for these alerts are shown in Table 1. The alerts have varying thresholds and the user of the service may specify at what level he wishes to be notified.

In addition to the alerts shown in Table 1, there is another category of service that is intermediate between prediction and alert. The prediction of flare consequences such as proton events and geomagnetic storms can be greatly improved once a flare has occurred by using measures of the flare energy such as optical, radio or x-ray emission. A proton prediction program by Smart (Smart, 1976) uses a combination of physical and heuristic models to predict the time, flux intensity and spectral parameters of a flare-induced proton event using any of several flare emission parameters. The Smart program is used as an operational prediction tool in the SESC. Similar models for flare-produced magnetic storms can also be used, though the model techniques for these are not as advanced as for the proton event modeling.

Real Time Data: SELDADS

Some customers, especially those doing scientific research or developmental studies, require more detailed data than are possible in alerts. The Space Environment Laboratory Data Acquisition and Display System (SELDADS) is a publicly accessible computer based data system that allows the user to log into the system with his own computer terminal and read out detailed data and complete event summaries. The types of data available in the SELDADS system are listed in Table 2 and described more thoroughly in Williams (1976).

After-the-Fact Data Summary

Summaries of solar and geomagnetic activity are distributed daily in the teletype messages "Joint AFGWC/SESC Primary Report of Solar and Geophysical Activity" and "Joint AFGWC/SESC Secondary Report of Solar and Geophysical Activity".

The Preliminary Report and Forecast of Solar Geophysical Activity, published and mailed weekly by SESC, contains lists of solar flares, solar active regions, radio flux levels, geomagnetic indices, maps of solar features and selected plots of proton events, x-ray flare profiles and magnetograms.

IV. TECHNICAL NEEDS

To provide improved service to the users of the SESC, several technical improvements are needed. Direct satellite monitoring of EUV fluxes rather than continued reliance on correlative parameters should improve the accuracy of both ionospheric and upper atmospheric density models.

There is at present no operational EUV satellite monitoring system and none is presently being built though as indicated in Section II, SOLRAD 11 could be the source of such data if it is operated for a sufficiently long period into the future. As indicated in Figure 4, the primary source of error in the class M flare forecasts is an inability to predict the time of the flare. This is generally true of all the flare predictions made by the SESC. Exceptions are for proton events and magnetic storm predictions, where the error in the predicted intensity or particle flux is generally as great or greater than the error in the predicted times. Monitoring of solar wind parameters outside the magnetosphere would give an improved prediction of magnetic storm times and intensities.

Better spatial resolution in the measurements or improved extrapolation from existing measurements for localized magnetospheric/

geomagnetic phenomena such as substorms would improve existing ionospheric models. If the real time space environment centers should be called upon to provide regular support for spacecraft operating anomalies such as charging, there is the same need for higher resolution information than is now available.

Finally, in the data currently monitored, specific indices could be developed that more closely relate the measurements to particular operating systems or models. The C-M-X flare classification system was developed to provide such an improvement. The geomagnetic A and K indices are examples of gross indices that continue to be used in a wide variety of applications with widely varying time dependencies. On the other hand, the continuation of gross indices such as A and K has enable the SESC to fulfill the important role of being a general information source to which operators of advanced technology systems can go to get a rapid first indication of whether solar and geomagnetic activity may be the source of some unexplained anomaly in their system. In several past instances, this has been the case and the result has led to further studies by the operators and enabled them to gain better understanding of environmental effects on their systems. Development of new systems in the future will presumably provide continued opportunities for the same service.

V. References

Anderson, C. W., Lanzerotti, L. J. and Mac Lennan, C. G., 1974, "Outage of the L-4 System and the Geomagnetic Disturbance of August 4, 1972", Bell System Technical Journal, Vol. 53, 1817.

Baker, D. M., 1970, "Flare Classification Based Upon X-ray Intensity", AIAA Paper No.70-1370.

Donnelly, R. F., 1977, The Solar Output and Its Variation, Colorado Associated University Press, Boulder, 320.

Jacobs, J. A., 1970, Geomagnetic Micropulsations, Springer-Verlag, Heidelberg, Berlin.

Pythe, T., McPherron, R. L., and Kokubun, S., 1976, "The Ground Signatures of the Expansion Phase During Multiple Onset Substorms", Planet. Space Sci., Vol.24, 1115-1132.

Smart, D. F., and Shea, M. A., 1977, "Application of Elementary Coronal Propagation and Co-Rotational Concepts to Solar Proton Event Prediction", Proceedings 15th International Cosmic Ray Conference, Vol. 5., 131.

Williams, D. J.,1976, SELDADS; An Operational Real-Time Solar-Terrestrial Environment Monitoring System, NOAA Technical Report ERL 357-SEL 37.

TABLE 1

SESC ALERT CATEGORIES AND THRESHOLDS FOR ISSUING ALERTS

SOLAR FLARES

X5 (1-8Å X-Ray Classification)
X1
M5
3B (Optical Classification
2B
1B

MAGNETIC DISTURBANCES

A ≥ 50 (real time A measured at Boulder)
A ≥ 30
A ≥ 20
K ≥ 6 (real time K measured at Boulder)
K ≥ 5 observed in successive three-hour intervals
K ≥ 5
K ≥ 4
SUDDEN COMMENCEMENT

RADIO BURSTS/NOISE STORMS

10CM RADIO BURST GREATER THAN 100 FLUX UNITS
245MHZ BURST OR NOISE STORM
TYPE II AND/OR TYPE IV DECAMETRIC EMISSION

PROTON EVENTS

PROTON FLUX (E > 10MEV) >10CM^{-2}-SEC^{-1}-STERAD^{-1}

TABLE 2

DATA AVAILABLE IN SELDADS

TYPE	WAVELENGTH	TIME RESOLUTION	PRIMARY SOURCES
X-ray Patrol	1-8Å	One minute values	SMS/GOES Satellites
Optical Patrol	H-alpha	Approximately one minute	U.S.Air Force Observatories at Teheran, Ramey, Palehua, Holloman.
			NOAA/IPS Joint Observatory at Culgoora
			Boulder (part time)
			International Ursigram and World Days Service (IUWDS) network.
Radio Patrol	Decametric through Centimetric	Approximately one minute	Quiet time flux levels and event reports from Sagamore Hill, Ottawa, Boulder, Palehua, Manila, Culgoora (CSIRO) and IUWDS.
Particle Patrol	0.6-500 MeV protons		SMS/GOES
	≥2 MeV electrons		SMS/GOES
	4.0-329 MeV alpha particles		SMS/GOES
	~10, ~30, ~60 MeV protons		NOAA 5
	Additional data from Vela and Meteor (USSR) satellites		
Geosynchronous Magnetometer Patrol	Three Field Components		SMS/GOES satellites
Geomagnetic Indices	A, K		18 stations
Total Electron Content		Hourly	
Auroral Back-scatter Radar		15 minutes	5 Alaskan stations plus Thule
High Latitude Riometers	30 MHZ	15 minutes	5 Alaskan stations plus Thule
Ionosonde Observations	FOF2, M3000, FMIN	One hour or six hours	Approximately 30 stations

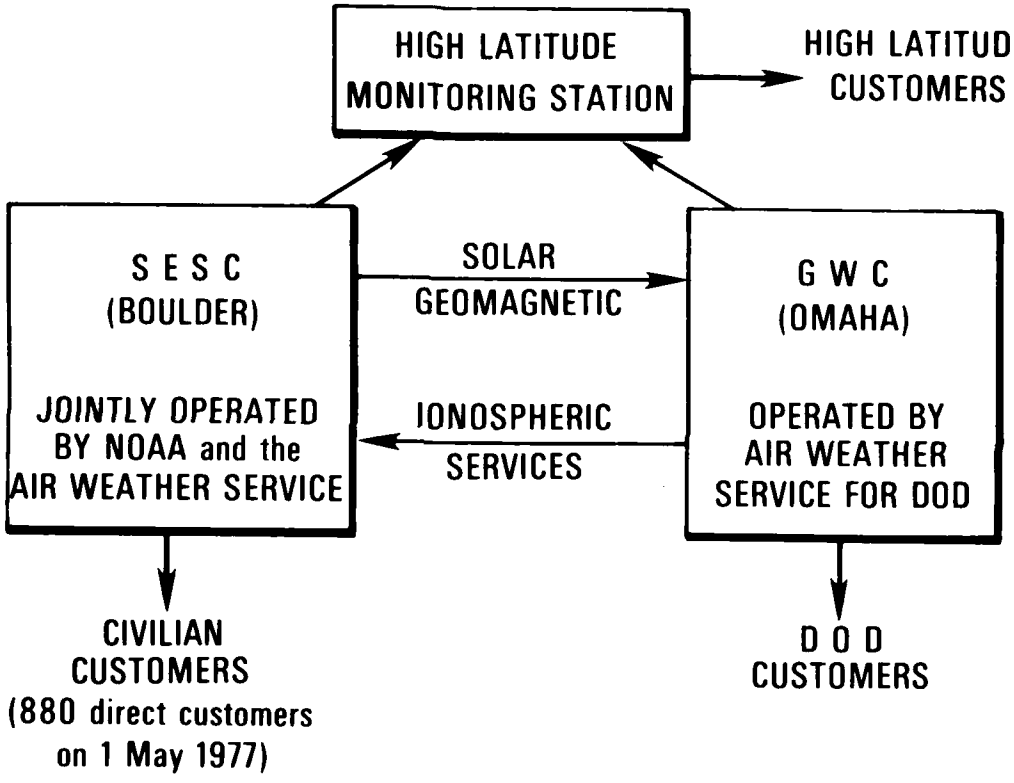
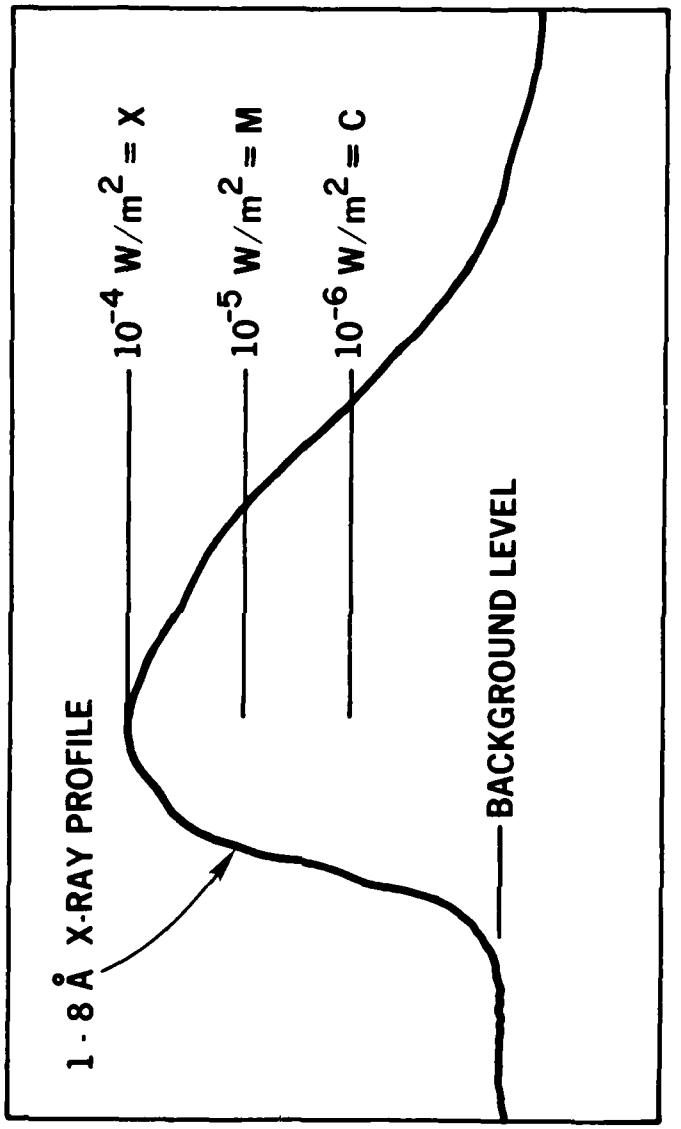

HIGH LATITUDE MONITORING STATION
HIGH LATITUDE CUSTOMERS
S E S C (BOULDER)
JOINTLY OPERATED BY NOAA and the AIR WEATHER SERVICE
SOLAR GEOMAGNETIC
IONOSPHERIC SERVICES
G W C (OMAHA)
OPERATED BY AIR WEATHER SERVICE FOR DOD
CIVILIAN CUSTOMERS (880 direct customers on 1 May 1977)
D O D CUSTOMERS

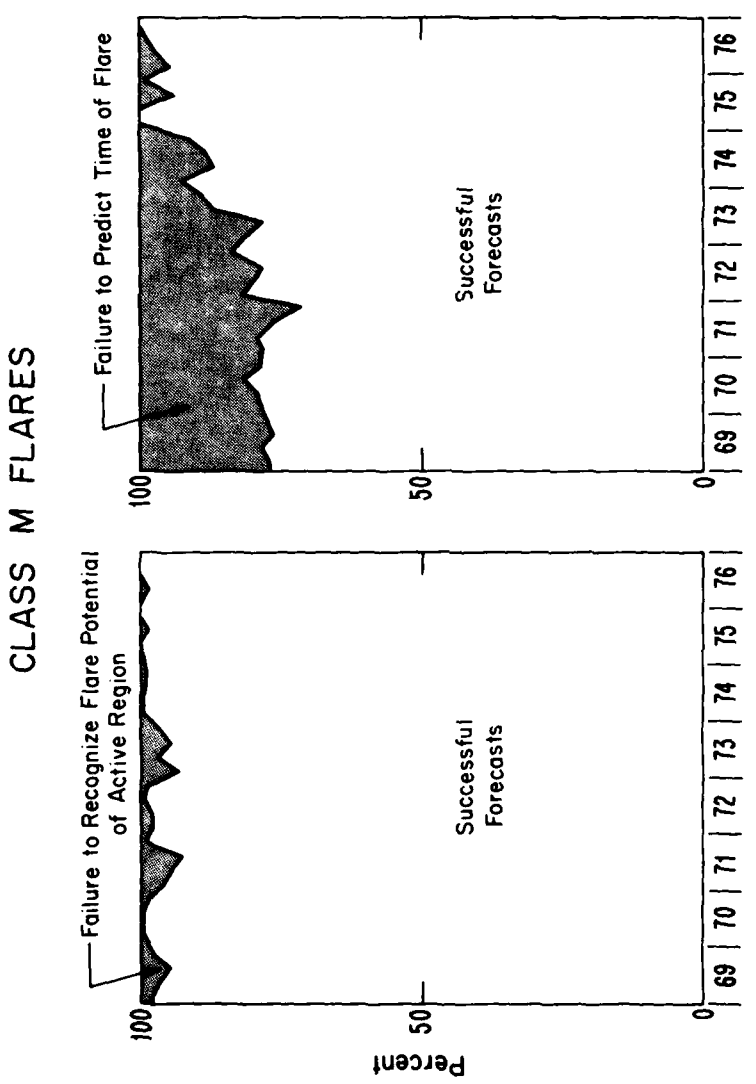
FIGURE 2


SESC FLARE CLASSES
1 - 8 Å X-RAY PROFILE
10^{-4} W/m^2 = X
10^{-5} W/m^2 = M
10^{-6} W/m^2 = C
BACKGROUND LEVEL
TIME

DISTRIBUTION OF SESC SERVICES

TYPE	TIME OF APPLICABILITY	PHENOMENA OF INTEREST
FORECAST	10 YEARS 6 MONTHS, 7 and 27 DAYS	SUNSPOT NUMBER GENERAL LEVEL OF SOLAR and GEOMAGNETIC ACTIVITY
FORECAST	1, 2, 3 DAYS FORECAST	FLARES, PROTON EVENTS, GEOMAGNETIC INDICES, SOLAR FLUX INDEX
FORECAST	30 MINUTES	FLARES
ALERT	IMMEDIATE	X-RAY, RADIO, OPTICAL EMISSION FROM FLARES
FORECAST	30 MINUTES - 100 HOURS	PROTON EVENT, GEOMAGNETIC STORM, SUBSTORMS
ALERT	IMMEDIATE	PROTON EVENT, GEOMAGNETIC STORM, SUBSTORMS
DATA (SELDADS)	REAL-TIME	ALL DATA
DATA (SUMMARY)	DAILY and WEEKLY	SELECTED DATA

FIGURE 1


CLASS M FLARES
Failure to Recognize Flare Potential of Active Region
Failure to Predict Time of Flare
Successful Forecasts
Successful Forecasts
100
50
0
Percent
69
70
71
72
73
74
75
76
Year
Year

AN OBJECTIVE APPROACH TO

REGION ANALYSIS FOR FLARE FORECASTING

Joseph W. Hirman
William E. Flowers

Space Environment Laboratory
Boulder, Colorado

I. INTRODUCTION

The Space Environment Services Center (SESC) is a part of the National Oceanic and Atmospheric Administration (NOAA) Environmental Research Laboratories in Boulder, Colorado. This center, which opened in 1964 and became fully operational in 1968, is jointly operated by NOAA and the USAF Air Weather Service. Through an elaborate real time communications system, SESC maintains a continuous watch of the sun by satellites and by optical and radio telescopes located at Boulder and around the world. The SESC is designated as the national and international focal point for real time solar-geophysical data collection, event warning, and forecasting (Mangis, 1975). Solar flare forecasts are a major product of this center and the subject of this report.

II. FLARE FORECASTING

The solar flare forecast, as issued three times daily by SESC, is a percentage forecast giving the probability that a solar flare producing significant geophysical effects will occur in the next 24, 48, and 72 hours. The close relationship between the flare-produced 1-8 Angstrom x-ray flux and subsequent ionospheric effects (Mitra, 1974) were studied. In 1970, SESC devised the C-M-X x-ray classification scheme for solar flares (Baker, 1970) (table 1) using satellite measurements from SOLRAD and more recently from SMS/GOES. For example, an M-class flare could typically produce significant short wave fade and up to 90^{0} phase shifts on VLF circuits, while an X-class flare results in major short wave fade and VLF phase shifts of up to 180^{0}. The smallest detectable flare effect on VLF circuits has been reported as C1.5 by Muraoka, et al, 1977. The SESC forecast gives the probability of the occurrence of each flare class.

The flare probability forecast should be indicative of the activity level of the sun. Since most near-earth environmental disciplines are interrelated with the activity level of the sun, the benefits of accurately predicting flares are numerous. As an example, the accurate prediction of a large flare could aid immeasurably in predicting the state of the geomagnetic field, in predicting the state of the ionosphere, and in computing satellite drag coefficients. The flare probability forecast is also vital to the research community. Various investigators depend on flare forecasts for decision making relative to detector attitudes and instrument configurations.

This probability forecast is currently based on subjective analysis of many pieces of data available at SESC in near real time (Williams, 1976). Information on this method has been published (McIntosh, 1970; Smith, 1970) and will only be described briefly here. The current SESC method involves the interpretation of photospheric (sunspots), chromospheric (H-alpha and Calcium), and magnetic field images and radio, x-ray, and coronal data as available. Every active region is analyzed according to its complexity, magnetic configuration, age, rate of growth, past history, and most important, its similarity to active regions observed in the past. It is this latter factor of experience that is the forecaster's most important asset.

Over the years, the more experienced forecasters have gained some insight as to how to interpret the data for the key parameters that best fit into the forecasting puzzle, and this skill has been passed on to the new forecasters. In this manner, "forecasting expertise" has developed. However, due to the subjectivity of the interpretation, quality control is difficult and the resulting forecasts are inconsistent. The need for an objective technique for flare forecasting has long been recognized; however, no objective studies exist that enable the duty forecaster to attach any quantitative significance to the many pieces of available data. With the advent of the new Solar Optical Observing Network (SOON) and its capability for objective observing, (Air Force Tech order 31M1-2 FM67-1, to be published), it was decided that the opportune

time had arrived to initiate procedures for developing objective forecasting techniques. All forecasters at the Space Environment Services Center were queried as to what parameters they found useful as inputs to their flare forecasts. At the same time, a literature search was made to determine which previously studied predictors were available to SESC in near real time. These parameters were then incorporated into a scheme that is designed to identify, in order of importance, those ingredients essential to flare occurrence. All the problems that impede the development of an objective flare forecasting technique still exist; however, it is believed that the project described in this paper will aid in reducing some of those problems.

There are many theories concerning which parameters should be monitored and measured when trying to forecast flares. The various links between flare production and certain active region characteristics have been discussed for years. The following are just a few: The Influence of Radio Brightness Temperature on Solar Flare Prediction (Reilly et al., 1967); The Relationship Between Chromospheric Flares and Magnetic Fields of Sunspot Groups (Bumba 1958); Solar Flare Occurrence as a Function of Sunspot Size (Enger et al., 1966); Some Peculiarities of Magnetic Fields Connected With Solar Flares (Severny 1962); and Inference of Solar Magnetic Polarities From H-alpha Observations (McIntosh 1970).

A scheme has been devised at SESC that will enable the forecaster to evaluate some of these theories, both collectively and individually. The procedure for analysis is to objectively evaluate certain designated parameters for each region on a daily basis. The results of these evaluations will be studied using various statistical methods, such as MVDA (Multi-Varient discriminent Analysis) and contingency tables. Through this method, it is hoped that significant parameters can be established and weighted in accordance with their importance. Parameters that are not significant can be disregarded. The purpose of this region analysis is not to find new predictors, but rather to take those predictors that are already known and integrate them into an analytical process that will yield a quantitative result. This scheme was initiated on 1 January 1977 and will be evaluated over the next two years or more with the immediate goal of improved forecasts, but with an ultimate goal of understanding solar flares. This real time analysis is similar to the study of historical data by Podsiadlo, 1973.

III. REGION ANALYSIS SCHEME

The prime consideration in developing this scheme was data availability; data not readily accessible to the duty forecaster in near real time on a daily basis was not included. With the exception of this limitation, all aspects of each active region are analyzed daily before the 2200UT forecast using observations reported for the previous 24 hours. To assist in the data reduction and to test the many theories associated with flare occurrence, it was decided to divide the types of observations into four main categories: White light, H-alpha, Radio, and Region History. These were further defined in order to reflect the kinds of data available from each type of observation. This also enables a hierarchy of importance to be established within each data set (table 2).

It has long been argued that the solar magnetic fields are the energy source for the production of flares (Bumba, 1958; Severny, 1958; Giovanelli, 1939; McIntosh, 1969). The White light and H-alpha categories in this scheme are basically concerned with magnetic fields. Spot groups are visual manifestations of magnetic loops of force. Therefore, spot behavior, spot growth, spot movement, and location of spot emergence are indicative of various magnetic field changes. Similar deductions can be made from observations in H-alpha. Close scrutiny of regions in H-alpha light will reveal certain changes in the fine structure, and the relationship of these changes to the large-scale features are dependent on magnetic changes and configurations which are not apparent from White light sunspot analysis alone. In short, this entire procedure is an attempt to quantitatively view magnetic structure and changes.

In the White light category, the parameters established are those used singularly in earlier flare studies and are generally values that can be measured objectively. The characteristic elements are: spot class (McIntosh classification), magnetic class, magnetic field strength, magnetic gradient, sunspot dynamics, region interaction, and stage of development. The data in the H-alpha category are those derived from analysis of H-alpha images and/or from observatory reports. These parameters are thought to be most significant by the forecasters, are difficult to measure objectively, are the least studied. The parameters included in this category are: where the region emerged with respect to the large scale fields and sector boundaries, plage compactness and activity of embedded filament, inferred neutral line orientation, complexity and temporal changes, activity of the associated filament, bright points and plage fluctuations, and the emergence of new flux.

The third data set consists of a single parameter: the radio activity of the region over the previous 24 hours. This characteristic is not expected to be significant for flare predictions, but is included to gather statistical information for particle event prediction. The last category is continuous history of the region. It includes the largest flare produced by the region, the number of transits this region has made, and

a history of the observed particle production of the region. In addition to the above four categories of active region characteristics, a category containing the forecaster's subjective prediction was included. It is recognized that this intuitive, subjective input of the forecaster cannot be quantized, yet it remains an important integral part of the forecast.

For the purpose of quantifying these characteristics, weight factors were assigned to each parameter in a hierarchy of increasing importance. However, our a-priori assumptions on some of these factors proved to be incorrect during an initial study and the actual measured value was used instead. For example, we originally broke down the 61 possible spot classifications into 6 groups based on forecaster intuition, only to find that some groups were not as productive as we thought.

The forecaster evaluates the data and chooses the factors as appropriate. These are then entered on a log (table 3) and finally into the computer. A factor analysis will be performed by the computer to determine if indeed the hierarchy assigned to the parameters is accurate and if not, just what it should be. Next, a multivarient discriminant analysis (MVDA) will be made to determine the most useful parameters and a probability function to be used. This function will then be applied to data in our real time operation to produce a flare probability forecast, and will be compared to the forecaster's prediction. It is hoped that the MVDA forecast will become a useful tool for the forecaster and will increase the accuracy of his predictions.

IV. CURRENT STATUS OF THE SCHEME

This project is currently in its infancy. Very limited preliminary results are available. Six months of data (January 1977-June 1977) have been analyzed. Taking into account the limited sample of the data and the phase of the cycle it represents, confidence factors relative to these results are guarded. The analysis indicates that the prime factor for whether or not a region will flare is whether or not the region has flared before, or persistence. Secondly, regions which produce bright points are likely to produce flares. However, again in light of the constraints, these results are what one would logically conclude. Also significant are the spot class and the product of the spot class and magnetic class. Three additional months of data are being prepared for analysis. These three months include periods of relatively higher levels of activity.

Data collection and computer storage are an ongoing process, and progress is being made toward familiarizing forecasters with what kinds of meaningful information can be extracted from the data and how to extract it. Also, a great deal of effort is being expended on investigating and refining the mechanics of the scheme. This, we feel, is important for two reasons. Firstly, the mechanics of the analysis are both time consuming and laborious. If we can incorporate the scheme into a useful work tool while concurrently gathering data for the actual testing, the duty forecaster will be more interested in doing an accurate analysis and will be more receptive to the program. Secondly, with sufficient refinement, the latitude for individual interpretation of data will be minimized, and the goal of an objective forecast will be approached.

ACKNOWLEDGEMENTS

A task of this magnitude required the aid of many and variously talented people. The authors would like to extend a special thanks to the following people:

Mr. Gary Heckman, Chief of The SESC;
All of the forecasters at the SESC;
Mr. Kurt Carran and all of the communication center operators;
Mr. Thomas Gray;
Ms Linda Bath;
Ms Cheryl Cruickshank;
Dr. Don Neidig, SPO;
Dr. Richard Donnelly, SEL;
Mr. Peter Tryon, NBS;
Ms Ginger Caldwell, NBS.

TABLE 1

C0	Peak flux less than 10^{-6} watts/m^2/sec
C1-C9	Peak flux from 1 - $9 * 10^{-6}$ watts/m^2/sec
M1-M9	Peak flux from 1 - $9 * 10^{-5}$ watts/m^2/sec
X1-X9	Peak flux from 1 - $9 * 10^{-4}$ watts/m^2/sec

TABLE 2

THIS ANALYSIS IS BASED ON DATA REPORTED OR OBSERVED DURING THE PAST 24 HOURS ONLY (0000UT TO 2400UT) AND SHOULD BE COMPLETED IN TIME FOR THE 2200Z FORECAST.

REGION ANALYSIS

REGION NUMBER.....
FIRST APPEARANCE LONGITUDE.....
CURRENT LONGITUDE.....
CURRENT LATITUDE.....
CARRINGTON LONGITUDE.....
AGE OF REGION IN DAYS THIS TRANSIT.....

CATEGORIES FOR INDEX CLASSIFICATION

PARAMETER	TYPE	CLASS
	WHITE LIGHT	
1. SPOT CLASS	NONE OBSERVED	0
	ENTER SPOT CLASS THREE LETTER CODE	---
	NO DATA	9
2. MAGNETIC CLASS	NO SPOTS	0
	ALPHA	1
	BETA	2
	BETA-GAMMA	3
	GAMMA	4
	DELTA	5
	NO DATA	9
3. MAGNETIC FIELD STRENGTHS (LARGEST)	NO SPOTS	0
	ENTER LETTER (P/N) AND TWO DIGIT VALUE	---
	(IF SAME USE LEADER POLARITY)	
	NO DATA	9
4. MAGNETIC GRADIENTS IN GAMMA/KP	NO SPOTS OR UNIPOLAR REGION	0
	ENTER THE GRADIENT AS N.NN	-.---
	NO DATA	9
5. SUNSPOT DYNAMICS	NO SPOTS OR NOT APPLICABLE	0
	COALESCING OF SPOTS	1
	SPOT ROTATION	2
	RELATIVE SPOT MOTION (OPPOSITE POLARITY SPOTS)	3
	NO DATA	9
6. INTERACTION WITH ANOTHER REGION	NONE OCCURRED	0
	STRONG SPOTS OF OPPOSITE POLARITY CONVERGE	
	(FROM LESS THAN 2 DEGREES APART)	1
	NO DATA	9
7. STAGE OF DEVELOPMENT	NO SPOTS	0
	MATURE GROUP (STABLE)	1
	DECAYING	2
	GROWING	3
	RAPID DECAY (SPOT OR AREA DECREASE BY > 50%)	4
	RAPID GROWTH (SPOT OR AREA INCREASE BY > 50%)	5
	RAPID GROWTH (SPOT OR AREA INCREASE BY > 100%)	6
	NO DATA	9
	H - ALPHA	
8. LEADER EMERGED IN LEADER OR TRAILER POLARITY FIELDS (FROM PREVIOUS SYNOPTIC MAP)	STRUCTURE-NOT DEFINITE	0
	RETURNING REGION - ENTER REGION NUMBER	---
	< 5 DEG OF NL AND OUT-OF-PHASE WITH NL	2
	> 5 DEG OF NL AND IN LEADER POLARITY FIELDS	3
	> 5 DEG OF NL AND IN TRAILER POLARITY FIELDS	4
	< 5 DEG OF NL AND IN-PHASE WITH NL	5
	NO DATA	9
9. RELATIONSHIP WITH NEAREST SECTOR BOUNDARY (MALE = REGION POLARITY MATCHES THE BOUNDARY)	SECTOR STRUCTURE NOT DEFINITE	0
	REGION IS > 30 DEGREES FROM NEAREST BOUNDARY	1
	NON-MALE AND 10 TO 30 DEG WEST OF BOUNDARY	2
	NON-MALE AND 10 TO 30 DEG EAST OF BOUNDARY	3
	NON-MALE AND < 10 DEG OF BOUNDARY	4
	MALE AND 10 TO 30 DEG WEST OF BOUNDARY	5
	MALE AND 10 TO 30 DEG EAST OF BOUNDARY	6
	MALE AND < 30 DEG OF BOUNDARY	7
	NO DATA	9
10. PLAGE COMPACTNESS AND EMBEDDED FILAMENT (COMPACT = NL CORRIDOR < 2 DEGREES WIDE)	NON-COMPACT PLAGE AND NO FILAMENT	0
	NON-COMPACT PLAGE WITH FILAMENT	1
	NON-COMPACT PLAGE WITH ACTIVE FILAMENT	2
	COMPACT PLAGE WITHOUT EMBEDDED FILAMENT	3
	COMPACT PLAGE WITH EMBEDDED FILAMENT	4
	COMPACT PLAGE WITH ACTIVE EMBEDDED FILAMENT	5
	NO DATA	9
11. MAIN NL ORIENTATION WITHIN PLAGE	WEAK STRUCTURE	0
	NORTH-SOUTH (+/- 45 DEGREES TO NS)	1
	EAST-WEST	2
	HAIRPIN (E-W)	3
	MOSTLY CIRCULAR	4
	REVERSE POLARITY REGION	*
	NO DATA	/

TABLE 2 (cont'd)

12. NEUTRAL LINE COMPLEXITY	NO KINKS OR WEAK STRUCTURE	0
	1-3 KINKS (VERY SIMPLE REGION)	1
	4-6 KINKS (SIMPLE REGION)	2
	7-12 KINKS (INTERMEDIATE REGION)	3
	> 12 KINKS (VERY COMPLEX)	4
	NO DATA	9
13. NEUTRAL LINE TEMPORAL CHANGES	NO DEFINITE TREND	0
	NEUTRAL LINE BECOMING SIMPLE	1
	NEUTRAL LINE BECOMING COMPLEX	2
	NO DATA	9
14. ASSOCIATED FILAMENT (EXTERNAL TO REGION BUT ALONG COMMON NEUTRAL LINE)	NO ASSOCIATED FILAMENT	0
	FILAMENT UNCHANGED	1
	FILAMENT GROWING	2
	FILAMENT DISAPPEARED WITHIN PAST 24 HOURS	3
	FILAMENT DARKENS OR IS ACTIVE	4
	NO DATA	9
15. BRIGHT POINTS AND/OR PLAGE FLUCTUATIONS	NONE OCCURRED	0
	OCCURRED BUT NOT ALONG NEUTRAL LINE	1
	OCCURRED ALONG THE NEUTRAL LINE	2
	PLAGE FLUCTUATIONS	3
	NO DATA	9
16. EMERGING FLUX AND/OR AFS	NONE OCCURRED OR REGION IS NEW	0
	ISOLATED POLE IN REGION	1
	NEW EFR EMERGES WITHIN EXISTING SPOT GROUP	2
	NEW EFR EMERGES NEAR REGION (WITHIN 5 DEGREES OF EXISTING SPOT GROUP)	3
	AFS PRESENT IN REGION	4
	NO DATA	9

RADIO

17. RADIO BURST AND/OR SWEEP	NONE OCCURRED OR SMALL EVENTS	0
	> 250 FLUX UNITS AT 10 CM	1
	>1000 FLUX UNITS AT 15 CM	2
MULTIPLE ENTRIES POSSIBLE	TYPE III SWEEP	3
	TYPE IV SWEEP	4
	TYPE II FOLLOWED BY TYPE IV SWEEP	5
	U BURST	6
	MAJOR AND COMPLEX 15 CM BURST	7
	NO DATA	9

REGION HISTORY THIS TRANSIT

18. LARGEST FLARE SINCE REGION APPEARED.	NONE OCCURRED OR FIRST DAY OBSERVED	0
	C CLASS FLARES HAVE OCCURRED	1
	M CLASS FLARES HAVE OCCURRED	2
	X CLASS FLARES HAVE OCCURRED	3
	NO DATA OR REGION APPEARED ON EAST LIMB	9
19. REGION FIRST APPEARED	FORMED ON DISK	0
	CAME AROUND EAST LIMB - FIRST TRANSIT	1
	SECOND TRANSIT	2
	THIRD TRANSIT (AND ETC)	3
	NO DATA	9
20. PROTON EVENT (THIS TRANSIT)	NO PARTICLE EVENT	0
	PROTON>10 EVENT (10P/CM*CM*SEC*STER # > 10PFU)	1
	GROUND LEVEL EVENT	2
	NO DATA	9

FORECASTS

21. C,M,X,P PROBABILITIES	C PROBABILITY	
	M PROBABILITY	
	X PROBABILITY	
	P PROBABILITY	

EVENTS FOR THE PAST 24 HOURS

22. LARGEST FLARE FOR THE PAST 24 HOURS MULTIPLE ENTRIES ARE POSSIBLE	NONE OCCURRED OR <C0	0
	CLASS C	1
	CLASS M	2
	CLASS X	3
	PROTON EVENT	4
	NO DATA	9

TOTAL SUN PARAMETERS (FROM 2800Z FORECAST)

23. 10 CM	10 CM FLUX VALUE FOR TODAY
24. FORECAST	C/M/X/P
LATEST UPDATE:	28 SEP 1977 REVISION #8 NH

TABLE 5

REGION ANALYSIS

PERFORMED BY: _____ (day fcstr.) CHECKED BY: _____ (swing fcstr.) DATE: _____ IN COMPUTER: _____ (comm. op.)

REGION NUMBER													
APPEARANCE LONGITUDE													
CURRENT LONGITUDE													
CURRENT LATITUDE													
CARRINGTON LONGITUDE													
AGE (DAYS) THIS TRANSIT													
WHITE LIGHT													
1. SPOT CLASS													
2. MAGNETIC CLASS													
3. MAGNETIC STRENGTH													
4. MAGNETIC GRADIENTS													
5. SUNSPOT DYNAMICS													
6. SUNSPOT INTERACTION													
7. STAGE OF DEVELOPMENT													
H-ALPHA													
8. LEADER OR TRAILER													
9. SECTOR EAST OR WEST													
10. PLAGE AND FILAMENT													
11. NEUTRAL LINE ORIENTATION													
12. NEUTRAL LINE COMPLEXITY													
13. NEUTRAL LINE CHANGES													
14. ASSOCIATED FILAMENT													
15. BRIGHT POINTS PLAGE FLUCT.													
16. EMERGING FLUX AND/OR AFS													
RADIO													
17. RADIO ACTIVITY													
HISTORY (THIS TRANSIT)													
18. FLARE HISTORY													
19. REGIONS FIRST APPEARANCE													
20. PROTON HISTORY													
FORECASTS (NEXT 24HRS)													
21. C CLASS M CLASS X CLASS PROTON													
EVENTS (CURRENT 24HRS)													
22. LARGEST FLARE													

TOTAL SUN PARAMETERS

23. 10 cm. FLUX = _____ (observed)
24. C/M/X/P = ___/___/___/___ from 2200Z forecast

NOTE: This form is to be completed daily as part of the 2200Z forecast and is a 24 hour summary and 24 hour forecast, 0000UT-2400UT.

COMMENTS

REFERENCES

Baker, D. M., (1970), "Flare Classification Based Upon X-ray Intensity", AIAA Paper No. 70-1370, AIAA Conference, Huntsville, Ala., U.S.A.

Bumba, V., (1958), "The Relationship Between Chromospheric Flares and Magnetic Fields of Sunspot Groups", *Izvestiia Krymskaia Astrofizicheskaia Observatoriia*, Vol. 19, U.S.S.R., pp 105.

Enger, I., Podsiadlo, T., Jensen, D. C., Ward, Fred, (1966), "Solar Flare Occurrence as a Function of Sunspot Size", *Air Force Surveys in Geophysics*, No. 178, Air Force Cambridge Research Laboratories, L. G. Hanscom Field, Bedford, Mass., U.S.A.

Giovanelli, R. G., (1939), "The Relation Between Eruptions and Sunspots", *Astrophys. J.*, Vol. 89, U.S.A., pp 555.

Mangis, S. J., (1975), "Introduction to Solar-Terrestrial Phenomena and the Space Environment Services Center", NOAA Technical Report ERL 315-SEL 32, Supt. of Documents, U.S. Govt. Printing Office, Washington, D. C. 20402.

McIntosh, P. S., (1969), "Birth and Development of the Sunspot Group Associated with the Proton Flare of July 1966", *Annals IQSY*, Vol. 3, MIT Press, Cambridge, Mass., U.S.A., pp 40.

McIntosh, P. S., (1972), "Inference of Solar Magnetic Polarities from H-Alpha Observations", *Solar Activity Observations and Predictions*, AIAA Progress in Aeronautics and Astronomics Series, Vol. 30, (P. S. McIntosh and M. Dryer, editors), MIT Press, Cambridge, Mass., U.S.A., pp 65-90.

Mitra, A. P., (1974), *Ionospheric Effects of Solar Flares*, Vol. 46, D. Reidel, Dordrecht, Netherlands.

Muraoka, Y., Murata, H., and Sato, T., (1977), "The Quantitative Relationship Between VLF Phase Deviations and 1 - 8 Å Solar X-ray Fluxes During Solar Flares", *J. Atmos. Terr. Phys.*, Vol. 39, U.S.A., pp 787-792.

Podsiadlo, R. T., (1973), "An Objective Baseline for Flare Prediction", *Space Research*, Vol. 13, Pt. 2, Proceedings of the Fifteenth Plenary Meeting of COSPAR, Akademie Verlag, Berlin, G.D.R., (M. I. Rycroft and S. K. Runcorn, editors), pp 817.

Reilly, A. E., and Enger, I., (1967), "The Influence of Radio Brightness Temperature on Solar Flares Prediction", *Air Force Surveys in Geophysics*, No. 200, Air Force Cambridge Research Laboratories, L. G. Hanscom Field, Bedford, Mass., U.S.A.

Severny, A. B., (1958), "The Appearance of Flares in Neutral Points of the Solar Magnetic Field and the Pinch-Effect", *Izvestiia Krymskaia Astrofizicheskaia Observatoriia*, Vol. 20, U.S.S.R., pp 22.

Severny, A. M., (1962), "Some Peculiarities of Magnetic Fields Connected with Solar Flares", *Astronomicheskii Zhurnal*, Vol. 39, U.S.S.R., pp 961.

Smith, J. B., (1972), "Predicting Activity Levels for Specific Locations Within Solar Active Regions", *Solar Activity Observations and Predictions*, AIAA Progress in Aeronautics and Astronomics Series, Vol. 30, (P. S. McIntosh and M. Dryer, editors), MIT Press, Cambridge, Mass., U.S.A., pp 429-442.

Williams, D. J., (1976), "SELDADS; An Operational Real-Time Solar-Terrestrial Environment Monitoring System", NOAA Technical Report ERL 357-SEL 37, Supt. of Documents, U.S. Govt. Printing Office, Washington, D. C. 20402.

AN UPDATE ON THE SOLAR CYCLE 21 FORECAST

Jay Roderick Hill
EM Propagation Division, Naval Ocean Systems Center
San Diego, CA 92152

The purpose of this paper is to examine the progress of the new solar cycle compared with a mathematical forecast model.

We expect the next solar cycle to be as large as cycle 19 based on the forecast presented in Nature (Hill, 1977). The new cycle is well under way, having started in August, 1976. The forecast was based on a high resolution spectral analysis of monthly sunspot numbers from 1749 through 1975. The rapid rise in monthly mean sunspot numbers is currently running at least as large as that of the forecast model (Figure 1).

The mathematical model consists of a sum of 42 sine waves whose frequencies are not harmonically related. That is, there exists no fundamental period of which the sine waves are multiples; the combination of them will never repeat itself.

The procedure used in deriving the model was developed by Adolf Paul (Paul, 1972) for determining frequencies, amplitudes and phases of unknown lines in a signal containing both broadband noise and pure tones. He has demonstrated its accuracy by analyzing San Francisco tide data. The tides are driven by gravitational forces from the moon and sun. These are known to high accuracy from observations and dynamical theory. The frequencies, amplitudes and phases derived by Paul from the tide data are in startlingly good agreement with those of the astronomical forces.

Paul's method differs from conventional FFT methods in that it does not assume that frequencies will be uniformly spaced. He uses the time shift theorem ($\Delta\phi = 2\pi f \Delta t$) to determine the unknown frequencies from the change of phase between periodogram analysis of segments of the data separated in time.

The first step of his method is to divide the data record into two or more possibly overlapping segments. For example, segment (a) may consist of the first 95% of the record centered at t_a, and segment (b) the last 95% of the record centered at t_b. To reduce the interference of neighboring lines and continuum noise, a data window is applied to each segment before performing the periodogram analysis. Paul uses a cosine (in frequency domain) window having a bandwidth parameter, δ, which reduces "side lobe" interference provided that $\delta \geq 2/L$, where L is the segment length.

Once the windowed segments are calculated, any number of test frequencies can be analyzed by periodograph. For each segment, Paul uses the FFT to efficiently scan the frequency domain in search for possible lines. Improvements in the frequency, amplitude and phase of the lines are made by repeating the periodogram analysis of the segments using as test frequencies those found in the scan. The analysis is performed using overlapping data segments with time shifts of less than about T/16, where T is the length of the data record.

An unknown frequency which is near a periodogram test frequency f_k is determined using the equation $f = (\phi_b - \phi_a)/[2\pi(t_a - t_b)]$, where ϕ_a and ϕ_b are the phases found with the periodogram analysis using segments (a) and (b) and the test frequency f_k. The 2π ambiguities in $\phi_a - \phi_b$ arising from large Δt can usually be resolved without difficulty. By using $(t_b - t_a) \leq 1/f_k$ in the early stages of the solution, the values of f will be determined uniquely. This is because the maximum phase shift of the wave at f_k will be 2π. For the sunspot data the highest frequency analyzed was 0.081468 month^{-1} (period of 12.27 months) so that $t_b - t_a$ could be as large as 222 months without 2π ambiguities arising in $\Delta\phi$. In later iterations when f_k was known more accuratley, larger Δt shifts were used in determining f.

The effects of unresolved closely spaced lines or white noise within the bandwidth δ near f show up as a change in the amplitude estimated in segment (a) compared to that of segment (b). The effects also show up as a change in the line amplitude when the filter bandwidth is changed. In order to determine these closely spaced frequencies, Paul's method was extended by adding steps consisting of subtracting the waveform corresponding to the approximately known lines

in the data, followed by iteratively solving for closely spaced lines using several bandwidths. When applied to the sunspot record, this method resolved closely spaced lines having a beat period ($1/\Delta f$) of 1045±200 years. The accuracy of each frequency is more a function of the white noise within the filter bandwidth used in the analysis than of the record length. The frequency uncertainty estimated for the sunspot data was $\pm 1/(20T)$ rather than the commonly quoted $\pm 1/(4T)$ for standard signal processing.

Although the model is currently tracking the sunspot data quite well, it will require at least another year before we can differentiate a large cycle from a small one. Most cycles, large and small, have a rapid onset with similar slopes. Thus we must wait until after the initial stages of the new cycle to test the model and the assumptions on which it is based.

REFERENCES

Hill, J. R., Nature 266, 151-153 (1977)
Paul, A. K., Maths Comput. 26, 437-447 (1972)

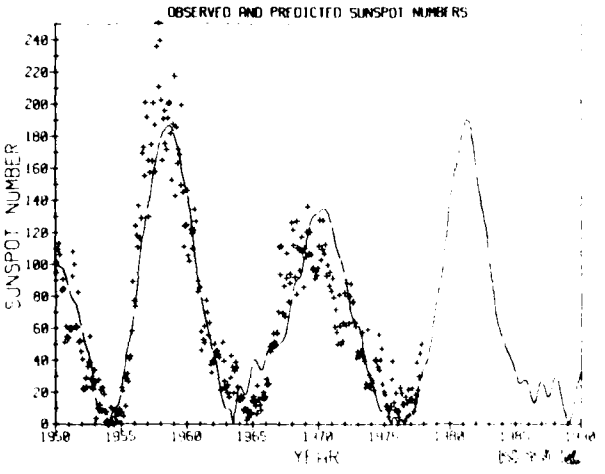


Figure 1. Observed monthly sunspot numbers (+) through October 1977 and Fourier series model (smooth curve) derived from observations from 1749 through 1975.

UHF SCINTILLATION AT HIGH LATITUDES FOR A BRIEF PERIOD DURING 1976

JOHN M. GOODMAN
NAVAL RESEARCH LABORATORY
WASHINGTON, D. C.

ABSTRACT

During two periods - the summer and autumn of 1976 - NRL and NADC collaborated in UHF amplitude scintillation measurements from Keflavik, Iceland using ATS-6 as a source of transmissions. The overall scintillation index S_4 was found to be 0.15 and the 1-99% fading range ~ 3.5dB. As expected the scintillation was found to be more strongly controlled by magnetic than solar activity. The amplitude statistics were found to be non-Rayleigh and the smaller scintillation levels observed in this study were roughly approximated by the Nakagami-m model. The power spectra exhibited a power-law roll-off from a "corner" frequency of 0.067 Hz with an index of -2 on the average. Generally 50% of the scintillation power was provided by frequencies less than 0.03 Hz and the fade-duration and fade-interval distributions were typically exponential.

INTRODUCTION

Under the sponsorship of NAVELEX, the Naval Research Laboratory (NRL), and the Naval Air Development Center (NADC) cooperated in a program on high-latitude scintillation research at UHF. Even though a considerable scintillation data set in the high latitude region has been obtained by workers at the Air Force Geophysics Laboratory (AFGL), principally at VHF, there was deemed to be an insufficient retrievable data base for the frequency domain near 400 MHz. Additionally the Navy studies were to emphasize design quantities derivable from the intrinsic signal statistics and spectra rather than phenomenological parameters such as the scintillation index, which are principally of geophysical interest.

Accordingly NADC set up a data-gathering system at the Keflavik, Iceland Naval Station recording the 360 MHz beacon of the Advanced Technology Satellite ATS-6. The testing intervals were during June and September-October 1976. During the summer interval, ATS-6 was geostationary, being positioned over Lake Victoria (Africa), yielding an elevation angle of approximately ~ 5° to the Keflavik installation. During the autumn interval the satellite was being repositioned to the west and the elevation angles were somewhat higher, ranging between about 14° to 16°.

The antenna system employed for the measurements was characterized by a beamwidth of approximately 13.5°. As a result the studies were potentially contaminated by multipath effects during the summer tests because of ground reflections. This realized at the outset and was recognized to be an observational constraint to be analyzed for the case of small disadvantaged users at high latitudes. Another constraint to the success of the measurement program was the aircraft traffic at the Naval station airfield, which produced unwanted receiver oscillations before adjustments were made to obviate the problem. Nevertheless a useful data set was obtained for processing by NRL personnel.

SCINTILLATION INDEX CONSIDERATIONS

During the overall observation periods the average scintillation index S_4, defined as the ratio of the standard deviation of the signal fluctuations to the average power, was 0.15. Also, the so-called SI scintillation index, employed by AFGL scientists and related to the full fading range, was about twice the value of the S_4 scintillation index on the

average. Other average statistics were an RMS signal fluctuation of 0.75 dB, a 1%-99% fading range of 3.5 dB, and a 1% fading depth of 1.8 dB. In addition, the peak observed fading range was 7.5 dB.

On the basis of a correlation analysis, the following relationship was found to exist between the scintillation index and the magnetic and solar activity indicies.

$$S_4 = 0.05 + 4.8X10^{-3}\Sigma k_p + 5.7X10^{-4} R_z \quad (1)$$

where Σk_p is the sum of eight 3-hourly values of the planetary magnetic index following the observation period, and R_z is the value of the smoothed Zurich sunspot number for the immediate day preceding the observations. This relation is approximate because of the limited number of data groups correlated but nevertheless indicates that magnetic activity is a stronger control than solar activity at these high latitudes, perhaps by an order of magnitude. That Magnetic activity was a stronger control since the nocturnal scintillation which was emphasized during this study is undoubtedly related to the nighttime location of the auroral oval which blankets Iceland and is a visible manifestation of magnetospheric influences. Indeed other workers have noted that magnetic activity is more important than solar activity in controlling the high-latitude scintillation depth at VHF.

Figure 1 shows the distribution of S_4 indicies obtained during the total observation period for which an analysis was conducted. Upon inspection, it is clear that scintillation was never particularly strong, with S_4 exceeding 0.25 less than 5% of the analysis time.

Amplitude Distributions

The amplitude statistics were observed to be non-Rayleigh at 360 MHz; however for the smaller scintillation levels observed in this study the distributions were closely approximated by the Nakagami-m model:

$$p(p) = m^m \left(\frac{p}{p_o}\right)^{m-1} \Gamma(m)^{-1} e^{-mp/p_o} \quad (2)$$

where $\Gamma(m)$ is the gamma function, m is a parameter and p and p_o are the instantaneous and average powers respectively. It is well known that

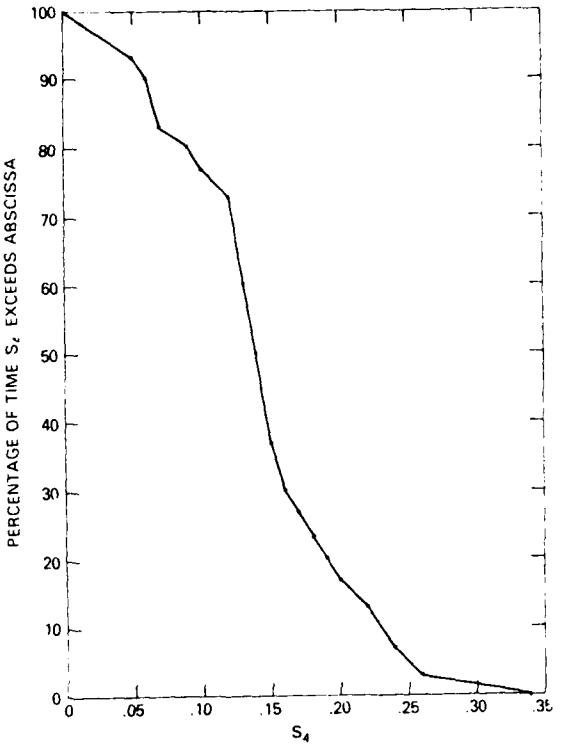


Figure 1 - Distribution of Scintillation Indicies

$m = S_4^{-2}$ and since $\langle S_4 \rangle \approx 0.15$ during the present observations then $\langle m \rangle$ is ≈ 44. An m value of unity constitutes Rayleigh fading. The extreme values of S_4 were 0.05 and 0.34, suggesting that m ranged between 400 (virtually zero scintillation) and 8.65 during the periods of observation. For the lower values of scintillation observed in the current investigation, a log-normal distribution would be an equally valid approximation to the observed situation. A typical distribution function is presented in Figure 2.

Scintillation Spectra

The power spectra in the current study were generally characterized by a high-frequency rolloff of f^{-2} rather than f^{-3} as was anticipated from our equatorial studies and some work of others at equatorial and midlatitudes. The low-frequency cutoff was not well defined but its approximation, called the corner frequency, was 0.06 Hz on the average. Typical spectra are displayed in Figure 3. Using the Fresnel Zone distance as the canonical scale

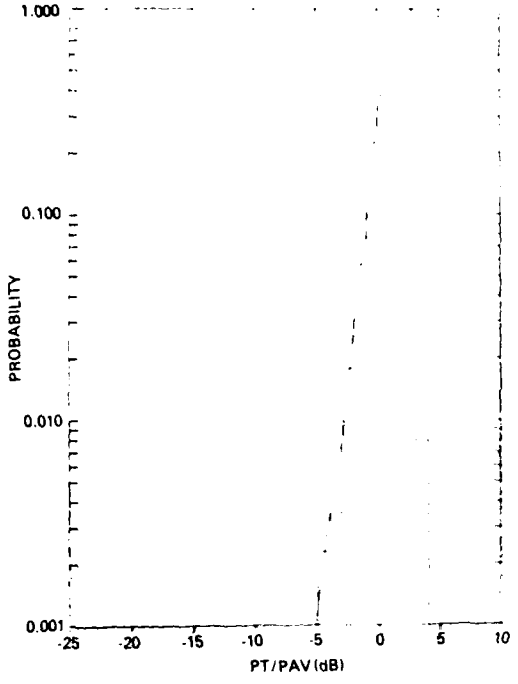


Fig. 2 - Probability density and Distribution functions for a 30 thirty minute period on the evening of 6/23/76. S_4 = 0.22

above which scintillation effects become filtered, we deduce the average drift velocity of the inhomogeneity field across the beam to be about 200 m/s. Correlation was insignificant between the rolloff exponent and either R_z or k_p. However for $k_p \geq 2$ the drift velocity of the ionosphere as determined from the spectral corner frequency appears to increase approximately 25 m/s per unity increase in the daily sum Σk_p.

From an analysis of the distribution of scintillation versus fluctuation frequency, we find that most of the scintillation power resides below the corner frequency ν_c , or for periodicities in excess of approximately 15 s. The S_4 index was found to obey a law of the form

$$S_4(\nu) = S_4 \ (1-e^{-\lambda\nu}) \qquad (3)$$

where ν is the fluctuation frequency S_4 is the observed total scintillation index, $S_4(\nu)$ is the scintillation index from DC to ν, and λ is a parameter. Typically λ is ν_c^{-1} or 15 s.

Figure 4 illustrates this ascertion.

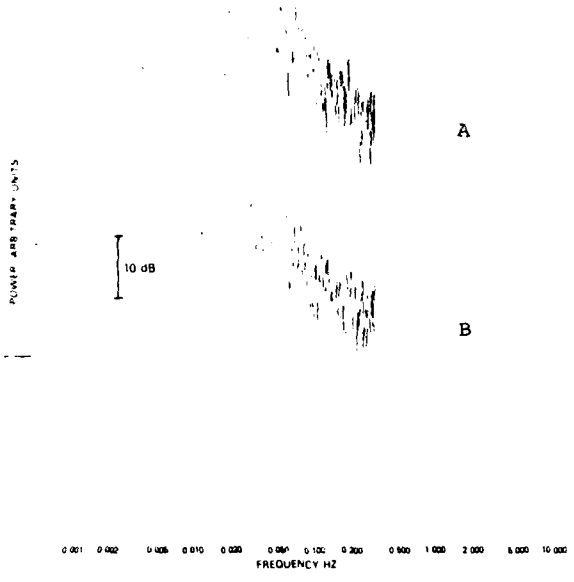


Fig.3 Power spectra for (A) 2105 - 2136Z on 9/29/76 and (B) 2137-2208 on 9/29/76. The S_4 values were 0.24 and 0.15 respectively.

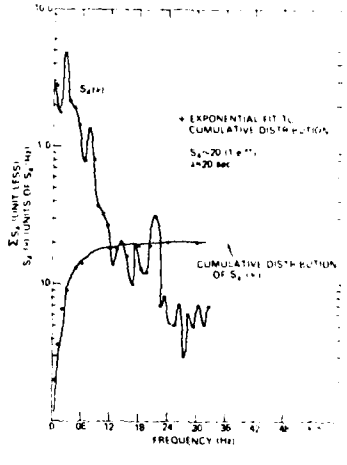


Fig. 4 - S_4 index per elementary frequency band versus ν. Note that the cumulative distribution of $S_4(\nu)$ asymptotically approaches S_4.

Fade Duration and Fade Interval Statistics

The Fade-duration and fade-inter-

val statistics were also shown to be exponential, although the slopes were different in the high-frequency and low-frequency domains. The fade-interval distributions indicate that there is a finite probability that severly disadvantaged users will benefit from scintillation, because constructive interference may increase the signal strength to a level which enhances signal detection.

Figure 5A is plot of the fractional number of fade intervals (i.e., periods between fades) as a function of the interval length with system detection threshold as parameter. The OdB case corresponds to a detection threshold at the rms signal level. Clearly as the system margin increases, (i.e., the detection threshold becomes more negative with respect to the rms signal level), the slope of the fade interval distribution decreases allowing a greater number of longer intervals during which fading is non-existent.

Figure 5B shows the fade duration distribution corresponding to the same date which was utilized to construct Figure 5A.

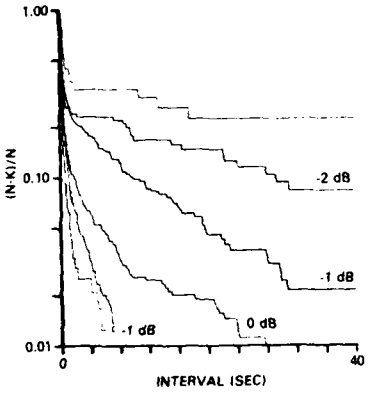


Fig. 5A - Fade interval distributions for 2149 - 2220Z on 6/23/76 during which time S_4 = 0.22.

Figure 6 is a plot of the number of fades as a function of fade duration for three periods of relatively high scintillation activity. A thirty (30) minute period is analyzed in each case.

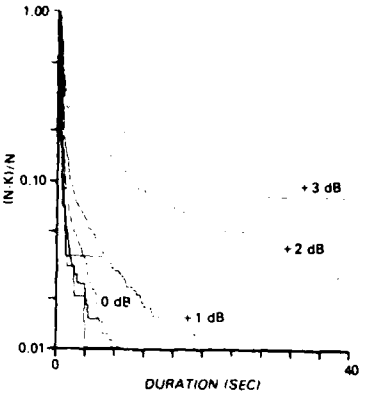


Fig. 5B - Fade Duration Distribution for 2149 - 2220Z on 6/23/76 during which time S_4 - 0.22.

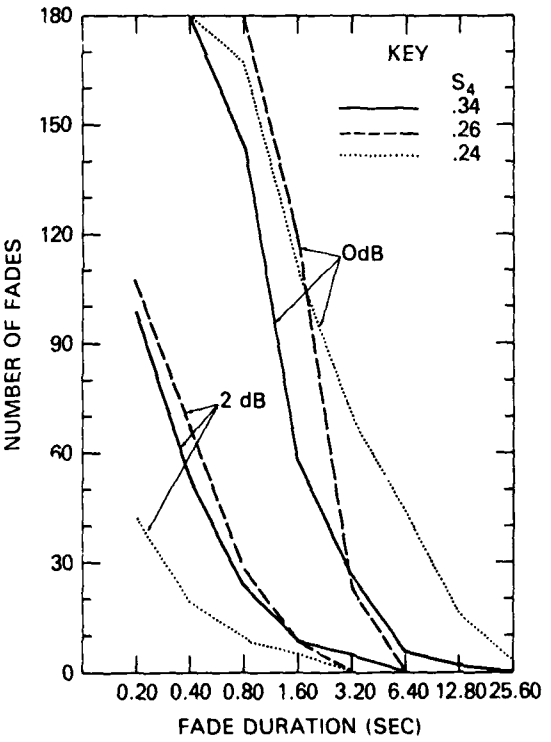


Fig. 6 - Number of fades occurring during a 30 minute interval having duration greater than the abscissa. Detection threshold is the parameter and is referenced to the rms signal level.

Concluding Remarks

It is emphasized that the current study is incomplete as of this writing and further analysis would clearly be helpful in increasing the validity of the conclusions drawn. Additionally the data described herein were obtained over a relatively short period of time during minimum solar activity, at a high latitude and for low elevation conditions. Extrapolations to other situations should be done cautiously.

Acknowledgements

The author would like to thank R. Hogg of NADC for providing the field measurements and P. L. Watkins and C. G. Myers of NRL for data processing assistance. This work was supported by NAVELEXSYSCOM PME 106 and ONR.

EFFECTS OF IONOSPHERIC SCINTILLATIONS ON SATELLITE COMMUNICATION

Banshidhar, N.M. Vadher, Hari Om Vats, M.R. Deshpande and R.G. Rastogi
Physical Research Laboratory
Ahmedabad 380009, India.

TEXT

The paper describes some results of the investigations carried out using amplitude recordings of radio beacons from ATS-6 (phase II) at an equatorial station Ootacamund, India. The effect of ionospheric irregularities has been studied in relation with the message length and its reliability for various fade margins at three carrier frequencies 40, 140 and 360 MHz. Out of various schemes to overcome the scintillation effect on communication, the time diversity scheme has been discussed.

From satellite to earth communication the radio waves have to pass through the ionosphere where ionisation irregularities might exist. These irregularities will scatter and distort the signal producing fading. This fading occurs primarily in the equatorial region of the earth, 20° of the magnetic equator. Similar fading occurs in the polar regions also. The fading is more severe during night than day. Nighttime scintillations are attributed to irregularities mostly present in F-region, whereas daytime scintillations are due to E-region irregularities. This morphology has been documented by numerous workers (Aarons et al. 1971, Koster 1972, Crane 1974, Aarons 1976, Rastogi et al. 1977). Here the effect of one specific example of severe scintillation simultaneously on 40, 140 and 360 MHz is discussed.

These amplitude records are shown in Fig.1. The scintillation activity is generally measured as normalized root mean square of power (Briggs and Parkin 1963) known as:

$$S_4 = \frac{1}{\overline{R^2}} \frac{}{(R^2 - \overline{R^2})^2}^{\frac{1}{2}}$$

where R is the instantaneous amplitude, and $\overline{R^2}$ is the average of squares of R.

The records shown in Fig.1 have these S_4 values as 0.97, 0.85 and 0.47 for 40, 140 and 360 MHz respectively. The effect of these scintillations on communication can be visualized by determining the message reliability. Message reliability is measured by determining the number of intervals that completely fit within the signal enhancements or increases above specified calibration levels compared with the total possible number in the sample under investigation (15 minutes) (Whitney and Basu 1977). Fig.2 shows the percentage of messages perfectly received as a function of message length for various fade margins. Here percentage of perfectly received message at 140 MHz reduced to less than 50% for message lengths exceeding 4 seconds and fade margin less than 6 db. Under the same condition, these percentages are 90 and 10 for 360 and 40 MHz respectively. A more realistic view of these effects can be simulated computationally as if a message transmitted from ATS-6 satellite and received by a receiver on the ground having a specified fade margin. One such simulation is shown in Fig.3, where each letter represents a message of 0.4 sec. Here only 140 MHz data from Fig.1 is used. Fig.3a shows the message which would have been received by a receiver having a fade margin of 15 db and Fig.3b shows the same for a fade margin of 4 db only. In Fig.3b 'X's represent non-receivable part of the message. Note that one way to combat scintillations is to increase fade margin.

Without increasing the fade margin, diversity reception also to some extent

help to get rid of scintillations. Diversity schemes attempt to reduce the effects of fading during a scintillation event by combining two signals that are fading independently. Out of three diversity schemes, viz., space, time and frequency, here only time diversity scheme would be discussed, wherein the same message be sent at two different times separated far enough such that the correlation coefficient to be less than 0.6. For this same sample was processed for correlation analysis. Fig.4 shows the auto-correlation curve for 40, 140 and 360 MHz. The cross-correlation coefficients for these records are 0.030, 0.127 and 0.029 for 40/140 MHz, 140/360 MHz and 40/360 MHz sets respectively. From auto-correlograms of 40, 140 and 360 MHz (Fig.4) the time lags for correlation coefficient 0.6 were obtained and using this time for diversity scheme message reliabilities versus message lengths were obtained for various fade margins. Fig.5 shows the time diversified message reliability plotts for 40 MHz. One can note that the reliability has improved to a considerable amount. These results are preliminary and detailed study is under progress.

Acknowledgements

Grateful thanks are due to Prof. G. Swarup, Radio Astronomy Centre, Ootacamund and Prof.K. Davies, SEL/NOAA, Boulder, for facilities and collaboration respectively for ATS-6 Ootacamund project. Thanks are also due to Miss Chhaya R. Shah for her sincere programming assistance during the course of this work. This project was supported by DOS and NASA.

References

Aarons J., H.E. Whitney and R.S. Allen, (1971), Global morphology of ionospheric scintillation, Proc. IEEE 59, 159-172.

Aarons J., (1976), Equatorial scintillation: A review, IEEE Trans. Antennas Propagat. (in press).

Briggs B.H. and I.A.Parkin, (1963), On the variation of radio star and satellite scintillations with zenith angle, J. Atmos. Terr. Phys., 25, 339.

Crane R.K., (1974), Morphology ionospheric scintillation, Tech. Note 1974-29, DDC AD-780522 19, MIT,Lincoln Laboratory, Leseington, MA. 02173.

Koster J.R., (1972), Equatorial scintillations, Planet. Space Sci., 20, 1977-2014.

Rastogi R.G., M.R.Deshpande, Hari Om Vats, K. Davies, R.N. Grubb and J.E. Jones, (1977), Amplitude scintillations of ATS-6 Radio beacon signals within the equatorial electrojet region (Ootacamund, dip 40°N), Pramana, 8, 1-13.

Whitney H.E. and Santimay Basu, (1977), The effect of ionospheric scintillation on VHF/UHF satellite communications, Rad. Sci., 12, 123-133.

ATS-6 OOTACAMUND 11 OCT 1975

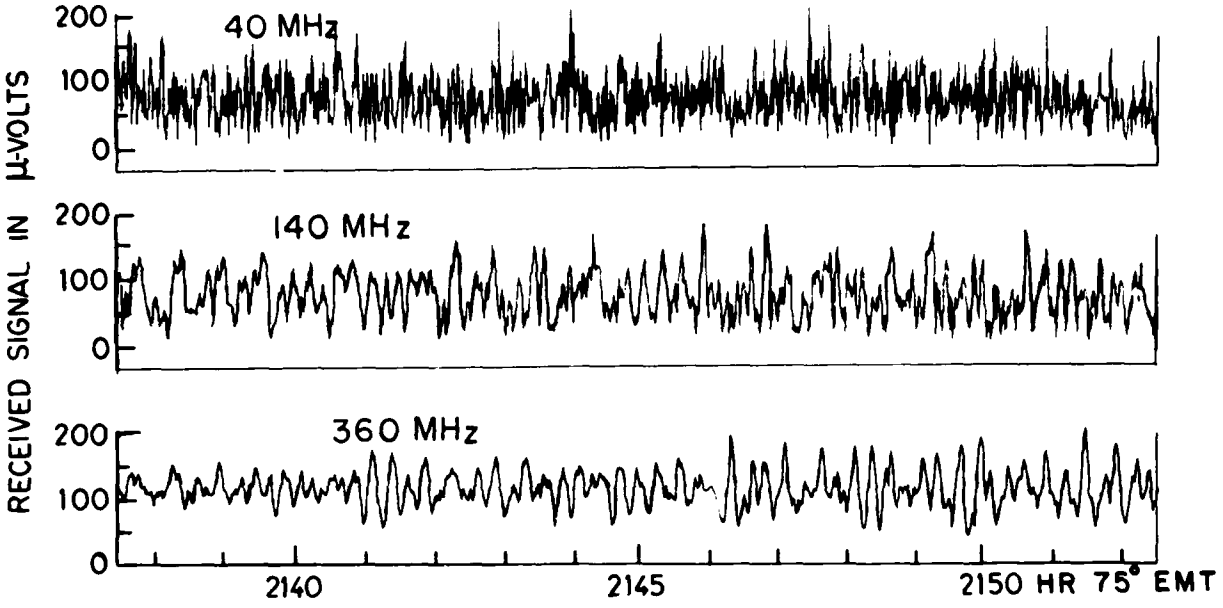


FIG. 1. AMPLITUDE RECORDS

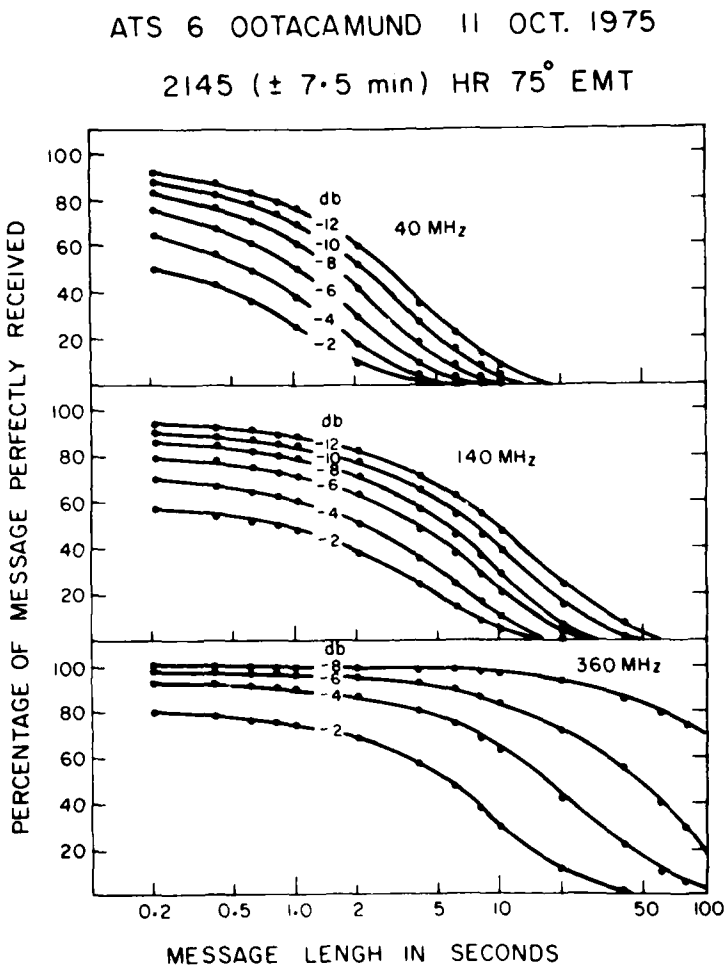


FIG. 2. MESSAGE RELIABLITY PLOTS FOR VARIOUS FADE MARGINS

(a)

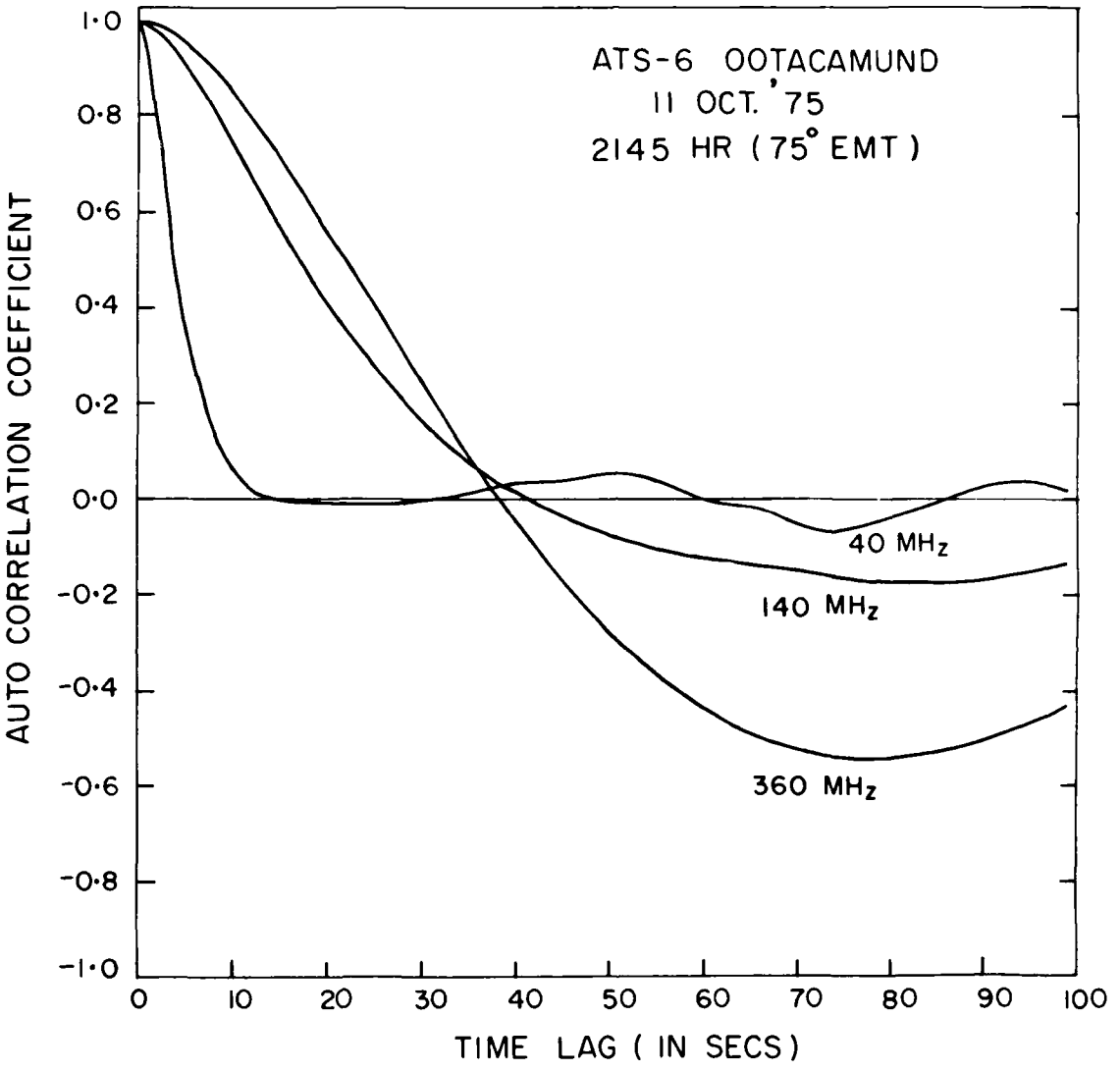
IONOSPHERIC SCINTILLATION IS PREDOMINANTLY A NIGHTTIME EFFECT NORMALLY OCCURRING A FEW HOURS AFTER LOCAL SUNSET AND DISSIPATING PRIOR TO SUNRISE THE EXTENT OF THIS PHENOMENA IS DEPEDENT UPON THE TIME OF YEAR USUALLY PEAKING DURING THE SPRING AND FALL EQUINOX PERIODS IT TENDS TO BE AT A MINIMUM DURING THE SUMMER AND WINTER IT IS ALSO EFFECTED BY THE SUNSPOT ACTIVITY, MAGNETIC STORMS AND A HOST OF OTHER MINOR FACTORS EVEN THOUGH IONOSPHERIC SCINTILLATION FADING IS A FUNCTION OF FREQUENCY GEOGRAPHICAL LOCATION, AND TIME OF OCCURRENCE, THERE ARE MANY MOBILE COMMUNICATION SYSTEMS WHICH ARE REQUIRED TO OPERATE IN THE VULNERABLE GEOGRAPHICAL REGIONS AND TIME ZONES

(b)

IONOXXXERXX SCINTILLATION XXXXXXXOMXXXXXXXXXXXXGHTTIME EFFECXXXXXXXXXX OXCURRING XXFEW HOURS AFTERXXXXXL SUNSET XXX DISSIPATING PRXXXXXXSUNRISEXXTHE EXTENXXXF XXIS PHENOMXXXXIS DEPXXXXXT UPON THE TIME OF YEARXXXXXLLY PEAKINXXXXRING XXXXSPRING XXXXXALL EQUINOX PERIODS IT TEXXXXXXXXX AT A MINIMXXXXURXXX THE SUMMEXXAND WINTER XXX ISXXXXO XXXXXTED BXXXXXXSUNSPOT XXTIVITYXXXXXXXXXXX SXORMSXXANDXXXHOST OF OTHER MINOR FACTORXX XXXX THXXGH IONOSPHEXXXXXXXXTILLATION XXXXXXX IS A FUNXXXON OXXFREQUENCYXX GXXXXXPHICALXXOCATXXN, AXXXXXMXXXXXXXXUXXXNCE. XXXERE XXXXXXXX MOBILEXXXOMXXXICATION SYSXXXXXXXXXH ARXXXEXUIRED TO OPERATEXXX THEXXXXXXRABLE GEXXAPHICAXXXXGIXXX XNXXXXXE ZONES

FIG. 3. COMPUTER SIMULATION OF MESSAGE RECEIVED BY A RECEIVER HAVING FADE MARGIN (a) 15 dB (b) 4 dB

FIG. 4. AUTO - CORRELOGRAMS OF AMPLITUDES



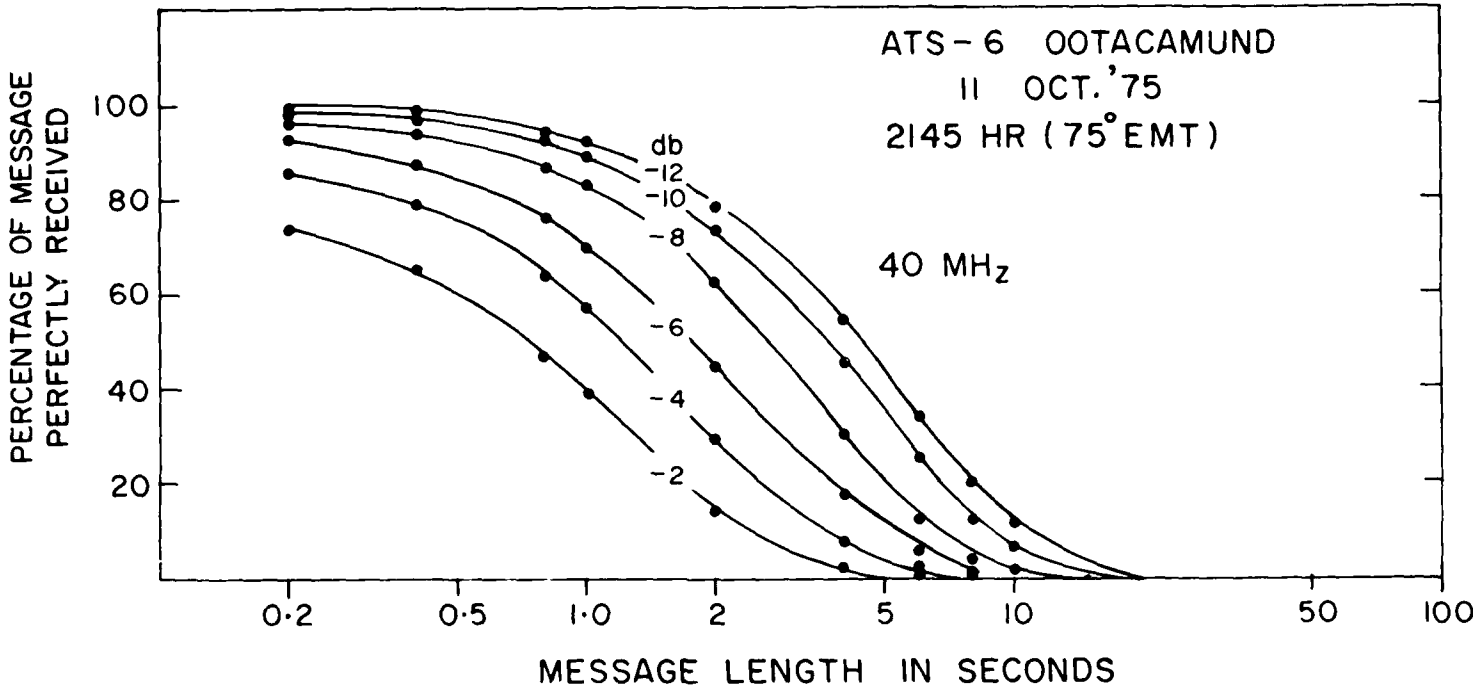


FIG. 5. MASSAGE RELIABLITY PLOTS USING TIME DIVERSITY SCHEME FOR 40MHz RECORD SHOWN IN FIG. 1.

EFFECTS OF THE JOVIAN IONOSPHERE ON ENTRY PROBE COMMUNICATIONS

Joseph H. Yuen
Richard Woo
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91103

Terry Grant
Ames Research Center, NASA
Moffett Field, California

Abstract

Although much progress has been made in understanding the earth ionosphere, relatively little is known about planetary ionosphere and its effects on planetary communication systems.

This paper (1) analyzes the scintillations of the radio signals received from Pioneer 10 during radio occultation by Jupiter and estimates from these results the scintillation of the signal from a direct entry probe, and (2) assesses the effects of Jovian ionospheric scintillation on entry probe communications.

Estimates indicate ionospheric scintillations are small at S-band, but could be substantial at 400 MHz. Its amplitude fluctuations are characterized by a log-normal distribution. The power spectra of this log-amplitude fluctuation indicates a bandwidth narrower than 1 Hz. The phase fluctuations can be modeled as a Gaussian random process. For the Jovian weak scintillation, the phase fluctuations can be readily tracked by a phase-locked loop receiver to maintain signal phase coherence.

The effects of Jovian ionospheric scintillation on Probe-to-Orbiter telemetry depends on signal frequency, modulation technique, coding scheme, and receiver design. For typical planetary communication system that employs convolutional code and phase coherent receiver, we find the additional signal-to-noise ratio per bit required to achieve the identical data quality in terms of bit error rate is in the order of 1 to 4 dB as compared with that for no fading, depending on the signal frequency.

Introduction

In order to properly design communication links, every factor contributing to the performance of the link must be considered. In addition to the usual receiver thermal noise disturbance, the Jupiter Probe-to-Orbiter link in the upcoming unmanned space flight to Jupiter (JOP) is expected to experience ionospheric scintillation in Jovian environment. Although much progress has been made in understanding the earth ionosphere, relatively little is known about the planetary ionosphere and its effects on planetary communication systems.

This paper first analyzes the scintillations of the radio signals received from Pioneers 10 and 11 during radio occultation by the atmosphere of Jupiter and estimates from these results the scintillation of the signal from a Jovian entry probe to orbiter spacecraft. Secondly, this paper assesses the effects of the Jovian ionospheric scintillation on entry probe telemetry.

In the scintillation analysis, four steps are necessary. First, the radio occultation data is processed and the frequency spectrum of the amplitude fluctuations is computed. Second, the theoretical spectrum is obtained by studying wave propagation in random media. Third, a turbulence model for the atmosphere is constructed by comparing the results from these first two steps. Finally, scintillation estimates for an entry probe link are estimated based on the turbulence model.

Ionospheric scintillation induces fading effects on communication signals. These must be taken into account in the selection of signal frequency, modulation technique, coding scheme and receiver design. In order to incorporate this fading effect in a telecommunication link design, we must first calculate the value of σ_χ^2 which characterizes the ionospheric effects based on specific probe environment. Then we find the additional signal-to-noise ratio per bit E_b/N_o required to achieve the identical data quality in terms of bit error rate P_e, for the particular telecommunication system used as compared with that for no fading. Also, adverse tolerance bounds must be given to minimize design risk.

Irregularity Model and Entry Probe Estimates

The study of the scintillation of the radio signals received from Pioneer 10 during radio occultation by the atmosphere of Jupiter is in principle similar to that carried out for the Pioneer Venus Multiprobe mission based on the Mariner 5 radio occltation data (Refs. 1 and 2). The Jovian electron density irregularities deduced from the Pioneer 10 radio occultation has been reported in Ref. 3. Briefly, Ref. 3 found that the ionospheric irregularities existed over an altitude range of 3000 km on the late afternoon side probe by Pioneer 10 and over an altitude range of 2000 km in the early morning side. The structure constant C_n of the refractive index fluctuations, which characterizes the strength of the fluctuations, lies in the range 3.3×10^{-9} to $4.9 \times 10^{-9} m^{-1/3}$ in the regions of ionosphere examined. The spatial wavenumber spectrum of the electron density irregularities is very close to the Kolmogorov spectrum and the outer scale size L_o is greater than 6 km.

The entry probe configuration is shown in Fig. 1. We will make scintillation estimates for a probe located at the zero altitude level, which corresponds to the level in the atmosphere of Jupiter where the refractivity is 10 (Refs. 3 and 4). The radial distance to this point on the late afternoon side is about 70,435 km. We will assume that the irregularities extend to 3000 km and that $C_n = 5 \times 10^{-9} m^{-1/3}$ and remains constant throughout this altitude range.

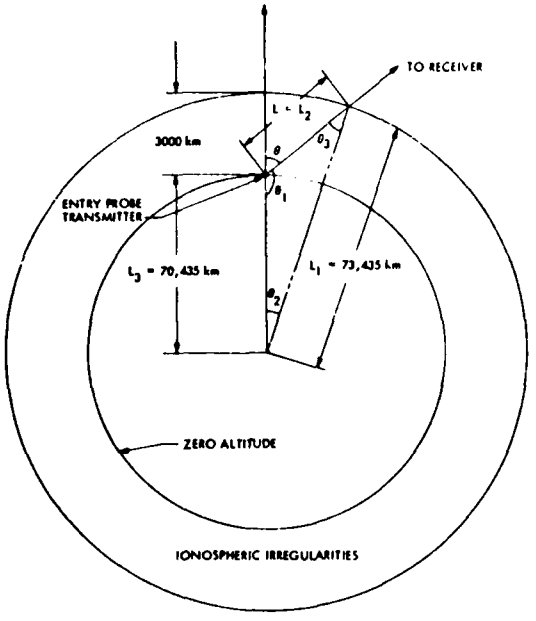


Fig. 1. Jupiter entry probe configuration

The raypath distance L through the ionospheric irregularities is an important parameter for estimating the scintillation of an entry probe. In terms of the communications angle θ, which is the angle the raypath makes with the zenith (see Fig. 1), L can be computed through the following relationships

$$\left.\begin{aligned} \theta_1 &= 180 - \theta \\ \theta_3 &= \sin^{-1}\left(\frac{L_3}{L_1}\sin\theta_1\right) \\ \theta_2 &= 180 - \theta_1 - \theta_3 \\ L &= L_1 \frac{\sin\theta_2}{\sin\theta_1} \end{aligned}\right\} \quad (1)$$

where θ, θ_1, θ_2 and θ_3 are in degrees. We have computed L for various values of θ and the results are shown in Table I.

The Fresnel size $\sqrt{\lambda L}$ corresponding to the θ = 90° case (L = 20775 km) is 1.646 km at S-band (2.3 GHz) and 3.95 km at 400 MHz. L_o is greater than these Fresnel sizes since from Pioneer 10 we have deduced that L_o exceeds 6 km. Under this condition and assuming the Kolmogorov spectrum, the variance σ_χ^2 of the log-amplitude fluctuations for the entry probe is (Ref. 1):

$$\sigma_\chi^2 = 0.308\, c_n^2\, k^{7/6}\, L^{11/6} \quad (2)$$

where k is the free space wavenumber. Because of Fresnel filtering, scale sizes larger than $\sqrt{\lambda L}$ do not contribute to σ_χ^2. Consequently, (2) is independent of L_o. For the phase fluctuations, Fresnel filtering does not occur and the variance σ_ϕ^2 of phase fluctuations depends on L_o (Ref. 1):

$$\sigma_\phi^2 = 0.782\, k^2\, L_o^{5/3}\, c_n^2\, L - \sigma_\chi^2 \quad (3)$$

Table I. Summary of the Calculations of L in Terms of θ

θ (deg)	L (km)
0	3000
30°	3441
60°	5671
90°	20775

where σ_ξ^2 is given by (2). It should be emphasized that (2) and (3) are only valid for weak fluctuations and the condition for weak fluctuations is that $\sigma_\chi^2 < 1$.

The refractive index n is $[1 - (\omega_p/\omega)^2]$ where ω_p is the plasma frequency and $\omega = kc$ (c is the speed of light). For $(\omega_p/\omega)^2 \ll 1$,

$$\sigma_n^2 \simeq \frac{1}{4}\left(\frac{\omega_p}{\omega}\right)^4 \frac{\sigma_{n_e}^2}{n_e^2} \qquad (4)$$

where n_e is the electron density, and σ_n^2 and $\sigma_{n_e}^2$ are the variances of the refractive index and electron density fluctuations, respectively. From the spatial wavenumber spectrum it can be shown that (Ref. 5)

$$c_n^2 \simeq 1.9\, \sigma_n^2\, L_o^{-2/3} \qquad (5)$$

It can be seen, therefore, that $c_n^2 \propto k^{-4}$.

The value of structure constant $c_n = 5 \times 10^{-9}\ m^{-1/3}$ derived from Pioneer 10 is for S-band only. For 400 MHz $c_n = 5 \times 10^{-9} \times F^2$ where F = 2300/400. We have computed σ_χ^2 and σ_ϕ^2 at S-band and 400 MHz for various values of θ. For the phase fluctuations we have made the calculations for two cases of L_o: 10 and 100 km. The results are summarized in Table II. σ_χ can be converted to the rms fluctuation in dB σ_{dB} by multiplying σ_χ by the factor 20/ln 10 = 8.7. As can be seen from Table II, although the amplitude fluctuations at S-band are small, they become substantial at 400 MHz. On the other hand, the phase fluctuations are large at both S-band and 400 MHz. It should be pointed out that even though the calculated results for σ_χ^2 at 400 MHz indicate that σ_χ^2 exceeds one for large θ, such values will not occur since σ_χ^2 saturates before reaching one (Ref. 6). When saturation occurs, (2) and (3), which are derived from weak fluctuation theory, are no longer valid and alternate results using strong fluctuation theory must be derived.

The frequency spectra of the log-amplitude and phase fluctuations [$W_\chi(f)$ and $W_S(f)$ respectively] are also important to the communication link design. The bandwidth of the log-amplitude fluctuations f_{3dB} is given by (Ref. 1)

$$f_{3dB} = 0.294\, v_\perp \sqrt{\frac{k}{L}} \qquad (6)$$

where $v_\perp$ is the wind velocity transverse to the line-of-sight path. Computed values of f_{3dB} assuming $v_\perp$ = 100 m/sec are summarized in Table III and indicate that the spectrum is rather narrow.

The complete spectra of the log-amplitude and phase fluctuations are given in Ref. 1 (Figs. 3 and 4 in Ref. 1, respectively). These spectra are normalized with respect to $f_\perp$ and β which are

$$f_\perp = \frac{v_\perp}{2}\sqrt{\frac{k}{L}}$$

$$\beta = \frac{1}{L_o}\sqrt{\frac{L}{k}}$$

so that the spectra corresponding to the different specific cases can readily be obtained. It is interesting to note that the spectrum of the rather large phase fluctuations shown in Table II is also very narrow.

Table II. Summary of the Calculations for σ_χ^2 and σ_ϕ^2

θ (deg)	σ_χ^2		σ_ϕ^2 (rad^2) L_o = 10 km		σ_ϕ^2 (rad^2) L_o = 100 km	
	S-band	400 MHz	S-band	400 MHz	S-band	400 MHz
0	5.28×10^{-4}	7.65×10^{-2}	0.625	20.9	29	969.4
30	6.8×10^{-4}	9.84×10^{-2}	0.72	23.95	33.4	1111.9
60	1.7×10^{-3}	0.246	1.185	39.48	55	1832.5
90	1.83×10^{-2}	2.66	4.34	144.63	201.6	6713

Table III. Summary of f_{3dB} ($v_\perp$ = 100 m/sec)

θ (deg)	f_{3dB} at S-band (Hz)	f_{3dB} at 400 MHz (Hz)
0	0.117	0.05
30	0.11	0.05
60	0.085	0.036
90	0.045	0.019

Communication Link Design Considerations

In order to properly design communication links, every factor contributing to the performance of the link must be considered. For the Jupiter Probe-to-Orbiter link in JOP, it is expected to experience ionospheric scintillation in the Jovian environment. This section presents some estimates of the communication link degradation due to Jovian ionospheric scintillations to aid the probe communication link design. We first calculate the value of σ_χ^2 which characterizes the ionospheric effects based on the specific probe environment. Then we find the additional signal-to-noise ratio per bit E_b/N_o required to achieve the identical data quality in terms of bit error rate P_e, for the particular telecommunication system used as compared with that for no fading. Also, adverse tolerance bounds are given to minimize design risk. We proceed to estimate these quantities.

For the weak scintillation assumption, an appropriate model in characterizing the ionosphere is the log-normal distribution which is completely characterized by the parameter σ_χ^2. The radio signal path distance L through the ionospheric irregularities is an important parameter for estimating the scintillation of an entry probe. Under the assumption that the turbulence is uniform and isotropic throughout and assuming the Kolmogorov spectrum, the variance of the log-amplitude fluctuations for the entry probe is given by eq. (2) expressed in terms of signal frequency f,

$$\sigma_\chi^2 = 0.308\left[5 \times 10^{-9}\left(\frac{2.3\times10^9}{f}\right)^2\right]^2 \left(\frac{2\pi f}{3\times10^8}\right)^{7/6} L^{11/6}$$

with f in Hz, L in meters.

Because the link design is still under study, several assumptions must be made to illustrate the range of possible degradation due to scintillation. Typical frequencies of 1 GHz and 1.7 GHz are chosen. For a design point that the orbiter is right at the zenith of the probe, L = 3000 km, we have

$$\sigma_\chi^2 = \begin{cases} 1.25 \times 10^{-3} & \text{for } f = 1.7 \text{ GHz} \\ 5.64 \times 10^{-3} & \text{for } f = 1 \text{ GHz} \end{cases}$$

The value of L is a function of the communication angle θ. Our scintillation estimates in the previous section must be regarded as order-of-magnitude only, since the irregularity model was constructed on the basis of the interpretation of just a portion of the Pioneer 10 data which were taken at a particular time of the day, a particular day of the month, a particular month of the year, etc. Allowing for the order-of-magnitude uncertainty of the estimate, the adverse values of σ_χ^2 are listed in Table IV.

The power spectra of this log-amplitude fluctuation indicates a bandwidth narrower than 1 Hz. The phase fluctuations can be modeled as a Gaussian random process. For the Jovian weak scintillation, the narrow band phase fluctuations can be readily tracked by a phase-locked loop receiver to maintain signal phase coherence.

A method to analyze the performance of phase-locked loop receivers in the presence of lognormal fading communication channels has been reported (Refs. 7 and 8). Degradations for coherent BPSK systems using convolutional codes with sequential and Viterbi decodings have been assessed (Refs. 9 and 10). These results indicate that the communication degradations due to weak fadings are tolerable in strong phase-locked loop signal-to-noise ratio.

The Jupiter entry probe is assumed to use a coherent BPSK system, K = 6, R = 1/3 convolutional code with Viterbi decoding. For the specified data quality $P_e = 10^{-3}$, we compute the signad degradations due to ionospheric scintillation in terms of signal-energy-to-noise-ratio per bit, E_b/N_o for various values of σ_χ^2 in Table IV. The results are summarized in Table V for a typical strong link with loop SNR = 15 dB. It is seen that for typical trajectory which limits $\theta < 30^o$, the adverse degradation is only 1.1 dB for signal frequency f = 1.7 GHz. This degradation is as large as 4.3 dB for f = 1 GHz.

Table IV. Adverse Values of σ_χ^2

	σ_χ^2	
θ (deg)	f = 1 GHz	f = 1.7 GHz
0^o	0.056	0.0125
30^o	0.072	0.016
60^o	0.18	0.04
90^o	1.94	0.43

Table V. Ionospheric Scintillation Degradation (for a coherent BPSK communication system using K = 6, R = 1/3 code with Viterbi Decoding, Loop SNR = 15 dB, $P_e = 10^{-3}$)

Signal frequency f		σ_χ^2	Ionospheric Scintillation Degradation (dB)
1.7 GHz	Design Point	1.25×10^{-3}	0.1
	Adverse $\theta = 0^o$	1.25×10^{-2}	0.9
	Adverse $\theta = 30^o$	1.6×10^{-2}	1.1
	Adverse $\theta = 60^o$	4.0×10^{-2}	2.7
1 GHz	Design Point	5.64×10^{-3}	0.3
	Adverse $\theta = 0^o$	5.64×10^{-2}	3.5
	Adverse $\theta = 30^o$	7.2×10^{-2}	4.3

Conclusions

The analysis of Pioneer 10 data indicates Jupiter ionospheric sintillations are small at S-band, but could be substantial at 400 MHz. Its amplitude fluctuations are characterized by a lognormal distribution. The power spectra of this lognormal fluctuation indicates a bandwidth narrower than 1 Hz. For this weak scintillation, the phase fluctuations can be readily tracked by phase-locked loop receiver to maintain signal phase coherence.

The effects of Jovian ionospheric scintillation on an entry Probe-to-Orbiter telemetry system employing coherent BPSK receiver and convolutional code, K = 6, R = 1/3, with Viterbi decoding have been assessed. For the expected adverse conditions, we find the degradation due to ionospheric scintillation in terms of signal-to-noise ratio per bit to achieve a bit error rate $P_e = 10^{-3}$, can be limited to 1.1 dB for one choice of signal frequency and mission trajectory design.

References

1. Woo, R., Kendall, W., Ishimaru, A., and Berwin, R., "Effects of Turbulence in the Atmosphere of Venus on Pioneer Venus Radio - Phase I", Jet Propulsion Laboratory Technical Memorandum 33-644, June 30, 1973.

2. Woo, R., and Kendall, W., "Effects of Turbulence in the Atmosphere of Venus on Pioneer Venus Radio - Phase II", Jet Propulsion Laboratory Technical Memorandum 33-702, August 15, 1974.

3. Woo, R., and Yang, F. C., "Measurements of Electron Density Irregularities in the Ionosphere of Jupiter by Pioneer 10", J of Geophys. Res., Vol. 81, pp 3417-3422, 1976.

4. Fjeldbo, G., Kliore, A., Seidel, B., Sweetnam, D., and Cain, D., "The Pioneer 10 Radio Occultation Measurements of the Ionosphere of Jupiter", Astron. Astrophys., Vol. 39, pp 91-96, 1975.

5. Strohbehn, J., "Line-of-Sight Wave Propagation Through the Turbulent Atmosphere", Proc. IEEE, Vol. 56, No. 8, pp. 1301-1318, August 1968.

6. Tatarski, V. I., "The Effects of the Turbulent Atmosphere on Wave Propagation", Nauka, Moscow, 1967, pg. 300. Translated and available from the U.S. Department of Commerce, Springfield, Va.

7. Weber, W. J. III, "Performance of Phase-Locked Loops in the Presence of Fading Communication Channels", IEEE Transactions on Communications, Vol. Com-24, No. 5, May 1976.

8. Weber, W. J. III, and Yuen, J. H., "Cascaded Coherent Tracking Systems with Time-Varying Channels", Proceedings, International Telemetering Conference, Los Angeles, Ca; December 1976.

9. Levitt, B. K., "Pioneer Venus 1978: Telemetry Performance Predicts", Deep Space Network Progress Report 42-24, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Ca., December 1974.

10. Modestino, J. W., Mui, S. Y., and Sher, J.S., "Performance of Convolutional Codes on Fading Channels Typical of Planetary Entry Missions", TR-77-2, Rensselaer Polytechnic Institute, Troy, N.Y., June 1977.

A MICROCOMPUTER-BASED SYSTEM FOR ACQUISITION AND ANALYSIS OF TRANSIONOSPHERIC COMMUNICATION SIGNALS

P. D. Stigall, R. E. Ziemer, & Bradford Williams
Elec. Engr. Dept., University of Missouri-Rolla
Rolla, Missouri 65401

ABSTRACT

Severe scintillation of radio signals propagating through the ionosphere in the polar and equatorial latitudes during local nighttime hours has been observed as a phenomenon which can cause signal loss on the order of seconds. In order to analyze link performance in the face of such perturbations it is important that accurate statistical channel models be constructed.

This paper describes a microcomputer based system with capabilities for data acquisition and calculation of the Cooley-Tukey FFT algorithm. The system is composed of a Southwest Technical Products Corporation (SWTPC) 6800 Computer System, an Analogic MP6812-D Data Acquisition Module, and dedicated hardware to calculate the butterfly network associated with the Cooley-Tukey FFT.

Among the software presently available are routines to computer power spectra and fade-duration distributions of the fading envelope. These may be used to characterize scintillation fading channel models introduced and discussed in an earlier paper.

INTRODUCTION

The system employed to analyze the date (Fig. 1) consists of an FM analog tape recorder, an A-D converter, and a computer system based upon an MC6800 microprocessor. The microprocessor controls the overall action of the system. Data from the FM analog tape is input to the A-D converter, and the digitized data from the A-D converter is input to the memory of the microprocessor system. This data is then processed to provide various statistics which are useful for channel modeling.

One such statistic is the fade-duration histogram (density). It is the number of fades below an established level greater than some duration but shorter than the next longest duration. The fade levels chosen for the fade-duration density are 180, 160, 140, 120, 100, 80, 60, 40, and 20 percent of the undisturbed root mean square (rms) signal level. Fade duration can be grouped into one of 13 different time intervals namely, fades greater than or equal to $(0.2)(2^i)$ seconds but less than $(0.2)(2^{i+1})$ seconds where i is an integer ranging from 0 to 12. Thus a particular fade is fully defined by the level it is below and the time duration that it is below that level.

In addition, the system includes hardware which in conjunction with the MC6800 microprocessor is capable of computing the fast Fourier transform (FFT) of data stored in memory. Such a capability allows the computation of statistical functions such as power spectral densities and correlation functions.

One of the chief advantages of the system is its portability. In order to gain the quick-look advantage of a highly portable system, accuracy and speed are sacrificed. It is felt that these are acceptable for the type of data analysis desired, however.

SYSTEM HARDWARE DESCRIPTION

The system employed to analyze the scintillation data is shown in block diagram form in Figure 1. The FM analog tape recorder is a Hewlett Packard 3960 Instrumentation Recorder. The data stored on the FM magnetic tape is a received beacon signal from the TASCAT I satellite (250 MHZ) recorded on Guam in 1972. Prior analysis by the Naval Electronics Laboratory Center (NEL) in San Diego, California, indicated that all significant frequency content of the scintillating signals is limited to frequencies below 1 HZ. Thus a 5 HZ sampling rate is employed after prefiltering to 1.25 HZ. The low pass filter used closely approximates a second-order Butterworth. The gain of the filter at dc is 1.075. Before the prefiltered signal is input to the ANALOGIC MP6812-D A-D CONVERTER, it is amplified by a factor of 4.19. The reason for the amplification is to fully utilize the input voltage range of the A-D converter. A higher input voltage to the A-D converter

increases the number of bits that must be used to represent the data point. The accuracy of the digitized signal increases with the number of bits used to represent it.

At this point in the article it is appropriate to elaborate on the Southwest Technical Products Corporation (SWTPC) 6800 Computer System, since it is the computer which controls the A-D converter. The block diagram in Figure 1 shows the basic components which compose the SWTPC 6800 Computer System. The major controlling unit of the computer system is the MC6800 Microprocessing Unit (MPU). The processor is a bi-directional bus-oriented 8-bit parallel machine with 16 bits of address. The processor has an interrupt request line IRQ. This level sensitive input requests that an interrupt sequence be generated within the machine when IRQ is low. The processor will wait until it completes the current instruction that is being executed before it recognizes the request. At this time, if the interrupt mask bit in the condition code register is not set, the machine will begin an interrupt sequence. It is through the interrupt system that the A-D converter tells the processor that valid data is ready for storage in system memory. The above is discussed in greater detail on the following pages.

The RAM is a static memory which interfaces directly to the MPU. The RAM is organized in an 8-bit-per-byte (word) fashion. For this particular application 4 K bytes of RAM, occupying hex addresses 0000 to OFFF, are used. The MCM6830L7 ROM used in the SWTPC 6800 Computer System is also organized in an 8-bit-per-byte fashion and interfaces directly to the MPU. The ROM contains the control program developed by Motorola known as MIKBUG. When the MPU is interfaced to a teletype, the MIKBUG program allows the user to examine and change memory, load and punch programs on paper tape, display the MPU registers and execute programs all from the teletype keyboard.

The SWTPC 6800 Computer System is interfaced to the Analogic MP6812-D A-D converter through the MC6820 Peripheral Interface Adapter (PIA). The PIA provides a universal means of interfacing peripheral equipment to the MPU through two 8-bit bi-directional peripheral data buses and four control lines. The PIA is composed of six working registers. The registers are Control Registers A and B, Data Direction Registers A and B, and Peripheral Data Registers A and B. The control registers allow the MPU to control the operation of the four peripheral control lines CA1, CA2, CB1, and CB2. In addition they allow the MPU to enable interrupt lines and monitor the status of the interrupt flags. The peripheral data registers are each composed of 8 bi-directional peripheral data lines. The data direction registers allow the MPU to control the direction of data through each corresponding peripheral data line.

FFT MECHANIZATION

In order to generate the FFT algorithm, analog data is input to the A-D converter where it is digitized and stored in the memory of the microprocessor based computer system. This digitized data is later input to the FFT hardware which calculates the butterfly network associated with the Cooley-Tukey FFT algorithm. The output data of the FFT hardware is then stored back into memory and the process continues until the algorithm is calculated. Referring to Figure 2, the system hardware can be divided into two main sections. One is the Peripheral Interface Adapter (PIA) - A-D converter interface and the other is the PIA-FFT hardware interface.

The SWTPC 6800 Computer System is interfaced to the FFT hardware through the MC6820 PIA. The PIA-FFT Hardware Interface calculates the butterfly operation associated with the Cooley-Tukey FFT algorithm.

The fast Fourier transform (FFT) is an algorithm (i.e., a particular method of performing a series of computations) that can compute the discrete Fourier transform much more rapidly than other available algorithms. The discrete Fourier transform can be stated as the sum

$$X(n) = \sum_{k=0}^{N-1} x(k)W_N^{kn}, \quad n=0,1,\ldots,N-1 \qquad (1)$$

where

$$W_N = e^{(-j2\pi/N)} \quad (N=\text{Number of Data Points}) \qquad (2)$$

Two approaches can be taken in the development of the FFT and they are decimation in time and decimation in frequency. The Cooley-Tukey FFT is based on the decimation in time approach. A signal flow graph of the Cooley-Tukey FFT for N = 4 is shown in Figure 3.

The signal flow graph in Figure 3 is interpreted as follows: Each node is entered by two solid lines representing transmission paths from previous nodes. A path transmits or brings a quantity from a node in one array, multiplies the quantity by W_N^P, and inputs the result into the node in the next array. Factor W_N^P appears near the arrowhead of the transmission path; absence of this factor implies that $W_N^P = 1$. Results entering a node from the two transmission paths are combined additively. The output at Array 2 is in bit reversed order. The dotted lines show the unscrambling necessary to obtain natural order at the output. Inspection of Figure 3

reveals that in every array there are always two nodes whose input paths stem from the same pair of nodes in the previous array. Two such nodes are defined as a dual node pair. The basic computation involving the dual node pair is given by the equations

$$X_{j+1}(k)=X_j(k)+W_N^P\ X_j(k+N/2) \qquad (3)$$

$$X_{j+1}(k+N/2)=X_j(k)-W_N^P\ X_j(k+N/2) \qquad (4)$$

where j is the particular computation array. Equations (3) and (4) are known as the butterfly of the Cooley-Tukey FFT algorithm.

The PIA-FFT hardware interface calculates the butterfly network of the Cooley-Tukey FFT defined by Equations (3) and (4).

Microcomputer (μC) program FFT calculates the Cooley-Tukey FFT algorithm. The number of input data points, N, is limited to integer powers of 2 less than or equal to 128. Input data points to each butterfly computation are divided by 2 to prevent overflow.

Upon execution of microcomputer program FFT, N transformed data points are stored in the same memory locations that the input data occupied. Calculation times are shown in Table I.

TABLE I

CALCULATION TIMES FOR MICROCOMPUTER PROGRAM FFT (12 Bits)

N	THEORETICAL BEST CASE (ms)	OBSERVED (ms)
4	10.03496	10.12497
8	27.68764	27.94681
16	70.90438	71.53526
32	174.6379	175.0609
64	426.8034	430.2698
128	1104.858	1105.696

There are three sources of error in the μC FFT calculations. One source of error is the divide by 2 operation performed on each input of every butterfly computation. The error incurred is zero if the LSB of the input data is a '0' before the divide by 2 operation. If the LSB of the input data is a '1' before the divide by 2 operation, the error is one-half the value associated with the LSB

The second source of error in the μC FFT calculations is the truncation performed on the outputs of the multipliers in the FFT hardware. A 12-bit multiplier and a 12-bit multiplicand produce a 24-bit result. This result is truncated to 12 bits by the FFT hardware.

A third source of error in computing the FFT results from representing SIN and COS by 12-bit numbers (coefficient quantization).

SOFTWARE FOR DATA ANALYSIS

Fade Duration Density

The fade duration density is the number of fades below an established level greater than some duration but shorter than the next longest duration (e.g., greater than 0.2 seconds but less than 0.4 seconds). The fade levels chosen are 180, 160, 140, 120, 100, 80, 60, 40 and 20 percent of the undisturbed signal level. The time periods chosen are the interger powers of 2 multiplied by 0.2 seconds. The range is from 0.2 seconds to 819.2 seconds. An analagous distribution, easily obtained from the fade duration density, is the fade duration distribution. The fade duration distribution is the number of fades per unit time below a given level for a given length of time or longer.

A flow chart for the calculation of the fade duration density is shown in Figure 4. In order to generate the fade duration density, the following steps are executed:

1. The undisturbed rms signal level, UNDX, is computed using the formula

$$UNDX=[(DATASIZ)^{-1}\sum_{I=1}^{DATASIZ} DATA(I)^2]^{\frac{1}{2}} \qquad (5)$$

   where DATA(I) represents the amplitude of a particular data point and DATASIZ equals the total number of data points.

2. The level, IX, of a particular data point is calculated using the formula,

$$IX=[(DATA(I)*5.0)/UNDX]+1.0 \qquad (6)$$

   where IX is an integer.

3. Depending upon the level calculated in Step 2, fade parameters are incremented. Each level has an associated fade parameter which keeps track of the fades below that level.

4. The current level is compared to the previous level (calculated from the previous data point). If the previous level is above or equal to the current level program flow branches to Step 5. If, on the other hand, the previous level is below the current level the duration of the fade is determined and the particular parameter associated with that level is incremented.

5. If all data points have been analyzed, the program is done; if not, control branches to Step 2.

The calculation of the fade duration density involves the use of basic arithmetic operations not defined in the instruction set of the MC6800 MPU. These operations include multiplication, division and squareroot. The above operations are implemented using the available instruction set. The fade duration density program, including the multiplication, division and squareroot routines, occupies 947 bytes of memory.

Power Spectral Density

It is well known that the periodogram estimate for power spectral density

$$S_N(\omega) = \frac{1}{N} |X_N(\omega)|^2 \qquad (7)$$

where $X_N(\omega)$ is the discrete Fourier transform for N points does not have a vanishingly small variance as $N \to \infty$[2]. In order to improve the periodogram as an estimator for power spectral density the technique of segmenting the data and averaging the periodograms of the separate segments may be employed. This technique is employed for power spectrum estimation here. The original data is broken up into 128-point segments and the separate periodograms found and then averaged. The variance of the resultant averaged periodogram is then decreased by a factor equal to the number of periodograms averaged over the variance of a single periodogram.

RESULTS

In this section the results of executing the fade duration density program are presented. A graph of the fade duration distribution is also presented.

The procedure for a typical run is outlined below:

1. Power up the system depicted in Fig. 1.
2. Enter and execute the PIA - A-D Converter Interface program. (An important point to note is that the input to the A-D converter is monitored on a scope to insure that the tape recorder has come up to speed before execution of the interface program.)
3. The fade duration density program is entered and executed.
4. The memory locations containing the fade duration parameters are examined.

The results of a typical run are shown in Fig. 5.

Figure 6 shows a typical power spectrum plot for the fading signal. Note that the dc (n=o line in the spectrum) has been removed in order to improve the dynamic range of the spectrum estimation. Figure 6 indicates that the spectral content of the fading signal is primarily below one Hertz.

CONCLUSIONS

The microcomputer data acquisition system has demonstrated the feasibility of using a microprocessor to analyze transionspheric communication channels. The microcomputer approach leads to several benefits including portability, field data analysis, and less resources involved in data acquisition.

REFERENCES

1. R. E. Ziemer and W. F. Deckelman, "Computer Simulation of Transionospheric Scintillation Communication Channels Based on Fade-Duration Statistics," Proceedings of the Ionospheric Effects Symposium, Jan. 20-22, 1975.
2. M. Schwartz and L. Shaw, "Signal Processing: Discrete Spectral Analysis, Detection, and Estimation," McGraw-Hill, 1975.

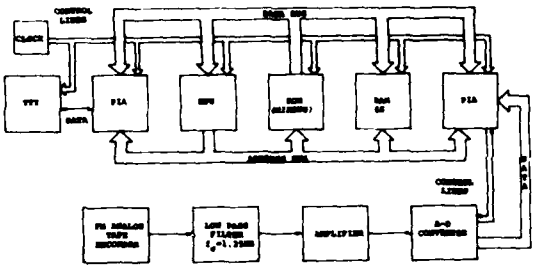


Figure 1. Block Diagram of Data Acquisition System.

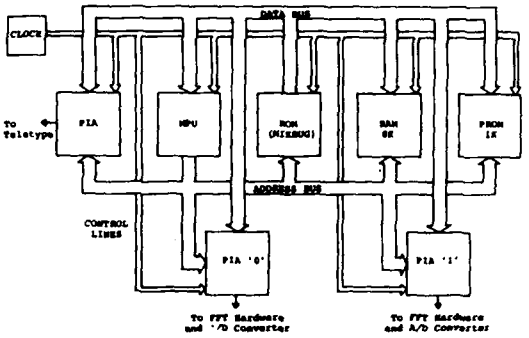


Figure 2. Expanded Block Diagram of SWTPC 6800 Computer System

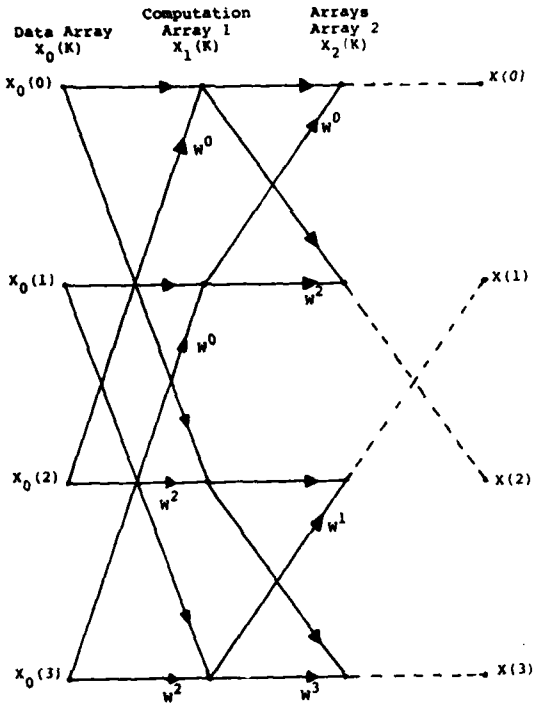
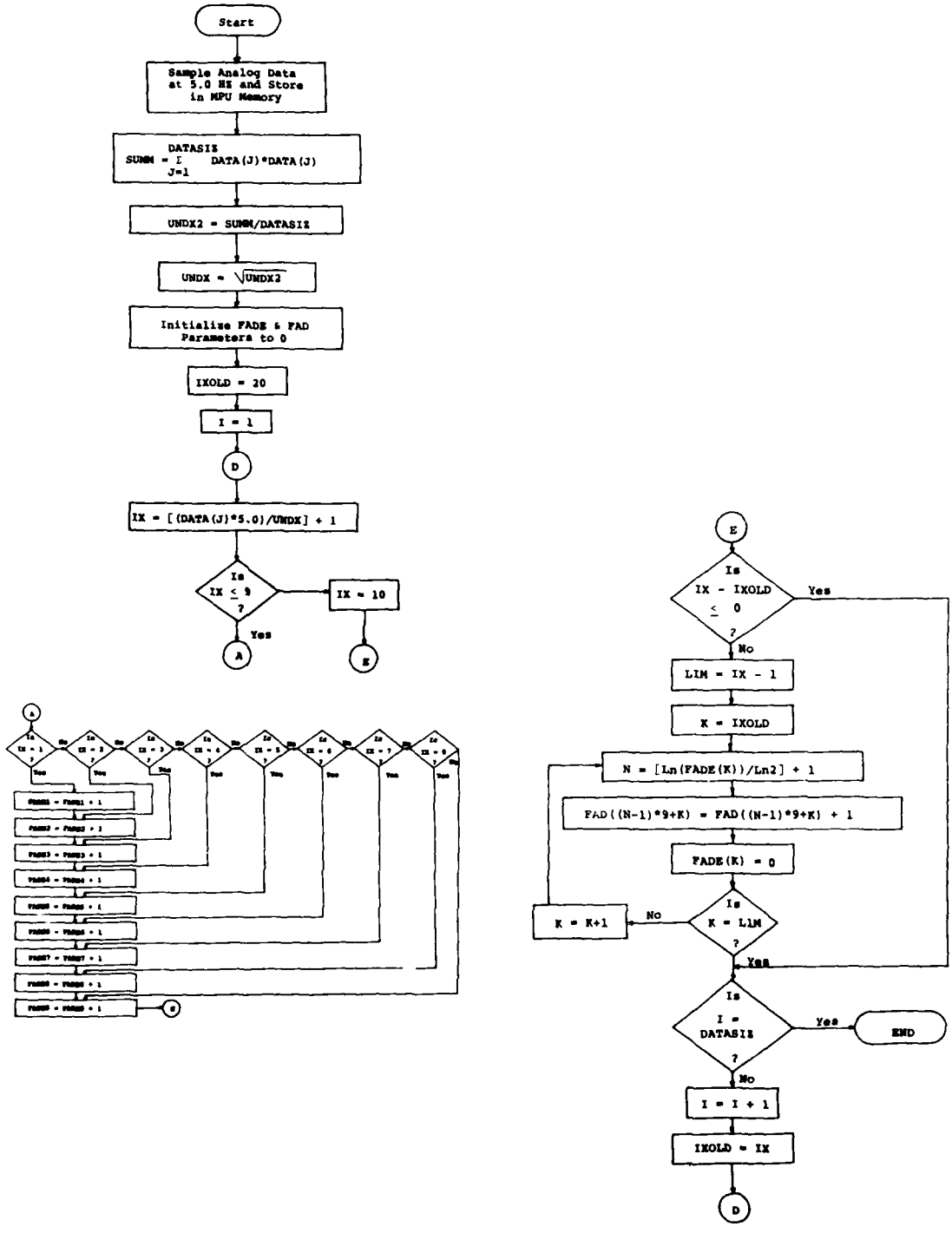


Figure 3. FFT Signal Flow Graph for N = 4

Figure 4. Flow Chart for Calculation of Fade Duration Density.



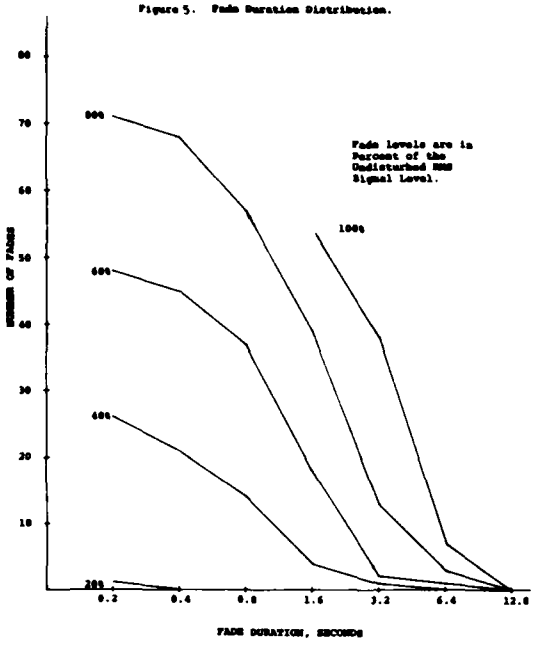


Figure 5. Fade Duration Distribution.

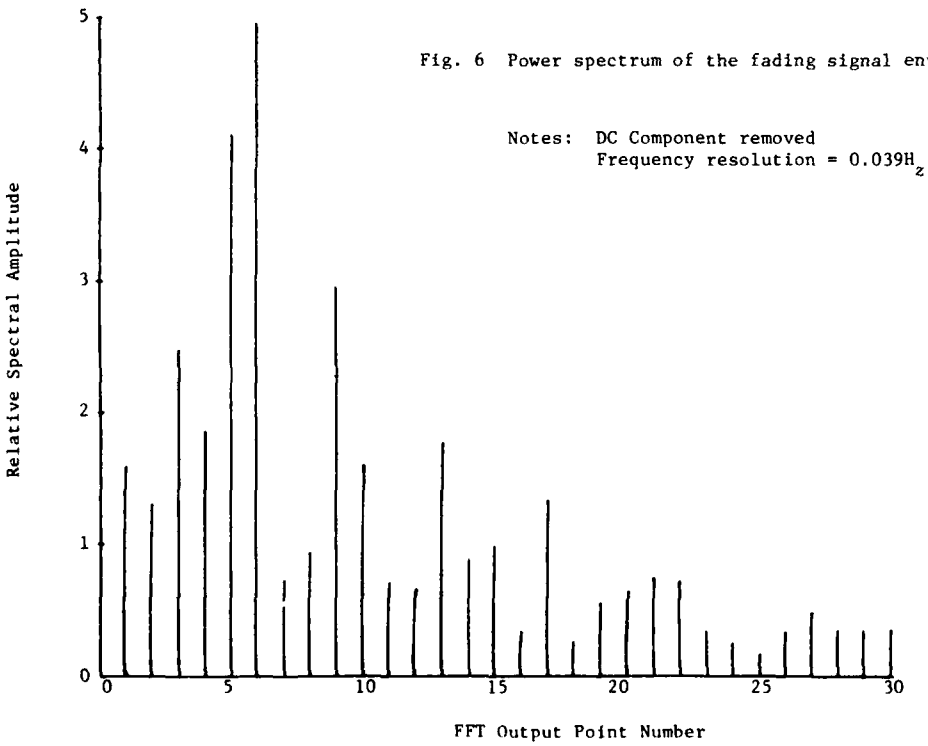


Fig. 6 Power spectrum of the fading signal envelope.

Notes: DC Component removed
Frequency resolution = $0.039H_z$

THE IONOSPHERIC LIMITATION TO COHERENT INTEGRATION IN TRANSIONOSPHERIC RADARS

C. L. Rino, C. H. Dawson, R. C. Livingston, and J. Petriceks
SRI International
Menlo Park, California 94025

ABSTRACT

The achievable coherent processing gain of transionospheric radars is ultimately limited by omnipresent ionospheric electron density variations that cause random but correlated pulse-to-pulse phase variations. To calculate the coherent integration loss, we have developed a model in which the phase variations are characterized by both linear and quadratic trend terms, and a stationary residual that is characterized by a power-law spectral density function.

Since a linear phase trend does not cause spectral smearing, only the quadratic term and the stationary residual contribute to the coherent integration loss. We have developed a simple analytic expression that gives the coherent integration loss in terms of the quadratic phase trend and of the strength and power-law index that characterize the power spectral density of the phase residual. Results are presented for typical ionospheric parameters.

A large amount of data from the synchronous satellite ATS-6 have been processed to determine the morphology of the quadratic phase trends and the residual. Our preliminary results indicate that the residual term is closely related to the scintillations and spread-F phenomena. However, quadratic phase trends with little associated amplitude scintillation can occur at any time.

I INTRODUCTION

The achievable coherent processing gain of transionospheric radars is ultimately limited by omnipresent ionospheric electron density variations. Since the signal phase is proportional to the integral of the electron density along the propagation path, the radar sees a random but correlated phase change from pulse to pulse. In addition, amplitude scintillation will be associated with the more rapid phase variations. We shall see, however, that the amplitude scintillation need not be considered explicitly in calculating the coherent integration loss.

To characterize the phase variations over time intervals of varying length, we must consider both trend-like and purely random variations. It is now generally accepted that the phase power spectral density (PSD) has the power-law form f^{-p} down to frequencies generated by electron density structures with dimensions greater than many tens of kilometers. Thus, for any reasonable drift rate, the phase PSD will admit a power-law characterization over periods greater than tens of minutes.

In such a power-law environment there is no clear-cut distinction between trends and purely random structures. Indeed, genuine trends or even wavelike (i.e., quasi-periodic) variations can be lost in the power-law continuum over short intervals. To deal with this problem, we have postulated the two-component model

$$\Phi(t) = \phi_o + w(t - t_o) + \dot{w}(t - t_o)^2 + \delta\Phi(t - t_c) \quad t_o \le t \le t_o + T . \qquad (1)$$

The quadratic polynomial accounts for the phase trends within the subinterval $t_o \le t \le t_o + T$. The purely random component $\delta\Phi(t)$ is characterized by a power-law PSD.

The coefficients ϕ_o, w, and $\dot{w}$ are determined by a least-squares fit over the data interval. The linear and quadratic phase trends have direct engineering interpretations. The linear term gives rise to a Doppler shift, which does not impair target detectability. The quadratic term, on the other hand, broadens the spectrum and ultimately causes a precipitous reduction in coherent processing gain as we shall show in Section IV.

II COHERENT PROCESSING GAIN

Consider an idealized transionospheric radar that transmits a pulse, P(t), every Δt seconds. Ignoring frequency dispersion, the pulse will emerge from the ionosphere as

$$P'(t) = P(t)h'(t;f_c) \exp\{2\pi i f_c t\} \tag{2}$$

where f_c is the carrier frequency, and $h(t;f_c)$ is the complex modulation induced by the ionosphere.

Upon interacting with a target and propagating back to the radar, the pulse will have the form

$$P''(t) = \sigma(t)\, P(t)\, h'(t;f_c) \times h''(t + t_d;f_c) \exp\{2\pi i f_c t\} \tag{3}$$

where $\sigma(t)$ represents the complex cross section of the target. Thus, if the target is moving at a constant rate, $\sigma(t)$ will contain a linear phase change. However, we shall assume that neither the target nor the ionosphere admit significant changes on a time scale comparable to the pulse duration, which we denote by τ.

In that case, there is no loss of generality in setting $\sigma(t) = 1$ and normalizing all calculations to unity (0 dB) average power at the receiver. Indeed, for large TW (time-bandwidth-product) signals we can take P(t) to be the "compressed" waveform, which is purely real. Finally, we define the two-way transfer function

$$h(t;f_c) = h'(t;f_c)\, h''(t - t_d;f_c) . \tag{4}$$

To continue, we consider a sequence of N pulses separated by Δt seconds. The radar processor forms the FFT spectral estimate

$$\hat{\Phi}_n = P^2(t_r) \left| \frac{1}{N} \sum_{k=0}^{N-1} h(k\Delta t;f_c) \times \exp\{-2\pi i n k/N\} \right|^2 \tag{5}$$

where t_r refers to the delay to a particular range gate. Since $h(t;f_c)$ is a random process, we must characterize $\hat{\Phi}_n$ by its expectation or average value $\langle\hat{\Phi}_n\rangle$ which we can compute directly from (5) as

$$\langle\hat{\Phi}_n\rangle = \frac{P^2(t_r)}{N^2} \sum_{k,k'=0}^{N-1} R_h(t,t') \times \exp\{-2\pi i n(k - k)/N\} \tag{6}$$

where

$$R_h(t,t') \triangleq \langle h(t;f_c)\, h^*(t';f_c)\rangle \tag{7}$$

is the mutual coherence function.

In Section III we shall show that $R_h(t,t')$ has the general form

$$R_h(t,t') = g(t^2 - t'^2)\, R_{\delta h}(\delta t) \tag{8}$$

where $\delta t = t - t'$. By substituting (8) into (6) and changing the summation variables to $m = k - k'$ and $k + k' = 2\ell + m$ so that $(k^2 - k'^2) = m(2\ell + m)$, (6) can be rewritten as

$$\langle\hat{\Phi}_n\rangle = \frac{P^2(t_r)}{N} \sum_{m=-N}^{N} G(m)\, R_{\delta h}(m\Delta t) \times \exp\{-2\pi i m n/N\} \tag{9}$$

where

$$G(m) = \frac{1}{N} \sum_{\ell=0}^{N-|m|} g[(2\ell + m)\, m\Delta t^2] . \tag{10}$$

We see that (9) has the form of a "weighted" Fourier series. If $g(x) = 1$ for all x, then $G(m) = 1 - |m|/N$--that is a triangular weighting.

The coherent integration loss, $\ell(T)$, [for small losses] is given to a good approximation as

$$\ell(T) = 10 \log_{10}\left[\frac{1}{N} + 2\sum_{m=1}^{N-1} \mathrm{Re}\{G(m)\} \times R_{\delta h}(m\Delta t)\right] \tag{11}$$

which achieves a maximum value of 0 dB when both G(m) and $R_{\delta h}(m\Delta t)$ are near unity over the entire coherent integration interval $T = N\Delta t$.

III THE CHANNEL MODEL

The mutual coherence function $R_{\delta h}(\delta t)$ admits the exact representation

$$R_{\delta h}(\delta t) = \exp\left\{-\frac{1}{2} D_{\delta\phi}(\delta t)\right\} \tag{12}$$

where

$$D_{\delta\phi}(\delta t) = \langle[\phi(t) - \phi(t')]^2\rangle \tag{13}$$

is the phase structure function [see, for example, Prokhorov et al., 1975]. For a power-law phase spectrum of the form $C_s f^{-p}$ with $1 < p \leq 3$, one can show that

$$D_{\delta\phi}(\delta t) = C_s 2\sqrt{\pi}\,\frac{\Gamma\left(\frac{3-p}{2}\right)}{\Gamma(p/2)2^{p-2}(p-1)}\,|\delta t|^{p-1} \tag{14}$$

Thus, the mutual coherence function is completely characterized in terms of two easily measured parameters C_s and the spectral index p.

For the trend terms we ignore all linear terms and take

$$g(t^2 - t'^2) = \exp\left\{i\bar{\dot{w}}(t^2 - t'^2)\right\} . \tag{15}$$

The reason is twofold. First, by using (15) we obtain the exact result for a purely deterministic quadratic phase trend. Secondly, it is the actual value of $\dot{w}$ in a given interval that contributes to the coherent integration loss. The dispersion is best taken into account by evaluating the coherent integration loss for the extreme values of $\dot{w}$.

By substituting (15) into (10) one can show that

$$\mathrm{Re}\{G(m)\} = \frac{\cos\,[am(N-1)]}{N}\,\frac{\sin\,[am(N-m)]}{\sin\,[am]} \tag{16}$$

where

$$a = \bar{\dot{w}}\,\Delta t^2 \tag{17}$$

It is easily shown that as $a \to 0$, (16) converges to the triangular weighting $1 - |m|/N$.

The coherent integration loss is now simply evaluated by substituting (16) and (14) into (11). The ionospheric effects are completely characterized by $\dot{w}$, C_s, and p. The parameters C_s and p are determined by a log-linear least-squares fit to the PSD of the phase residual $\delta\phi(t)$. The Doppler spectrum computed by Fourier transforming (12) has been found to follow the measured Doppler spectrum very closely.

IV RESULTS AND CONCLUSIONS

We first consider the residual term separately. In Figure 1, $\ell(T)$ is evaluated for p = 2 and p = 2.5. As one should expect, the larger p value, which implies more intense large-scale phase perturbations for the same C_s value, produces the larger loss. The general character of the loss due to the residual term is a smooth, monotonic decrease (increasing loss) with increasing T.

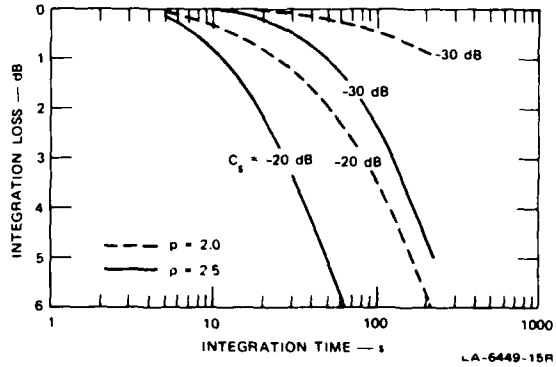


FIGURE 1 RESIDUAL TERM INTEGRATION LOSS

For comparison, in Figure 2 we show the coherent integration loss for the quadratic term alone. The loss due to the quadratic term tends to be small initially and then increases abruptly. When both terms are included, the quadratic term ultimately dominates the loss. This can be seen in Figure 3 where the combined loss is plotted for C_s = -20 dB with p = 2 and p = 2.5 for different values of $\dot{w}$.

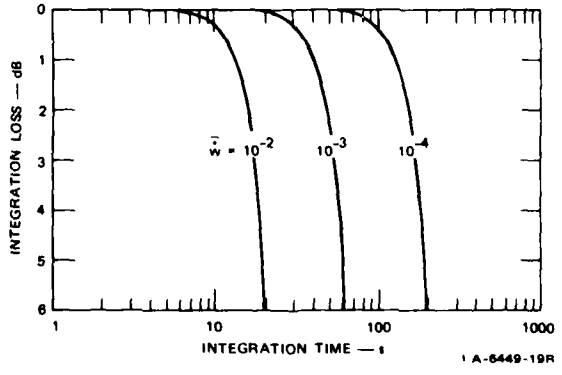


FIGURE 2 QUADRATIC TERM INTEGRATION LOSS

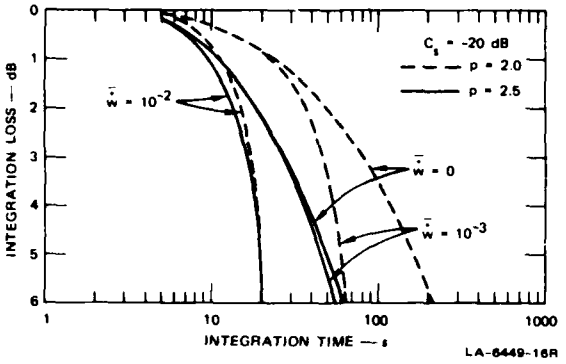


FIGURE 3 COMBINED COHERENT INTEGRATION LOSS

The results presented here are for a one-way propagation path. For a two-way propagation path both $\dot{w}$ and C_s must be doubled. With regard to frequency scaling, C_s varies as f_c^{-2} whereas $\dot{w}$ varies as f_c^{-1}. The spectral index does not vary with frequency.

To determine appropriate C_s, $\dot{\bar{w}}$, and p values for real ionospheric data, we have processed a large quantity of differential phase data from the synchronous satellite ATS-6 recorded at the equatorial station at Ootacamund, India. Only preliminary results were available at the time of writing of this paper. The data have been divided into four 6-hour time segments, covering midnight, sunrise, noon, and sunset periods.

Histograms from a midnight period and a noon period are shown in Figures 4 and 5, respectively. The midnight period shows a high probability of occurrence of large C_s and $\dot{w}$ values. We believe these large values to be associated with ordinary scintillation and spread-F type phenomena, which are most often observed during the evening.

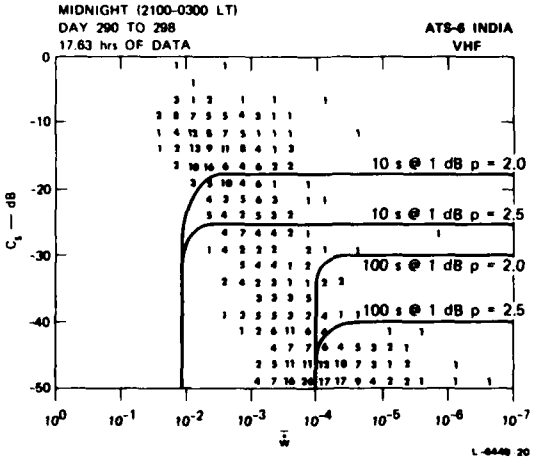


FIGURE 4 HISTOGRAM OF C_s AND $\dot{\bar{w}}$ FOR MIDNIGHT SECTION

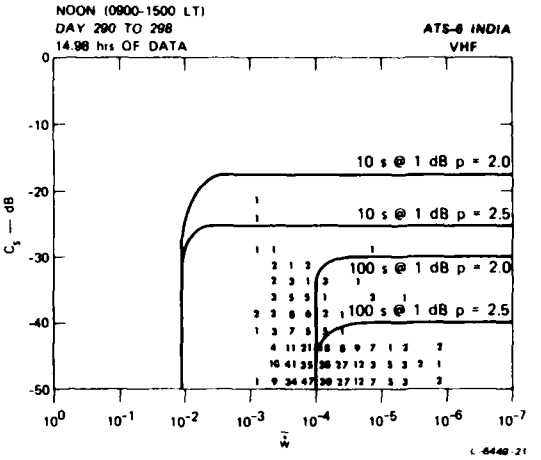


FIGURE 5 HISTOGRAM OF C_s AND $\dot{\bar{w}}$ FOR NOON SECTION

However, moderate w ($\dot{w} \geq 10^{-4}$) values persist throughout the day as can be seen in Figure 5. This component is evidently not associated with scintillation activity. That is, there is an omnipresent level of phase structure in the ionosphere that can give rise to large enough $\dot{w}$ values to limit the coherent integration time to ~1 minute at VHF.

The morphology of this component is under study. In Figures 4 and 5 we have indicated rough boundaries for the 2-way coherent integration time at the 1-dB level. The curves are for an effective frequency of 165 MHz. The contours can be scaled with frequency as described at the beginning of this section.

Acknowledgments

This work was sponsored by the Electronic Systems Division, Air Force Systems Command under subcontract to the General Electric Company [Contract No. F19628-76-C-0251.]

References

Prokhorov, F. V., K. S. Bunkin, K. S. Gochelashvily, and V. I. Shishov, "Laser Irradiance Propagation in Turbulent Media," Proc. IEEE, 63(5), 790-811, 1975.

CHARACTERISTICS OF LATE-TIME STRIATIONS OBSERVED DURING OPERATION STRESS

John H. Wolcott
University of California, Los Alamos Scientific Laboratory
David J. Simons, Timothy E. Eastman, T. Joseph Fitzgerald
EG&G, Inc.
Los Alamos, New Mexico

Introduction

The Los Alamos Scientific Laboratory (LASL) participated in project STRESS, a series of rocket-borne, barium thermite release experiments sponsored by the Defense Nuclear Agency (DNA). For DNA, the principal objective of STRESS (Satellite Transmission Effects Simulation) was the study of actual satellite communication links through structured barium plasma clouds intended to simulate many aspects of the highly disturbed, post-nuclear-explosion, upper atmosphere. For LASL, the principal objective was the investigation of the late time spatial decay of plasma striations, or filaments, imbedded within the barium ion cloud. Most operational aspects of this experimental project were well suited to address both principal objectives.

The release, or injection, of barium has now become a standard technique for simulation of the highly structured plasmas that follow a high-altitude nuclear detonation. Complex computer models which predict nuclear weapons effects are validated by application to such simulation events. From a scientific viewpoint, barium experiments aid greatly in the understanding of the dynamics of plasma processes which ultimately affect a number of planned or operational systems. The performance of those systems which depend on electromagnetic propagation can be severely degraded when a transit through highly disturbed environments is necessary.

Numerous observations of the behavior of ionized barium clouds and jets have resulted in a relatively thorough understanding of the plasma process leading to the formation and growth of striations within a plasma cloud. Little experimentation to date has been directed to those processes which result in striation decay.

Project STRESS, conducted at Eglin AFB, Florida, actually included two phases with all experiments involving a 48 kg barium thermite release at 185 km altitude. The first phase, Pre-STRESS, was completed on December 1, 1976 when a single release occurred at 2312 UT, and ∿ 29°55'N, 86°30'W; background and signal level data were obtained by LASL during Pre-STRESS. The second and more extensive phase occurred between February 26 and March 14, 1977 when five releases were conducted at widely varying local times.

Pre-STRESS Results

The Pre-STRESS experiment was conducted in order to exercise the operational interfaces that would be involved later during STRESS and to provide LASL with the opportunity to perform optical measurements during a local time period and lunar phase similar to those to be encountered during the STRESS operation.

Sky background measurements obtained on the evenings of November 30 and December 1, 1976 are presented in Fig. 1. Look angles were those to a nominal STRESS event location and a 3/4 moon was above the horizon. Comparable data obtained June 6, 1976 are also included. These data were sufficient for design of the STRESS instrument plan and an operational program for temporal adjustments of instrument sensitivities.

Averaged over a 2° diam field-of-view, the radiance of the fully solar-illuminated Pre-STRESS release was ∿ 2.5×10^{-9} W $cm^{-2} sr^{-1}$ at 455.4 nm. Assuming an optically thin source, this radiance corresponds to a barium ion column density of ∿ 2×10^{11} cm^{-2}. Based on these results and a minimum signal-to-background ratio of 0.5 with 1-nm-wide spectral filtering, photographic recording of the release appeared feasible at solar depression angles of ∿ 6° and larger. Signal-to-background ratios would improve with time until the cloud was no longer illuminated by the sun at a depression angle of ∿ 13°.

STRESS Operation

Realization of the LASL objectives for STRESS depended on the ability to identify and

observe optically thin regions of the structured barium ion cloud. Comparisons with current models for decay of striated plasmas require observation of the spatial inhomogeneities within the total ion cloud with no obscuration by intervening optically thick regions.

One approach, observation of a barium ion spectral-line shape, was utilized by Lockheed, a DNA contractor. Here, an imaging Fabry-Perot interferometer was spectrally scanned across the line profile at 455.4 nm. The LASL approach involved a search for optically thin regions of the cloud by realtime TV observations. A series of filtered (455.4 nm) cameras with high time and space resolution were then pointed toward those volumes. Kodak 2484 film was used and was force processed to an ASA of $\sim$ 2000. All camera systems were radiometrically calibrated and the film processing was controlled so that absolute cloud radiance could be obtained. Since Lockheed selected the Tyndall AFB location early in the STRESS planning and direct comparison of the two optical techniques was highly desirable, the LASL optical site was also located at Tyndall AFB (29°58'N, 85°28'W), southeast of Panama City, Florida.

The tracking TV was an ISIT, a single-stage intensified and cooled, SIT TV camera. Two of the camera systems employed three-stage electrostatically focused intensifiers, while the third contained a three-stage, magnetically focused tube. Focal lengths used were 50 mm, 400 mm, and 1250 mm, with the 1250 mm lens used with the magnetic intensifier. With this selection of optics and cameras, fields-of-view ranged from 1.2° to 29°, and spatial resolution at the barium cloud location ranged from about 20 to 480 m. To achieve these spatial resolutions realizing that cloud motion is expected to be of order 10 m s^{-1}, the necessity for fraction-of-a-second exposures possible only with intensified systems is recognized. Exposures as short as 0.1 s were used.

Table 1 gives a summary of the five STRESS events. Esther provided the best optical data and analysis has been concentrated on this event. The cloud was highly striated when it was first observed. It had drifted southeast and passed south of the local magnetic zenith. Figure 2 shows the ion cloud at 00:35 UT as viewed by the intensified camera systems. The fields-of-view of the 400 mm and 1250 mm lenses are superimposed on the 50 mm and 400 mm images.

Analysis Technique

Optical systems degrade spatial frequency information in a fashion analogous to electronic circuits degrading the frequency content of electrical signals. Using techniques drawn from the analysis of linear electronic system it is possible to reconstruct the spatial frequency content of an image field out to a frequency determined by system noise. We have attempted to do this with the stress data.

The system response of the optics can be characterized by its Modulation Transfer Function (MTF) which is determined by measuring the degradation in images of targets of known spatial frequency. Each component of the system has an MTF, from the objective lens to the recording film and the digitizing microdensitometer. We have determined the MTF for each component of the stress image recording and analysis system. The MTF of the entire system is the product of the individual MTF's. Assuming a shift-invariant linear system response functions the image and object fields are related by the integral equation

$$\hat{F}(x,y) = \iint F(x',y')\, T(x-x',y-y')\, dx'dy',$$

in which $\hat{F}(x,y)$ is the image field, $F(x,y)$ is the object field and $T(x,y)$ is the system response function (also called the point spread function for optical systems). Taking the Fourier Transform of this equation we find

$$\hat{F}(f_x,f_y) = F(f_x,f_y)\, T(f_x,f_y) \quad .$$

For systems with cylindrical symmetry (which is approximately true for most optical systems), $T(f_x,f_y)$ is real and cylindrically symmetric and is identically equal to the MTF. The object field is then reconstructed in the following manner.

$$F(x,y) = \iint \frac{\hat{F}(f_x,f_y)}{T(f_x,f_y)}\, e^{-i2\pi f_x x}\, e^{-i2\pi f_y y}\, df_x df_y$$

There are practical difficulties which arise when one attempts to carry out these calculations and we can only touch upon them here briefly. For real data one always measures the true signal, $F(x,y)$, plus a noise component, $N(x,y)$. In general the noise component will contribute a term to the Fourier transform which is flat across the frequency spectrum. At high frequency the noise will then dominate the signal and the method prescribed above tends to diverge. It becomes necessary to incorporate low pass filters into the unfolding process to roll off these amplified high frequency noise components.

Another difficulty arises from using truncated Fourier Series representations (called discrete Fourier transforms) to perform Fourier Transformations on digital computers. This introduces both aliasing and leakage distortion into the calculated transformation. These effects can be reduced to acceptable levels by using specially designed window function which reduce side lobes in the frequency domain. We have chosen the "Hanning Window" as being most appropriate to our data. This window reduces side lobe leakage at relatively small cost in resolution.

An intensified camera system also exhibits pin cushion distortion and non-linear radial gain. These effects are calibrated and then removed during the analysis procedure. In Fig. 3 we show a flow diagram of this procedure.

Data

In this section we will present the results of the power spectral analysis of data taken at 00:30 UT during Event Esther. Both one-dimensional and two-dimensional power spectra are presented. The one-dimensional power spectra have been derived from raster scans across the images perpendicular to the striations.

The fields-of-view of the frames selected are in the region of the barium cloud that has the lowest apparent elevation from the camera station. The striations in this region, as depicted in Fig. 2, are distinct and show a great deal of structure. From radiometric analysis the barium cloud in the upper portion of this region is optically thick. In particular, the 1250 mm camera image, restricted to a small field-of-view near the center of the 400 mm camera field-of-view, is optically thick. However, the lower portion of the 400 mm camera image extends down to a region that appears to be optically thin.

Figure 4 gives an example of a one-dimensional profile and shows a scan from the 400 mm camera. The scan is in a region that has been determined to be optically thick. The effect of the scanning aperture has been removed from the scan. Figure 5 shows the amplitude spectrum of the spatial frequency components derived from the profile. A profile from a 1250 mm camera frame at the same time is shown in Fig. 6. The amplitude spectrum derived from the profile is shown in Fig. 7.

Profiles from the two-dimensional power spectra of these frames are shown in Figs. 8 and 9 for the 400 mm and 1250 cameras, respectively. The upper profile in each figure show the power spectra in a direction perpendicular to the striations while the lower profiles show the power spectra parallel to the striations.

Data analysis is continuing and further results will be compiled in an addendum which will be available at the NRL Symposium on Ionospheric Effects. These results will contain further spatial power spectra, and the time behavior of striation power spectra decay, as observed from Tyndall AFB.

Conclusions

An analysis technique has been developed to carefully analyze photographs of emissions from plasmas. This technique removes instrumental and data reduction artifacts and gives the power spectrum for an optically thin cloud. In an optically thick cloud, the derived power spectrum will represent a lower limit to the true modulation which would affect a radio wave traversing the cloud.

Acknowledgements

We wish to express our appreciation to Technology International Corporation for their assistance during the experimental phases of this experiment. Funding for data analysis was provided by DNA under IACRO #77-829.

Table I

Event Date (Local)	Betty 26 Feb 77	Carolyn 2 Mar 77	Dianne 7 Mar 77	Esther 13 Mar 77	Fern 14 Mar 77
Release Time (UT)	2352	2354	0001	2301	2246
Altitude of Release (km)	179	191	186	189	186
Δ T (min)	14	15	11	75	90

($\Delta T = T_{6^\circ} - T_R$ where T_{6° is the time when the sun reaches a 6° depression angle at the nominal sub-event point and T_R is the release time)

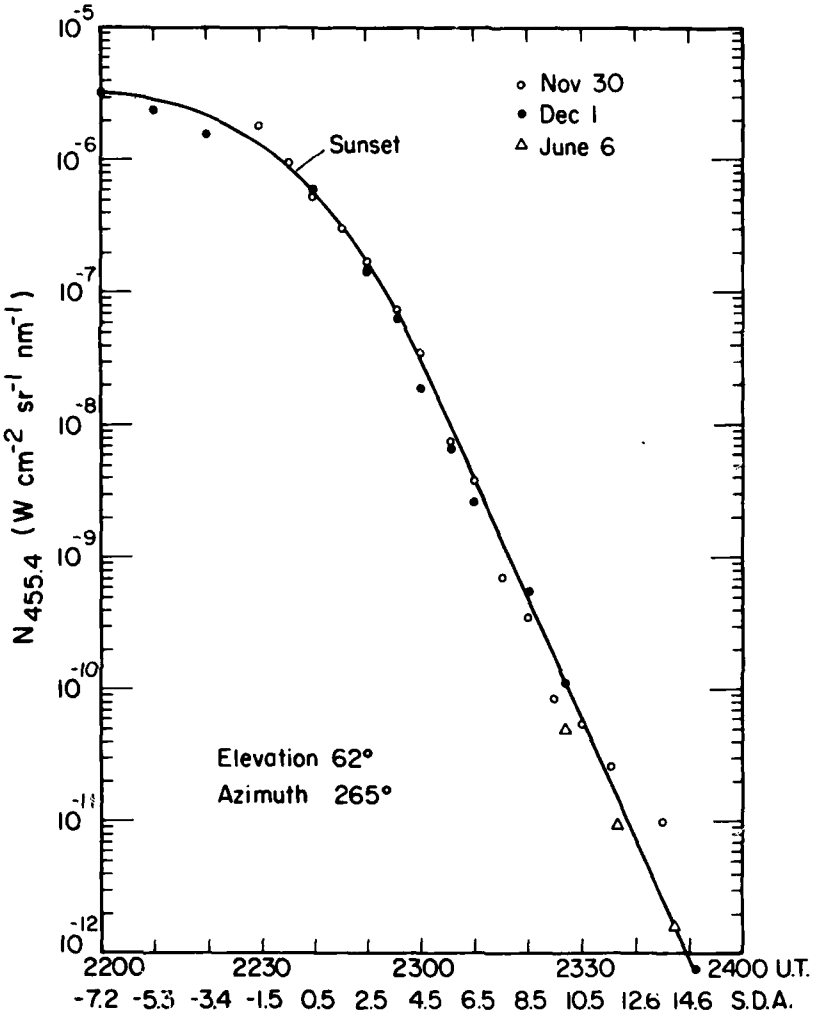


Figure 1. Sky background measurements made on November 30 and December 1, 1976 and June 6, 1976.

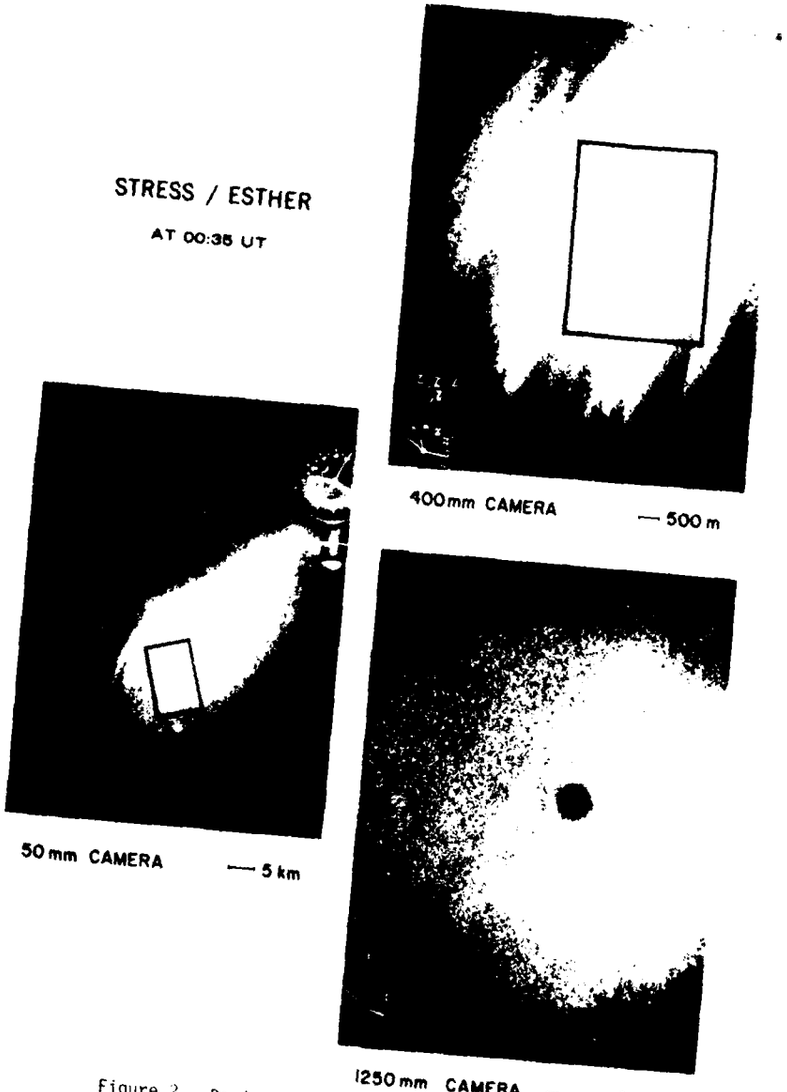


Figure 2. Barium ion cloud at 00:35 UT - Event Esther.

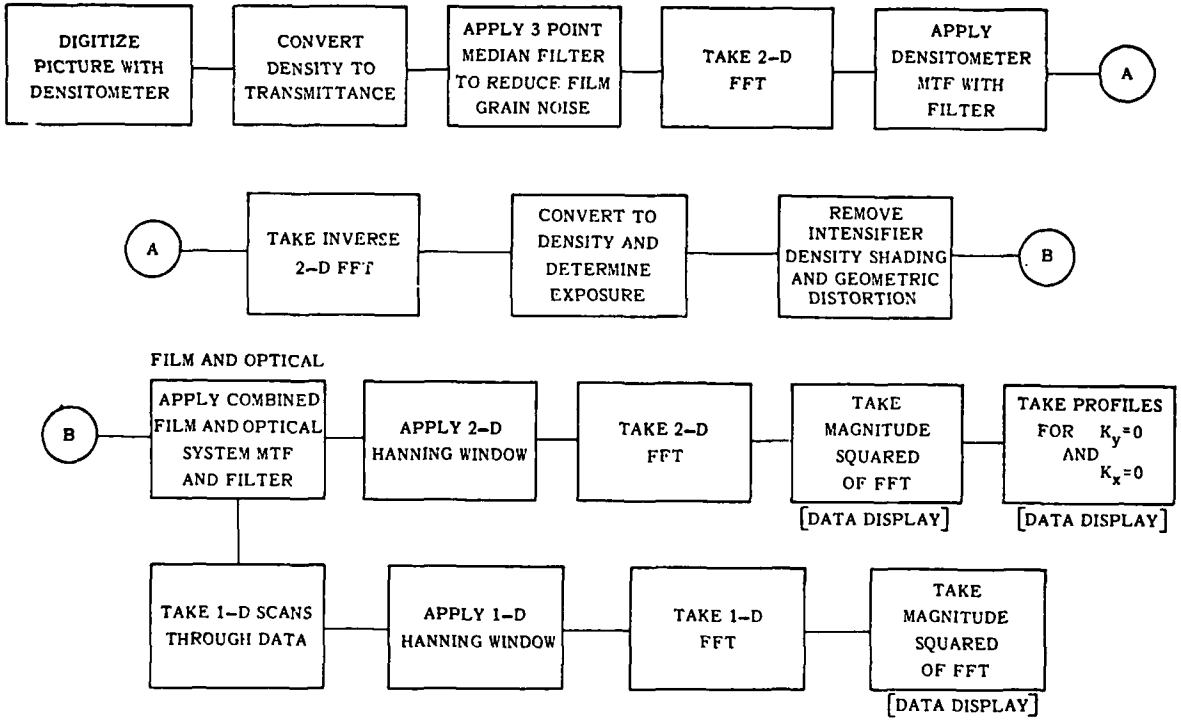


Figure 3. Flow diagram of photographic data reduction.

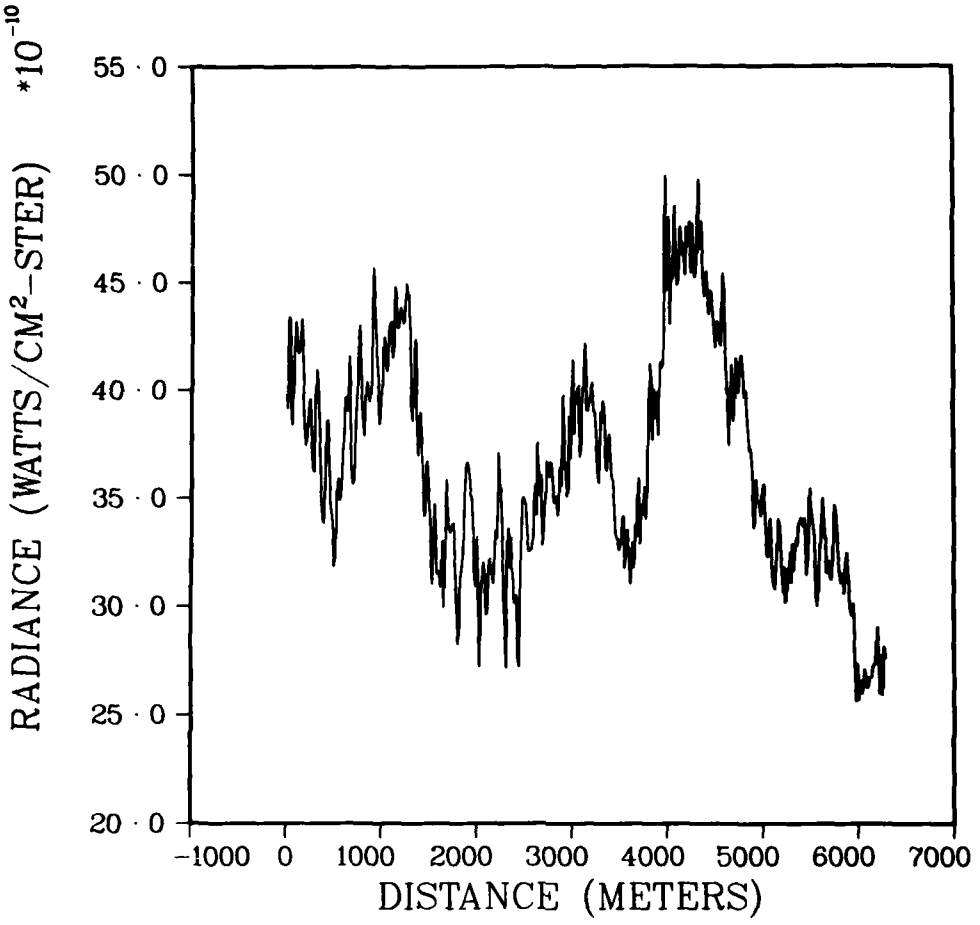


Figure 4. Profile across the striations (400 mm camera).

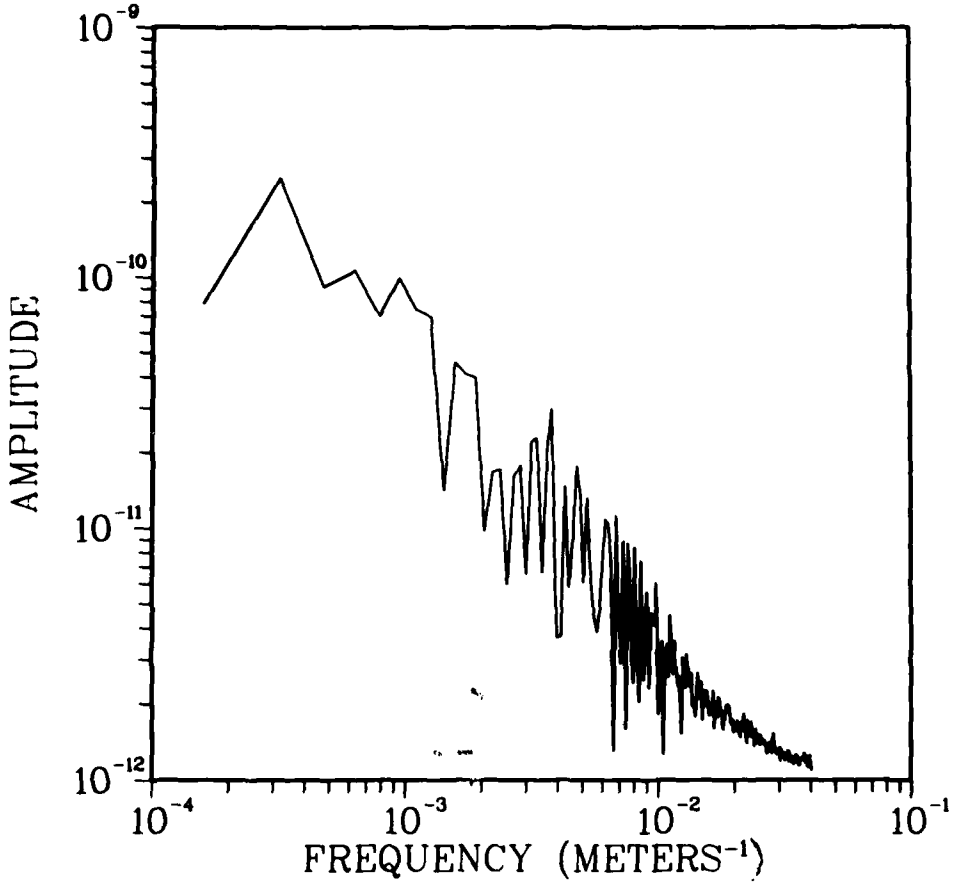


Figure 5. One-dimensional amplitude spectrum (400 mm camera).

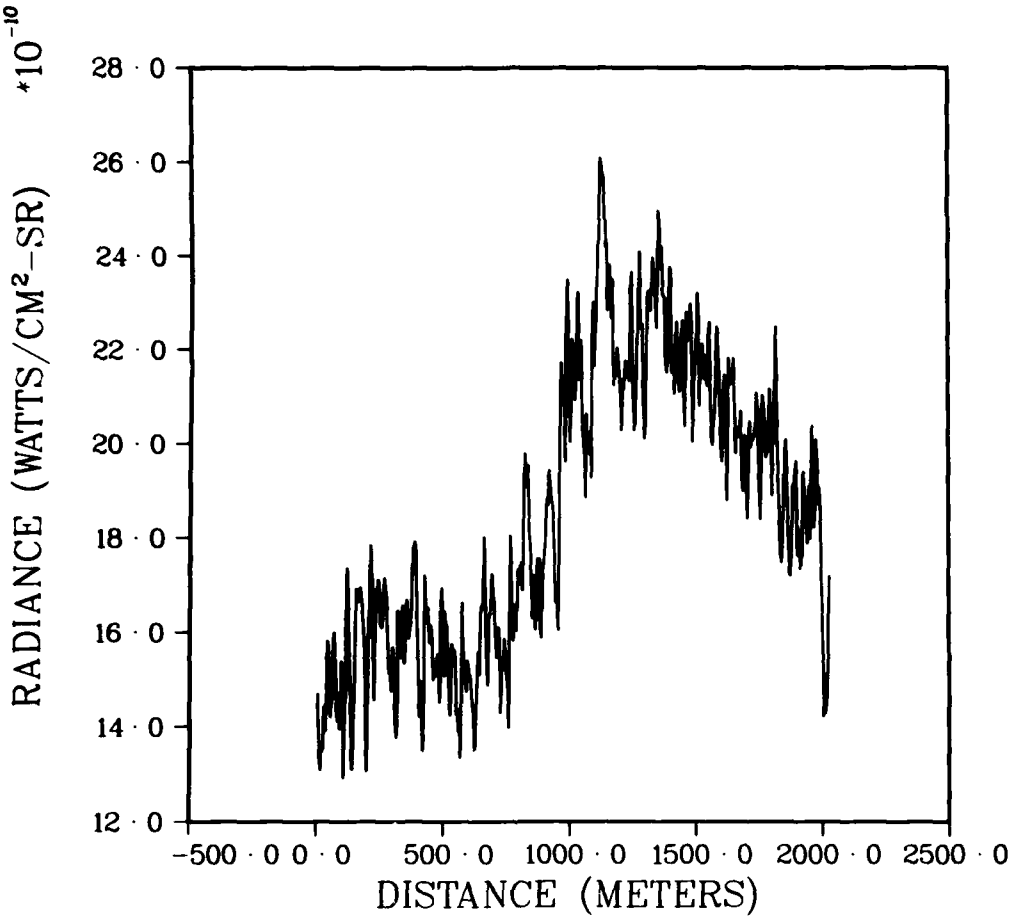


Figure 6. Profile across the striations (1250 mm camera).

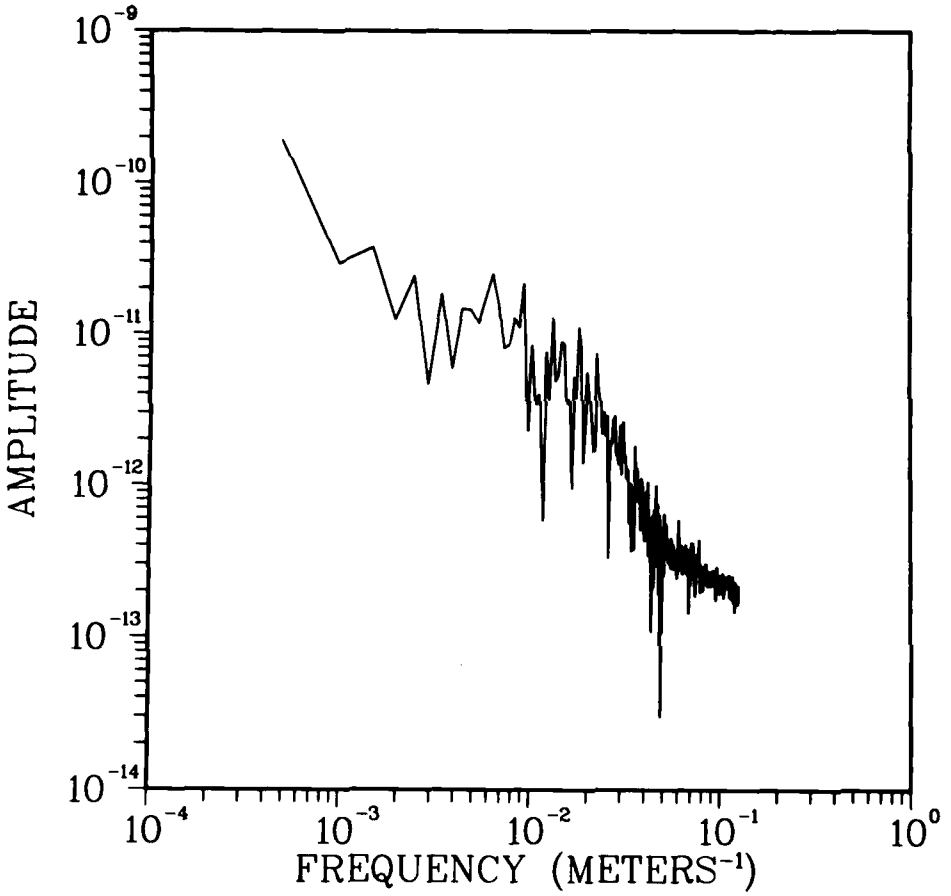


Figure 7. One-dimensional amplitude spectrum (1250 mm camera).

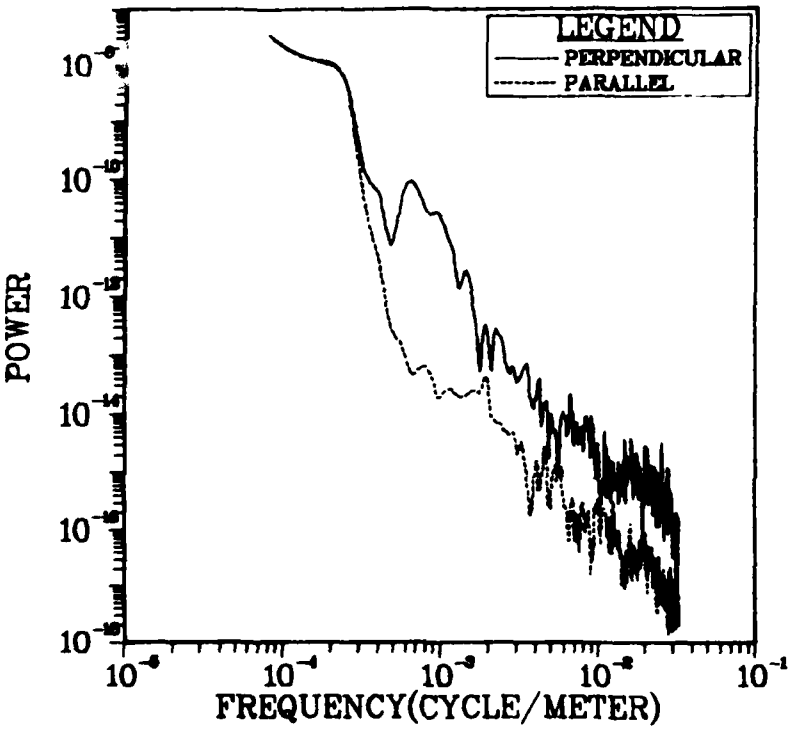


Figure 8. Two-dimensional power spectrum (400 mm camera).

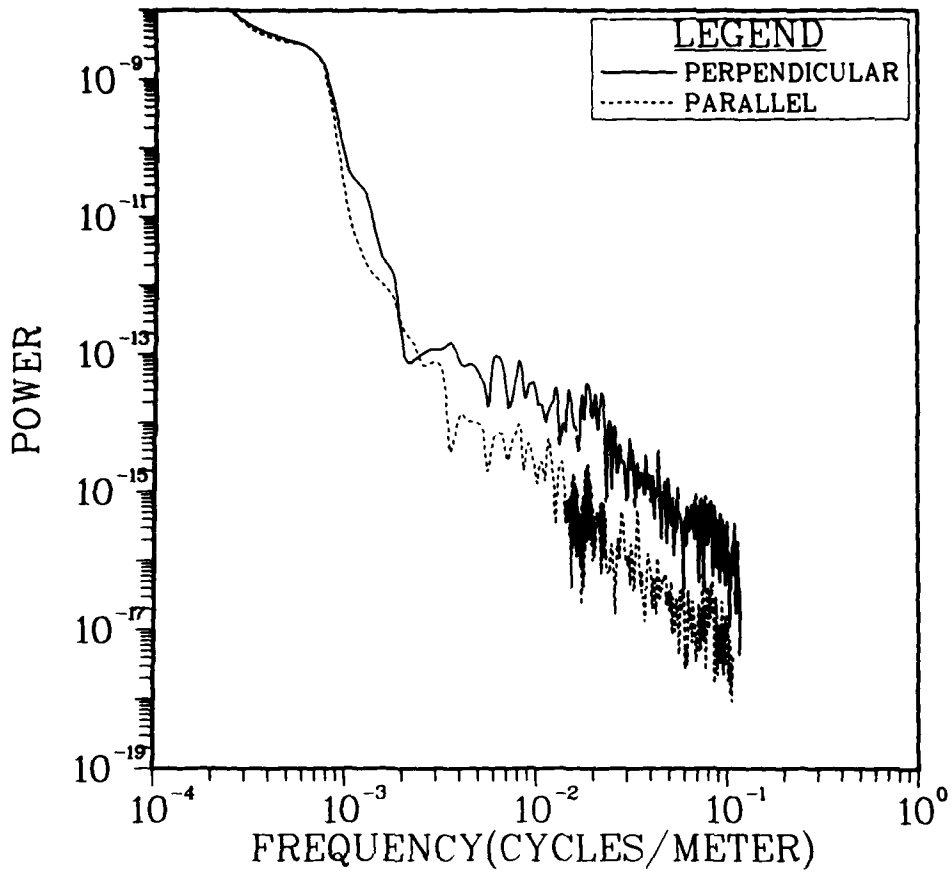


Figure 9. Two-dimensional power spectrum (1250 mm camera).

EFFECTS OF LOCAL SITE SCATTERING ON HF ANTENNA PERFORMANCE*

F. H. EDDY

MITRE CORPORATION

Bedford, MA 01730

INTRODUCTION

Highly-directive low-sidelobe antennas provide unique advantages for HF radar and communications systems. These advantages result from both the smaller ionospheric area illuminated and the increased azimuthal discrimination available for excluding undesired signals. The ability to achieve the desired low-sidelobe response is ultimately limited by site scattering. This paper seeks to establish theoretical limits upon that scattering as a function of aperture size and desired sidelobe level. Instrumentation was developed for the purpose of ensuring that potential antenna sites would satisfy the required scattering criteria. Experimental results obtained with that equipment are briefly described.

ADVANTAGES OF HIGHLY DIRECTIVE ANTENNAS

Interference

In most HF applications involving information bandwidths of 3 kHz or greater, performance sensitivity is frequently limited by external interference. This interference can be minimized through processing in any of the available domains: time, frequency, and/or angle. The more domains in which discrimination is employed, the greater the attainable interference rejection. Owing to the log-normal distribution of interferers, the potential gains (and losses) are greater than would be computed for Gaussian noise.

It is instructive to consider the effects of angular processing upon interference. The interference probability distribution exhibits log-normal behavior of the form plotted in Figure 1. In the absence of specific siting data, assume that the probability of an interferer is uniformly distributed in azimuth. It is possible to convolve the antenna gain function with the (possibly azimuthally weighted) interference probability distribution so as to obtain the probability distribution at the antenna terminals. This resulting distribution is highly dependent upon antenna sidelobe structure and beamwidth. Low-resolution, high sidelobe-level antennas yield high-probability of high interference, as shown in Figure 1, curve A. Conversely, high-resolution, low-sidelobe antennas yield a distribution of the form shown in Figure 1, curve B. It is seen that with narrow-beam antennas, the interference peak has been lowered by roughly the sidelobe suppression ratio. An interesting unsolved problem is that of determining the optimal aperture illumination function for maximizing SNR for a given aperture in the presence of log-normal noise.

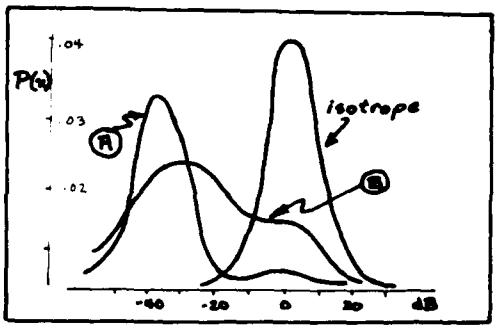


Figure 1 -- Log-Normal Distributions Observed with Various Antennas

Ionosphere

It is well known that the ionosphere fine-structure serves to scatter and distort signals passing through it. These ionospheric inhomogeneities become of greater significance as the size of the illuminated area is increased. For small angles, the lateral extent of the illuminated ionosphere is proportional to

$$D(1+\tau\lambda/D^2) \qquad (1a)$$

*The above effort was carried out in support of the USAF Systems Command Electronics System Division CONUS OTH-B program under contract F19628-77-C-0001.

for an aperture focused at infinity, and

$$r\lambda/D \qquad (1b)$$

for focused beams. Here, D is the aperture extent, r is the range to the ionosphere, and λ is the operating wavelength. In most practical applications, it is seen that the illuminated ionospheric area is minimized through utilization of large apertures.

Capture Area and Resolution

Increased aperture area permits greater resolution for a given sidelobe level. Specifically, the received power is proportional to

$$A_r / 4\pi r^2 \qquad (2)$$

where A_r is the effective receiving aperture area. Angular resolution is, is of course, proportional to

$$\lambda/D \qquad (3)$$

It is therefore seen that large apertures confer important benefits to both radar and communications systems operating in the HF band. These advantages, however, are contingent upon the ability to regulate sidelobe level. Independent of element-excitation errors, site-induced scattering and reflections may serve to limit these sidelobes, and to limit the performance of adaptive systems. It is therefore of considerable practical concern to develop means for estimating site limitations to antenna gain and sidelobes.

THEORETICAL BACKGROUND

Introduction

Scatterers re-radiate incident fields in a complex manner. Such re-radiation from scatterers near antennas can cause an apparent degradation to sidelobe performance. In the case of the transmit antenna, some fraction of the energy initially within the mainlobe is re-radiated at undesired which vectorially combines with the free space sidelobe radiation, thereby modifying (generally degrading) the free-space sidelobes in the manner illustrated in Figure 2.

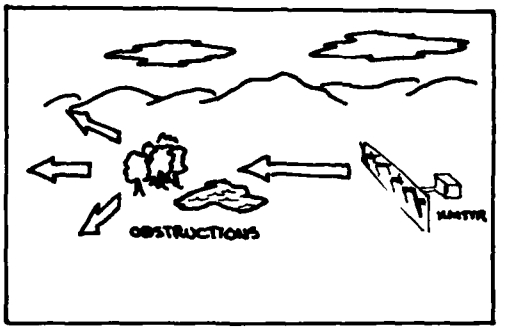


Figure 2 -- Transmit Sidelobe Degradation

In the case of the receive antenna, scatterers in front of the array may re-radiate signals from undesired azimuth angles. These signals, generally in the near field of the array, likewise combine with the free-space response function. The net result cannot be distinguished from a degraded sidelobe response as shown in Figure 3.

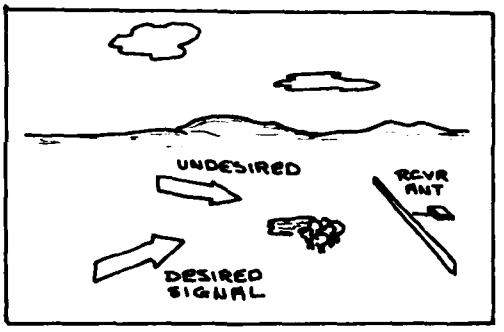


Figure 3 -- Receive Sidelobe Degradation

In the following, reciprocity applies. The discussion is framed in terms of a receiving antenna, but the results apply equally to a transmitting antenna.

Cross-Section Definition

It is appropriate to consider definitions of scattering cross-section. Radar cross-section σ is normally defined in terms of the radar equation,

$$\sigma = (4\pi)^3 r^4 P_r / P_T (G\lambda)^3 \qquad (4)$$

and therefore has the dimensions of area. Equivalently, it can be defined (see Kerr, Chapters 2 and 6) as the limit

$$\sigma = 4\pi r^2 \lim_{r \to \infty} (E_s/E_i) \qquad (5)$$

where E_i is the incident electric field and E_s is the scattered field measured with the same probe at distance r. (The limit is extended to infinity so as to eliminate possible near-field limitations.)

For a target of physical area A it is often useful to introduce a dimensionless scattering parameter σ° such that

$$\sigma^\circ = \sigma/A \qquad (6)$$

where σ° may be viewed as a loss (gain) function normalized to the power-density isotropically scattered by an "ideal" scatterer of area A. Imbedded within σ° is an implicit angular dependence, i.e., $\sigma^\circ = \sigma^\circ$ (θ,φ). Since scattering from certain rough surfaces is greatest near normal incidence, and least parallel to the surface, a further dimensionless scattering coefficient γ is occasionally introduced, where

$$\gamma = \sigma^\circ \sec\theta \qquad (7)$$

and where θ is defined in Figure 4.

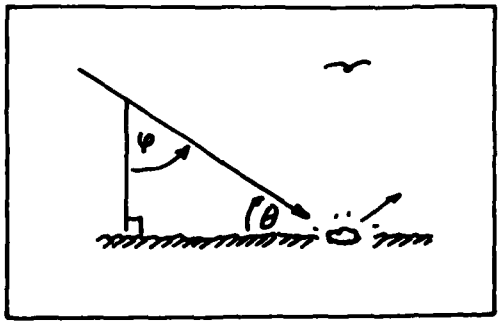


Figure 4 -- Scattering Geometry

Hagn has further introduced a function (φ) which differs from γ by essentially a factor of two. In the literature, all the foregoing (and some additional) definitions of scattering coefficient have been used. For the sake of internal consistency, cross-section will be used here, however, only in the form

$$\sigma^\circ = P_R r^4 (4\pi)^3 / P_T A_R (G\lambda)^2 \qquad (8)$$

Propagation over Lossy Ground-plane

Scattering objects, such as trees, etc., are often positioned upon a lossy ground plane. Propagation of signals scattered from a phase center above that lossy ground to a receiving aperture will be subject to interference (Lloyd's Mirror) as shown in Figure 5.

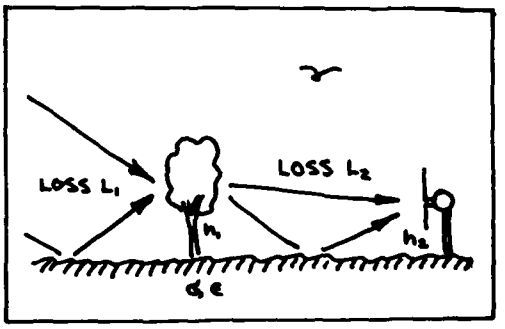


Figure 5 --Ground Interference Geometry

Normally the two rays (geometric optics assumed to apply for distances large compared to a wavelength) will destructively interfere at small angles (i.e., $r/h \gg 1$). In the literature, the effect of this propagation loss has sometimes been included within the angular dependence of the cross-section. This has the obvious advantage that the experimental measurement is greatly simplified: it is not necessary to measure the ground parameters. The corresponding disadvantage follows as a direct corollary: it is difficult to apply the results obtained from scatterers on one unspecified surface to those on another.

Computation of ground losses is dependent upon the shape of the ground in the region of the reflection, but decreasingly so as the incidence angle becomes smaller. In the following, the theoretical flat-earth propagation is presumed based upon best-fit ground parameters computed using that model.

From Bullington, the propagation loss L can be written in the form

$$L = P_R / P_T [\lambda/4\pi r]^2 \qquad (9)$$

$$= (1 + Re^{j\Delta} + [1-R]Ae^{j\Delta})$$

where R = (sin θ - z)/(sin θ + z), and θ is the incidence angle

and where $z = \begin{cases} \sqrt{\epsilon_0 - \cos^2\theta}\,/\,\epsilon_0, & \text{vert.} \\ \sqrt{\epsilon_0 - \cos^2\theta}, & \text{horiz.} \end{cases}$

$$\epsilon_0 = \epsilon - j\,60\sigma\lambda \quad (mks)$$

σ = conductivity (mhos/m)

ϵ = relative dielectric const.

$$\Delta = (2\pi r/\lambda)\left\{\left[\left(\frac{h_T + h_e}{r}\right)^2 + 1\right]^{1/2} - \left[\left(\frac{h_T - h_e}{r}\right)^2 + 1\right]^{1/2}\right\}$$

$$\simeq 4\pi h_e h_T, \quad h_T + h_e \ll r$$

$$A = -1/(1 + j[2\pi r/\lambda] \cdot [\sin\theta + z]^2)$$

where A is the surface-wave attenuation factor with it understood that this relationship holds only for $|A| \ll 1$.

A computer program was written to evaluate propagation loss for horizontal, vertical and free-space propagation for transmit antenna height of 10 m and with receive heights of 2 and 4 m (corresponding to presumed scatterer height). These results have been plotted in Figure 6 with h = 4 m for representative values of both dielectric constant and ground conductivity.

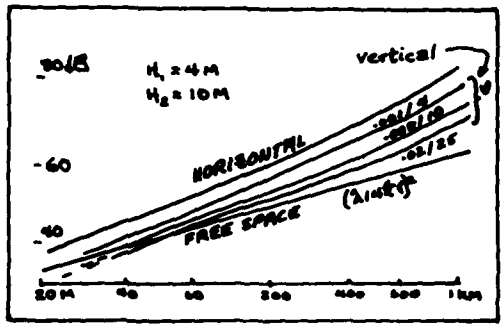


Figure 6 -- Ground Losses

The spread in loss for vertical polarization is less than 10 dB at 1-km range. In experimental measurements to be discussed later, the attempt has been made to infer ground parameters from the excess loss measured as a function of range. But since the experimental error (reproducibility) is estimated to be +/- 0.5 dB with fixed calibration error of up to 1.0 dB for a maximum range under 1-km, it can be appreciated that the experimental measurement of field-strength is only sufficient to confirm "poor" vs. "good" earth. Conversely, the actual field strength (and hence, measured cross section) is only weakly dependent upon ground parameters.

From these foregoing results it is possible to introduce an excess loss factor L within the range equation. This one-way loss is plotted in Figure 7 for 4 and 10 m scatterer and antenna heights and from 20 m to 1 km range. The loss can be considered to be sensibly constant over restricted ranges and heights.

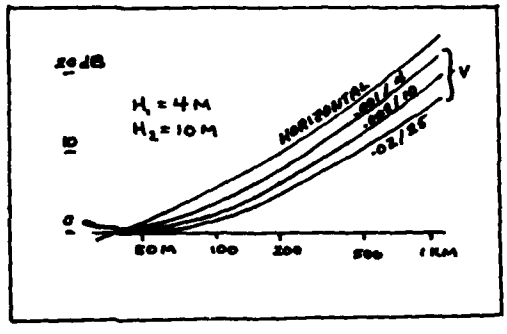


Figure 7

Additional Sources of Loss

Additional loss may result from such factors as terrain irregularity, vegetation, and diffraction (screening). These losses are not readily computed except for highly idealized circumstances. Consider diffraction. The diffractive loss for a knife-edge screen is readily computed. Unfortunately, the loss introduced by natural (rounded) obstacles is not: the rule of thumb normally employed is merely to add 10 to 20 dB to the theoretical result for the knife edge (cf. ITT Handbook, IV-th Edition, 26-23). The rule-of-thumb correction uncertainty for obstacle rounding is large and must be applied twice. Experimentally observed losses tend to confirm that the diffraction loss from extended bodies is large, but has not permitted experimental verification of any of the various theoretical analyses in the literature.

When the antenna is located in a forested region, then propagation through the trees is of significant interest for computing effective scattering cross section of those areas screened by the trees. It has been shown (cf. Pounds and LaGrone, Lippmann, Taylor, Sachs and Wyatt, and Tamir) that propagation through heavily wooded forest can be viewed as being equivalent to propagation over the surface of a lossy dielectric slab. The theoretical complexities of the supporting analysis are formidable. Nonetheless, reasonable agreement between theoretical predictions and experimental data has been obtained for effective dielectric constant and conductivity as determined by Hagn and Parker. Specifically, for forest at HF frequencies, the slab parameters range between 1.05 and 1.15 for equivalent dielectric constant and between 1.5E-4 and 5E-5 (mhos/m) for equivalent conductivity yielding an attenuation in the order of 0.5 dB/m over the HF band (with variations due to polarization and frequency ranging from 0.2 to 1.0 dB/m) and are consistent with our measurements.

Azimuthal Dependence of Scattering

Both for the purposes of measurement and extrapolation it is necessary to consider the possible variation in cross-section as a function of azimuth angle. As shown by the work of Steele, trees can be represented by equivalent conducting cylinders with dimensions comparable to that of the trunk. The scattering from vertical cylinders will be truly isotropic in azimuth and with a cosinusoidal envelope in elevation. The elevation factor will be dominated by ground loss. If the trees were to grow at an oblique angle, say as the result of pressure from prevailing winds, or if the branches grew in asymmetric fashion, a weak angular dependence might be observed. If scatterer dimensions are large compared to a wavelength, substantial azimuthal variation could be anticipated by, for example, preferential scattering from a cliff or perhaps a periodic corrugation of the terrain.

Sidelobe Degradation From Scatter

Assume a distant (far-field) signal. The power output of the antenna for the direct signal is

$$P_0 = P_T G_T A_R / 4\pi r_D^2 L_1 \qquad (10)$$

where A_R is the effective receiving aperture, $P_T G_T$ is the effective radiated power, and r_D is the range. The loss L is as previously defined in Eqn (6). For typical over-the-horizon elevation angles and receive aperture heights, L_1 will have values near unity (near zero dB excess loss).

The full gain of the aperture will not normally be realized for near-field scatters. The effective aperture area for a near-field scatterer (allowing for plus-or-minus 45-degree phase variation over the central zone and assuming the integral over the outer zones to be zero) is, from Figure 8,

$$A_R \simeq \sqrt{2\lambda r_0}\, h \qquad (11)$$

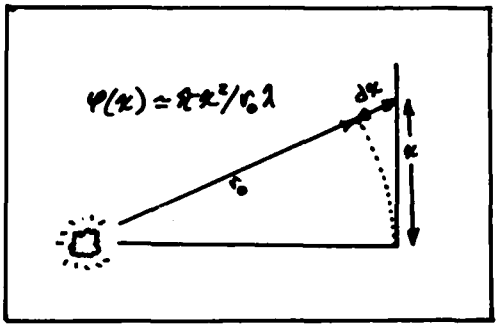


Figure 8 -- Near-Field Aperture

where h is the effective height of the aperture. Note the inherent advantages of large aperture antennas. They reduce the relative response to near-field scatterers while enabling looser element tolerances (presuming random phases for both scatterers and errors). Hence the scattered power at the receive antenna output is

$$P_s = \frac{P_T G_T h\sqrt{2\lambda r_s}}{4\pi r_D^2} \int \frac{\sigma^\circ \, dA_s}{L_1 L_2 \, 4\pi r_s^2} \qquad (12)$$

and the normalized output is

$$P_s/P_D = \frac{h\sqrt{2\lambda}}{4\pi L_2 A_r} \int \frac{\sigma^\circ \, dA_s}{r_s^{3/2}} \qquad (13)$$

or, for $A_R = hD$

$$P_s/P_D = \frac{\sqrt{2\lambda}}{4\pi D L_2} \int \frac{\sigma^\circ \, dA_s}{r_s^{3/2}} \qquad (14)$$

This scattered signal must be integrated (in power as each scatterer's phase is presumed to be random) over an area bounded by that parabola with differential time delay corresponding to the width of one range bin. For example, for bandwidth W ≃ 10 kHz and

$$\Delta r = c/2W \qquad (15)$$

$$= 3E8 \text{ m/s} \;//\; 2E4/\text{s}$$

$$\simeq 15 \text{ km}$$

For example, if

$\lambda = 30$ m
$r_s = 1.5 \times 10^3$ m
$dA_s = 10^6$ m
$D = 10^3$ m
$SL = 10^{-5}$

where SL is the required (normalized) sidelobe level, then

$$P_s/P_D < 1E\text{-}5 \qquad (16)$$

$$\sigma^\circ \simeq (1E\text{-}3) L_2 \qquad (17)$$

Thus for L_2 = +10 dB (neglecting diffraction losses), the effective cross-section must be less than 20-dB if 50-dB sidelobe suppression is to be obtained.

EQUIPMENT

Equipment was developed to permit direct measurement of site scattering and is shown in block diagram form in Figure 9. A Fluke synthesizer is used to generate the carrier frequency. The carrier frequency (29.6 MHz) can be quadratically phase modulated at a rate not coherent with the 2-us pulse-width (1/4-usec in the high-resolution mode) gate modulation (nominal 100-Hz, but operator selectable). In practice, this feature was not found to be necessary; the video return does not appear to be highly coherent when examined on a logarithmic scale. Modulator leakage level is less than receiver noise. The pulse-modulated signal is amplified up to an approximate 100-Watt peak power level, then applied to an 8.5-dB gain Yagi antenna through a hybrid-T.

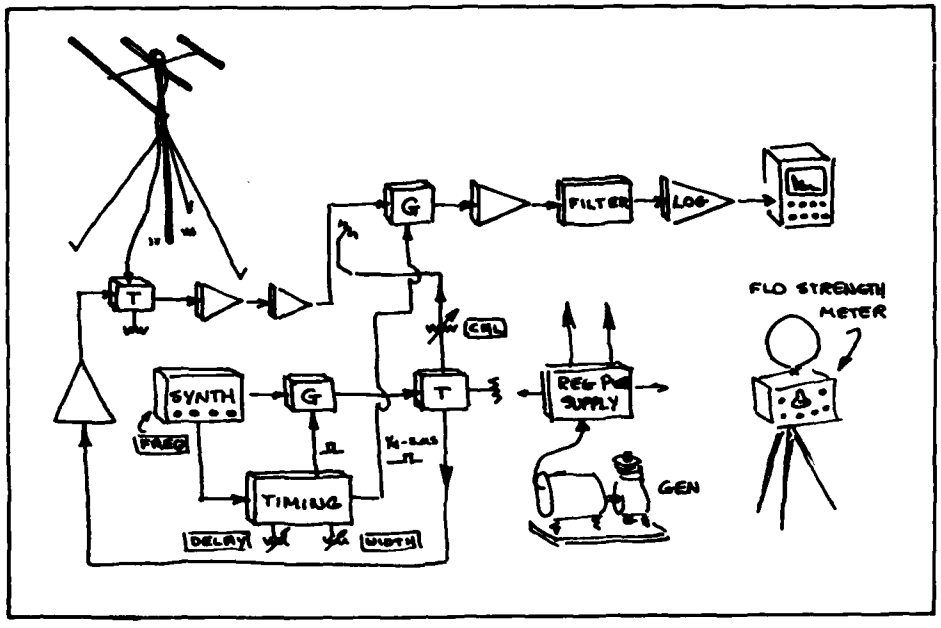


Figure 9 -- Equipment Block Diagram

The Yagi is tuned to the operating frequency and has a swept frequency VSWR of 1.2:1 or less over the 3-dB bandwidth (approximately 300 kHz narrow band and up to 2 MHz in the wide-band mode). The antenna is mounted on a 30-ft pole which can be manually steered in aziumth. Polarization can be selected by rotating the elements on the boom to lie in either the horizontal or vertical plane. By mounting the elements on a five-foot fiberglass tube, which itself extends from a 25-foot metal tube, and restricting operation to no more than 10-degrees from the vertical, antenna VSWR is independent of polarization. The isolated port of the hybrid is connected to preamplifiers followed by an approximate matched-filter (linear-phase 300-kHz bandwidth filter with 30-dB insertion loss for the wide pulse; about 2 MHz and with less loss in the wide-band mode).

Preamplifier gain is sufficient to offset filter loss. Overall noise figure is substantially less than 5-dB. The filter input is dunked during the main bang. The filter output is then applied to a logarithmic amplifier whose detected-video output is linear over at least 70-dB and with useful output over approximately 35-dB. The output is displayed on a rack-mounted oscilloscope with calibrated graticule (calibrations based upon 1/r-cubed propagation).

Equipment calibration has been performed before and usually after operations at each site. No discernible variation in equipment sensitivity has been observed. Calibration is accomplished by injecting an attenuated sample of the transmitted signal into the receiver with a suitable time delay. Oscilloscope gain is varied until a 20-dB change in injection level causes a 20-dB change in display level. The d-c offset of the oscilloscope is then adjusted until the apparent cross-section lines up with the indicated cross-section. Calibration accuracy is estimated at about +/- 1 dB. Since the rms antenna mismatch (less than 1.2:1 VSWR) introduces less than 0.5-dB loss (each way), and since the hybrid is estimated to introduce 0.6-dB (each way), the total system loss is estimated at 2.2 dB. During operation, the gain of the final three stages of power amplification are not directly monitored. The antenna 17.0-dB two-way gain was obtained from manufacturer data and has not been independently calibrated. The total uncertainty in uncalibrated system gain is believed to be less than 5 dB and is probably somewhat lower.

TEST RESULTS

Scattering cross-section measurements obtained with this equipment, when corrected in accord with the theoretical procedures outlined in previous sections (with that corroboration checked using a commercially available field-strength meter) have yielded scattering cross-sections in general agreement with the literature. Specifically, densely forested regions exhibit backscatter cross section in the order of -20 dB.

This equipment has been successfully operated at two prospective HF radar sites for the purpose of determining site-induced sidelobe limitations. After compensation for propagation loss and ultimately planned antenna aperture dimensions, these site limits were computed to be between -30 and 35 dB at one site, and below -60 dB at the other.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the support of D. J. Marino, the design effort of C. J. Beanland, the general assistance rendered by R. G. Hansen, and the invaluable field efforts of J. O. Runkle.

BIBLIOGRAPHY

1. Bachynski, M. P. and Kingsmill, M. G. Effect of Obstacle Profile on Knife Edge Diffraction, IEEE Trans., AP-10 (March 1962), p. 201

2. Barnum, J. R., High Frequency Backscatter From Terrain with Cement Block Walls, IEEE Trans. AP-19, 3, (May 1971), pp. 343-347.

3. Beckmann, P. and Spizzichino, A., The Scattering of Electromagnetic Waves From Rough Surfaces, Pergamon Press, 1963.

4. Bullington, K., Radio Propagation at Frequencies Above 30 MHz, Proc.IRE, 35 (October 1947), p. 1122

5. Bullington, K. Radio Propagation Fundamentals, BSTJ 36 (May 1957), p. 593

6. Dougherty, H. T. and Maloney, L. J., Application of Diffraction by Convex Surfaces to Irregular Terrain Situations, Radio Phone, 68B, No. 2 (February 1964).

7. Egan, K., McCullough, W., and Clancy, J., Beverage Element Planar Array Test and Analysis Final Report, Rome Air Development Center, Long Range Surveillance Section, Griffiss AFB, NY, November 1976.

8. Green, J. A., Dufresne, D., Effective Antenna Array Sidelobe Levels, Proc. of the OHD Technical Review Meeting of 25-27 March 1975, Volume 1, Stanford Research Institute, Menlo Park, CA, May 1975.

9. Hagn, G., An Investigation of the Direct Backscatter of High-Frequency Radio Waves From Land, Sea Water, and Ice Surfaces, Final Report II, SRI, Menlo Park, CA May 1962.

10. Head, H. T., The Influence of Trees on Television Field-strengths at Ultra-high Frequencies, Proc. IRE, 48, pp. 1016-1020, June 1960.

11. Kerr, D. E., Propagation of Short Radio Waves, McGraw-Hill, 1951

12. Millington, G., A Note on Diffraction Around a Sphere or Cylinder, Marconi Rev., 23, (1960), p. 170

13. Steele, J. G. High Frequency Measurementsof Radar Cross-section Using the Standing-Wave Method, IEEE Trans. AP-15, 4, (July 1967), pp. 557-563.

14. Steele, J. G., High Frequency Backscatter from Terrain with Trees, Proc. IEEE, 55 9 (September 1967) p. 1583.

15. Tamir, T., On Radio Wave Propagation in Forest Environments, IEEE Trans. AP-15, 6 (Nov 1967), pp. 806-817.

16. Tamir, T., Radio Wave Propagation Along Mixed Paths in Forest Environments, IEEE Trans. AP-25, 4, (July 1977) p. 471 ff.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1978 O - 277-182