AGARD-CP-574

AGARD-CP-574

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

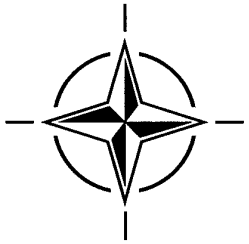
AGARD CONFERENCE PROCEEDINGS 574

Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design

(Systèmes de propagation numériques: effets de la propagation, solutions techniques, conception des systèmes)

Papers presented at the Sensor and Propagation Panel Symposium, held in Athens, Greece, 18-21 September 1995.

19960719 016



NORTH ATLANTIC TREATY ORGANIZATION

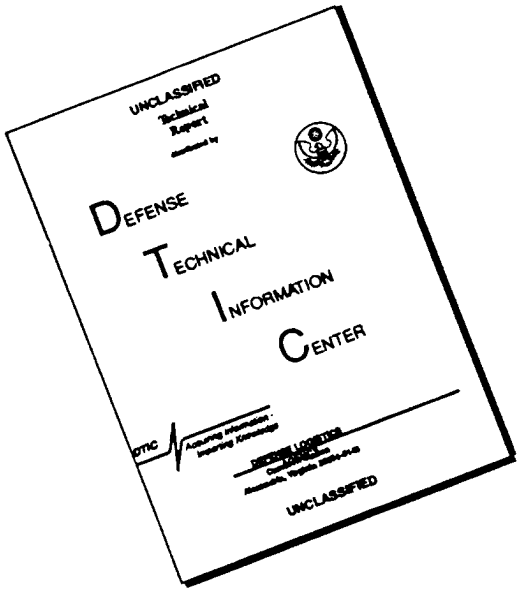
DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

Published April 1996

Distribution and Availability on Back Cover

QUALITY INSPECTED

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

The Mission of AGARD

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

- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;
- Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Exchange of scientific and technical information;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

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

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

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

Published April 1996

Copyright © AGARD 1996
All Rights Reserved

ISBN 92-836-0023-1

Printed by Canada Communication Group
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design

(AGARD CP-574)

Executive Summary

Digital communications systems are important elements in military systems because of the security and discretion which they provide in the transmission of information.

In addition to the great many advances which have been made in the last decade, this symposium showed that very considerable progress will be achieved in the near future.

The studies being carried out on transmission media in all the frequency ranges, from decametric waves (HF waves) to submillimetric waves (EHF waves) enable us to model them better, to simulate them better and then to use them better.

The spin-off for transmission systems is immediate. New advances have been made, using more complex signal processing, data encoding and processing procedures, which now appear as the vectors of future progress. The symposium showed that we can expect considerable improvements in the future:

— in the quality of data transmission by the use of more complex techniques and better matching to transmission channels;

— in the security and discretion of communications, in particular by the use of more powerful encoding;

— in the adaptivity of systems to transmission channels by the use of new methods, such as the real time evaluation of these channels and the use of new methods such as passive evaluation, which introduces a new argument into the debate about the discretion of communication links.

In conclusion, it would appear that digital data links are still in full progress, driven by major technical developments.

Systèmes de transmission numérique: effets de la propagation, solutions techniques, conception des systèmes

(AGARD CP-574)

Synthèse

Les systèmes de transmissions numériques sont importants dans les systèmes militaires par la sécurité et la discrétion qu'ils assurent dans la transmission des informations.

Si de nombreuses avancées ont été faites durant la dernière décennie, le symposium a montré que de très importants progrès sont à attendre dans un avenir proche.

Les études menées dans toutes les gammes de fréquences, des ondes décamétriques (ondes HF) aux ondes submillimétriques (ondes EHF), sur les milieux de transmission permettent de mieux les modéliser, de mieux les simuler et, par suite, de mieux les utiliser.

Les retombées sur les systèmes de transmission sont immédiates. De nouveaux progrès ont été révélés, faisant appel, notamment, à des procédures plus complexes de traitement de signal, de codage et de traitement de l'information qui apparaissent comme les vecteurs porteurs des progrès futurs.

Le symposium a montré qu'il faut s'attendre pour l'avenir à des améliorations importantes:

— Sur la qualité des transmissions de données par l'emploi de techniques plus complexes et de meilleures adaptations aux canaux de transmission;

— Sur la sécurité et la discrétion des communications, notamment par la mise en œuvre de codages plus puissants;

— Sur l'adaptativité des systèmes aux canaux de transmission par l'évaluation en temps réel de ces derniers et par de nouvelles méthodes, telles que les évaluations passives qui ajoutent une discussion supplémentaire à la discrétion des liaisons.

Il apparaît finalement, que les transmissions numériques de données sont encore en pleine progression soutenues par les importants développements de la technique.

Contents

	Page
Executive Summary	iii
Synthèse	iv
Theme/Thème	ix
Sensor and Propagation Panel	x
Foreword	xi
Avant-propos	xiii
SESSION I: RECENT ADVANCES IN PROPAGATION **Chairmen: B. Ghicopoulos, E. Schweicher and D. Yavuz**	
Flow of Information in Future Air War by D. Heliotis	1
An Analytic Ray Tracing Model for HF Ionospheric Propagation by R.J. Norman, I.G. Platt and P.S. Cannon	2
An Improved Model of Ionospheric Absorption at HF High Latitude Paths by V. Jodalen and E.V. Thrane	3
PAPER 4 WITHDRAWN	4
Passive HF Propagation Evaluation Technique by H. Soicher and Z. Houminer	5
A Ray-based Propagation Tool for Digital Systems by P. Charrière and K.H. Craig	6
PAPER 7 WITHDRAWN	7
UHF and Microwave Propagation Prediction in an Urban Environment by T.M. Willis, B.J. Guarino, J.D. Moore, R. Luebbers, J. Schuster, H.L. Bertoni, G. Liang, N. Wo, T.S. Rappaport and P.M. Koushik	8
Military Applications of Site-Specific Radio Propagation Modeling and Simulation by R.M. Bauman	9
PAPER 10 WITHDRAWN	10
A Model for Estimating Electromagnetic Wave Attenuation in a Forest (EWAF) Environment by C. Welch, C. Lemak and L. Corrington	11
Atmospheric Path Loss and Scintillations affecting Communications in the EHF Frequency Range by H.H. Fuchs and R. Makaruschka	12

Characterisation of the Transfer Function of the Clear Atmosphere at Millimetre Wavelengths 13
by A.D. Papatsoris

MILSATCOM – Where Are We Now? 14
by J.B. Cairns and J.R. Watson

Earth-Space Links and Fade-Duration Statistics 15
by F. Davarian

Optogeometrical Characterisation of Erbium Doped Fibres 16
by F. Ravet, B. Heens, E. Jaunart, P. Mégret and M. Blondel

SESSION II: PROPAGATION LIMITATIONS
Chairmen: A. Lanusse and J. Harvey

Measurements of Doppler Spread on High Latitude HF Paths 17
by M.J. Angling, P.S. Cannon, N.C. Davies, B. Lundborg, V. Jodalen and K.W. Moreland

Observations of Doppler Spreading and FSK Signalling Errors on HF Signals Propagating over High Latitude Paths 18
by B.S. Dhanda, E.M. Warrington and T.B. Jones

Comparaison entre les Formes d'Onde Série et Parallèle pour les Transmissions HF 19
by D. Pirez and D. Merel

Channel Evaluation From Predicted Zero-Crossing Analysis 20
by P.W. Piggin and M. Gallagher

New Technologies to Improve HF Modems Performance for Standardized Waveforms 21
by A. Brakemeier and A. Kotlowski

Aide à la Décision pour la Recherche de Canaux Clairs dans la Bande Décamétrique 22
by J. Wolf and Y.M. Le Roux

Compatibilité Electromagnétique dans la Bande Décamétrique: Modélisation, Prévision, Evaluation Passive de l'Ionosphère 23
by J. Caratori and C. Goutelard

Evaluation de la bande de cohérence du canal de transmissions VHF/UHF longues distances 24
by G. Aignel, O. Ravard, D. Sorais and L. Bertel

Communications and the Global Positioning System 25
by J. Aarons, M. Mendillo and R. Yantosca

Modélisation et Mesure de la Propagation Radioélectrique sur les Liaisons Terre-Satellite dans la Bande Millimétrique 26
by L. Castanet, N. Douchin, J. Lemorton, R. Colas des Francs, D. Le Boulc'h and S. Grely

L'Atténuation des Effets de Trajets Multiples et de l'Interférence dans les Communications Radio Militaires par le Traitement du Signal 27
by G. Multedo and G. Goudezeune

SESSION III: SIGNAL PROCESSING
Chairmen: A. Lanusse and J. Harvey

Les Séquences GQ Séquences Q-Aire Orthogonales à Corrélation Parfaite 28
by C. Goutelard

Discrete Hilbert Transform 29
by A. Marguinaud and S. Gienger

Effects of Bandwidth Reduction of Transmitted Motion Picture Sequences on Human Recognition Performance 30
by K.-P. Gärtner and F.E. Schneider

Analysis of Propagation Effects and Non-Linear Distortions on Digital Modulation Schemes 31
by K. Metzger and R. Valentin

Integrated Data and Channel Estimation for Selective Fading Channels 32
by L. D'Ambrosio, R. Marchesani and M. Ruggieri

Link Level Protocols for Radio Channels 33
by D. Yavuz

Definition of a Communication System Using the Tropospheric Channel 34
by G. Vivier, P. Brélivet, J. Richard and P. Sehier

PAPER 35 WITHDRAWN 35

Comparison of RLS Adaptive Array Algorithms in a Strong Multipath Propagation Environment 36
by D. Peeters and E. Van Lil

SESSION IV: SIMULATION CHANNELS
Chairmen: J. Harvey and L.B. Stotts

Testing HF-Modems: DSP's Open New Possibilities and Make New Demands on Real-Time HF-Simulators 37
by L. Van der Perre and A. Van de Capelle

A Flexibly Configurable Statistical Channel Model for Mobile Radio Systems with Directional Diversity 38
by J.J. Blanz, P. Jung and P.W. Baier

Error Probability Computation in Optical Fiber Links 39
by M. Vandroogenbroek, P. Mégret and M. Blondel

Gain-Switched DFB Lasers as Soliton Sources for High Bit Rate, Long Distance Transmissions 40
by L. Meuleman, X. Wang and M. Blondel

Site-Specific Radio Propagation Prediction Methods for Urban Environments 41
by R. Luebbers and J. Schuster

SESSION V: METEOR COMMUNICATIONS
Chairman: L.B. Stotts

An Impulse Response Measurement System and Some Experimental Results from a Forward Scatter Meteor Burst Link 42
by K.J. Ellis, A.R. Webster, J. Jones and S. Chow

Description expérimentale et modélisations du canal météorique 43
by O. Ravard and D. Sorais

Beam Forming Techniques in Meteor Scatter Communications Systems 44
by A. Akram and P.S. Cannon

SESSION VI: COMMUNICATION SYSTEMS
Chairmen: L.B. Stotts and K.H. Craig

Coverage Area Computations: An Approach with Diffuse Reflections 45
by E. Van Hoof, E. Van Lil, F.J. Rodriguez Blanco and F. Perez Fontan

Computer Aided Frequency Planning for the Radio and TV Broadcasts 46
by A. Altintaş, O. Ocali, S. Topçu, S.G. Tanyer and H. Köymen

Introduction to Rural Areas Telecommunications and Development of a Pilot Terminal Unit 47
by R. Makri, M. Gargalakos and N.K. Uzunoglu

Communications And Situation Awareness On Hostile Borders 48
by R.W. Kocher and J.G. Allen

Communications Realism with Integrated Terrain-Environment-Multipath Model 49
by C. Sheth, S. Barone, M.R. Lambert and L.R. Kennedy

Système de Transmission Numérique Multi-Porteuses par Voie Ionosphérique 50
by S. Rigaudeau and Y.M. Le Roux

PAPER 51 WITHDRAWN 51

Influence des Antennes et de la Propagation sur le Comportement d'un Système de Transmissions Numériques en H.F. 52
by L. Bertel, O. Lebaillif, Y. Le Roux and R. Fleury

SISP — A Software Tool for Propagation Prediction 53
by P.M. Koushik, T.S. Rappaport, M. Ahmed and N. Zhang

Hybrid Direct Sequence – Frequency Hopping Cellular System 54
by K. Kehagias

Theme

The increase of digital system data rates, their rapid extension to ever-growing applications make these systems more dependent on propagation effects. Much work is being carried out in NATO countries on systems of considerable interest for military applications. In the high-frequency domain, studies on reliable modems are still of current interest. In the metric and decimetric wave bands, propagation effects are well known, especially those related to communication range variations, electronic countermeasures, and urban zone propagation. In the microwave band, diffusion phenomena are of greater importance. The problems caused by propagation in digital communications systems are far from being mastered, and the optimum technical solutions have not yet clearly appeared, partly because of the rapid emergence of new signal and data processing techniques.

The symposium enabled the scientific and technical communities to exchange their expert knowledge, in particular in the areas of the physics of propagation, modelling and systems.

The symposium covered the following topics:

— Limitations imposed by propagation on digital communications systems. All frequency bands;

— Recent advances in propagation evaluation and in propagation models;

— Simulation of communication channels;

— Signal processing methods for digital communication systems;

— Data digital communication systems: urban zones, satellites, networks, adaptive systems;

— New concepts for systems design.

Thème

L'accroissement des débits de données des systèmes numériques, leur généralisation dans des applications de plus en plus nombreuses, rendent les systèmes de plus en plus dépendants des effets de la propagation. Une somme considérable de recherches se développe dans les pays de l'OTAN sur des systèmes qui présentent une importance extrême dans les applications militaires. Dans le domaine des ondes décamétriques, la recherche de modems fiables est toujours d'actualité. Dans les gammes métriques et décimétriques, les effets de la propagation sont notoires, notamment dans les problèmes de portée, de contre-mesures électroniques et de propagation en zone urbaine. Dans le domaine des micro-ondes, les phénomènes de diffusion prennent encore une plus grande importance. Les problèmes que pose la propagation dans les systèmes numériques sont bien loin d'être maîtrisés, et les solutions techniques les meilleures n'apparaissent pas encore clairement, compte tenu notamment de l'évolution rapide des procédures de traitement susceptibles d'être utilisées.

Le symposium proposé a permis à une double communauté scientifique et technique de s'exprimer, d'apporter des informations nouvelles sur la physique de la propagation, la modélisation et les systèmes.

Questions examinées:

— Contraintes de la propagation sur les systèmes de communication numérique. Toutes les gammes de fréquences;

— Progrès récents sur les mesures de la propagation et les modèles de propagation;

— Simulation des canaux de propagation;

— Procédés de traitement du signal pour les systèmes de transmission numérique;

— Systèmes de transmissions numériques de l'information: zones urbaines, satellites, réseaux, systèmes adaptatifs;

— Conception des systèmes.

Sensor and Propagation Panel

Chairman: Prof. D.H. HÖHN
FGAN
Forschungsinstitut für Optik – FfO
Schloss Kressbach
72072 TÜBINGEN
Germany

Deputy Chairman: Mr. F. CHRISTOPHE
Dept Micro-Ondes
ONERA-CERT Toulouse
BP 4025
2, Avenue E. Belin
31055 Toulouse Cedex
France

SYMPOSIUM
TECHNICAL PROGRAMME COMMITTEE

Co-Chairmen: Prof. C. GOUTELARD (France)
and
Dr. P.A. KOSSEY (United States)

Members: Dipl. Ing. E.P. BAARS (Germany)
Dr. P.S. CANNON (United Kingdom)
Dr. K.H. CRAIG (United Kingdom)
Mr. M. ELLIOTT (NATO Brussels)
Dr. B. GHICOPOULOS (Greece)
Dr. J. HARVEY (United States)
L'IPA A. LANUSSE (France)
Lt-Col. P.-L. MANCINI (Italy)
Mr. J. ROGGE (Netherlands)
Dr. B. STOTTS (United States)
Prof. Dr. C. TOKER (Turkey)
Dr. D. YAVUZ (SHAPE, The Hague)

PANEL EXECUTIVE

Lt-Col. G. DEL DUCA (Italy)

from Europe
AGARD-NATO
ATTN: SPP Executive
7, rue Ancelle
92200 Neuilly-sur-Seine
France

Phone: 33-1-47.38.57.68

from North America
AGARD NATO/SPP
PSC 116
APO AE 09777

Fax: 33-1-47.38.57.99

Foreword

The need to communicate has been essential to man ever since he first started to live in society with his fellows. Telecommunications are simply the extension of the kind of transmission embodied by the legendary messanger of Marathon. The Emperor Tiberius, a figure from the beginning of the modern era, was probably the precursor of digital optical communications. His ideas were relayed to us by the Chappe telegraph, which marked the first scientific invention of an organised telecommunications system. More recent developments in the last decade have seen greater progress in this field than in all the previous centuries.

This symposium was marked by the recent discoveries made in this discipline and characterised by a high number of presentations and a particularly enriching series of subsequent discussions.

An initial overall analysis shows that many new ideas were proposed, producing animated discussion and serious consideration by the participants.

Section 1, devoted to recent advances in propagation phenomena, centred on the physics of the earth's environment. Although the basic phenomena have been understood for a long time, their modelling has led to important advances. Random terms are being replaced by the more precise ones of Fractal, Chaos and Butterfly effect, which demonstrate the concern to achieve more accurate modelling, with more determinist approaches tending to take over from statistical ones. The following salient points emerge from a closer examination of the presentations given:

- First of all, new techniques were proposed for reducing radio wave propagation model computation time by the introduction of codes used in real time in the frequency ranges HF to EHF. Real time evaluation methods for transmission channels are also being developed and there is a desire to introduce passive evaluation methods which are particularly well suited to discrete transmissions and perfectly complementary to existing methods. These passive methods are useful for extreme situations such as those in which combatants are isolated in enemy territory.

- Secondly, the growing interest in propagation, and therefore communications, in urban environments was clearly evident. There are many underlying applications in these studies: the cellular telephone, GPS systems, direct satellite radio and TV broadcasting.... There are a great many problems with propagation in urban environments and not all of them have yet been solved, but it emerged that complete simulation codes have been developed. However, they do not yet seem to provide all the solutions to this difficult question of propagation in urban environments; rather they should be considered as another step in the right direction.

The first set of papers dealt with the characterisation of highly diversified channels working in the entire radio spectrum. A considerable amount of work has been carried out in recent years in the upper part of the spectrum used, from 35 to 90 GHz, and technology developments have produced components at prices which encourage further experimentation.

The results presented show that these channels have special characteristics which require new techniques in order to fully exploit the new possibilities which have been opened up.

Another very important point was underlined by the papers on propagation in polar, auroral and equatorial regions. These parts of the globe are particularly hostile but militarily vital. There are two reasons for studying them, one geophysical, the other technical. The presentations clearly illustrated these two aspects. Although the papers given during this second part of the symposium described undeniable progress, almost all the authors mentioned that they needed to continue their work in order to improve their knowledge, showing the difficulties posed by the hostility to telecommunications channels in these geographical regions.

The third part of the symposium concerned signal processing. This session is often the one where new theories are brought forward. This was the case and so the tradition was maintained. We can now say that complexity of computation is no longer an obstacle for signal and data processing. This session proposed adaptive wave forms, encodings and signal processing techniques in which imagination was the common denominator between the speakers.

Part four dealt with channel simulation. This is a familiar subject in symposia on electromagnetic wave propagation. The papers presented could be seen as an extension of previous studies, introducing as they did, major improvements aimed at a more faithful reproduction of true channels. It is nonetheless true that the choice of model remains the major problem.

Telecommunications systems were discussed in Parts five and six. A great many systems were presented and the question of meteor trail communications systems was raised and as usual, provoked some impassioned discussion. These are, in fact, very old systems, which are nevertheless of great interest for strategic links, as they can be used under extreme limit conditions. Generally speaking, it can be said that the theoretical studies carried out in this field have important spin-offs for system design and operating procedures. This is a comforting thought for the researchers who are upstream of this work. It was shown that civilian systems are either usable or transposable to military systems. This is certainly of increasing interest to the NATO community and is a point for future consideration.

Finally, this symposium shed new light on the important problem of digital communications between points on the earth's surface as well as with airborne and satellite systems. The progress made in recent years is probably one of the reasons for the interest which this symposium generated and it is safe to say that the studies presented and the systems proposed will be overtaken in the future by new studies and new systems with increased performances.

Avant-propos

Le besoin de communiquer s'est imposé à l'homme dès qu'il a voulu se constituer en société. Les télécommunications sont le prolongement de ces communications dont le légendaire messager de Marathon a souligné l'importance. Au début de notre ère, l'empereur Tibère a probablement été le précurseur des communications numériques optiques. Il a été relayé par le télégraphe Chappe qui a marqué la première implantation scientifique d'un système de télécommunication organisé. Les plus récentes évolutions effectuées dans ces 10 dernières années ont fait faire à cette discipline des progrès supérieurs à ceux qui ont existé durant les siècles passés.

Ce symposium a été marqué par les récentes découvertes faites dans ce domaine et caractérisé par un nombre important de présentations et une richesse particulièrement grande des discussions qui ont été engagées.

Dans une première analyse globale, on peut noter que beaucoup d'idées nouvelles ont été proposées et qu'elles ont entraîné de la part de l'auditoire des discussions animées et des réflexions profondes.

Dans la section 1 consacrée aux avancées récentes dans les phénomènes de propagation, la physique de l'environnement terrestre a été au centre des communications. On peut remarquer que, si les phénomènes de base sont pour la plupart connus depuis longtemps, leur modélisation a donné lieu à des progrès relativement importants. On commence à substituer aux termes aléatoires des termes plus précis de fractal, de chaos, d'effet papillon qui traduisent le souci d'une modélisation plus précise tendant à substituer aux approches statistiques des approches plus déterministes. Si on examine plus en détails les communications présentées, quelques points émergeants doivent être signalés.

- Tout d'abord, de nouvelles techniques ont été proposées pour réduire les temps de calcul des modèles de propagation des ondes radioélectriques par l'introduction de codes utilisables en temps réel dans les gammes de fréquences s'étendant du domaine HF au domaine EHF. Des méthodes d'évaluation en temps réel des canaux de transmission sont également développées et il est apparu un souci d'implanter des méthodes d'évaluations passives particulièrement bien adaptées aux transmissions discrètes et parfaitement complémentaires aux méthodes existantes. Ces méthodes passives présentent un intérêt dans des situations extrêmes telles que celles que connaissent des personnes isolées en période conflictuelle sur un territoire ennemi.

- Il est apparu enfin clairement l'intérêt grandissant des propagations, donc des communications, en milieu urbain. Beaucoup d'applications sont sous-jacentes à ces études: le téléphone cellulaire, les systèmes GPS, la radio et la télédiffusion directes par satellite... Les problèmes que posent ces propagations en milieu urbain sont nombreux et encore non résolus en totalité mais il est apparu que des codes de simulation complets avaient été développés. Cependant, ils ne semblent pas, pour l'instant, apporter toutes les solutions souhaitées au difficile problème de la propagation en milieu urbain mais ils doivent être considérés comme un pas supplémentaire effectué dans cette direction.

Un premier groupe de communications a été consacré à la caractérisation des canaux très diversifiés travaillant dans tout le spectre radioélectrique. Dans la partie haute du spectre utilisé, 35 à 90GHz, des efforts ont été effectués ces dernières années et le développement de la technologie permet actuellement de disposer des composants dont les prix d'acquisition autorisent davantage d'expérimentations.

Les résultats présentés révèlent que ces canaux ont des caractéristiques particulières qui nécessiteront des techniques nouvelles pour exploiter pleinement les possibilités qui se trouvent ainsi ouvertes.

Un second point particulièrement important a été souligné par l'apport d'études de propagation dans les zones polaires, aurorales et équatoriales. Ces régions sont particulièrement hostiles mais militairement capitales. L'intérêt de leur étude est double, d'une part géophysique, d'autre part technique. Les présentations effectuées ont clairement révélé ces deux aspects. Si les communications présentées dans cette seconde partie ont révélé des progrès indéniables, la quasi-totalité des orateurs a cependant mentionné qu'ils devaient poursuivre leurs études pour approfondir leurs connaissances révélant ainsi les difficultés que pose l'hostilité des canaux de télécommunication dans ces régions géographiques.

La troisième section a été consacrée au traitement de signal. Cette session est souvent celle où apparaissent des théories nouvelles. Ce fut, selon la tradition, le cas. On peut constater que la complexité des calculs ne paraît plus être un obstacle pour le traitement de signal et de l'information. On a donc proposé des formes d'ondes, des codages et des traitements de signaux souvent adaptatifs, dans lesquels on a pu juger que l'imagination était un dénominateur commun à tous les orateurs.

La section 4 portait sur la simulation des canaux. Ce thème apparaît très souvent dans les symposia qui traitent de la propagation des ondes électromagnétiques. Les communications présentées se sont inscrites dans le prolongement des études antérieures en introduisant des perfectionnements importants visant à une reproduction plus fidèle des canaux réels. Il n'en demeure pas moins vrai que le choix du modèle inséré demeure encore le problème majeur.

Les systèmes de télécommunication étaient abordés dans les sections 5 et 6. De nombreux systèmes ont été présentés et les systèmes de communication par traînées météoritiques ont soulevé, comme il est usuel, des discussions passionnées. Il s'agit d'un système très ancien qui, cependant, offre toujours un intérêt majeur pour les liaisons stratégiques demeurant utilisable dans des conditions extrêmes. D'une façon générale, on a pu noter que les études théoriques ont des retombées importantes vers la conception des systèmes et les procédures d'utilisation. Cette constatation est réconfortante pour les chercheurs qui se situent en amont des ces réalisations. Il est apparu que les systèmes civils étaient utilisables ou transposables aux systèmes militaires. Ce point est sûrement d'un intérêt croissant pour la communauté de l'Organisation du Traité de l'Atlantique Nord et constitue probablement un point sur lequel il y aura lieu, à l'avenir, de se préoccuper.

Finalement, on peut considérer que ce symposium a apporté des éclairages nouveaux sur l'important problème des télécommunications numériques entre des points situés au sol aussi bien que sur les liaisons avec les systèmes embarqués et les systèmes satellitaires. Les progrès réalisés ces dernières années sont probablement l'une des raisons des intérêts qu'a suscité ce symposium et on doit constater qu'à l'avenir les études présentées, les systèmes proposés, verront de nouvelles études et de nouveaux systèmes leur succéder avec des performances accrues.

'FLOW OF INFORMATION IN FUTURE AIR WAR'

by
Colonel D. HELIOTIS,
Commander of HAF Air Tactics Center
Adravida AFB,
Greece.

1. ACKNOWLEDGEMENTS

It is a great pleasure for me to be invited to speak at this Symposium, and I personally thank the organizers for this opportunity.

2. AIR TACTICS CENTER

Before getting into the subject of this tutorial, I am going to say a few things about the Hellenic Air Tactics Center (AIRTAC), where I am honoured to be the Commander; namely about its history, organization and purpose.

The AIRTAC was founded in July 1983 as an independent unit in Andravida AFB and it was the result of the continuously changing operational environment and the subsequent requirements for advanced training in air tactics and electronic warfare.

Until then, the training of personnel was carried out by schools in different locations all over Greece, under the command of Hellenic Tactical Airforce (HTAF) or other branches of the armed forces.

In 1987 a three-branch school, specialised in joint electronic warfare, was created in AIRTAC in order to apply the new developments in this sector of activity. This shows the dynamic evolution of the Center and the uniqueness of its mission.

The Center covers a wide spectrum of tasks (activities) which are referred to, as missions and are outlined below:

a. It provides advanced academic and flight training in all kinds of missions and air tactics, for experienced pilots, according to our doctrine.

b. Under the supervision of HTAF, the Center carries out operational tests and evaluation of weapons and weapon systems.

c. It observes both the technological evolution and the advances in equipment and it elaborates studies concerning the increase of HAF combat capability.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

d. Finally, the Center is responsible for transfering specific operational knowledge in the fields of air-support and co-ordination, to the other two branches of the Hellenic Armed Forces and to the intelligence community as well.

AIRTAC offers several training courses in yearly basis, which are provided by its separate Schools; namely, the Weapon and Tactics School, the HAF Electronic Warfare School, the Air To Ground Joint Operations School, and the Joint Electronic Warfare School.

Having outlined the purpose and the function of the AIRTAC, where I am honoured to be the Commander, it is now time to pass into the main part of this tutorial which is the 'Flow of Information in Future Air War'.

3. INTRODUCTION

However, before proceeding further, it would be useful to say a few words about the reasons that inspired me to select this subject.

Information technology has advanced rapidly during the last two or three decades not only in the aspects of transfer speed and capacity, but also in versatility and robustness. At the same time the cost of the electronic material has been reduced to a tiny fraction of the cost of the first machines which were built in `60s. It has been said [Sedgewick] that '*if this tremendous advance had been achieved in commercial passenger airplanes, an airplane should carry 2.000 passengers round the world in one hour, consuming only 1 lt. of petrol*'.

Despite this obviously advantageous technological infrastructure, which we, today, have the privilege to use, it is regretfully true, that the amount of messages to and from various departments of the modern armed forces, have multiplied by such a factor, that they threaten to fill up the capacity of the existing communication systems. When this happens, new, more powerful communication systems will be developed and used in accordance with the famous Parkinson's Law: '*The use of communication will expand to fill the capacity of the system*'. A more serious concern is the number of people required to execute or take some action on them. The ratio of the staff officers to the soldiers is, already, big and increasing and it is not an exaggeration to say that in some years, there will be less soldiers than staff officers.

I will start this introducrtory tutorial with the, always necessary, Historic Backround, and then, the Description of a War Scenario will follow. After that, I will give a general overview of the requirements and the capabilities of all the systems involved in the modern Air Battle communications; namely the Intelligent Information Systems, the Information Fusion Systems, the Information Processing Systems, and the Distribution Information Systems. Finally, I will close my speech with some Future Trends for the Electronic Transfer of Information.

4. HISTORIC BACKGROUND

The flow of information for military purposes is traced back in the ancient years. A characteristic example can be extracted from the Marathon Battle, which took place in 490 B.C. and where the Greeks defeated the Persians despite their disadvantage in the number of troops. After the battle, a messager, known until today as the Marathon Runner, ran continuously for 40 km to convey the victory word '*NENIKIKAMEN*' to the Athenians and then dropped dead. More recently Napoleon had his orders transmitted at the speed of about 10 Km/h at Jean - Auerstadt. With his corps separated by an average distance of 60 km, he could order and manoeuvre his individual corps about three times a day.

Since then, great strides have been made. Even in the unsophisticated days of the First World War, a single field army required daily an amazingly high number of telegrams, telephone calls and dispatches carried by courier.

Today the rapid advances in communication have changed completely the role of information in the modern battlefield. It can be said, that the air battle, where the actions are fast and the environment is changing second after second, has most benefitted from these advances. The increasing need for fast decisions and reactions of the pilots, combined with the need for real time communications with the command and control centers, have been satisfied by the capabilities of the modern communication equipment.

Computer networks permit the communication between headquarters separated by long distances which, now, can share information in real time.

Satellite communications opened new horizons, but also posed problems since a single transmission covers an enormous area and the national boundaries are no barriers anymore. Therefore, the issue of security must be taken more seriously into account.

Optical fibre communications present probably the most challenging aspects of modern systems. Their low cost, enormous bandwidth, very small attenuation, high flexibility and, especially, very good security against external interference make them ideally suited for military applications.

All these, and other technological advances, have given new meaning to the term 'flow of information' in the modern air war.

However all this vast amount of information introduces two big problems.

The first one is the correct transfer of the data and its security against external factors. In order to solve this problem, different kinds of codes are used, either for security, or error detection and correction.

The other problem is focused in the amount of information available today. It is obvious that only a small percentage of it is really useful. In one recent large scale exercise, 10,000 messages were lost: There was no noticeable effect on the exercise. Therefore, there is an increasing need for filtration, prioritisation and training in the process of data handling.

5. DESCRIPTION OF A WAR SCENARIO

A typical air defence scenario which shows the importance of rapid and reliable flow of information between the command centers and the fighter pilots is shown in Figure 1.

The scenario includes three main types of enemy raid: attack on airbone early warning, enemy fighter sweep, and an escorted deep strike raid. In addition to the primary airbone targets and threats, the tactical aid must also deal with numerous other enemy and friendly aircraft in the scene, as well as threats from enemy ground forces.

In a situation like this, the importance of an appropriate man machine interface, with particular emphasis on display formats and crew interaction, is self-evident. Also, for the tactical aid to be of maximum benefit, it is necessary to demonstrate and assess it in a realistic environment.

The first function of a system, with the purpose to assist the crew in understanding the environment into which they will 'play' their role, should be the Situation Assessment. Figure 2 shows some of the factors which are related with the term 'Situation Assessment'. The main purpose of it is to reduce the whole of the outside scene, referred to as the alpha scene, to a smaller selected number of the most important targets and threats, known as the beta scene. This process includes features like deletion of friendly tracks from the threat list, determination of which enemy aircraft will come within engagement range first, and addressing of target behaviour.

The next function as it is shown in Figure 3 should be the Attack Planning. Here, a range of different attack options are computed against each aircraft in the beta scene. These include various tactics in terms of aircraft approach paths and missile launch points.

On each attack option, full missile firing brackets, from maximum to minimum range, should be computed against primary and secondary targets. These include representative performance of the missile in each of its critical phases; an example to illustrate those for a mid-course guided active homing missile is shown in Figure 4.

The third major system function should be the Enemy Counter - Attack Assessment where the risks associated with each of the attack options is assessed. This is done by examining the attack paths and missile launch opportunities of the enemy threat aircraft (Figure 5).

By now, sufficient data should have been derived to allow the next process to Defence Planing. As indicated in Figure 6, this process includes defensive measures to increase own survival probability

The final system function should be the Option Analysis and Ranking, which decides the best option to go for, and ranks the alternatives in a preferred order. This is done by analysing all the data acquired and calculated from the previous stages in order to determine which option maximises a special tactical value function.

6. INTELLIGENT INFORMATION SYSTEMS

An intelligent information collection system should be able to manage with the co-ordinated employment of various types of sensors as well as with other information sources and to deal in real time with the collected data in order to provide timely the users at each level with secure, detailed and reliable information.

The information should cover all aspects and factors concerning the development of the battle i.e. from the enemy forces to the atmospheric conditions.

Furthermore, one of the most important operational requirements is that the correlation, processing and distribution of data and information should concern all the interested centers and command posts wherever they operate.

The above are, of course, applicable not only in the air war, but also in the battlefield area in its three components in total (land, airspace and sea). However, the importance of the operational role of an information collection system in the Airforce stands on a completely different basis. That is because the air battle is different from the other two Armed Force branches in two important aspects. First, is that the flexibility of the air power, which derives from the range and speed of aircraft, means that the air battle reaches places far beyond those covered by others. Secondly, the air power differs in the intensity of the battle. Whereas it can take days to manoeuvre land corps, the air power cannot only generate on line very quickly, but it can also operate at a rate of many sorties a day. These differences mean that an information collection system customised in the needs of the Air Force should provide the extra speed and analytical ability which are important in order to cope with the rapid decision making which characterises the air power.

An intelligent information collection system should provide a global picture of the battlefield in real-time, to allow the performance of the command and control functions. Therefore, it should be capable of responding fully and immediately to any variation of the threat of an enemy attack.

In order to achieve this, the system should have the feature of combining the knowledge of the enemy deployment and capabilities, the general environment and the evaluation of enemy intentions.

There are already enough information collection systems in existence (CATRIN, SIACCON etc.) and their technical aspects are not the subject of this introductory tutorial.

However, all of them have two common elements: The basic architecture structure and the pre-requisites in order for the system to be efficient.

6.1 Pre-requisites for efficiency

Elements of success of an efficient system are:

- Continuity in time and space of the coverage. Information on the enemy must be collected by mean of highly sensitive instruments utilising diversified technologies capable of operating in different environmental conditions.

- Capacity to conceal, or to contain at a low level the enemy's knowledge of its own collection facilities. That could induce the enemy to risk incomplete preparation in certain areas to achieve total surprise.

- Time response of the system sufficiently small to satisfy the need of the decision maker.

- Intelligence conception. The limit of intelligence to offer full proof advance warning, is dependant not only upon the limitation of information collection, but also upon the limitation of analysis.

6.2 System Architecture

To meet the previously mentioned strategic demands and to satisfy, in the meantime, requirements like operational flexibility, high survivability, reconfigurability and mobility, an efficient system should be marked by intelligence distribution and redundancies of functions and structures.

The basic architecture structure of a typical system presents the following main elements:

- Acquisition Sensors
- Processing Network

The Acquisition Sensors, depending on each independent configuration, comprise a range of elements such as different kinds of telemeters, radars, RPV's etc..

On the other hand, the Processing Network, in general terms, is a network of computing facilities associated to the Command Centers at various levels. The Acquisition Sensors are attached on this network and provide the computer infrastructure with the necessary data. It is evident that the products of the data processing should be easily accessible to each level of command

Special attention to the architecture of an information collection system should be devoted to the problem of interoperability between ground and air operations in order to achieve the best battle management.

6.3 Performance of an efficient information system

As a conclusion for the information systems it can be said that the basic demand which should be fulfilled by them is the prompt acquisition of suitable, complete and updated information. This enables the rapid decisional process on the battle field and optimises the employment of the available resources and the exploitation of the new weapons as well.

The above has to be guaranteed for all weather conditions and in extremely complex electromagnetic environment taking also into account all the possible countermeasures adopted by the enemy.

7. INFORMATION FUSION SYSTEMS

The fusion of the data collected from various sources as mentioned above, is a complex problem. The principal characteristics, apart from the amount of data, which make this procedure difficult are:

- information is incomplete.
- uncertainty and error are inherent in the information
- information is from numerous sources of many different characteristics, relating to various levels of the forces
- information is both temporally and spatially different
- information may be conflicting

Thus, there is a clear requirement for computer based data fusion aids to assist the intelligent officer in understanding what is actually happening, in recognising the enemy's intentions and the threats they pose, and in the effective deployment of own forces.

A first level of the analysis of the incoming data is its correlation with existing data to determine whether the new data is previously known to the system, or it is new data on its own right.

Secondly, it is necessary to cluster the data into divisions, regiments etc., in order to utilise all the information regarding a single object, such as a battle force.

Finally, this information can be used to identify the force structure and posture.

The process is continuous with new data utilised to update each of the levels, and more complex because there are feedback loops and interaction with the human intelligence officers.

These functions can be implemented using Artificial Intelligence techniques, as the Knowledge Based Systems, each of which refers to a different object. The different Knowledge Based Systems must be organised and communicate in an effective manner. Two representation paradigms can be utilised:

- The first, Blackboard System, is analogous to a number of human experts who observe and modify the contents of a blackboard according to their individual specialisation. The Blackboard System contains various Knowledge Sources, which modify and reason about the data, and they are not restricted to real world information but they also refer to inferred data, and scheduling and control information.

- The second, Message Passing System, treats Knowledge Sources as autonomous objects, which interact with each other only when it is needed.

Combining the two paradigms a third possibility arises, the Distributed Blackboards System. Here we have multiple Blackboard Systems, which interact with each other when needed, using a special Blackboard Knowledge Source, which is treated like any other Knowledge Source.

Underlying in the whole data fusion process, is the necessity of the human - computer interaction. Beyond the man - machine interface, interfacing with the inference system must be possible. The user must be able to modify both the final (inferred from the data) hypotheses, and the underlying hypotheses (the inference rules). This ability is critical to the acceptance and the performance of the data fusion process.

8. INFORMATION PROCESSING SYSTEMS

The operational effectiveness of air command and control systems is likely to be severely reduced in wartime by the problems inherent in information collection systems involving intermittent data of poor quality. In war time, attack aircraft flying offensive missions will face a wide variety of electromagnetic threats in a quickly changing environment.

Therefore, the incoming information requires to be processed and the corresponding systems should be designed to operate in a very dense and complex EM environment in order to perform automatic enemy radar detection, analysis, identification and very accurate localisation in real time.

The critical questions which need to be answered are how significantly will a wartime environment degrade the air picture, how much will this degradation reduce the operational effectiveness of the command and control system, and what can be done to present what limited information that may be available in order to be exploited in a more useful way.

The incorporation of one or more expert systems is a key element in this processing. An expert system "learns" while it operates and incorporates this "knowledge" in its future

decisions. By building up knowledge on the environment, tactics and the air situation, it will be possible to improve the air picture created by the incoming data.

The expert systems have already found many applications in the field of information processing. As an example we could mention the use of expert systems to improve multi-sensor tracking of air targets. Prior multi-sensor tracking methods suffered from intermittent tracks, track identity switches and false tracks. The use of expert systems made possible to improve the air picture by "tuning" the multi-sensor tracker and resolving track ambiguities

However, the technology is evolving rapidly and the expert systems or other technical innovations will continue to play an important role on the processing of the incoming data during wartime.

9. DISTRIBUTION INFORMATION SYSTEMS

A distribution information system should comprise personnel, equipment, facilities and communications to provide authorities at all levels with adequate data to plan, direct, co-ordinate and control conventional air operations.

The whole process of distribution of information consists of data processing equipment, support and application software and digital communications equipment.

Tactical air operations comprise both offensive and defensive operations, complemented by supporting air operations. Each operation has an impact on the other. Because of their inter-relationship, they cannot considered in isolation, but should always be viewed as a synergistic whole.

An efficient system should support the operations conducted in times of peace, tension and war.

In peace, the available information supports routine training operations, exercises and air policing.

Times of tension are characterised by a change in procedures and a gradual build up of the flow of information. In these circumstances, military activities are likely to be controlled from relatively high command levels which will require timely and more detailed information to a far greater degree than in peacetime. Furthermore, extended coverage and airspace surveillance will be required into areas of increased interest.

During war state, heavy demands will be made on the information systems. These systems will be attacked both physically and electronically by the enemy. Rapid changes in situation requiring more dynamic reaction, increased co-ordination at all levels, and a multiple increase in information flow at the cost of decreasing detail as the conflict intensifies, are the main priorities in wartime.

The command and control activities performed during wartime are grouped into six operational functions. These functions which should be supported by a distribution information system and for which the information exchange is crucial, are:

1. Force Management: This includes the planning and tasking of air missions and the weapon or mission preparation activities which occur at the wing or unit level to prepare the weapon system for launch.

2. Command and Control Resource Management: It includes those activities associated with the employment and use of command control resources and the provision of logistics to support planned air operations.

3. Airspace Management: The goals here are to ensure maximum freedom for own forces in the use of airspace and, at the same time, to reduce the risk of fratricide.

4. Surveillance: It provides the users with detailed Air, Land and Sea pictures.

5. Air Mission Control: It comprises those activities that control and support manned and unmanned tactical air missions to achieve their objectives. Its goal is to enable all air missions to be ordered, controlled, supported, co-ordinated, monitored and directed without delay.

6. Air Traffic Control: It consists of different services with the aim to prevent collisions between aircraft and other aircraft or obstructions, to provide advice and info on the safe and efficient conduct of flights and to notify agencies about aircraft in need of Search and Resque, assisting them as required.

10. FUTURE TRENDS FOR ELECTRONIC TRANSFER OF INFORMATION

The emergence of new data storage and transfer media is having a growing effect on the electronic information industry. Today, the mass storage possibilities of new technologies, combined with the rapid advances in personal computers and telecommunication networks, have a great impact on the flow of information for military reasons.

Telecommunications is the life support system for the electronic transfer of information. While telecommunications did not always grow at such a rapid rate, in near future, the, long ago anticipated, integration of services, with the roles of telephone companies, cable operators and power suppliers being overlapped, is going to change completely the field.

Further noticeable advances include the database supply market with continuously increasing number of databases, different kinds of technology base, different kind of content, origin and access mode.

The use of optical storage media for mass information storage and retrieval, combined with the recent innovation of the re-writable CD, has given rise to a spate of recent market developments.

Electronic mail, is today available on, virtually, every online or timesharing system that reaches a searcher. Its importance for the elimination of the paper syndrome in every public or military service, is self-proven. However, especially for the latter ones, a matter of security arises since the usage of cryptographic codes is still under debate. It is obvious that the Armed Forces should be excluded from such a prohibition.

Generally speaking, there is an increasing trend in military communications for greater compatibility with civilian infrastructure which reduces development costs and increases equipment compatibility between allies.

As an example on this, we can mention the mobile cellular digital radio telephone system which, apart from the security, has the advantages of reliable and robust mobile communications. The infrastructure is relatively simple with the Base Stations connected to the Mobile Telecommunications Switching Offices, which, in their turn, are connected to the Public Switched Telephone Network. Moreover, the cellurar concept permits the partition of the limited frequency spectrum dedicated to the mobile communications, into frequency sub-groups which, then, can be re-used in symmetric cells. This re-usability of the few available frequencies permits the coverage of an enormous number of users.

This flexibility of cellular telephone communications systems makes them ideal for military applications and they have already been used in operations Desert Shield and Just Cause (Panama), as well as, in Former Yugoslavia.

11. CONCLUSION

In conclusion, we can say that the rapid advances in communication technology have brought major changes in all the aspects of the modern battlefield.

If today's achievements are good examples of what can happen in the near future, then the next years will see major developments in all aspects of information systems, which, in their turn, will have a major impact on the planning and execution of Future Air War.

CLOSING

Before I finish, I would like to express, for one more time, my thanks to the Sensor and Propagation Panel of the AGARD which gave me the chance to open this Symposium.

Closing this speech, I am at your disposal for any questions you might have.

REFERENCES

[1] Association Momentanee ACCSCO-AMS (1992). *Draft Overall System Specification for the ACCS System Specification Contract - Volume I.* NACMA, Brussels, Belgium.

[2] Beauvais, T. (1989). *Means to Know the External Situation: A Solution to Gather and to Process Information Regarding EM Threat.* AGARD Conference Proceedings No.440: Near Real-Time Aspects in Air Battle Management.

[3] Bravin, J. and Alston, A.J. (1989). *Means of Knowledge Exploitation: The Use of Expert Systems to Improve Multi-Sensor Tracking.* AGARD Conference Proceedings No.440: Near Real-Time Aspects in Air Battle Management.

[4] Byrne, C.D., Miles, J.A.H. and Lakin, W.L. (1989). *Means of Knowledge Exploitation: Data Fusion and Situation Assessment for Naval Battle Management.* AGARD Conference Proceedings No.440: Near Real-Time Aspects in Air Battle Management.

[5] Edwards, P.H. (1989). *Means of Knowledge Exploitation: Data Fusion in the Land Battle Environment.* AGARD Conference Proceedings No.440: Near Real-Time Aspects in Air Battle Management.

[6] Fiorina, G. and Chiusano, P. (1989). *Means of Knowledge Exploitation: A Concept of Intelligent Information Collection and Diffusion in Battle Management.* AGARD Conference Proceedings No.440: Near Real-Time Aspects in Air Battle Management.

[7] Mastroddi, F. (1990). *Technologies for Electronic Transfer of Information: The Present State and Trends in Europe.* AGARD Conference Proceedings No.446: Electronic Transfer of Information and its Impact on Aerospace and Defence Research and Development.

AIR DEFENCE SCENARIO

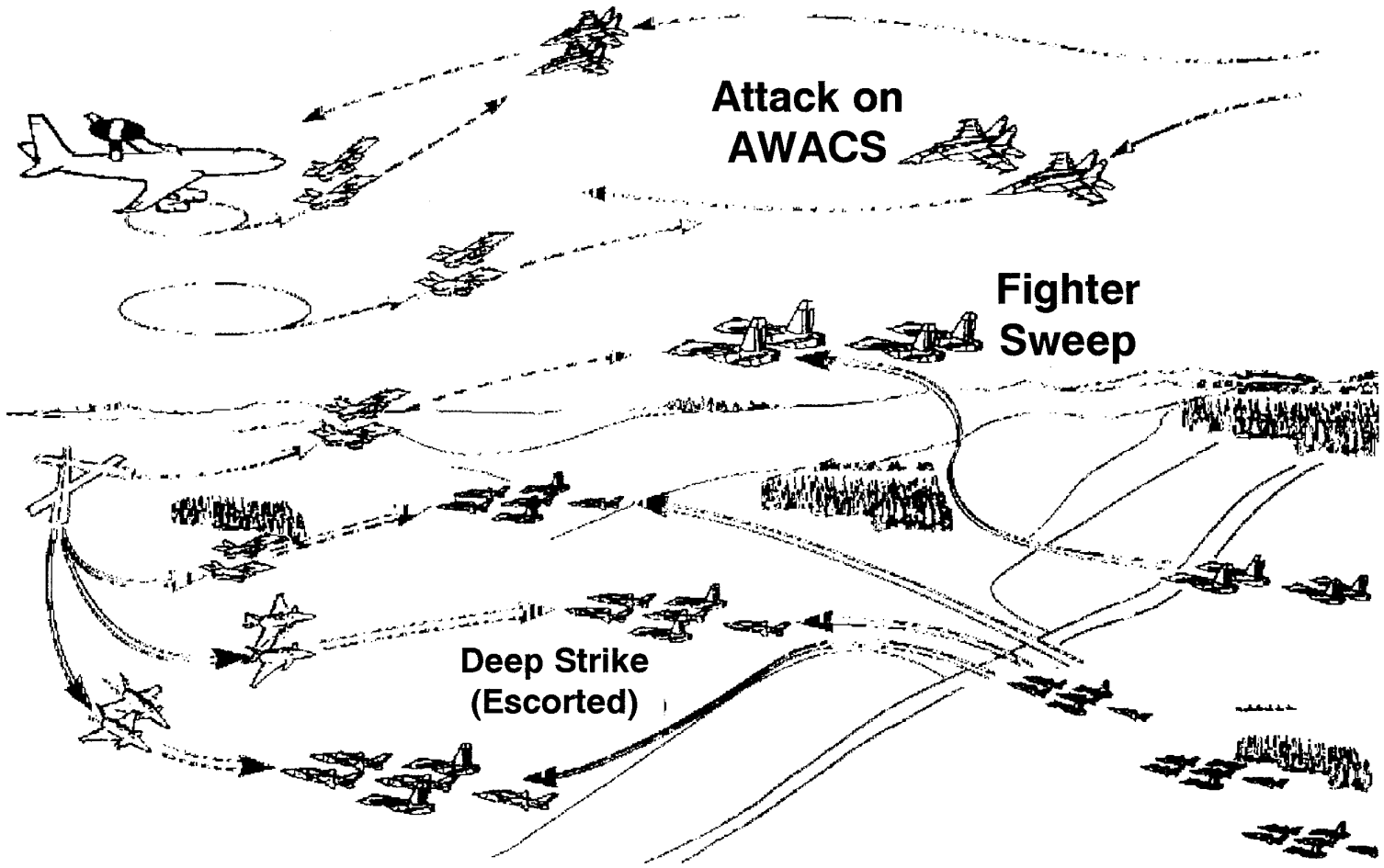


Figure 1

SITUATION ASSESSMENT

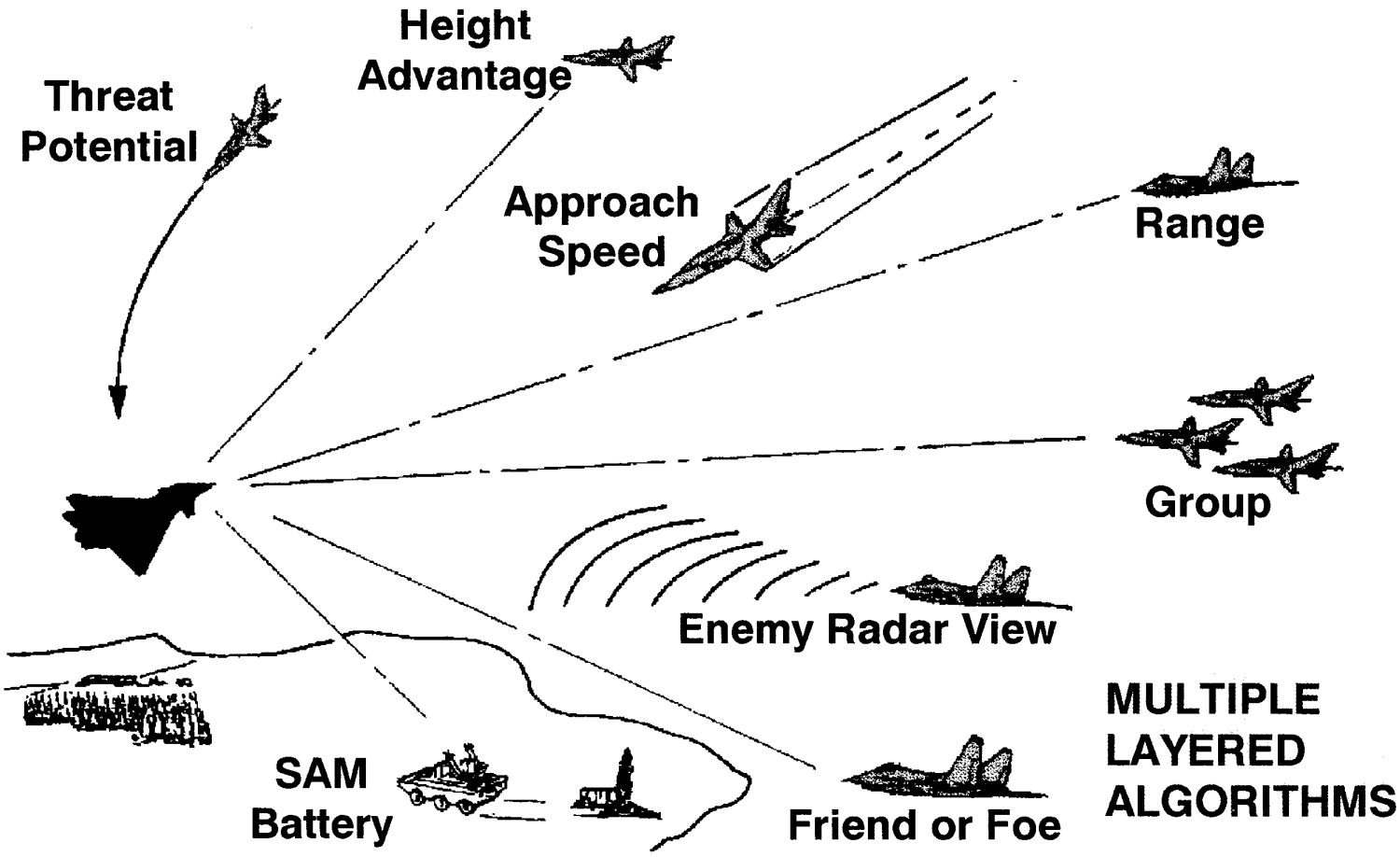


Figure 2

ATTACK PLANNING

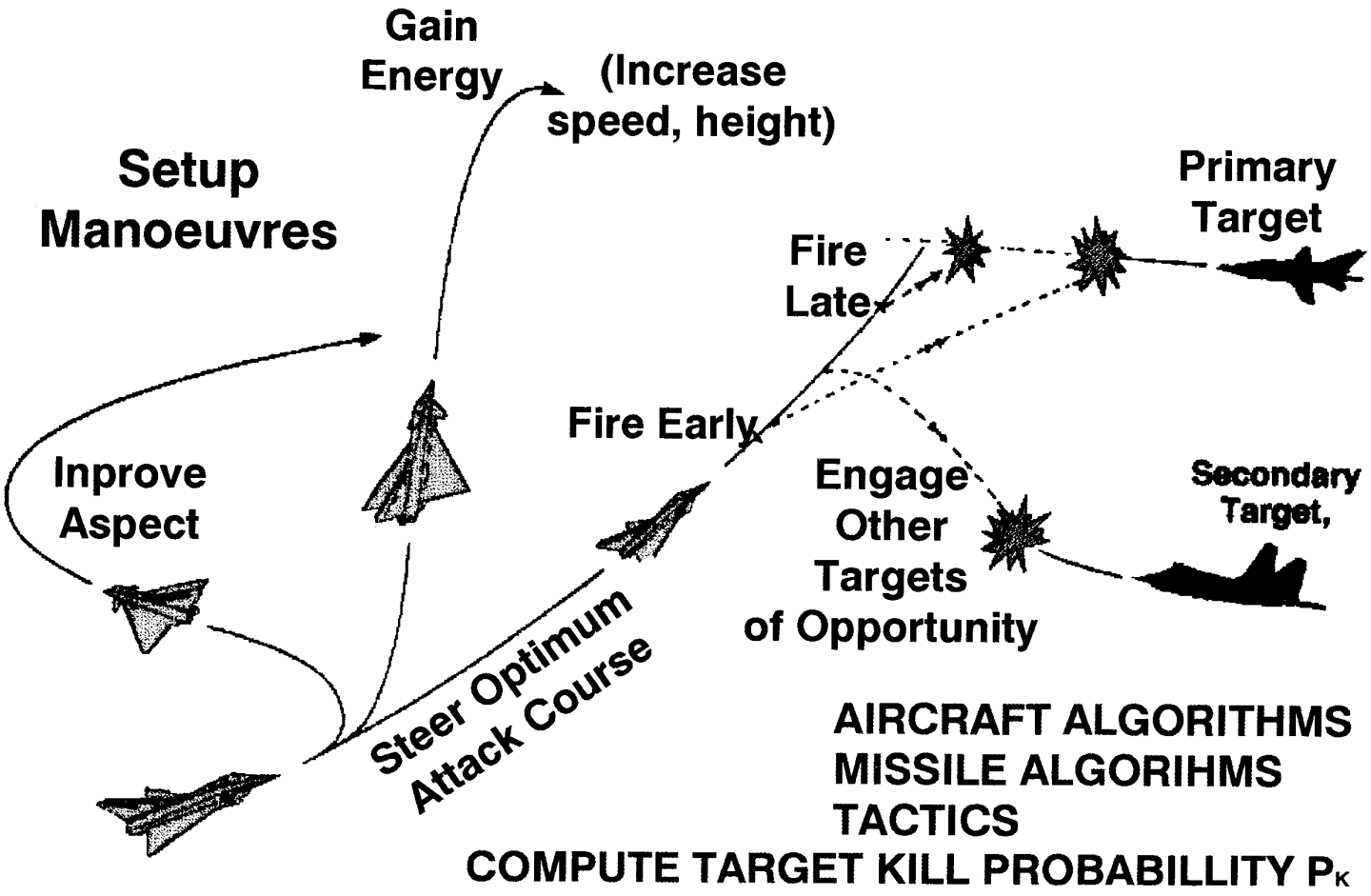


Figure 3

MISSILE EFFECTIVENESS (P_K)

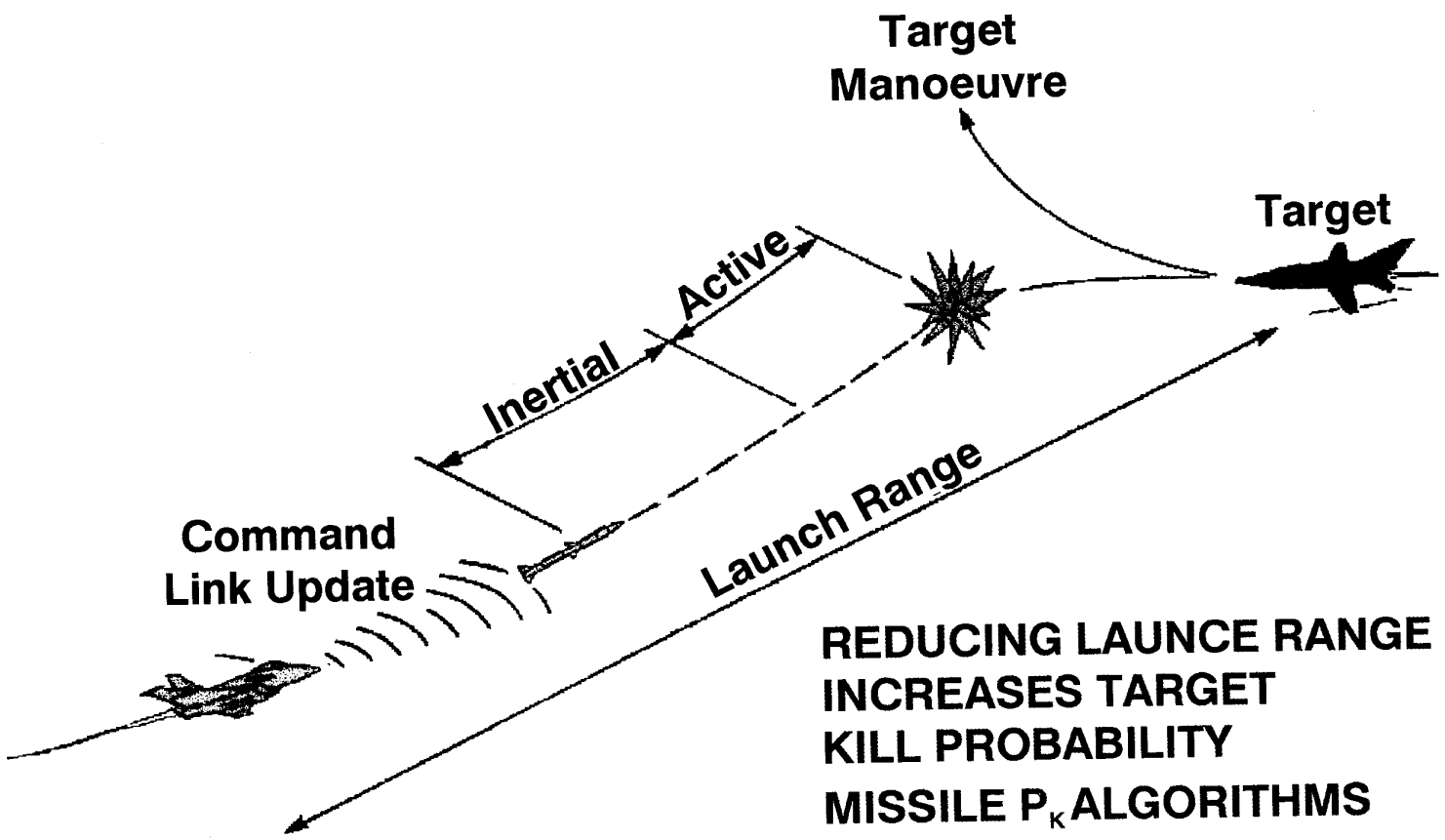
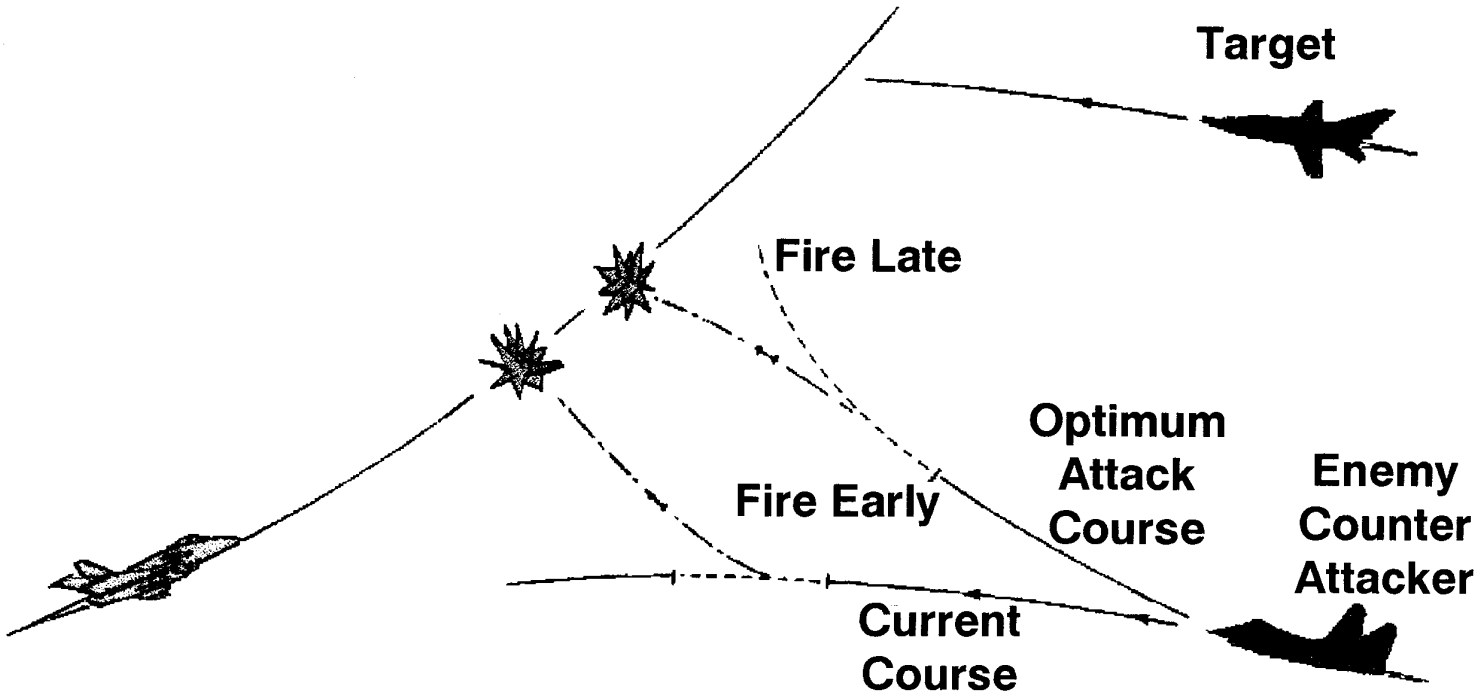


Figure 4

ENEMY COUNTER - ATTACK ASSESSMENT



COMPUTE OWN SURVIVAL PROBABILITY P_s

Figure 5

DEFENCE PLANNING

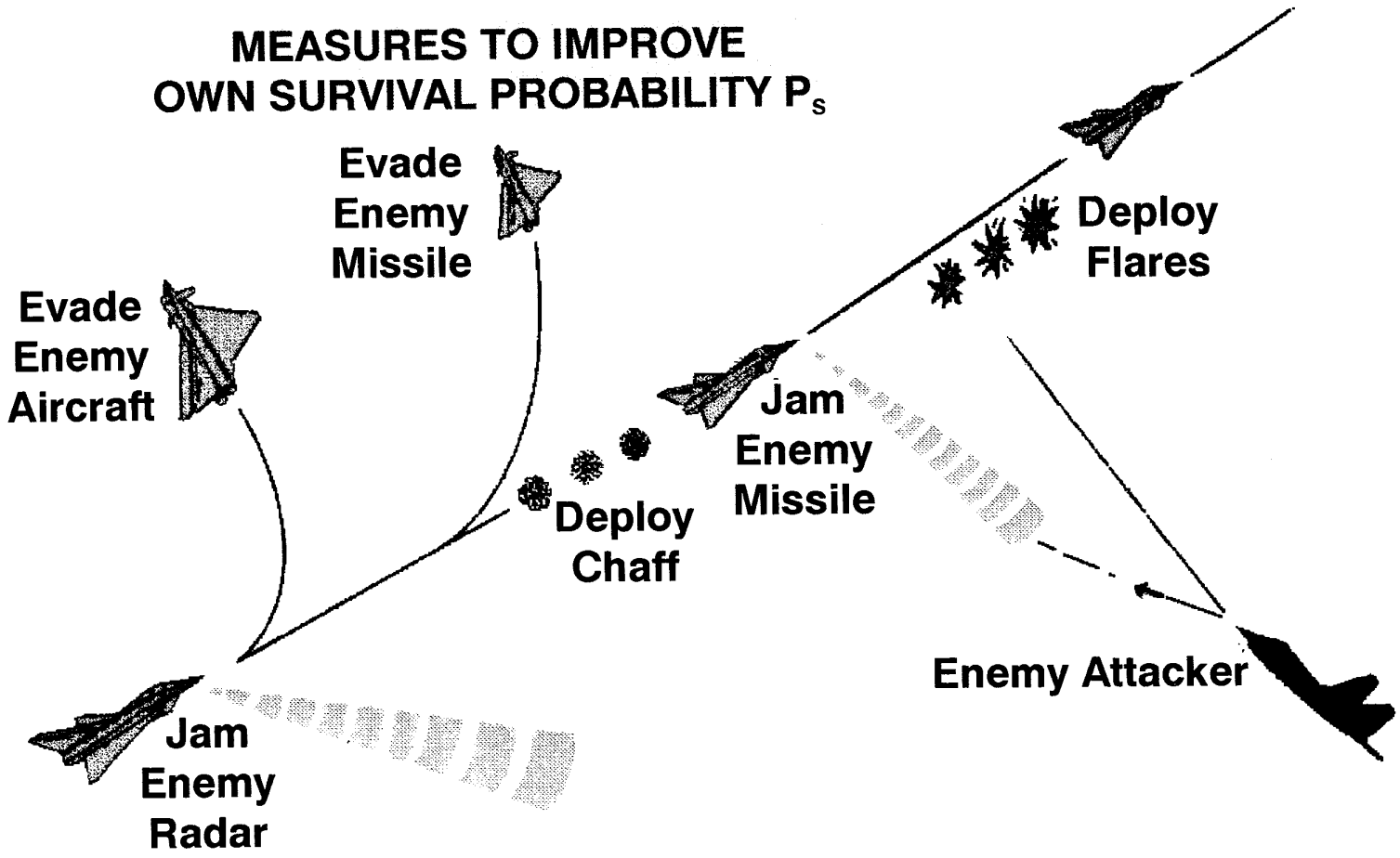


Figure 6

AN ANALYTIC RAY TRACING MODEL FOR HF IONOSPHERIC PROPAGATION

Robert J Norman
Ian G Platt
Paul S Cannon

Radio Science and Propagation Group
Defence Research Agency
Malvern, Worcs.
WR14 3PS
United Kingdom

Introduction

Ray tracing is a powerful tool and is especially useful in applications requiring a detailed knowledge of radio wave propagation through the ionosphere. There are a number of modern radio wave applications which make use of ionospheric propagation such as over-the-horizon radar systems, single station location and HF direction finding systems. These radio wave systems depend critically on realistic ionospheric modelling and accurate ray tracing through these ionospheric models.

Accurate ray tracing is normally carried out using numerical techniques. These are very accurate but the required computational time is high. This is exacerbated in many applications, when it is necessary to trace a vast number of individual rays. Thus, when dealing with near real-time applications, which is increasingly the case, it is far more desirable, if not essential, to make use of much faster analytic ray tracing techniques.

As its name suggests analytic ray tracing uses explicit equations to define the ionosphere and to determine ray parameters such as ground range, reflection height, phase path, group path and divergent power loss. Thus, analytic ray tracing is considerably less time consuming than numerical ray tracing. Analytic ray tracing has some disadvantages for example there are no explicit equations which include the effects on the ray path due to the earth's magnetic field (however a first order correction has been developed by *Dyson* and *Bennett* [1991]) and there is the major difficulty of including horizontal gradients. Thus, analytic ray tracing has up until now been limited to simple and unrealistic ionospheric models.

Another important factor influencing the ability to effectively describe HF propagation is an accurate and up-to-date representation of the ionosphere. To achieve this level of ionospheric specification requires an ionospheric map that can be updated from real-time sounding networks. In practice such networks may only consist of a small number of sounders, so that a technique specifically designed for data sparse regions needs to be used. Such a technique has been developed in the field of meteorology by *Cressman* [1959] and successfully applied to the ionospheric case by *Rush* and *Edwards* [1976] and *Platt* and *Cannon* [1994] producing a grid point ionospheric model.

In this paper a method to approximate horizontal gradients when using analytic ray tracing will be presented. This method involves automatically tilting the modelled ionospheric profile by displacing the earth's centre. This method is applied to a grid point ionospheric model which consists of a known ionospheric model coupled with real-time ionospheric measurements at a number of locations. Thus, the grid point ionospheric model is a real-time ionospheric model and together with the tilting method, is useful in a number of applications. Analytic ray tracing results incorporating the tilting method and the grid point ionospheric model are presented in this paper. In order to demonstrate the effectiveness of this analytic approach the results will also be compared with those obtained using the numerical ray tracing package HIRT (Homing-In Ray Tracing) *Norman* et al., [1994] over the same region.

Analytic ray tracing model

A fully automatic analytic ray tracing model has been developed which uses a grid point ionospheric model. This grid point ionospheric model consists of a known ionospheric model, in our case FAIM (Fully Analytic Ionospheric Model) [*Anderson* et al., 1989], which can be coupled with real-time ionospheric measurements from either one, or a network of ionospheric sounders. Weighting functions and a linear correlation function are applied to the predicted and measured values producing a smoothed data set at specific grid point locations over a defined coverage set [*Platt* and *Cannon*, 1994].

Thus, this grid point model consists of a number of grid points in the latitudinal and longitudinal directions and a number of layers of grid points in the vertical direction, starting at the base of the ionosphere right up to the height of the peak electron density of the ionosphere at that grid point location.

Once the grid point procedure constructs a reasonable representation of the ionosphere, we are in a position to ray trace through the newly constructed ionosphere. In order to undertake analytic ray tracing the approximate midpoint, of the propagation path, of each ray traced must be found. Initially the midpoint is estimated by a straight line, with the same elevation as the ray to be traced, from a given transmitter to a height of 300 km. This height is used because it is a good first order representation of the height of reflection.

The electron density profile at this particular midpoint location is then fitted with a multi-quasi-parabolic segment, QPS, model [*Chen* et al., 1990]. The major asset of the QPS model is that it can be used to obtain explicit analytic equations, for a number of ray parameters, for a ray travelling from a transmitter up into the QPS ionosphere and back to the earth. This method starts at the highest peak of the layer (i.e., foF2) and fits a quasi-parabolic segment to the entire profile. If the attempt fails within a specified r.m.s error, one point is

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

discarded from the base and the process is repeated. The discarding of points continues until a QPS has been fitted to part of the real profile. The procedure then proceeds to a point just inside the lower boundary of this QPS and another QPS is smoothly attached. This continues until the entire profile has been fitted.

Whilst our analytic ray tracing, combined with the grid point model, does not yet include the effects caused by the earth's magnetic field, it does include an approximation to simple horizontal gradients. This is achieved by tilting the ionosphere with respect to the earth.

In order to determine the horizontal gradients in the ionospheric map, analytic ray tracing is carried out using the midpoint QPS ionospheric profile. From the resulting ray, three points are chosen corresponding to the ray entry point (P1) into the ionosphere, the ray apogee (P2) and the ray exit point (P3). The height of reflection is then determined and the corresponding plasma frequencies at this reflection height at the three locations P1, P2 and P3 are determined. The lowest plasma frequency of these three locations is found and the height at which the other two locations have this value of plasma frequency are determined. For example, suppose at the three ray points the plasma frequencies at the reflection height are:

Location P1 (ray entry) Plasma Frequency = 9.0 MHz,
Ref. Height=300km
Location P2 (ray apogee) Plasma Frequency = 8.5 MHz,
Ref. Height=300 km
Location P3 (ray exit) Plasma Frequency = 8.0 MHz,
Ref. Height=300 km

Clearly the lowest plasma frequency is at location P3. By linear interpolation (using values of the electron density profile generated by the map) it is found that the heights at which locations P1 and P2 have the value of 8.0 MHz are 280 km and 290 km respectively. The 8.0 MHz curve passing through 280 km, 290 km and 300 km is then used to determine the tilt of the ionosphere over the propagation path. This is achieved using the height and ground range of each point from the transmitter and solving for the quadratic equation which passes through the three points. The angle of tilt is then determined from the difference, at the apogee of the ray path, between the normal of the parabolic surface and the radial projection from the earth's centre.

It should be noted that the QPS ionospheric profile at the ray apogee is now used to represent the ionosphere.

Once we have worked out the angle of tilt, η, we are in a position to trace the ray. This method is simple and the results that will be shown in the following section are encouraging. Figure 1 helps to explain the tilting procedure where the curve in Figure 1(a) represents the expected propagation path of a ray with an initial elevation, β_o, when ionospheric tilts are ignored. The arrow on the diagram shows the actual tilt of the ionosphere at the apogee of the trace. In order to allow for the tilt we keep that part of the ray, from the ground to the base of the ionosphere, which is a straight line. The next step is to rotate the earth and the base of the ionosphere, about the ray entry point to the ionosphere, so that they are parallel to the arrow which represents the angle of tilt, as shown in Figure 1(b). Ray tracing is then carried out as if no tilt were present with a new elevation of $\beta = \beta_o + \eta$.
The resultant ground range, GR, that we require is given by

GR=R1+R2

where R1, represents the ground range of the ray in Figure 1(a) from the transmitter to the point where the ray enters the ionosphere. R2, represents the ground range of the ray in Figure 1(b) from the point where the ray enters the ionosphere to the end of the trace where the ray hits the earth's surface.
It should be noted that Figures 1(a) and 1(b) are a little misleading in that the earth's surface and base of the ionosphere are represented by straight lines. The analytic ray tracing package does in fact take into account the curvature of the earth and the ionosphere.

This tilting method is simple and even appears a little crude. However, the results obtained from its use are encouraging. The tilting method works well over the region we are interested in because the horizontal gradients increase with increasing altitude up to the peak height of electron density. The horizontal gradients are significantly greater at altitudes of 300 km than at the base height (approximately 90 km) of the ionosphere. However, if ray tracing in the region of the equatorial anomaly or a known ionospheric trough, then a tilting method which tilts the ionosphere with respect to the earth's surface may produce better results.

Propagation Results

The example (Figure 2), shows a contour plot of the plasma frequencies at a height of 300 km. This contour plot was obtained from the grid point ionospheric model, where the FAIM model was used to determine the plasma frequency at each grid point location. The FAIM model had the following settings

F10.7 Value of Solar Flux = 170,
Month = March,
Local Time = 1700.

The plasma frequencies in Figure 2 range from 5.0 to 13.0 MHz. The steady rate of increase in electron density towards the south and the decrease towards the north makes this a good ionosphere to compare the analytic ray tracing with numerical ray tracing.

In the analytic case a 43 equally spaced, layered, grid system was used starting at an altitude of 90 km and finishing at a height of 300 km. The FAIM model was used to produce the relevant electron density at each grid point location. The grid points start at geographic co-ordinates 28.0°N, -12.0°E, with 26 steps of 2° in latitude and longitude. In both the numerical and analytic ray tracing cases, ray tracing was performed with elevations ranging from 1° to 90°, in elevation steps of 1°. All azimuth angles were covered with a 1° resolution. At each grid point location the profile was represented by a QPS model with around 15 segments.

Figures 3 and 4 show the resultant divergent power loss footprints for numerical and analytic ray tracing respectively, when the operating frequency of the rays traced is 14.0 MHz and when using the ionospheric model which produced Figure 2. The transmitter is placed at the centre of the diagram (55°N, 15°E). Figure 3 represents the footprint obtained when using numerical ray tracing through a continuous (not grid point) FAIM model and Figure 4 represents the footprint obtained from the tilt compensated analytic ray tracing through the ionospheric grid point model. The centre white section of the plots represents a no propagation zone, the boundary of which is the skip distance.

The propagation footprints given in Figures 3 and 4 are similar. The skip distance from the transmitter is approximately the same in both cases. The region of the -130

dB power loss is slightly larger in the numerical ray tracing case, however closer inspection of the actual power loss values indicates a difference of less than two dB. Notice that the skip distance to the south is closer to the transmitter, in both figures, which is consistent with our expectations.

The error in the analytic ray tracing model, in terms of resultant ground range, was calculated to be approximately 2.9%. This value was determined by comparing the resultant ground range when using the analytic ray tracing model with the ground range calculated using numerical ray tracing through the FAIM model, for the rays traced in Figures 3 and 4 in the north, south, east and west directions. When the tilting method was not used, the percentage error in ground range was calculated to be 4.0%. Thus using the tilting method increases the accuracy of the rays being traced.

The example (Figure 5), shows the plasma frequencies at a height of 300 km when a synthesized terminator is added. Towards the west of the plot the plasma frequency is similar to that shown in Figure 2 (i.e., using FAIM). To the east of the plot we have introduced a terminator. The terminator is constructed so that the plasma frequency increases linearly starting at the geomagnetic longitude 106.0^{o} and finishing with a 70.0% increase at 108.0^{o}. Any geomagnetic longitude greater than 108.0^{o}, has been given a 70.0% increase in plasma frequency. This example was chosen to illustrate the capabilities of both the mapping technique and the analytic ray tracing, by allowing for such large horizontal gradients. The transmitter is again operating from the centre of the plot.

Figures 6 and 7, show the respective analytic and numerical, divergent power loss footprints, when ray tracing through the ionospheric model which produced Figure 5. The operating frequency is 14.0 MHz. The skip regions and power loss values of the two footprints are in good agreement. Note the asymmetry of the footprints, especially in terms of the skip which has a greater range from the transmitter on the eastern side than on the western side which is consistent with the contour plot in Figure 5.

The example (Figure 8), shows a contour plot of the plasma frequencies at a height of 300 km when three artificial ionosonde data sets are placed at a longitude of 30.0^{o} east and latitudes of 31.2^{o}, 55.3^{o} and 75.2^{o} north. This example was chosen to show that ionospheric data can be added to the mapping technique. Figure 9, shows the resultant divergent power loss footprint using analytic ray tracing, when ray tracing through the ionospheric model which produced the profile in Figure 8. The operating frequency is 14.0 MHz. The Footprint appears to be consistent with the contour plot in Figure 8. At present we have not implemented a method to incorporate measured ionospheric data into the numerical ray tracing model. Thus we are unable to compare Figure 9 with a corresponding numerical ray propagation footprint.

Summary

The analytic ray tracing model described here is a near real-time ionospheric model which includes an approximation to the ionospheric profile which allows the inclusion of horizontal gradients.

The results in this paper are encouraging, especially when we compare the propagation results when using the analytic ray tracing model (i.e., Figure 4), with the numerical ray tracing (i.e., Figure 5), which traces through the synthesized ionospheric profile produced using the FAIM model. The percentage difference in the resultant ground range between the numerical ray tracing and the analytic ray tracing model described here is only 2.9%. Even less difference would be observed using smaller steps in grid size and using more grid layers. This will be investigated in a later paper.

The analytic ray tracing model produces a fast approximation to the entire propagation footprint. In fact the analytic ray tracing technique described here together with the tilting method is more than 10 times faster than the numerical ray tracing package. Higher resolution regions of the footprint will still need to rely on, the slower, numerical ray tracing techniques. Thus, it is more than likely, in any particular application, that a propagation footprint may require a combination of both the analytic ray tracing model and numerical ray tracing techniques.

References

Anderson, D. N., Forbes, J. M. and Codrescu, M., (1989) 'A Fully Analytic Ionospheric Model (FAIM),' J. Geophys. Res., **94**, 1520.

Bennett, J. A., Chen, J. and Dyson, P. L., (1991) 'Analytic Ray Tracing for the Study of HF Magneto-ionic Radio Propagation in the Ionosphere,' *ACES* (Applied Computational Electromagnetic Society) J. **6**, 192.

Chen, J., Bennett, J. A. and Dyson, P. L., (1990) 'Automatic Fitting of Quasi-Parabolic Segments to Ionospheric Profiles with Application to Ground Range Estimation for Single-Station Location,' J. of Atmos. and Terr. Phys. **52**, 277.

Cressman, G. P., (1954) 'An operational Objective Analysis System,' Mon. Weather Rev., **87**, 367.

Norman, R. J., Bennett, J. A., Dyson, P. L. and Nguyen. L., (1994) 'HIRT: Homing-In Ray Tracing Program,' Research Report, School of Physics, La Trobe University.

Platt, I. G. and Cannon, P. S., (1994) 'A Propagation Model for the Mid and High Latitude Ionosphere Over Europe,' IEE HF Conference, York, UK, July.

Rush, C. M. and Edwards, W . R., (1976) 'An Automatic Mapping Technique for Representing the Hourly Behaviour of the Ionosphere,' Radio Sci., **11**, 931.

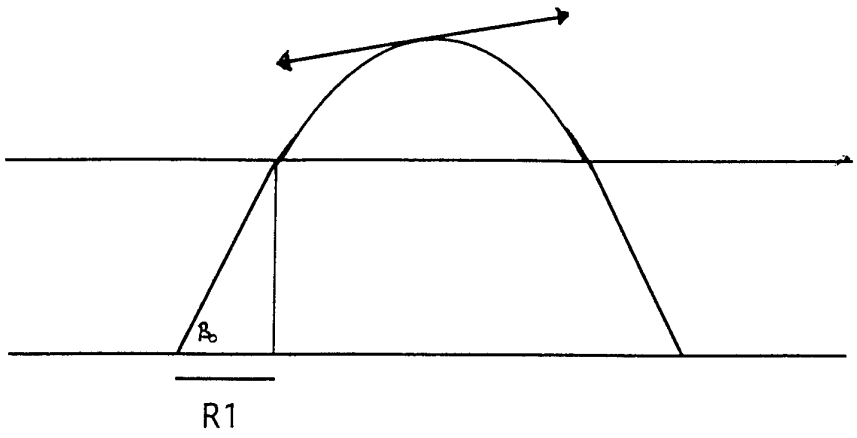


Figure 1(a) A propagated ray with elevation, β_o, ignoring the ionospheric tilt represented by the arrow.

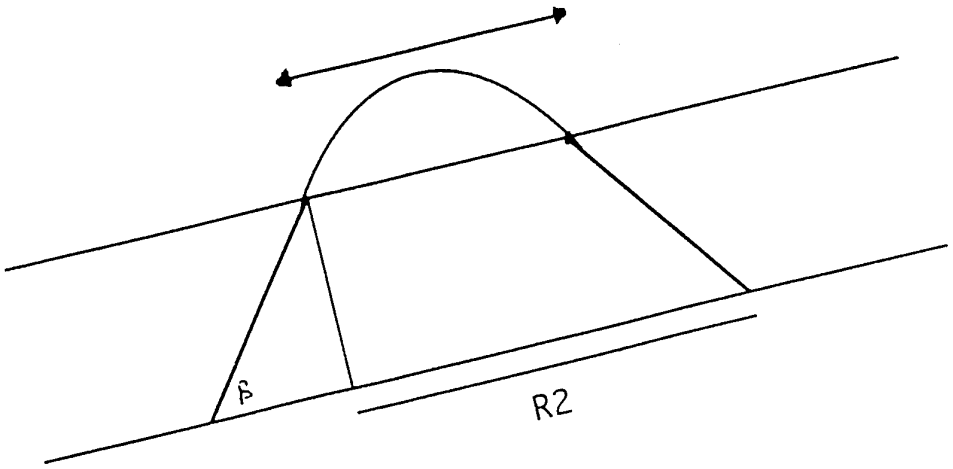


Figure 1(b) A propagated ray with elevation, β. The earth and the base of the ionosphere have been rotated, about the ray entry point to the ionosphere, by the angle of tilt.

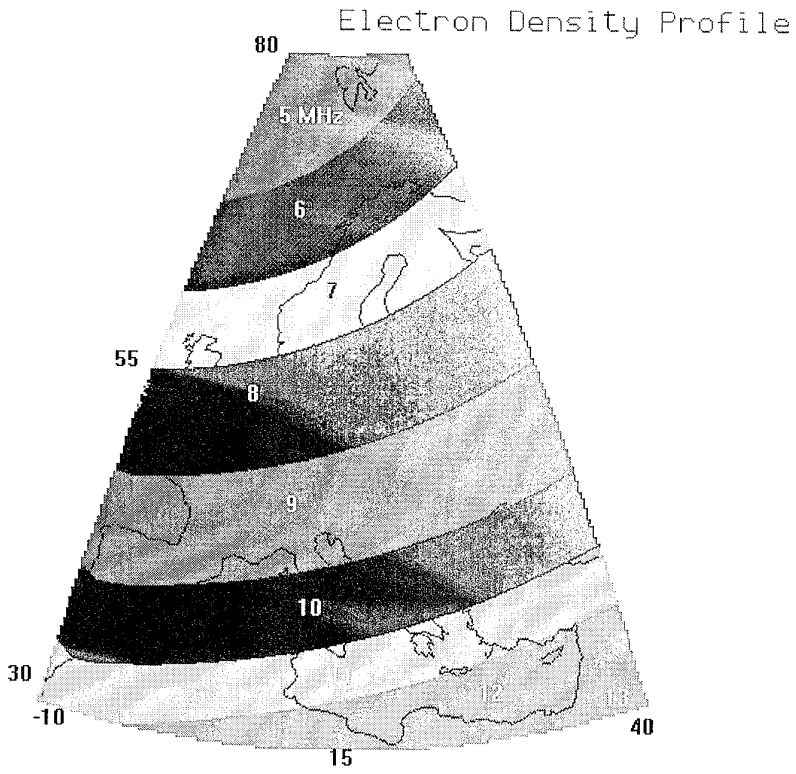


Figure 2 A contour plot showing the plasma frequency at a height of 300 km.

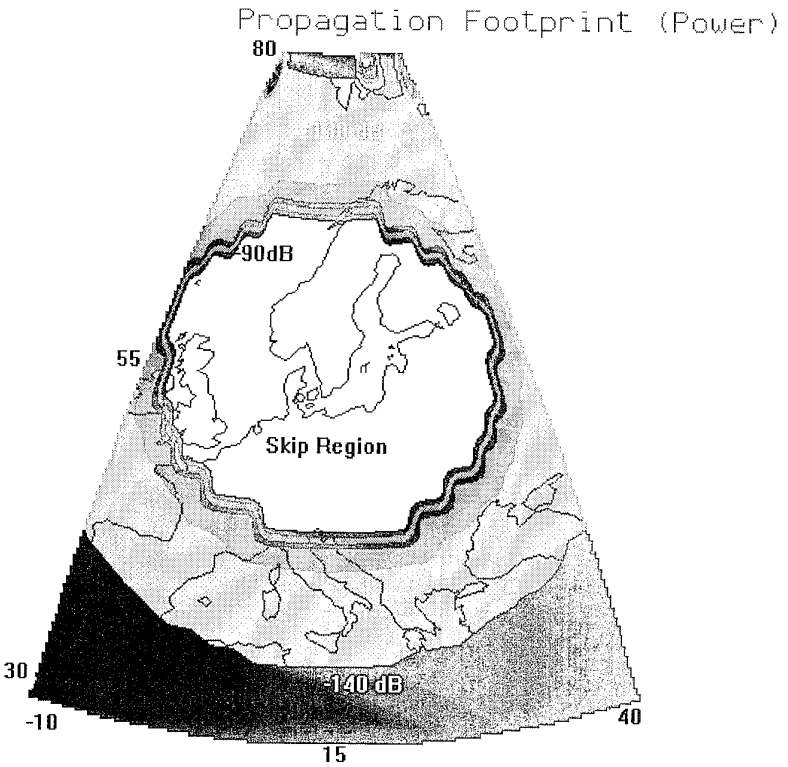


Figure 3 Numerical ray propagation footprint showing the divergent power loss.

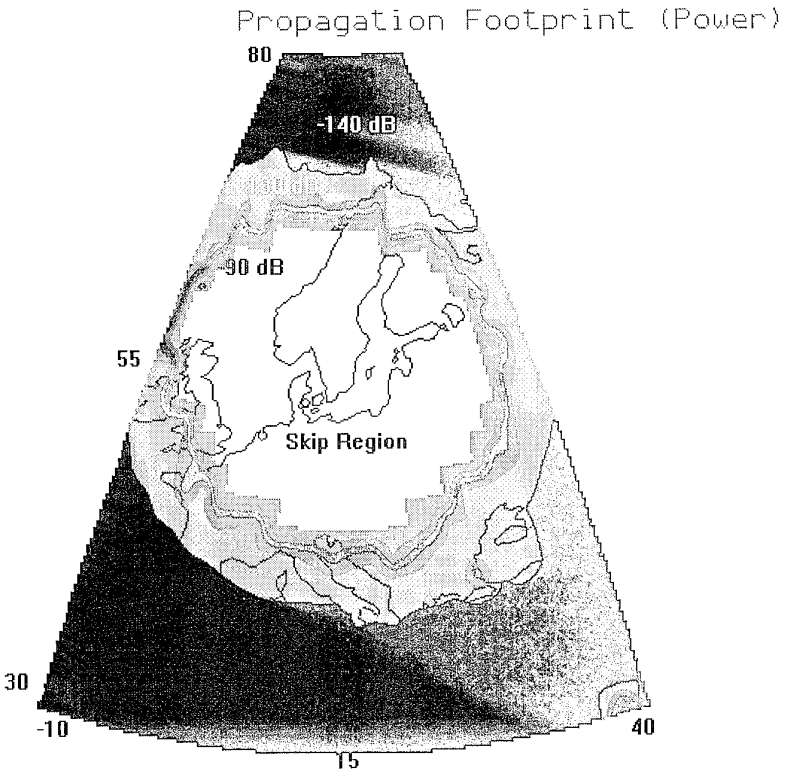


Figure 4 Analytic ray propagation footprint showing the divergent power loss.

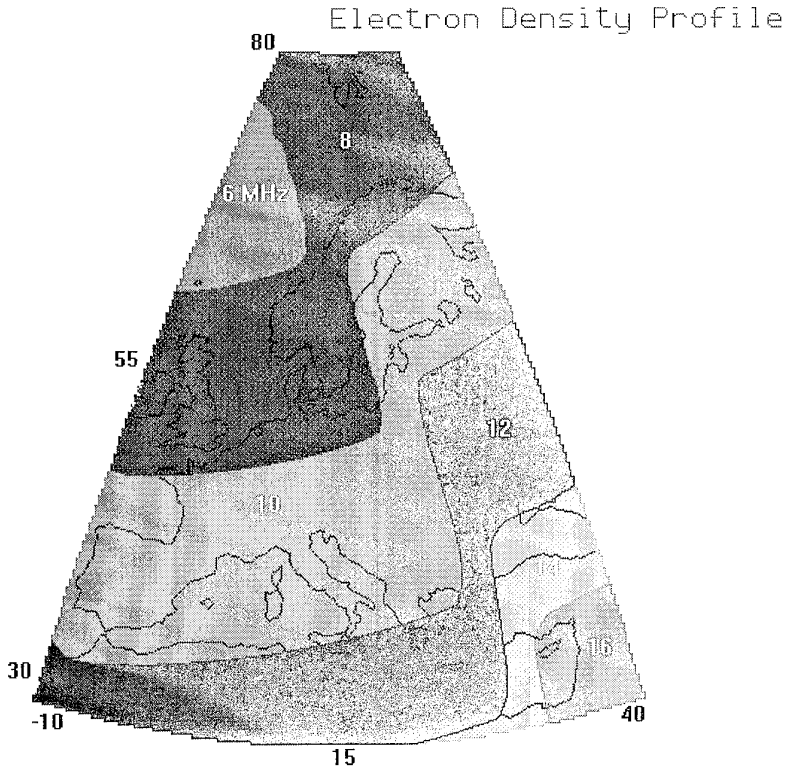


Figure 5 A contour plot showing the plasma frequency, with a terminator present in the east, at a height of 300 km.

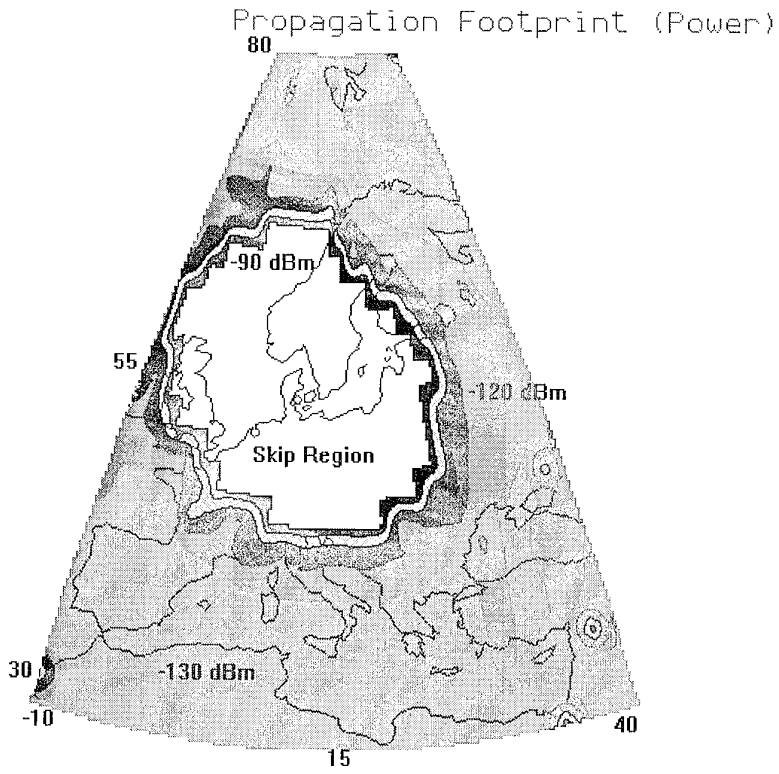


Figure 6 Analytic ray propagation footprint showing the divergent power loss when a terminator is placed east of the transmitter.

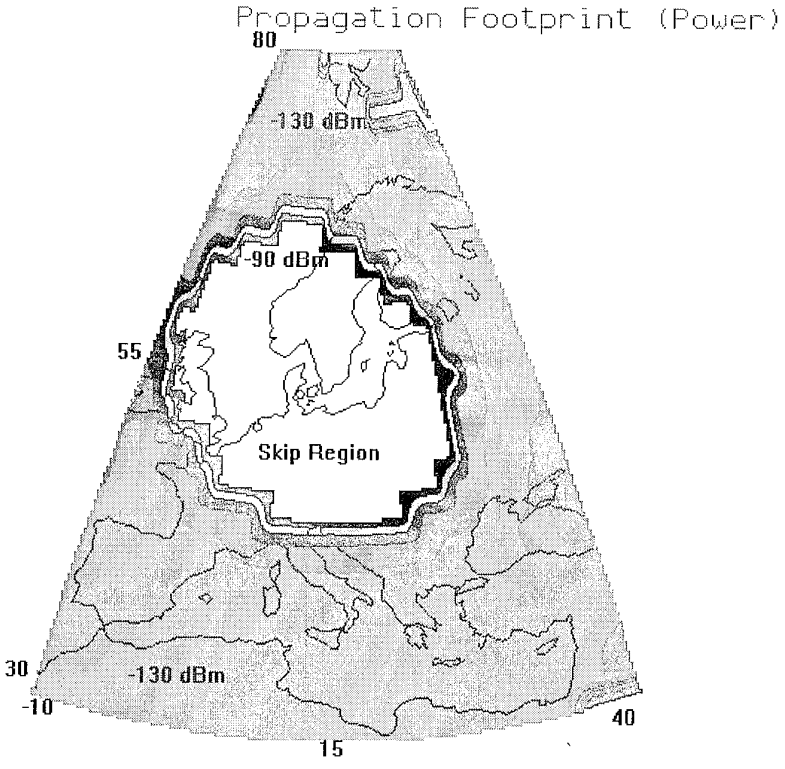


Figure 7 Numerical ray propagation footprint showing the divergent power loss when a terminator is placed east of the transmitter.

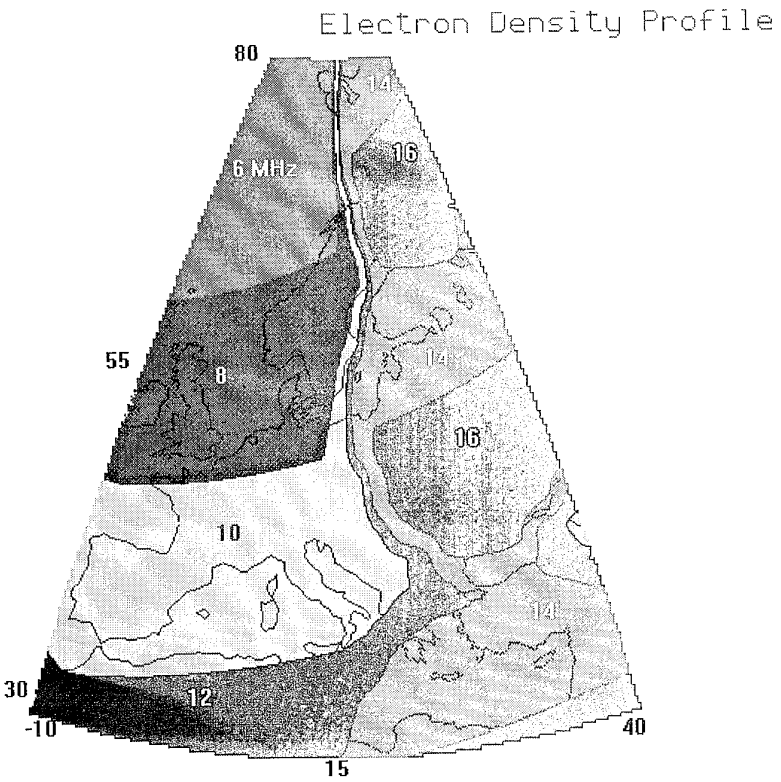


Figure 8 A contour plot showing the plasma frequency using data from three artificial ionospheric sounders placed east of the transmitter.

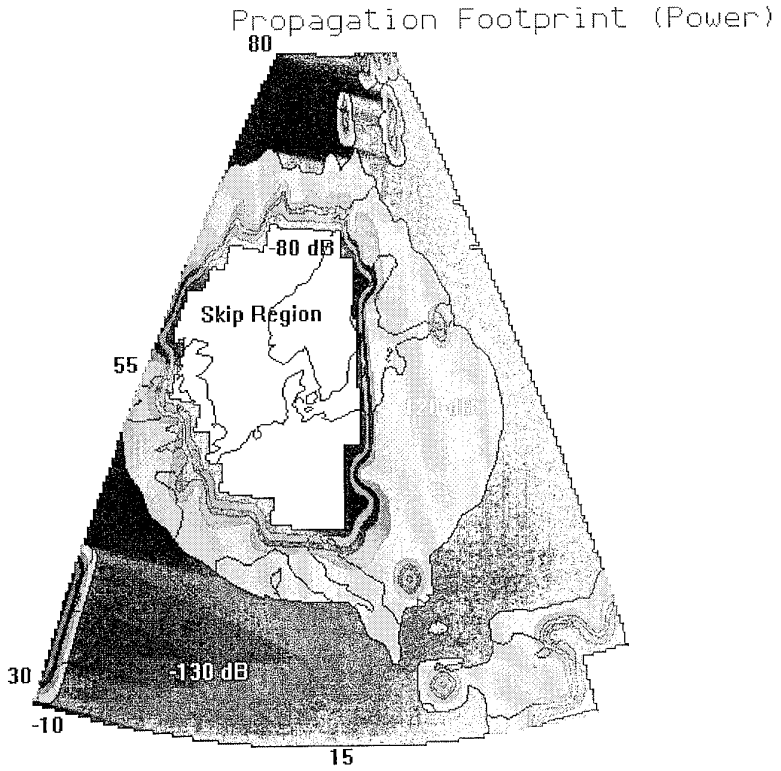


Figure 9 Analytical ray propagation footprint showing the divergent power loss when tracing through the ionospheric model which produced Figure 8.

DISCUSSION

Discussor's name: H. J. Strangeways

Comment/Question:

In the case of large horizontal gradients, numerical ray-tracing shows that a horizontal as well as vertical "reflection" can occur. This can reduce the skip distance and for a sufficiently large gradient even cause the ray to be refracted backwards with respect to its original propagation direction. I think that your analytical model may not be able to reproduce such paths.

Author/Presenter's Reply:

Yes, I agree. The analytic model at present would not be able to produce such paths.

Discussor's name: R. E. McIntosh

Comment/Question:

What computer equipment did you use to perform these simulations and how long did it take to simulate the results with numerical ray tracing, as opposed to the analytical ray tracing?

Author/Presenter's reply:

A 586 90MHz computer was used to produce the propagation footprints. Ray tracing was performed in both the analytic and numerical propagation footprints with 1.0 degree steps in elevation and azimuth. To produce the analytic footprint it took approximately 40 minutes and around 6-7 hours to produce the numerical propagation footprint.

Discussors's name: H. Soicher

Comment/Question:

How do you account for varying gradients, e.g. as those created by traveling atmospheric disturbances (TID's), along any one path?

Author/Presenter's reply:

At present we have only taken into account simple horizontal gradients. We hope to be able to apply the analytical ray tracing model to more complicated horizontal gradients in the near future, either by considering the curvature of the horizontal gradient and then adjusting the curvature of the model, or, using a new technique where the ionospheric model along the ray path is divided up into a number of vertical segments, where each vertical segment represents a new QPS ionospheric model.

DISCUSSION

Discussor's name: C. Goutelard

Comment/Question:

Dans votre présentation vous ne mentionnez qu'une seule fréquence critique pour représenter l'ionosphère. Représentez-vous une seule couche (Region F de l'Ionosphère)?

Par ailleurs, avez-vous comparé votre modèle à un calcul par ray tracing ou des profils réels ou représentatifs?

Translation

In your presentation you only spoke of a single critical frequency to represent the ionosphere. Do you represent only one layer/Region F of the ionosphere? Have you compared your model to a ray tracing computation or to real or representative profiles?

Author/Presenter's reply:

I spoke of a single operating frequency, and not a single critical frequency.

The ionospheric model, described here, represents all layers. Analytic ray tracing was carried out using a 43 equally spaced layered grid system ionospheric model, starting at a height of 90 km and finishing at a height of 300 km.

Yes, I compared results determined from the analytic ray tracing model with results from the Numerical Homing-In ray tracing package, HIRT, coupled with the FAIM model.

Discussor's name: L. Bertel

Comment/Question:

Avez-vous, à partir de votre logiciel, comparé des simulations de déviation azimuthales de rayons à des mesures issues de systèmes de localisation?

Translation

Have you compared simulations of ray azimuth deviations to measurements obtained from locating systems using your software?

Author/Presenter's reply:

No, I have not as yet looked at any azimuth variations.

AN IMPROVED MODEL OF IONOSPHERIC ABSORPTION AT HF HIGH LATITUDE PATHS

V. Jodalen
E.V. Thrane
Norwegian Defence Research Establishment
P.O.Box 25
N-2007 Kjeller
Norway

1. SUMMARY

The performance of current HF prediction programmes at high latitudes is often poor due to a lack of observations at these latitudes, but also because of incomplete understanding of the complex processes occuring in the ionosphere. Earlier work have shown that the prediction of signal strength and propagation using IONCAP/ICEPAC is too optimistic at high latitudes. This work introduces into ICEPAC a new model of electron densities in the high latitude D-region developed by Friedrich and Torkar,1983. The model uses a statistical analysis to relate electron density profiles (60-120 km) to time of day and measured riometer absorption. We suggest that ionospheric loss on oblique incidence skywave paths be calculated in ICEPAC by integrating through the model using the Appleton-Hartree equation. Further we implement mode availability as an independent statistical factor to be multiplied with the probability of achieving a certain signal-to-noise ratio. The new predictions are compared with HF-measurements on a high-latitude path within Norway. The modifications have improved the agreement between data and predictions, particularly for disturbed geomagnetic conditions. For example, the observed diurnal variation of absorption is now reflected in the predictions.

2. INTRODUCTION

Current developement of HF communications include Automatic Radio Communication Systems (ARCS) and Real-Time Channel Evaluation (RTCE) that will ease the work of the radio operators. Knowledge of the ionosphere and models of its behaviour is nevertheless very important in the design of HF equipment and in frequency management. Efficient use of the ionosphere depends on good models both of the ionosphere and of the propagation of radio waves. These models are combined in HF prediction programmes such as IONCAP, CCIR-252, and FTZ. A large, reliable database of observations exist at low and middle latitudes, and this database together with physical models of the ionospheric layers have made it possible to construct quite efficient prediction programmes at these latitudes. Developing adequate models has been difficult in polar regions where the medium is extremely variable in time and space, and few observations exist. Predictions at these latitudes have relied on extrapolation from lower latitudes. ICEPAC (Ionospheric Communications Enhanced Profile Analysis and Circuit prediction program) (1) is a further develpoment of IONCAP (IONnospheric Communications Analysis and Prediction program) (2). ICEPAC contains a high latitude model that includes the dependence of the ionosphere on magnetic activity.

The NDRE (Norwegian Defence Research Establishment) in collaboration with the NUWC (Naval Underwater Warfare Center), Newport, deployed in 1987 an ionospheric oblique incidence sounder in the northern part of Norway, in the auroral zone. The aims of this experiment were to obtain a statistical description of the properties of ionospheric communication channels at high latitudes, and to use these results in the evaluation and possible improvement of existing prediction programmes. IONCAP/ICEPAC was chosen as the prediction program to be evaluated, and a collaboration with ITS initiated. This work reports the results from the evaluation, and suggests further improvements to the high latitude model that incorporates features seen in the experimental data.

The first results from this study (3), (4) and (5) showed that large discrepancies between data and predictions exist in the modelling of radio wave absorption, and that absorption should be modelled as a function of geomagnetic activity. Jodalen and Thrane (5) incorporated in ICEPAC a new method for calculating ionospheric absorption appropriate for high latitudes and the present article describes an extension of this work. Section 3 describes the HF test experiment, Section 4 give an example of comparison between data and predictions and Section 5 describes current formulas and methods used in ICEPAC. In Section 6, a high-latitude D-region model is introduced as well as a revision of the propagation modelling. New comparisons of data and modified predictions are made in Section 7 and conclusions are drawn in Section 8.

3. THE HF TEST EXPERIMENT

The first part of the tests took place within the auroral oval. The transmitter was located at Andøya (69.30 N, 16.02 E), and the receiver was in Alta (69.90 N, 23.20 E), a distance of 285 km. The tests were conducted between January 1987 and December 1988 on this path. The HF test transmitter was then moved to Kløfta (60.07 N, 11.12 E), and data collected on the longer path (1230 km) from January 1990 to June 1992. The two paths are shown in Figure 1. The right part of the figure show two possible propagation modes on the long and short path, respectively. The observations were organized in an hourly schedule and ran 24 hours a day. A standard message was transmitted every hour on nine frequencies using five different bandwidths for each frequency. The Bit Error Rate (BER), relative amplitude and time delay of up to five multipaths, as well as signal and noise level were recorded automatically for each transmission. The database contains observations recorded over several years, but, unfortunately, technical and logistic problems have caused gaps in the sequence of recordings. Nevertheless, the database covers diurnal and seasonal variations as well as different levels of geomagnetic activity and sunspot numbers ranging from 35 to 158.

The oblique incidence sounder used both conventional and direct sequence, spread-spectrum (DSSS) radio techniques. The HF test transmitter and test receiver were designed and built at the NDRE to control the timing, the scheduling of

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

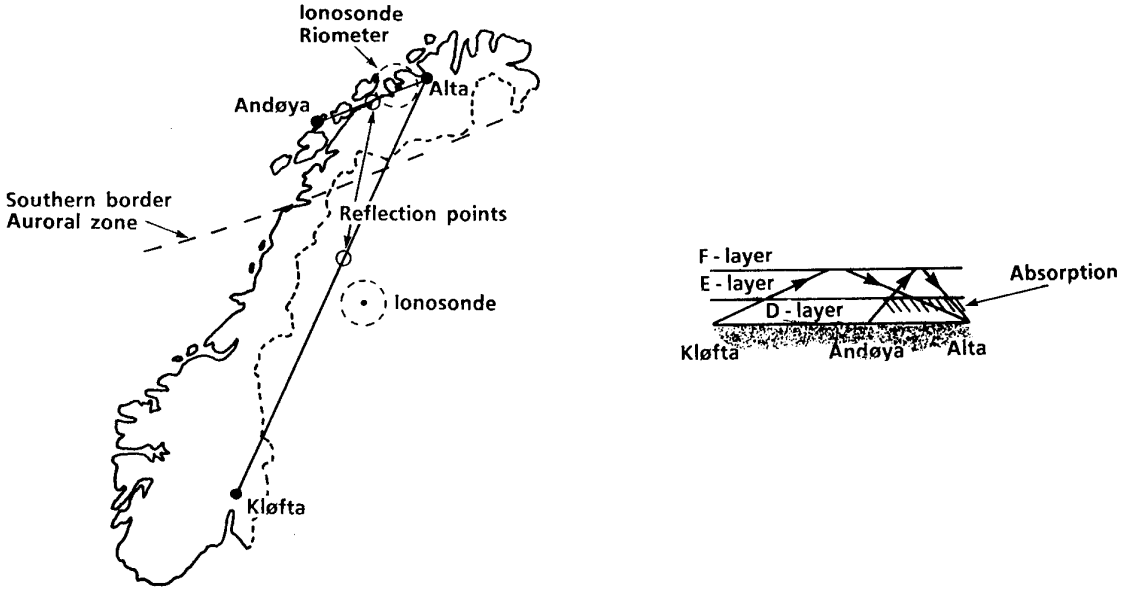


Figure 1. Location of transmitter and receiver for the short and long path.

different frequencies and bandwidths, the modulation /de-modulation and coding/decoding of the standard message. For each hour the message is sent on nine different frequencies; 3.0, 4.5, 6.0, 8.5, 10.5, 12.5, 14.5, 17.0 and 19.0 MHz, and for each frequency, using five different bandwidths; 0.5 kHz (conventional signaling), 5, 20 80 and 160 kHz (spread spectrum). The modulation method used is differentially encoded PSK (DPSK), and the signals are detected non-coherently. When synchronization is obtained, a counter starts, and the minimum BER that can be measured is $6.7 \cdot 10^{-4}$. From the bit error number we derive reliability defined as the number of messages received with a bit error rate less than 10 % for the actual period of study. The choice of 10 % as a limit for the "acceptable" BER is somewhat arbitrary, but when comparing this reliability with *predicted* reliability, the same limit has been used for the predictions by specifying the appropriate required signal to noise ratio for the modulation scheme in use. Reliability has been derived from the 0.5 kHz "conventional" messages.

Transmitter and receiver were synchronized to better than 1 μs. To derive the virtual height of reflection from these measurements, a one- hop mode is assumed for both the short and the long path, which is in good agreement with predictions. Using the 5 kHz spread spectrum messages, the height resolution obtained is approximately 30 km. In Figure 2 we have only displayed the two strongest multipaths. A larger number of modes were very unlikely to occur.

In this work we compare virtual height of reflection and reliability with predictions. We show only data from the short path, but data from the long path can be found in (3) and (6).

4. COMPARISON OF DATA AND PREDICTIONS

We focus on the influence of geomagnetic activity on propagation, and select two groups of days with low ($Q \leq 4$) and high ($Q \geq 5$) geomagnetic activity. The selected days are not necessarily consecutive, but they are all from the same season, within a period of two-three months. The number of days for each dataset is between 10 and 40. Each dataset is compared with predictions for Q=3 or Q=6, respectively. The prediction is made for the month within the period that contains the largest number of measurements.

Figure 2 shows the measured (X) and predicted (Δ) virtual heights of reflection for March/April 1988 on the short path. The days with low geomagnetic activity (16 days) are displayed in the left column and high geomagnetic activity (17 days) in the right column. Only the four lowest frequencies in the measurement are shown, since transmission above 8.5 MHz on this path only occur sporadically. The number of points reflects the number of messages received. The median of the data is drawn as a continuous line. The predicted virtual height is calculated for the most reliable mode. Note that a virtual height is predicted even for zero predicted reliability.

From the figure we can draw the following conclusions: As expected, more messages have been received for low than for high geomagnetic activity, and the quiet periods show a more well defined regular diurnal variation. The data of the disturbed period shows a wide spread of reflection height during daytime, but fairly well defined auroral E-layer reflections at night. There are more night time auroral E-layer reflections during the disturbed period, strong enough to reflect 6.0 MHz. The normal daytime E-layer is only observed at 3.0 MHz. The most prominent difference between the predictions for Q=3 and Q=6 is more E-layer reflections in the hours after midnight for Q=6. Compared to the data, ICEPAC performs reasonably well after midnight and during daytime. The prediction performance is poor for pre-midnight hours and at f=3.0 MHz.

Measured and predicted reliability for the same two periods are displayed as separate three-dimensional plots in Figures 3 and 4. The reliability is plotted as a function of hour of the day (UT) and frequency (MHz). A notable difference between the two plots of data, Q<5 and Q>4, is the low reliability observed during the disturbed period for the hours 6-10 UT. This feature is observed in all datasets studied. This minimum is most likely caused by pronounced absorption due to precipitating particles from the magnetosphere. ICEPAC does not predict this feature, and the predictions are in both cases much too optimistic, both in reliability and in predicting the MUF. The two predictions are very similar, but with a slightly enhanced MUF during disturbed conditions around midnight and 17 hours.

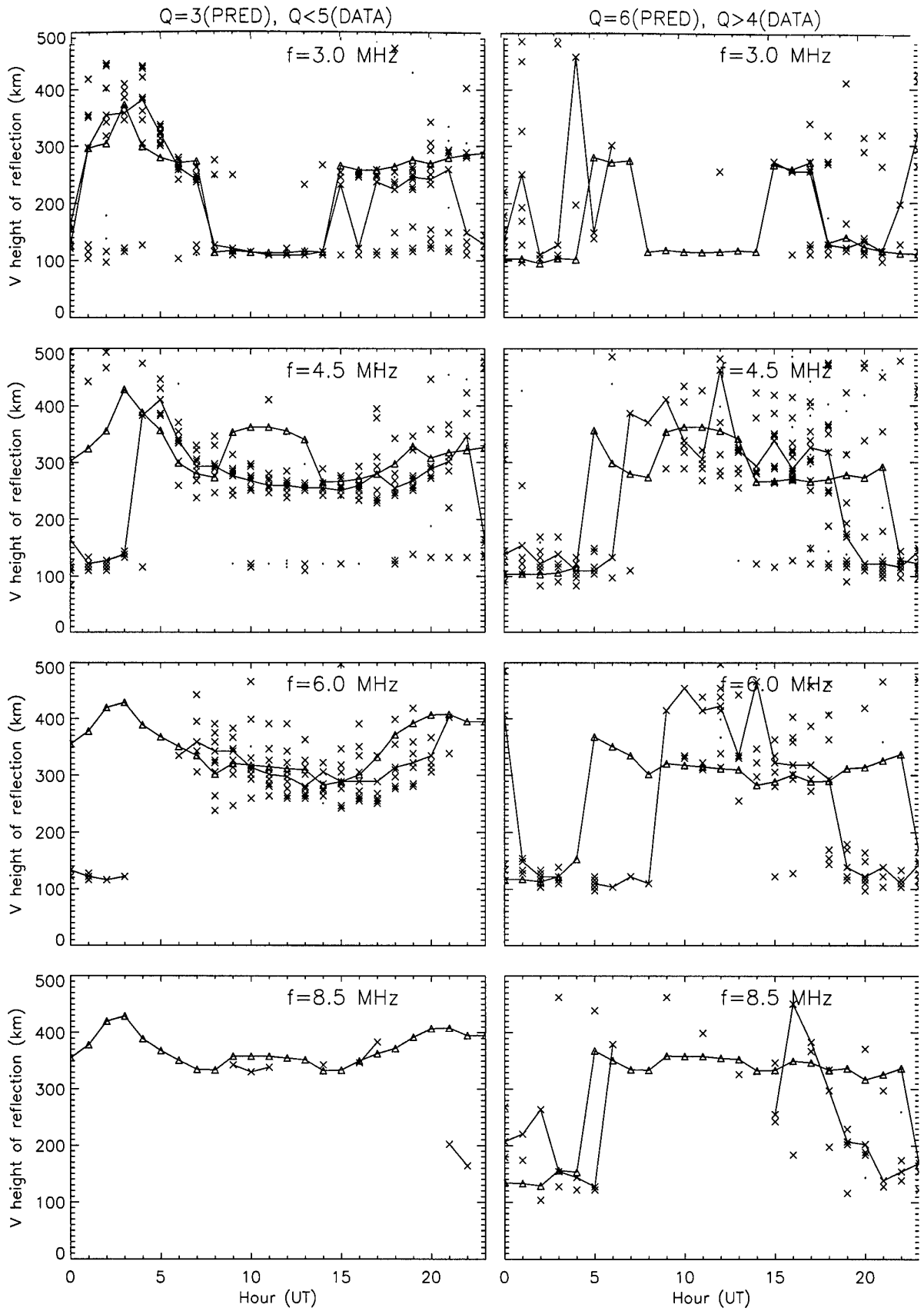


Figure 2. Virtual height of reflection, Andøya-Alta (285 km), March/April 1988. The median of data drawn as a continous line. Small dots indicate multipath. Predicted virtual height shown as triangles. SSN=109.

Q<5:

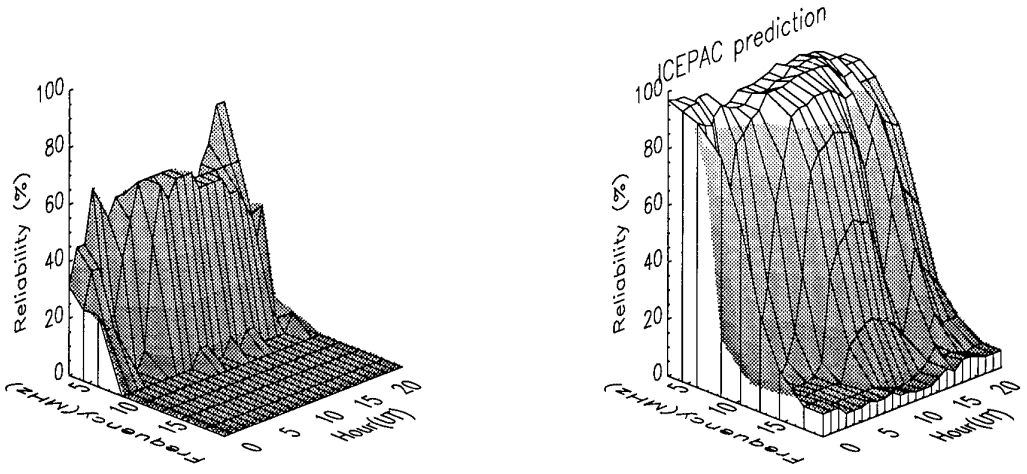


Figure 3. Reliability of data for March/April 1988, left, prediction for March, right (SSN=104). Andøya-Alta (285 km).

Q>4:

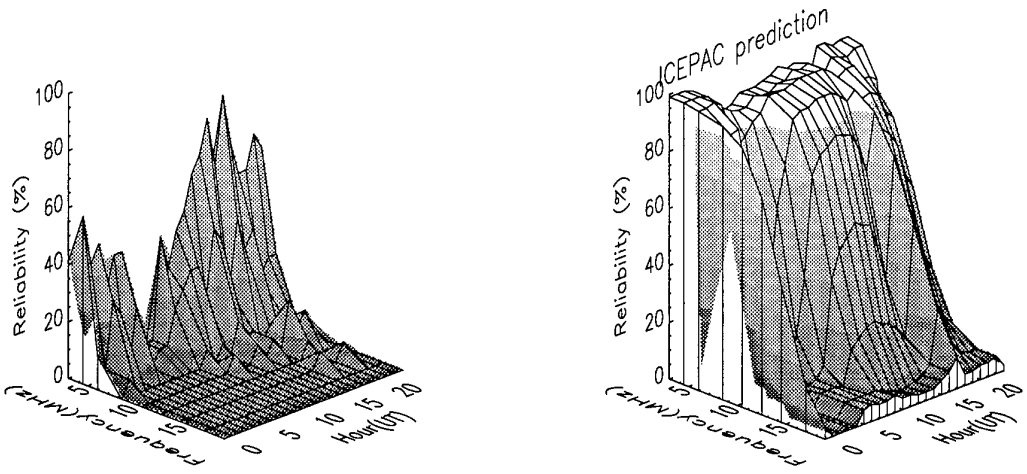


Figure 4. Reliability of data for March/April 1988, left, prediction for March, right (SSN=104). Andøya-Alta (285 km).

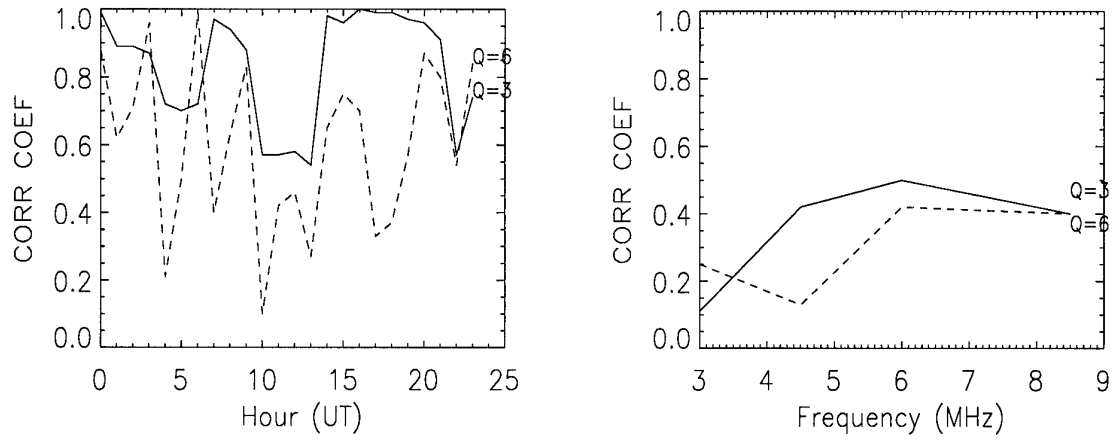


Figure 5. Normalized correlation coefficients between data and unmodified predictions. Q=3 continous line, Q=6 dashed line.

There are uncertainties associated with the estimated required signal to noise ratio that we use as input to ICEPAC. An increase of 3 dB of this parameter corresponds to a decrease of the predicted reliability of 4-5 %. So the absolute levels of predicted and observed reliability should be compared with caution. We therefore emphasize a comparison of the relative variations of observed and predicted reliabilities and MUFs rather than the absolute levels. We have chosen to use a normalized correlation coefficient k(N) between data and predictions as a quantitative measure of relative variation:

$$k(N) = \frac{\sum_{n=1}^{N}(x(n)-\bar{x})\cdot(y(n)-\bar{y})}{\sqrt{\sum_{n=1}^{N}(x(n)-\bar{x})^2}\cdot\sqrt{\sum_{n=1}^{N}(y(n)-\bar{y})^2}} \qquad (1)$$

$|k(N)| \le 1$. N is the number of points to be correlated, x(n) is in our case the observed reliabilities, $\bar{x}$ their mean value, y(n) is the predicted reliabilities and $\bar{y}$ their mean value. The normalization makes this measure independent of absolute levels.

Figure 5 shows the correlation coefficients for the March/April time period, for each frequency across all hours (N=24) in the right panel and for each hour across the four lowest frequencies (N=4) in the left panel. The correlations are in general lower for the disturbed period.

In Table 1 we have calculated the correlation coefficient over all datapoints (N=96) for four periods from different seasons. For the period we are considering (March/April 1988), the total correlation coefficients for all hours and the four lowest frequencies are 0.71 for Q=3 and 0.54 for Q=6.

k(N) for time period:	Q=3	Q=6
September/October/November 1987, SSN=39	0,66	0,56
March/April 1988, SSN=71	0,71	0,54
June/July/August 1988, SSN=109	0,65	0,43
October/November 1988, SSN=125	0,63	0,61

Table 1. Total correlation coefficients between data and predictions for different time periods.

5. PREDICTIONS

Long-term predictions of the average state of the ionosphere are actively used in daily operations of communication circuits, in design of equipment and in frequency mangement. Noting the large discrepancies between predictions and data described in the last section, we shall here discuss the features of ICEPAC that lead to these differences. The two main contributors to the discrepancies are ionospheric absorption and Above-the-MUF propagation.

5.1 Ionospheric absorption

The Ionospheric Conductivity and Electron Density (ICED)-model (7) have been included in ICEPAC to improve predictions at high latitudes. The ICED-model specifies the location of the auroral oval using the effective Q-index as a measure of the global geomagnetic activity. A large effective Q-index expands the auroral oval towards middle latitudes. The ICED model specifies the electron densities from 90 km to 1000 km altitude as a function of solar activity, solar zenith angle, geomagnetic activity and magnetic local time. It divides the ionosphere into four latitudinal regions: the polar cap, the auroral zone, the sub-auroral trough and low/mid-latitudes. It does not contain any specific model of the high latitude D-region (60-90 km) where strong and variable absorption of HF radio waves occurs. ICEPAC creates complete electron density profiles in a certain number of controlpoints along the great-circle path using the ICED-model. The number of control points is dependent on the path length, for our particular example (<2000 km), only one control point is used. Possible modes are found for each frequency, and delay, losses, signal strength and reliability are predicted for each mode. The most reliable mode is then selected.

When the ICED-model was implemented in ICEPAC, ionospheric loss in the special absorption conditions found at high latitudes was only taken into account by adding an auroral loss term (5-7 dB) to the total loss. Without this loss term the formula (in dB) used in ICEPAC is:

$$L_i = \frac{677.2\cdot n\cdot \sec\varphi\cdot(-0.04+\exp(-2.937+0.8445\cdot f_0E)\cdot(\cos^{1.3}0.881\chi)}{(f+f_H)^{1.98}+10.2} \qquad (2)$$

where n is number of hops, φ is angle of incidence of the wave on the ionosphere, f_0E is critical frequency of the E-layer, χ is solar zenith angle and f_H the gyrofrequency at 100 km altitude. This is a semi-empirical equation derived for middle latitude conditions where a dependence of the critical frequency f_0E of the normal E-layer is included. It was pointed out in (4) that this dependence will result in excessive absorption in those cases where the ICED-model predicts a large auroral E-layer (f_0E large).

5.2 Above-the-MUF propagation

Above-the-MUF propagation has been extensively discussed in the literature. The problem is the following: Even if the instantaneous MUF (MOF) has been determined using ionosondes, observations have shown that as the frequency increases from just below to just above the MUF, the signal does not drop instantly to zero. The propagation mechanism at these frequencies is not by ionospheric refraction, but rather by scatter from ionospheric inhomogenities and two-hop backscatter (8). Other authors (9) claims that this Above-the-MUF propagation is not well documented.

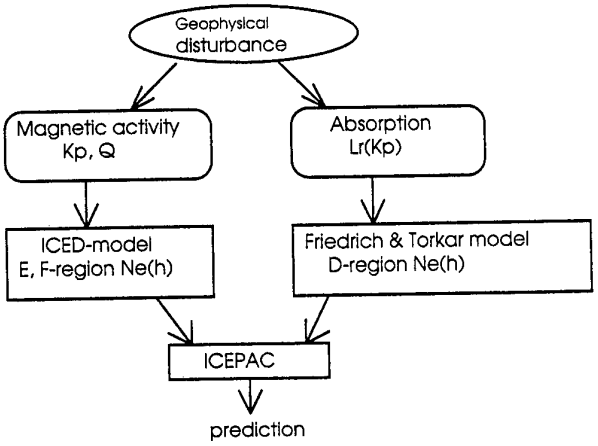
Instead of modelling the scatter as a separate propagation phenomenon, ICEPAC and some other prediction programmes have added a loss term, Above-the-MUF loss (ABM), to the total loss. This results in large losses as the frequency is increased above the MUF. The probability of a mode actually being present is not dealt with explicitly. Using this approach, the calculated reliability expresses the total time a specific grade of service can be expected, independently of the available modes. An approach formerly used is perhaps phenomenologically more correct. It calculates the reliability as the joint probability of the mode availability and the probability of the S/N exceeding the required S/N *provided* that the mode exists. The two approaches yield very different reliabilities at frequencies around the MUF.

We noted in the previous section that ICEPAC produced far too optimistic predictions at higher frequencies. We therefore choose to remove the ABM-losses and instead

implement the former approach. This means that the reliability is computed as a product of the mode availability and the probability of S/N exceeding the required signal-to-noise ratio.

6. FURTHER DEVELOPMENTS OF ICEPAC

In addition to the implementation of mode availability described in the last section, we would like to suggest an approach to calculate ionospheric absorption that is more consistent with the physical process of absorption than the current method. The idea is to incorporate a model of the D-region in ICEPAC, and can be illustrated by Figure 6.



Figur 6. Basic principles of incorporating a D-region model in ICEPAC.

Geomagnetic activity is often associated with increased radio wave absorption in the D-region. The ICED-model selects electron density profiles $N_e(h)$ for the E and F-region according to the degree of disturbance expressed by the index Q, but does not include an explicit Q dependence of the electron density in the D- and lower E-region. Jodalen and Thrane (5) introduced such a dependence in ICEPAC by adopting the model of electron densities in the auroral D-and E-regions developed by Friedrich and Torkar (10). The Friedrich and Torkar statistical model relates the electron density profile to riometer absorption and Jodalen and Thrane introduced a statistical relation between riometer absorption and Q. The D-and lower E-region profiles are then combined with the ICED-profile for the E and F-regions. Instead of the semi-empirical formula previously used for calculating ionospheric loss (Equation 2), we find the absorption by integration of the Appleton Hartree equation along the raypath in the lower ionosphere.

Figure 7 shows the day time F&T-model as a function of riometer absorption and solar zenith angle.

The electron densities found in the high-latitude ionosphere are produced both by regular solar radiation that varies throughout the day, and by irregularly precipitating particles from the magnetosphere. The F&T-model profiles are believed to be representative for all degrees of auroral disturbances at high latitudes. Riometer absorption between 0 and 2.5 dB occurs more than 90% of the time in the auroral zone, and consequently, N_e -profiles for these conditions will also cover the situation for 90% of all cases.

M. Friedrich and K. Torkar have also developed a similar model of the low latitude D-region, and it has been adopted, but not yet published, as the Interim Model of the D-region in the IRI (International Reference Ionosphere).

6.1 Implementation of the D-region model in ICEPAC

In this work we have implemented the F&T-model taking one month as an example. The month of March (as in Section 3) was chosen. Since the F&T-model is driven by the amount of absorption measured, L_r, and the ICED-model is controlled by

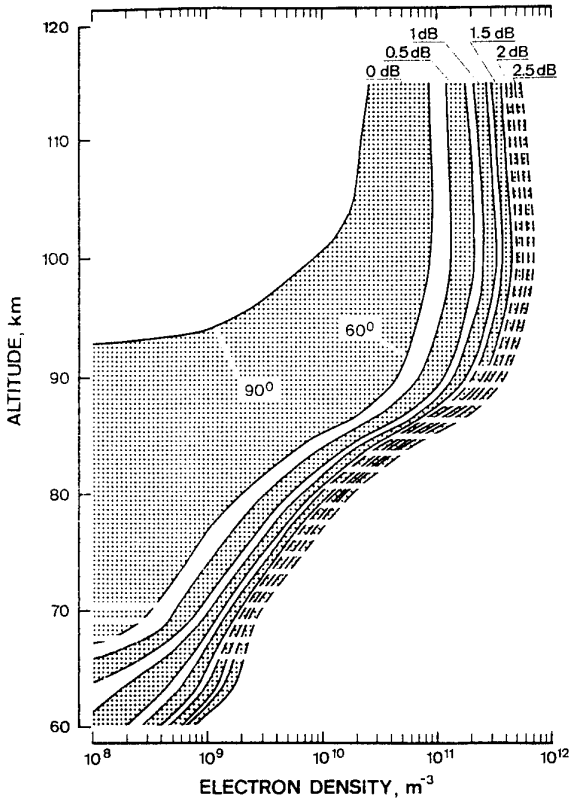


Figure 7. The day time statistical model of D-region electron densities

the measured geomagnetic activity, Q, a relationship between the two parameters had to be found in order to combine the two electron density profiles. Q-data and L_r-data for the month of March 1988 were correlated, and a least-square first order polynomial fit was made to the data. A linear relationship was found for each of eight time intervals during the day. The steepest linear relationship can be found in the morning (3-11 UT). Now a certain value of Q to be given as input to ICEPAC, can be associated with a certain L_r (depending on time of day), and L_r will again select the right D-region electron density to be combined with the ICED-profile. The profiles should in principle be combined smoothly in order to construct a realistic profile, but a discontinuity is not important since the collision frequency, and hence the contribution to the ionospheric absorption, is small around 120 km.

A computerized version of the Appleton-Hartree equation expresses the ionospheric non-deviative absorption in dB on an oblique incidence path:

$$L_i = 4.6 \cdot 10^{-2} \cdot \frac{1}{\cos\varphi} \cdot 2 \cdot \sum_{h_1}^{h_2} \frac{N(h) \cdot \nu(h)}{4\pi^2 (f + f_L)^2 + \nu(h)^2} \qquad (3)$$

where ν is the electron-neutral collision frequency, N the electron density, φ is the angle of incidence of the radio wave, the factor 2 accounts for the radio wave traversing the ionosphere twice, h1 is the lowest altitude (where N approaches zero) and h2 is the altitude of the reflection of the ray. ICEPAC predictions for the path Andøya-Alta provides h2 from the predicted virtual height assuming a one hop path.

The D-region profiles have then been used to calculate ionospheric losses as a function of frequency and time as shown in Figures 8 and 9. Unmodified ICEPAC predictions of ionospheric loss are presented in the right hand panels for comparison. Note that the scaling of the loss axis is different in the figures, but the gray scaling is the same. The MUF for this path is never above 8 MHz, so propagation above 8 MHz is unlikely. In the prediction plots we have therefore shown frequencies only up to 8.5 MHz.

For Q=0, the model calculations (left) show a smooth loss surface with a maximum ionospheric absorption of 28 dB at 11 UT (12 local time), and a night absorption of 12 dB. The statistical electron density profiles for each hour have been chosen based on a constant riometer absorption L_r=0. The only varying parameters are the solar zenith angle for day hours and the virtual height of reflection. An increased height of reflection reduces the angle of incidence and thereby the absorption. Because of this obliquity factor, an E-layer mode may experience an additional loss of up to 15 dB compared to an F2-layer mode. This is the case for the lowest frequencies around noon. The transition regions between night and daytime loss, where the loss is minimum, is due to a discontinuity in the F&T-model electron density profiles, which has no physical basis.

Q=0:

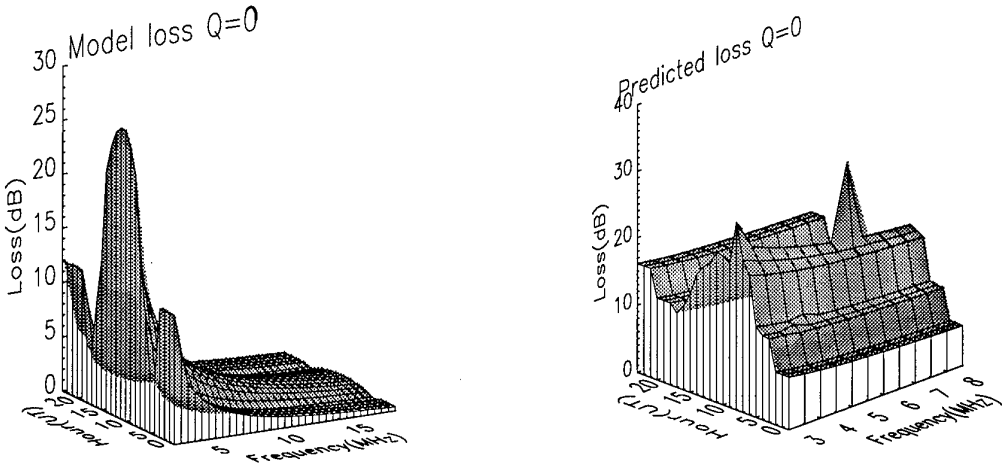


Figure 8. Ionospheric absorption using the D-region electron density model and Equation 3 left, unmodified ICEPAC predictions right.

Q=6:

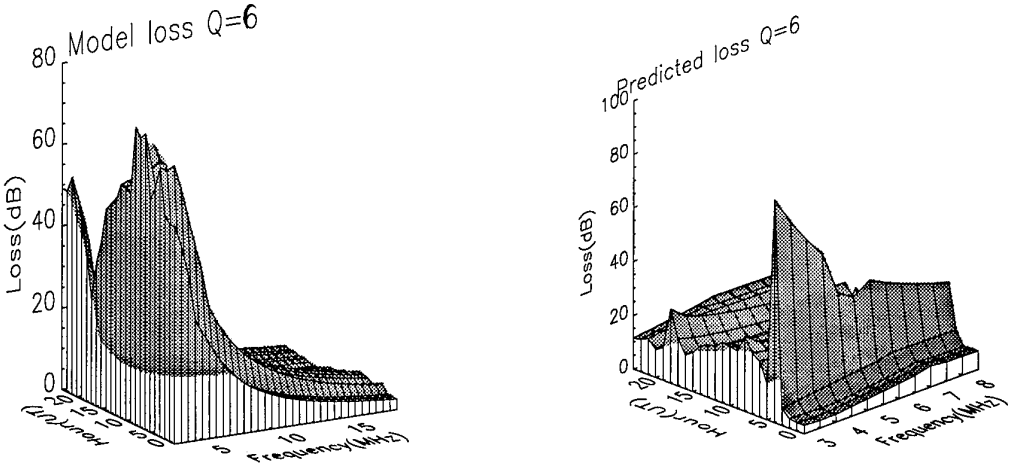


Figure 9. Ionospheric absorption using the D-region electron density model and Equation 3 left, unmodified ICEPAC predictions right.

The unmodified absorption predicted for Q=0 is of the same order as our F&T-model calculation. Note that the unmodified predictions includes a small deviative loss term as well as the auroral loss correction term. The deviative loss is the cause of the two peaks seen in the unmodified predictions for Q=0. The unmodified predictions and model calculations should therefore be compared with some caution.

For Q=6, the model loss values are more irregular during daytime hours. The reason is that the L_r-values corresponding to Q=6 that drives the choice of electron density profiles, vary for the different hours. This "peaky" nature is probably not realistic. Note however that the model calculations, in contrast to the unmodified predictions, now reflect an important characteristic of high latitude radio communications: An absorption maximum in the morning (6-8 UT). The maximum absorption is 70-80 dB, far above the maximum absorption during quiet conditions.

The most characteristic features of the unmodified predictions for Q=6 are the two peaks, one of 90 dB at 4 UT and the other one of 30 dB at 18 UT. They are purely a result of an erroneous implementation of Equation 2 as pointed out in (4). Neglecting these two peaks, the absorption level for Q=6 is close to that for Q=0, much smaller than the model calculations.

In the following, we have substituted in ICEPAC Equation 2 with these model calculations.

7. DISCUSSION: COMPARISON OF DATA AND MODIFIED PREDICTIONS

The purpose of our modifications to ICEPAC is to improve the predictions of signal strength by incorporating a new loss model. As a test of our suggestions in the last sections, we will compare the modified predictions with our high-latitude data, using the same methods of comparison as in Section 4. With both the D-region model and mode availability implemented in ICEPAC as described in Section 6, predicted virtual height of reflection is displayed in Figure 10. Again the month of March 1988 is chosen as an example. Only the median of the data is shown (crosses) together with old predictions (triangles) and modified predictions (squares).

The statistical D-region profiles have large plasma densities, particularly for disturbed conditions, and have caused E-region reflections for the lowest frequency. The data show that this is partly true, but also that reflections from F2 are frequent. The modified predictions have not picked up the pre-midnight E-layer reflections for Q>4. The conclusion is that the modifications to ICEPAC have not improved predictions of virtual height.

Figures 11 and 12 show the modified predictions of reliability (right) together with the HF test data (left). Only the four lowest frequencies are shown since no data exist above 8.5 MHz. For the higher frequencies the inclusion of mode availability has caused a large change. For both levels of geomagnetic activity, the reliability has been drastically reduced. For instance, for Q=3, the reliability at 8.5 MHz has been reduced from 80 % to 0 % at night and from 90 % to less than 30 % around noon. The effect of including mode availability are almost the same for Q=3 and Q=6.

The influence of the F&T-model is strongest for the lowest frequencies. It is negligible above 4.5 MHz for Q<5 and above 6 MHz for Q>4. Because f^2 appears in the denominators in the equations for ionospheric loss, this is to be expected. For Q<5, the predicted reliability has been reduced from 90% to about 80% for f=3.0 MHz, which is still far above the measured reliability (20-70%). For Q>4, the reliability has been reduced from about 90% to 10% in the morning hours and by minimum 10 % during the late afternoon. The predictions are still optimistic, but the marked diurnal variation now seen for Q>4 is similar to the observed variation. Because of the uncertainty of the absolute level of the predictions, we will again compare the normalized correlation coefficients between predictions and data, see Figure 13.

Compared to Figure 5 the correlations in Figure 13 for all individual frequencies for Q=3 have increased whereas that is the case only for the two lowest frequencies for Q=6. Individual correlations for most of the hours have increased for both levels of geomagnetic activity. Since the ionospheric medium is much more variable and difficult to predict during disturbances, one must expect less correlation between data and predictions than for quiet conditions. The correlations across all frequencies for each hour are in general larger than the correlations across all hours for each frequency.

The impression of a much better agreement between data and modified predictions is also documented by an overall correlation coefficient for Q=3 of 0.91 and for Q=6 of 0.78. This is an increase of 20 % for Q=3 and 24 % for Q=6 compared with Table 1.

Tests show that including mode availability is the main reason for the improved correlations for Q=3. This is an expected result since Equation 2 is an empirical fit to low latitude data, and does not perfom badly at low latitudes (=quiet conditions). The F&T-model have little to contribute in these situations, but provides an alternative way of calculating ionospheric absorption. For Q=6, both the F&T-model and mode availability are contributing to the increased correlations in Figure 13.

Based on our data and the comparisons made here, we support those authors that claim that the probability for over-the-MUF propagation is overestimated. Including mode availability as an independent factor in the reliability calculations, has given a much better agreement between data and predictions.

One comment should be made about the method of comparison. Unfortunately, our signal strength measurements did not turn out well. Since signal strength data was not available, we have been using bit error rates as a measure of the signal to noise ratio. This measure includes in addition to absorption, also fading, interference and white noise that all will degrade the measured BER. The *predicted* signal to noise ratio includes absorption and different sources of noise (man-made, atmospheric, galactic) that together have a white

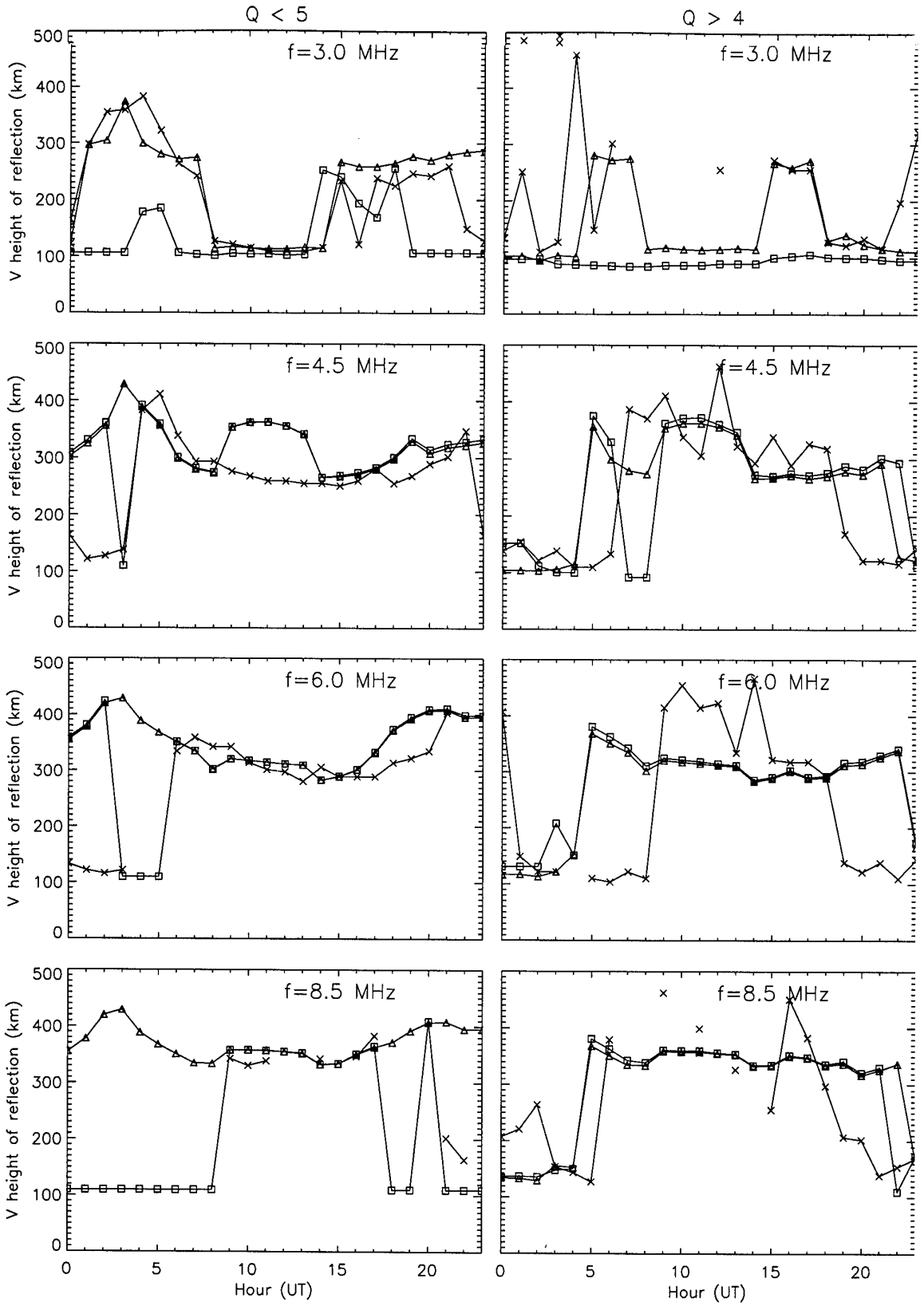


Figure 10. Virtual height of reflection for median of measured data (X), unmodified predictions (Δ) and prediction with both the F&T-model and mode availability included ( ). Andøya-Alta (285 km), March/April 1988.

Q<5:

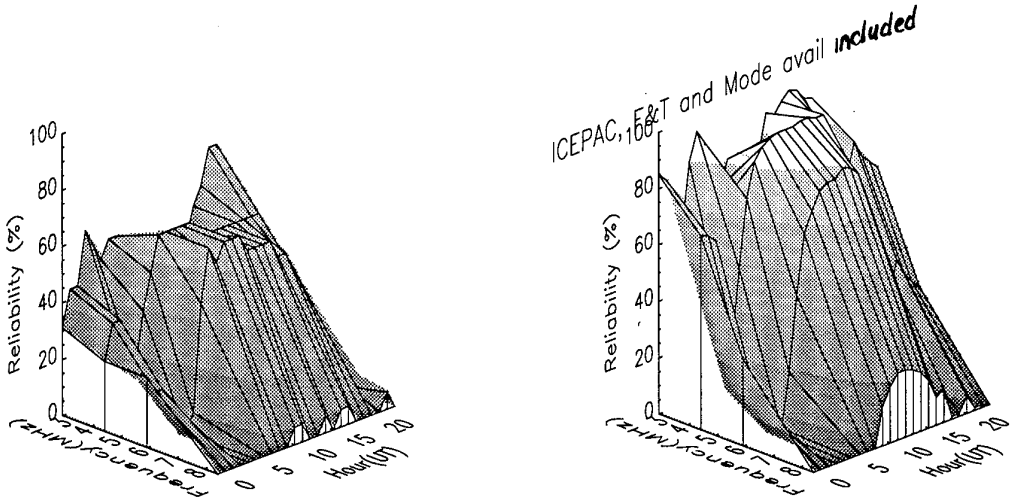


Figure 11. Reliability of data for March/April 1988, left, prediction for March with both the F&T-model and mode availability included, right. Andøya-Alta (285 km).

Q>4:

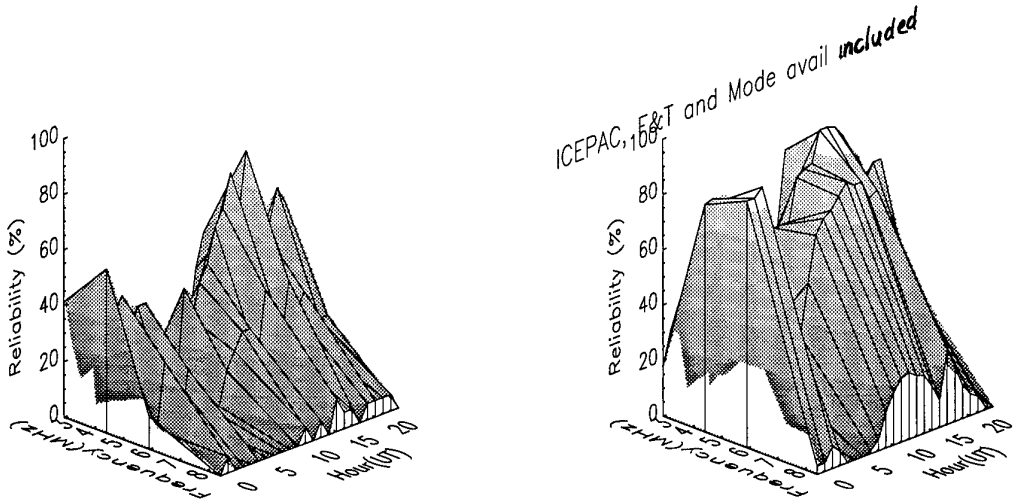


Figure 12. Reliability of data for March/April 1988, left, prediction for March with both the F&T-model and mode availability included, right. Andøya-Alta (285 km).

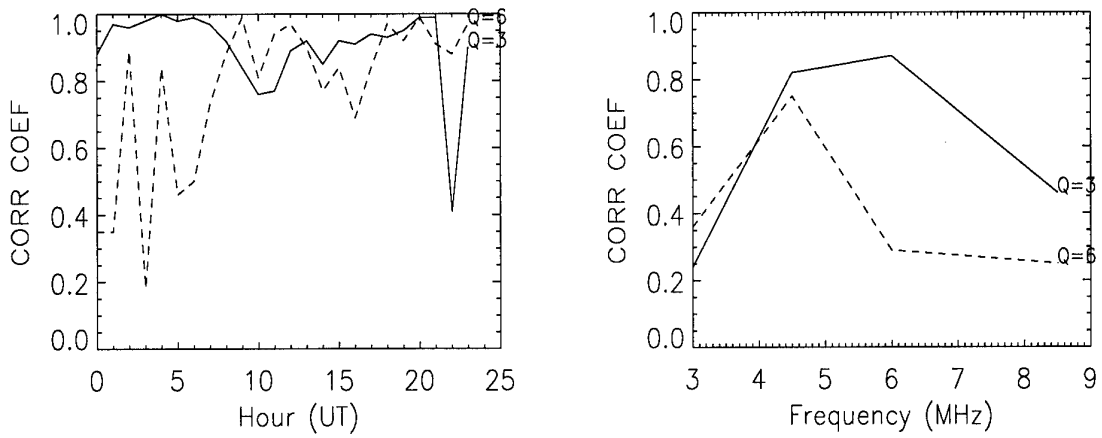


Figure 13. Normalized correlation coefficients between data and modified predictions, Q=3 continous line, Q=6 dashed line.

noise distribution. It is therefore difficult to compare directly signal characteristics of the two since the data also contain other sources of impairment. Anyway, measurements have shown (11) that the amount of interference in the northern part of Norway is very small, and should therefore not be a major source of signal degradation. Fading will probably be a problem. In spite of the difficulties encountered when comparing measured BER to predicted signal-to-noise ratioes, we feel that our correlation analysis has to some extent overcome these problems and provided a meaningful comparison of data and predictions.

8. CONCLUSIONS

By introducing a D-region model of the high latitude ionosphere and also implementing mode availability as an independent statistical parameter in ICEPAC, the agreement between observations and predictions of reliability on our high latitude path has increased significantly. In particular, the modified predictions now reflect the high latitude diurnal variation and the dependence of absorption on geomagnetic activity. The prediction of virtual height has not improved by our modifications.

The users of the HF spectrum should be aware of the large absorption occuring at high latitudes at certain times of day, and possibly lessen its effect by relaying their signals via HF stations outside the auroral oval.

9. REFERENCES

1. Stewart F., G. Hand, Institute of Telecommunications Sciences, Boulder, Colorado, USA -private communication 1991-1994.

2. Teters L.R., J.L. Lloyd, G.W. Haydon, D.L. Lucas, "Estimating the performance of telecommunication systems using the ionospheric transmission channel", *NTIA Rep.*, 83-127, National Telecommunications and Information Administration, Boulder, Colorado, 1983.

3. Jodalen V., E.V. Thrane, I. Koltveit, T. Hellum, "Ionospheric HF radio propagation at high latitudes using conventional and spread spectrum modulation", *FFI/Rap.* 91/7003, Forsvarets forskningsinstitutt, Kjeller, Norway, 1991.

4. Thrane E.V., V. Jodalen, F Stewart, D Saleem, J Katan "Study of measured and predicted reliability of the ionospheric HF communication channel at high atitudes", *Radio Science,* 29, 5, September/October 1994, pp 1293-1309.

5. Jodalen V., E.V. Thrane, "A study of the relation between ionospheric absorption and predicted HF propagation parameters at high latitudes", *IEE Conference Publication,* 392, Sixth Int. Conf. HF Radio Systems and Techniques, York, 1994.

6. Jodalen V., PhD under preparation (1995).

7. Tascione T.F., H.W. Kroehl, R. Creiger, J.W. Freeman Jr., R.A. Wolf, R.W. Spiro, R.V. Hilmer, J.W. Shade, B.A. Hausman, "New ionospheric and magnetospheric specification models", *Radio Science,* 23, 1988, pp 211-222.

8. Hagn G., A.J. Gibson, P.A. Bradley, "Propagation on frequencies above the basic maximum usable frequency", *Proceedings of the Ionospheric Effects Symposium, IES-93,* Washington, USA, 1993.

9. Argo P., "Reliability Calculations and Above-the-MUF propagation", *Proceedings of the Ionospheric Effects Symposium, IES-93,* Washington, USA, 1993.

10. Friedrich M., K.M. Torkar, "Typical behaviour of the high latitude lower ionosphere", *Adv. Space Res.* 16, 1, 1995, pp 73-81.

11. Clutterbuck C., Defence Resarch Agency, Malvern, UK, -private communication, 1995.

DISCUSSION

Discussor's name: G. S. Brown

Comment/Question:

Have you thought about painting the roadway to reduce the contrast between it and the netting? It is the contrast that the operators seem to be keying on.

Author/Presenter's reply:

You are right, the good driving performance of the test persons is caused by the contrast between net and roadway, especially for the lowest picture quality. In answer to your question relating to the painting of the roadway, we will not go in that direction in the future. In addition, we think that the light endless course is quite easy to drive. Our attempts are first to go with the vehicle in the field and secondly to drive a simulated vehicle in a simulated world. The goal will be to get familiar with the new picture compression techniques on the base of motion - JPEG in real time, and the related theory.

Discussor's name: P. S. Cannon

Comment/Question:

Absorption is more of a problem to European auroral ionosondes than it is to North American auroral ionosondes. This may, therefore, indicate that absorption will be more of a problem on oblique European paths. Your model seems to provide a substantial improvement in Europe. Given the possible difference between Europe and North America do you have any plans to test your new model in North America (or any other location)?

Author/Presenter's reply:

I was not aware of the difference between ionospheric absorption in Europe and in North America. Assuming that the spectrum of precipitating particles is the same in North America and in Europe, the electron density profiles will be the same, and the model can be applied independent of longitude. The statistics of riometer absorption in North America will cause the model to calculate less absorption than on European paths. It could certainly be an interesting exercise to compare the model with a North American data base.

Discussor's name: N. Farsaris

Comment/Question:

What about the M(3000) propagation factor (or any other factor considered) in high latiitudes?

Author/Presenter's reply:

The M(3000) factor has not been studied here. We assume that the high latitude M(3000) factor has been properly modelled in the ICED-model which is a part of the ICEPAC prediction program.

PASSIVE HF PROPAGATION EVALUATION TECHNIQUE

Haim Soicher
US Army Communications-Electronic Command
Space and Terrestrial Communications Directorate
Fort Monmouth, NJ 07703-5203

Zwi Houminer
Asher Space Research Institute
Technion-Israel Institute of Technology
Haifa, 32000, Israel

ABSTRACT

Reliable communications among mobile as well as fixed HF skywave nodes along short, medium and long-range paths require propagation assessment. Such assessment could be facilitated with the monitoring of ionospheric characteristics by continuously available passive means i.e., measurement of the total electron content (TEC) using satellite-emitted signals without a need for burdening the electromagnetic spectrum. With the availability of the ubiquitous Global Positioning System (GPS) to provide instantaneous time-delay, or equivalently TEC, values when needed, an assessment of HF propagation conditions may be available on a near-real-time basis.

Both TEC and the peak electron density of the ionosphere, which determines the ordinary upper frequency limit (f_0F_2) for HF skywave vertical propagation, vary strongly with solar and geomagnetic parameters. Their ratio, the equivalent slab thickness, may vary to a much lesser degree and as such may be modeled with greater accuracy. A slab thickness model combined with real-time TEC measurement anywhere on the globe may thus yield an improved HF parameter prediction algorithm.

To test the efficacy of the hypothesis, one has to ascertain the correlation between the TEC daily variability about the monthly mean and the f_0F_2 variability. To determine such correlation a study was conducted using Faraday TEC data as well as GPS generated TEC data collected in Israel and comparing it to f_0F_2 measurements near the appropriate subionospheric location in Cyprus. The analysis shows that for large percentages of the time very good correlation exists between TEC and f_0F_2 short-term variations. The correlation coefficient varies between 0.7 or better during winter and summer months to about 0.5 - 0.6 during equinox months. A study of the diurnal dependence of the correlation indicates that a better correlation exists during daytime than nighttime. There was no indication that the coefficient is dependent on geomagnetic activity during the period of this study.

INTRODUCTION

HF radio communication depends on the ability of the ionosphere to return the radio signal incident on it back to earth. Prediction of ionization levels in the various ionospheric regions are derived from models and are used as a basis for planning and frequency management of HF radio systems worldwide. The models permit the calculation of system parameters such as operating frequencies, signal strengths, signal-to-noise ratios and multipath probability that can be used to describe the performance of HF radio systems. Uncertainties or inaccuracies in the models of the ionosphere have long been known to be one of the major, if not the major, cause for inaccuracies in the calculated propagation characteristics. This is particularly true for those applications of ionospheric predictions involving time scales that are less than the monthly average or monthly median.

To reduce average monthly RMS errors in predictions, adaptive techniques that use real-time observations to correct model biases have been devised. Examples of such are sounding the desired communication path prior to information transmittal or sounding the vertical local ionosphere to determine a real-time model reference point and then to adjust the model for the future until the next reference sounding. While the first is appropriate for fixed point communication paths it is cumbersome and causes EM interference over wide geographic areas, whereas the second technique, while not causing interference, may not yield values representative of ionospheric conditions at the reflection point of an oblique path.

A method that is potentially global in nature involves the monitoring of satellite-emitted signals which yields information on the ionospheric parameters along the propagation path and converting such information into the HF propagation parameters of interest. The advantage of monitoring satellite-emitted signals is the fact that it is passive for the potential user, and the existence of a global network of satellites (e.g. GPS) affords the possibility of global coverage. The problem at hand is the conversion of integrated ionospheric parameters along the transionospheric path experienced by the satellite-emitted signal to the parameters along the path experienced by the HF skywave up to the point of reflection from the ionospheric layers. A parameter of great importance in HF propagation is f_0F_2, the upper frequency limit for HF vertical propagation, whose square is proportional to N_{max}, the maximum electron density in the ionosphere. The transionospheric parameter of importance is the TEC which is the integrated electron density along the propagation path of a satellite-emitted signal to the observer ($TEC = \int_o^s N ds$ where N is the electron density and ds is an element of distance along the path from observer (o) to satellite(S)). Since both TEC and N_{max} vary diurnally, seasonally, geographically and in response to magnetic activity, it is expected that their ratio - the so called slab thickness, τ (ref. 1) - will vary to a lesser degree and hence can be modeled more easily. Global models of slab thickness updated with real- time measurements of TEC (e.g., by using the GPS network) might thus yield improved values of N_{max} and hence f_0F_2 for HF propagation.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

THE DATA

To assess the day-to-day variability of slab thickness, it is of great interest to ascertain whether good correlation exists between TEC daily variability about the monthly mean and f_0F_2 variability. To determine such correlation, a pilot study was conducted using several months of TEC data taken in Haifa, Israel during 1980, 1981 (ref. 2), as well as GPS time-delay measurement taken during the summers of 1992, 1994 in Jerusalem, Israel. The corresponding f_0F_2 measurements were from Cape Zevgari, Cyprus. The TEC was determined from Faraday rotation observations using the signal of the Sirio satellite. The geographic subionospheric point corresponding to a mean ionospheric height of 350km is 30.3°N and 28.9°E. The geographic coordinates of Cape Zevgari are 34.6°N and 32.9°E. The measurements are thus separated by 4.3° in latitude and 4° in longitude. The GPS observations were taken at the National Physical Laboratory in Jerusalem using a Ionospheric Measurement System developed at the National Institute of Standards and Technology (NIST) at Boulder, Colorado for accurate time transfer. This system provides 15-minute averages of time-delay data to all GPS satellites above the horizon. The equivalent TEC from the time-delay measurements was determined only for satellites at elevations larger than 30° and having a subionospheric point along the line of sight within 5° of the latitude and longitude of the ionosonde. The observed time delays were then corrected for satellite biases using the JPL table of corrections (ref. 3), converted into vertical TEC (ref. 4) and averaged to obtain hourly values.

EXPERIMENTAL RESULTS

Figure 1 shows the hourly values of the variability of f_0F_2 and TEC for the period 2 - 31 July 1981. That period is near the maximum of solar cycle 21 (monthly mean sunspot number 155). The variability is determined by subtracting the monthly average value from each hourly value and dividing by the monthly average value. The results of cross correlation analysis on the f_0F_2 and TEC depicted in Figure 1 are shown in Figure 2. A maximum cross correlation coefficient of 0.73 occurs at zero time lag while the correlation reduces very quickly with time lag. It is thus shown that the correlation between f_0F_2 and TEC is very good. Autocorrelation results for both f_0F_2 and TEC for the same time period (Figure 3) shows that the autocorrelation is similar in character to the cross correlation function and drops to 0.5 after a time lag of 3 - 4 hours. This is consistent with results obtained elsewhere indicating that short-term predictions based on real-time observations can only be useful for a few hours. Figure 4 shows the hourly values of the variability of f_0F_2 and TEC for the period 2 - 31 July 1992. That period is near the maximum of solar cycle 22 (monthly mean sunspot number 91). The results of cross correlation analysis on the f_0F_2 and TEC depicted in Figure 4 are shown in Figure 5. A maximum cross-correlation coefficient of 0.778 occurs at zero time lag while the correlation reduces very quickly with time lag. Again the correlation between f_0F_2 and TEC is very good. Auto-correlation results for both f_0F_2 and TEC for the same time period (Figure 6) shows that the auto-correlation is similar in character to the cross correlation function and drops to 0.5 after a time lag of 3 - 4 hours.

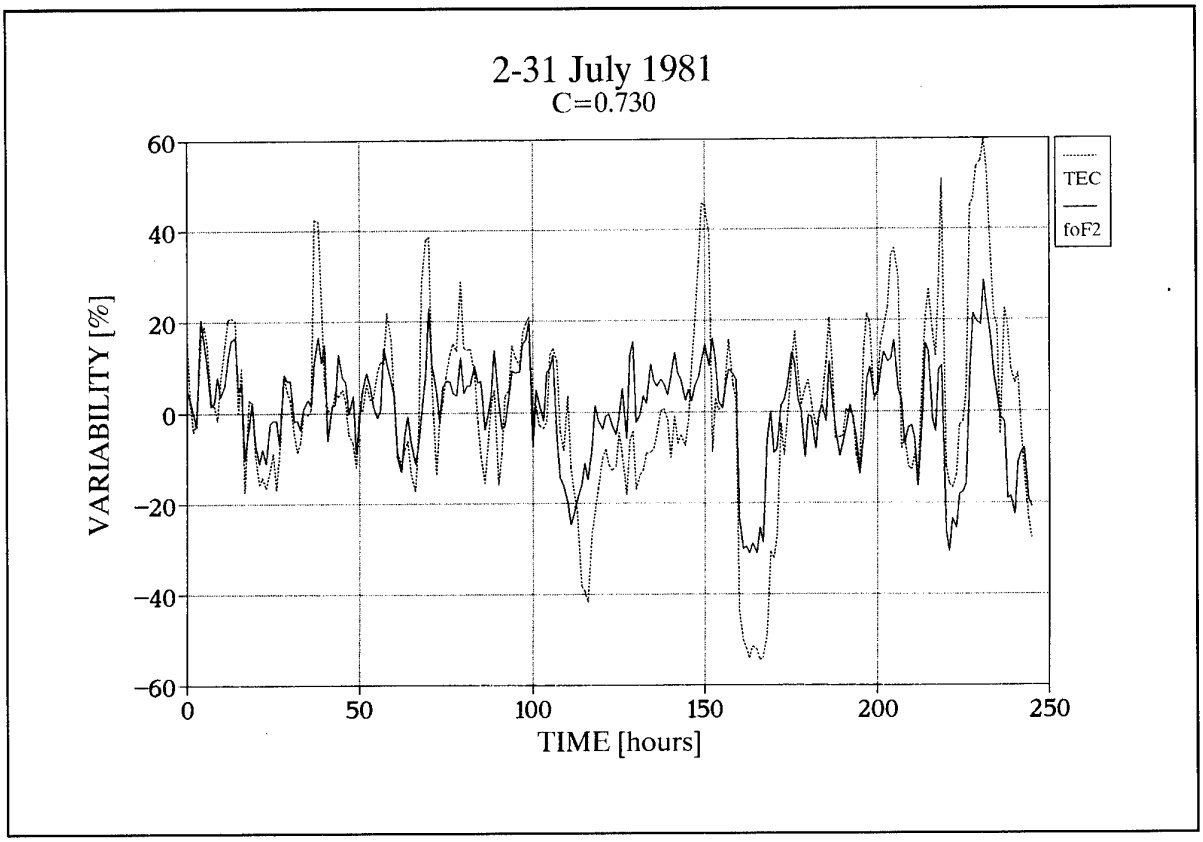


FIGURE 1

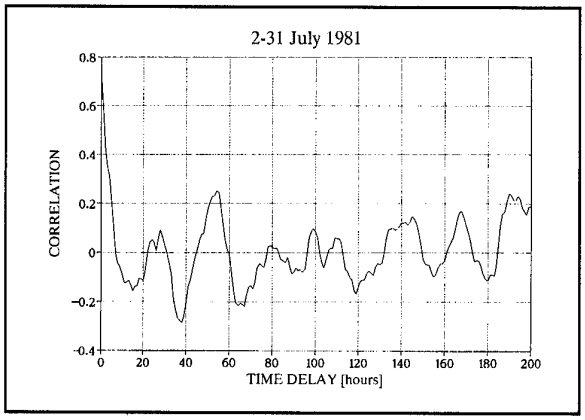


FIGURE 2

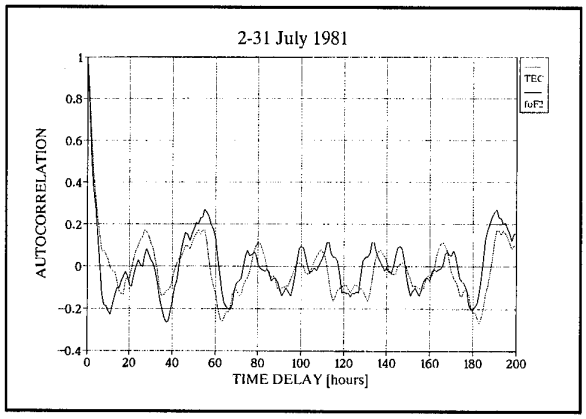


FIGURE 3

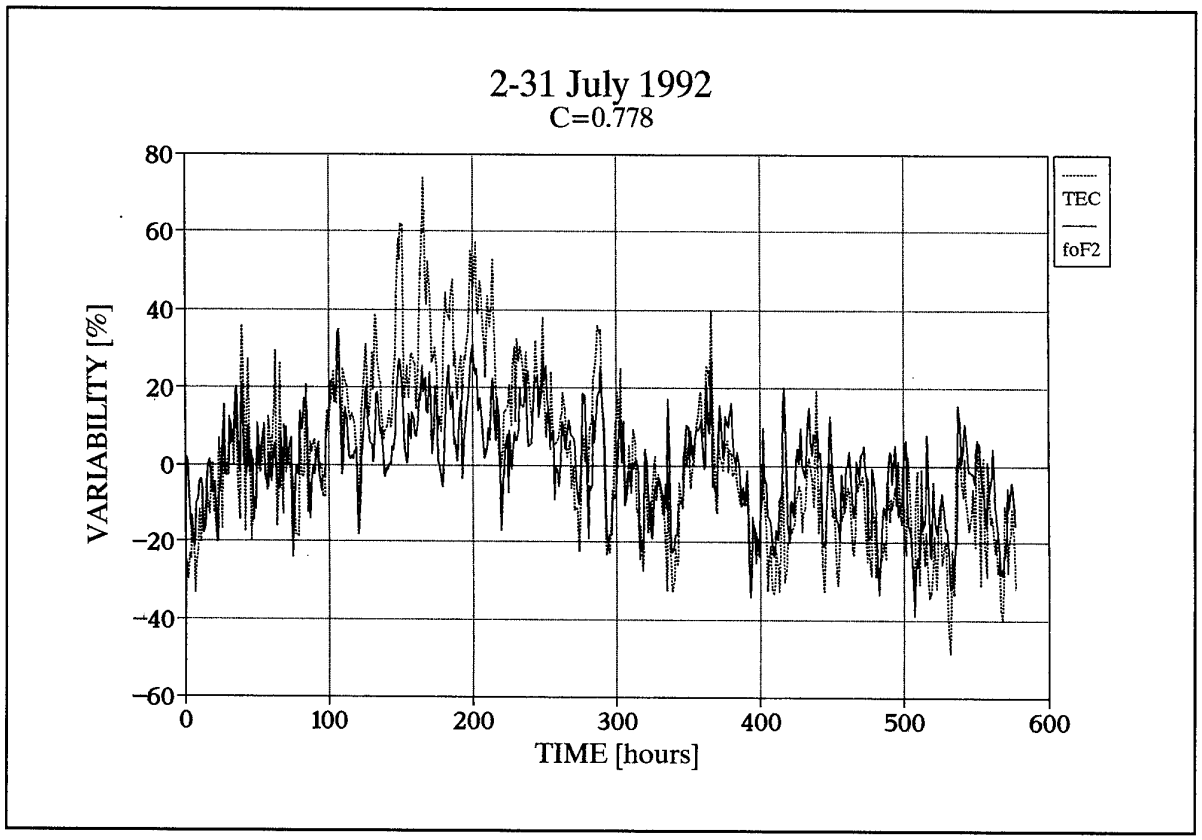


FIGURE 4

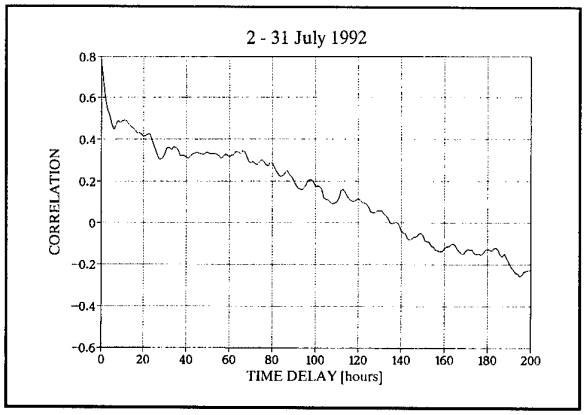


FIGURE 5

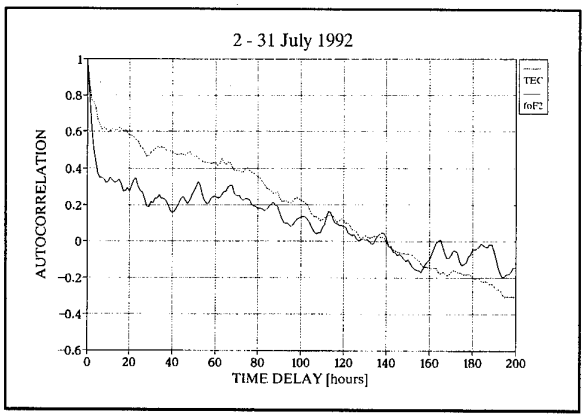


FIGURE 6

The rapid fluctuations of the variability (as compared to the 1981 data) are probably due to data noise which depends on the motion and choice of a particular GPS satellite for the analysis, rather than due to actual physical phenomena. The results of smoothing the variations in TEC and f_0F_2 by a 3-hour running mean is shown in Figure 7. It can be seen that the correlation is remarkably improved (correlation coefficient of 0.835). This shows that, indeed, the data are noisy and that the correlation between TEC and f_0F_2 is actually better than what the raw data indicate.

The variabilities of the equivalent slab thickness τ (=TEC/N_{max}) for the above two time periods are shown in Figures 8, 9 respectively. It is seen that the slab thickness variabilities are of the order of the variabilities of its constituent parameters variabilities. This is most likely due to the separation, and its change, of the satellite subionospheric point with respect to the probing sounder.

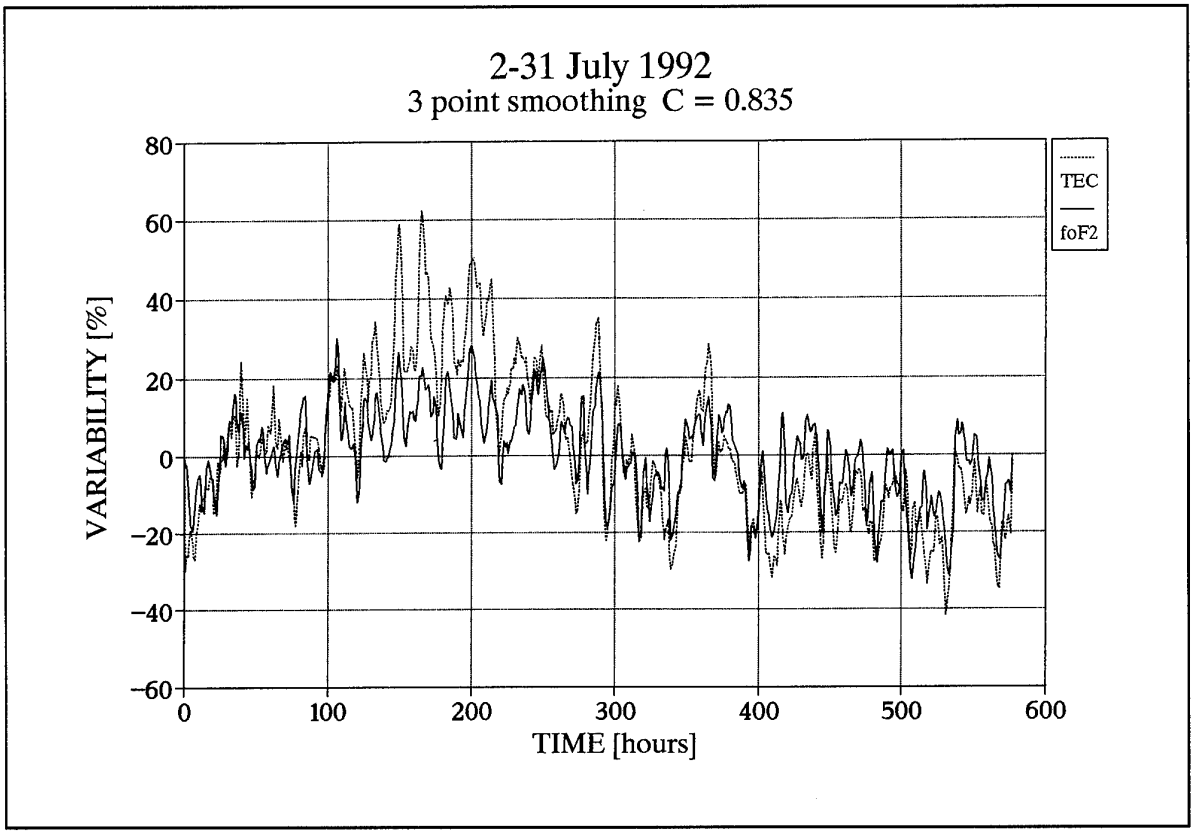


FIGURE 7

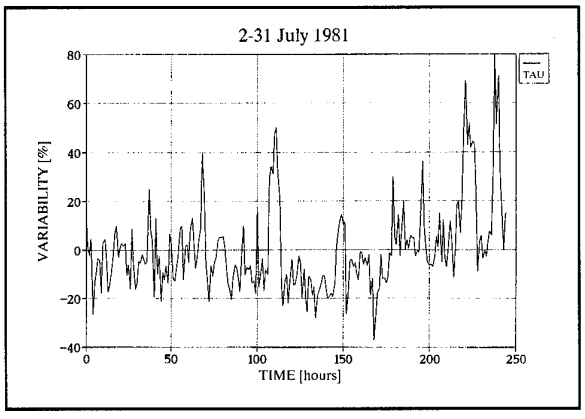


FIGURE 8

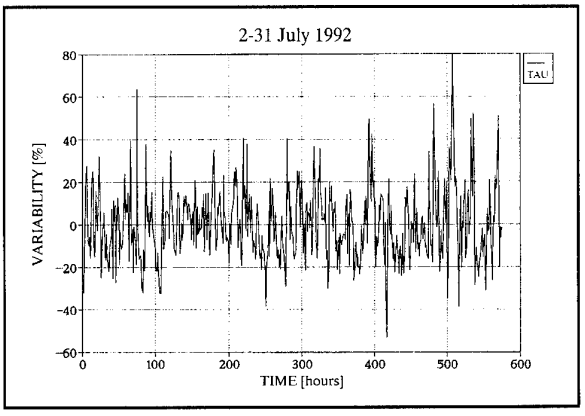


FIGURE 9

The seasonal dependence of the correlation coefficient between the variability in TEC and f_0F_2 is shown in Table 1. It can be seen that the correlation coefficient is better than 0.7 in winter and summer. However, during equinox, especially in April and May 1980, the correlation drops to about 0.5. The reason for the low correlation is the large post sunset enhancements in TEC which were not seen in f_0F_2. These post sunset enhancements, observed mainly in equinox months, are attributed to electron fluxes arriving from the equatorial regions along the magnetic lines of force. They are latitude-dependent (ref. 5) and did not affect the f_0F_2 observations.

SEASON	MONTH	SSN	R
Winter	Jan 80	163.9	0.74
	Feb 80	162.6	0.73
Equinox	Mar 80	160.9	0.66
	Apr 80	158.7	0.50
	May 80	156.3	0.47
Summer	Jun 92	97.1	0.67
	Jun 94	32.0	0.65
	Jul 81	140.3	0.73
	Jul 92	90.7	0.78
	Aug 92	84.0	0.67

TABLE 1

The diurnal dependence of the correlation in 1980 - 81 is shown in Table 2. It can be seen that higher correlation occurs during daytime than nighttime. Between 10 - 18 hours local time, the correlation coefficient is about 0.7; while between 22 - 06 hours local time, the correlation drops to between 0.55 - 0.6.

HOURS LT	CORRELATION COEFFICEINT
02-06	0.55
06-10	0.63
10-14	0.70
14-18	0.71
18-22	0.60
22-02	0.61

TABLE 2

The ionosphere is known to vary substantially with magnetic activity (ref. 1). To ascertain whether magnetic activity has any impact on the cross correlation of f_0F_2 and TEC, two time periods – one quiet (18 - 26 January 1980) and one active (14 - 19 February 1980) – were examined. The former with index Ap<6 is depicted in Figure 10 with maximum correlation coefficient of 0.823 and the latter with index Ap<40 is depicted in Figure 11 with maximum correlation coefficient 0.774. Thus it appears that magnetic changes do not have a marked impact on the correlation of the two ionospheric parameters.

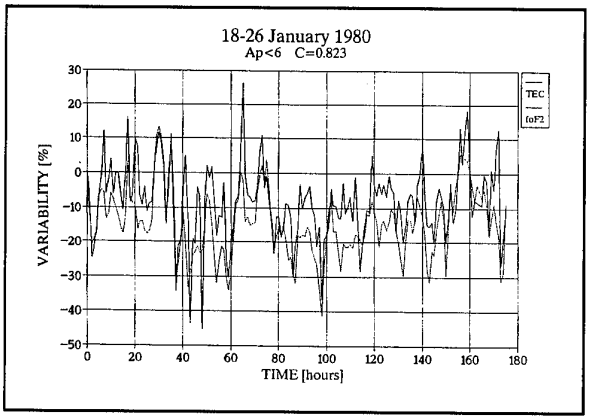


FIGURE 10

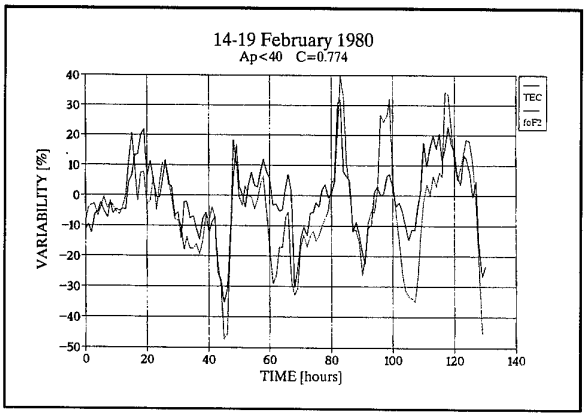


FIGURE 11

The GPS generated time-delay method of calculating TEC measures the columnar integrated electron density from observer to the satellite, i.e., the ionospheric (to about 1500 km attitude) contribution as well as the protonospheric (from ~ 1500 km to the height of the satellite). The Faraday rotation method, which is heavily weighted near the earth, measures only the ionospheric contribution. The protonospheric electron content varies diurnally very little and is often assumed to be ~.15 of the daytime total electron content. The ratio of the protonospheric content to the total content at night is quite large and it often maximizes at >.5 of the total content just prior to dawn (ref. 6). Klobuchar et al (ref. 7) claim that for geomagnetic L-shells greater than 4, the geomagnetic lines are considered open and consequently do not contain any protonospheric electrons. For L-shells less than 4, the protonospheric electrons exist and must be accounted for. In the present case where Faraday data are compared to GPS-generated data, it is important to ascertain the protonospheric impact. While the GPS data has not been separated for L-shell distribution, a comparison of the correlations for the Faraday and the GPS-generated TEC data was done (Figure 12). At night the correlation for the GPS-generated TEC vs f_0F_2 was slightly higher than that of the Faraday TEC, but at dawn – when the maximum impact of the protonospheric content occurs – the correlations are similar. Thus, it appears that the protonospheric content has little influence on the variabilities of TEC and f_0F_2 – at least during solar maximum, when the ionospheric contribution is much higher than the protonospheric one.

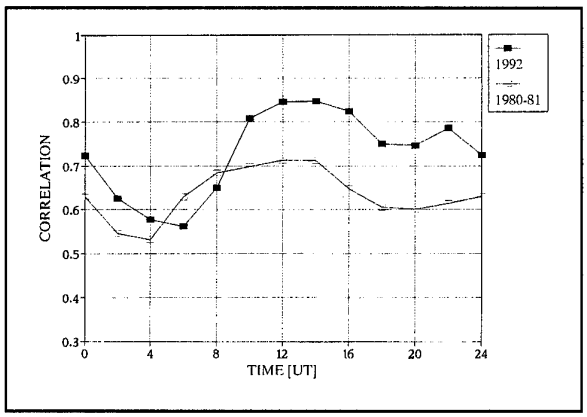


FIGURE 12

CONCLUSIONS

The high cross-correlation (>0.7) coefficient for f_0F_2 and TEC in the limited data presented in this paper raises the possibility that real-time TEC measurements may be used to update f_0F_2 value determinations. The cross-correlation may even be higher if the geographic subionospheric point of the TEC measurement is closer to the geographic point of the f_0F_2 measurement, which introduces an error, in addition to the possible inherent measurement uncertainties. TEC measurements utilizing satellite emitted signals are passive in nature and do not burden the electromagnetic spectrum. In addition, the availability of the global GPS constellation to provide instantaneous time-delay, or equivalently TEC, values could provide an instantaneous updating of f_0F_2 models on a global basis as well as on a regional basis. Such capability is important for HF communication along short, medium and long range paths.

REFERENCES

1. Goodman, J. M., HF Communication Science and Technology, Van Nostrand Reinhold, New York, NY, pp631, 1992.

2. Soicher, H., Z. Houminer and A. Shuval, Total Electron content in the Middle East, Radio Science, 17, 1623 - 1631, 1982.

3. Wilson, B. D. and A. J. Mannucci, Instrumental Biases in Ionospheric Measurements Derived from GPS Data, Proceedings of the Institute of Navigation GPS-93, Salt Lake City, Utah, September 1993.

4. Klobuchar, J. A., Ionospheric Time-Delay Algorithm for Single Frequency GPS Users, IEEE Trans AES, AES-23, 325 - 331, 1987.

5. Soicher, H., J. A. Klobuchar and P. H. Doherty, Spatial Variability of Total Electron Content in the Eastern Mediterranean Region, Radio Science, 19, 757-764, 1984.

6. Soicher, H., Ionospheric and Plasmaspheric Effects in Satellite Navigation Systems, IEEE Trans. A&P, 25, 5 , 705 - 708, 1977.

7. Klobuchar, J. A., P. H. Doherty, G. J. Bailey and K. Davies, Limitations in Determining Absolute Total Electron Content from Dual Frequency GPS Group Delay Measurements, Proceedings-International Beacon Satellite Symposium, Ed. L. Kersley, Univ of Wales, 1 - 4 , 11 - 15 July 1994.

FIGURE CAPTIONS

1. Hourly values of the variability of f_0F_2 (at Cyprus) and TEC (at Israel) for the time period 2 - 31 July 1981.

2. Cross-correlation function for the f_0F_2 and TEC values of Figure 1.

3. Auto-correlation functions for the f_0F_2 and TEC values of Figure 1.

4. Hourly values of the variability of f_0F_2 (at Cyprus) and TEC (at Israel) for the time period 2 - 31 July 1992.

5. Cross-correlation function for the f_0F_2 and TEC values of Figure 4.

6. Auto-correlation functions for the f_0F_2 and TEC values of Figure 4.

7. Three hour smoothing of the values of Figure 4.

8. Hourly values of the variability of τ for the period 2 - 31 July 1981.

9. Hourly values of he variability of τ for the period 2 - 31 July 1992.

10. The variability (in percent) of f_0F_2 and TEC for the geomagnetically quiet period of 18 - 26 January 1980.

11. The variability (in percent) of f_0F_2 and TEC for the geo magnetically active period of 14 - 19 February 1980.

12. Diurnal variation of the correlation for the f_oF_2 and TEC values for two periods – 1980-81 (only Faraday Data) and 1992 (GPS Data). [Each point represents correlation results for a four-hour period.`]

DISCUSSION

Discussor's name: C. Goutelard

Comment/Question:

Votre méthode est intéressante, notamment parce qu'elle ne dépend pas des émissions d'utilisateurs.

Avez-vous en projet d'étendre vos études auz zones polaires et aurorales et aux zones equatoriales?

Translation

Yours is an interesting method, in particular since it does not depend on user transmissions.

Do you intend to study the auroral, polar, equatorial regions?

Author/Presenter's reply:

The experiment will be extended to Central Europe where sounder facilities are available to us (see Soicher et al. RADIO SCIENCE, 1995). Other researchers/institutions are encouraged to carry out the studies in the high latitude and equatorial regions.

Discussor's name: P. Cannon

Comment/Question:

The cross and autocorrelation function for Faraday measurements of TEC variability show a periodicity which appears to be close to a day. The GPS based measurements do not show this periodicity. Can you explain this?

Author/Presenter's reply:

The Faraday rotation TEC/foF2 variability is centred about 0% variability for the overall time period considered (see Figure 1), whereas the GPS-generated TEC/foF2 variability is not, but itself exhibits a periodic behaviour (see Figure 4). This may be the cause for the different behaviour of the correlation/autocorrelation curves in the two cases.

DISCUSSION

Discussor's name: D. Yavuz

Comment:

It is also possible to have a very effective passive HF propagation prediction system by monitoring thousands of transmissions that can be located quite accurately; various broadcast, teletype transmissions and AX.25 packet radio transmissions. We at STC find this a very effective tool for setting up HF links quickly. The system can be put together with a database of transmitter locations and frequencies and a scanning receiver and some simple software. One could add map displays and various other niceties but we have not yet included these.

A Ray-based Propagation Tool for Digital Systems

P. Charrière and K.H. Craig
Radio Communications Research Unit
Rutherford Appleton Laboratory
Chilton, Didcot, OX11 0QX
U.K.

SUMMARY

The paper describes the development of a practical, deterministic ray-based propagation tool for the prediction of the performance of wideband digital systems. Applications over a wide frequency range from VHF to millimetric waves are considered, with emphasis on propagation in the urban environment.

The discussion covers the propagation mechanisms, database and computational issues, system applications and model validation.

1 INTRODUCTION

The increasing demand for digital services and higher traffic capacities has resulted in a requirement for larger bandwidths, and a move to higher frequencies. The use of digital systems in an urban environment is problematic due to the multipath nature of the channel caused by scatter from buildings and other man-made and natural obstacles.

In many countries, civil digital Personal Communication Systems (PCS) have developed considerably in the last few years. This has resulted in much effort being expended on developing models for service area and interference prediction (particularly important as the density of systems increases). These PCS use frequencies in the UHF range (for the mobile system) and in the millimetric range (for the high capacity backbone connections).

Statistical methods are of limited value at millimetric frequencies since the siting of individual buildings or hilly terrain can produce strong shielding or provide a source of high reflected field levels which could affect the communication channel. In such conditions, a propagation tool needs to be deterministic, based on detailed information concerning the terrain and/or building databases.

The propagation mechanisms of interest in the urban or cluttered rural environment are:

- scatter (specular and non-specular);
- diffraction (at low frequencies);
- penetration through obstacles such as buildings.

Any propagation tool used must support these mechanisms and also:

- be three-dimensional;
- be computationally efficient (since a practical tool will be required to make area coverage predictions);
- have a wideband capability (for modelling of multipath effects on digital systems).

Two fundamentally different approaches have been considered for this problem. Full-wave parabolic equation (PE) models have been used for several years to model the effects of refraction [1] and diffraction by terrain [2]. More recently, the PE has been applied to the building scatter/diffraction problem, both in two and three dimensions [3]. However PE development for urban propagation problems is still at an early stage. The PE is intrinsically a forward scatter method, and modelling of high angle reflections and non-specular effects are not straightforward. Computation times are still high, although much effort has gone into improving this in recent years. In particular, although wideband effects can be modelled [4], this is done by an explicit calculation of the channel transfer function in the frequency domain, multiplying computation times by a factor corresponding to the number of frequency components of interest. One of the biggest advantages of PE methods is that they can deal in a uniform way with the effects of the atmosphere as well as diffraction from terrain and terrain cover.

An alternative approach is the use of ray methods. This is by now a well-established technique for modelling the effects of multipath. In the last few years, several groups have been developing tools for urban propagation based on ray methods. Most of these have been aimed at the mobile communications application in the UHF band, and there are now commercial packages available [5].

Ray tracing is very simple in principle. All ray paths connecting transmitter and a receiver location taking account of the relevant propagation mechanisms between these are found, and the parameters of interest (for example, field strength, delay spread, channel transfer function), found by suitably combining the ray parameters of all the rays at the receiver.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

The task then is:

- to develop a computationally efficient algorithm for finding all the ray paths;
- to decide the kind of "rays" required to model the physical mechanisms required for the application.

The simplest ray trace model will include only specularly-reflected rays, whose properties are governed by the laws of geometrical optics (GO). At frequencies for which diffraction is important, edge diffracted rays must be included, whose properties are governed by the laws of the geometrical theory of diffraction (GTD) or more precisely, the uniform theory of diffraction. Most existing mobile communications urban propagation tools at UHF are based on GTD methods. The modelling of non-specular scatter requires the incorporation of rays that do not follow GO/GTD paths, and a suitable model for the non-specular reflection process. At lower VHF frequencies where diffraction effects are important, there is an alternative to GTD methods: this is to model the reflections according to the laws of physical optics (PO), and to represent the propagation between reflectors by means of (non-specular) rays connecting the phase centre of the reflecting facets.

This paper discusses some of these issues. The propagation tool was developed with two, very different, applications in mind. The first was point-to-point (wideband) digital links at 38 GHz in an environment where building reflections and obstructions dominated. The second was an air-to-ground application at VHF where multipath from buildings and terrain dominated. In both cases, an efficient ray trace "engine" was used to calculate the ray paths. The millimetre wave application used a GO specular reflection model, while the VHF application used a PO non-specular reflection model.

Any deterministic urban propagation model requires a building database. The availability of these varies from country to country, but satellite and air survey methods are improving the availability, and reducing the costs.

The main interest here is the application to digital systems. Some examples of the types of model "product", and subsequent processing of the propagation results, are given.

2 THE MODEL

2.1 The Ray Trace "Engine"

The model is based on a three-dimensional ray-tracing technique. All geometrical ray paths connecting transmitter and receiver are found, up to a given order of reflection. (We do not consider here a "shooting" method of ray tracing, where rays are emitted from the transmitter at predefined separations over a given solid angle: in a highly reflection-dominated geometry, it is extremely difficult to ensure that all relevant rays are included in the field computation at the receiver).

While the basic techniques are well known, the computational complexity of their naïve implementation is very high, since 3D building databases with several hundreds of facets are required, and reflections up to 5th or 6th order may be necessary for prediction of field strength in the "shadow" of a building. A straightforward implementation of a ray search would have a computation time proportional to T_1:

$$T_1 = n\sum_{i=1}^{i=d}(n-1)^{i-1}$$

where n is the number of facets included in the database and d is the order of reflection. The run time increases exponentially with the order of reflection!

In this implementation, we use an "image" method to represent specular reflections and diffraction: each interaction with a reflecting or diffracting surface is replaced by a virtual image whose characteristics model both the geometry and the electrical properties of the object. Within this framework, the computational complexity is greatly improved by pre-calculating a connectivity matrix for the images. In the "average" urban situation, building obstruction ensures that the matrix is sparse, and the run time is now proportional to T_2:

$$T_2 = n\sum_{i=1}^{i=d} f(i,n)$$

Here $f(i,n)$ is a decreasing function of i which tends toward a positive constant value as i increases and becomes independent of n for large n. The computation is still intensive but now the run time tends to be proportional to the order of reflection d instead of exponential.

The images required depend on the physical interaction to be modelled.

2.1.1 Specular reflections

This is the simplest case. Each reflecting surface is replaced by a single optical image whose location is determined by GO and whose (complex) amplitude is determined by the electrical properties of the surface. Since the specular loss on reflection at a building surface is of the order of 10 dB at millimetric wavelengths [6], a practical model for digital system performance does not normally need to include rays beyond second order reflection.

2.1.2 Diffraction

Edge and corner diffraction at the corner of a building or at the edge of a roof can be modelled by a set of images whose locations and amplitudes are determined by GTD.

The use of diffraction in the propagation model involves a much higher computational cost (since many more images are required, and the connectivity matrix is less sparse). However, the effects of diffraction decrease as the frequency increases, and in our application to 38 GHz point-to-point links, its effects are much less significant than the effects of direct and/or reflected

power. The practical effects on digital system performance are usually negligible compared to reflected multipath, and in such conditions one can neglect diffraction, or only consider low order effects (one reflection plus one diffraction at most).

2.1.3 Building penetration

Practical observations show that building penetration can be a dominant propagation mode and in some cases modelling its effects can be necessary even if it is expensive in term of computation. Each surface then requires two images—one for the reflected ray and one for the transmitted ray. It is often difficult to predict the effects of building penetration because the materials used in building construction are not always very well known and the way the inside of a building is furnished is very variable as well. The impact on the computational workload is also extremely significant: every facet becomes connected to every other facet (i.e. the connectivity matrix is dense), and other *ad hoc* methods must be devised to keep the problem tractable.

2.2 The Database

The 3D database contains information on the studied site which is stored in the form of a list of facets. The simplest representation of many buildings would be as a cuboid defined by 5 facets. Similarly, facets can be used for the representation of the ground.

Every facet includes information on:

- its dimensions and position.
- its electrical properties.
- its surface roughness.

In an urban situation, because of the mostly cuboidal shape of buildings, one can use rectangular reflectors to give an acceptable description of the site of interest; for other type of terrain triangular shaped reflectors should be used in order to describe the area topology, although other shapes may be more convenient for propagation models (for example in the PO approach described below).

A good balance must be found between a high precision database for high accuracy prediction of radio wave propagation on the one hand, and the computational cost of such a prediction and the availability of the necessary databases on the other. Further discussion is found in [7].

From a practical point of view, site databases are often unavailable. One way of proceeding is to start from a 2D digital database and then to carry out an additional survey to get information on building heights, material used for the construction, etc. An example of such a database is shown in Figure 1. This was created for architectural purposes [8] and contains much more detail than is necessary for propagation models, but does show the kind of database that can be created. Figure 2 gives an idea on the degree of detail used in simulation at RAL.

Alternatively, some companies now provide survey services for urban areas using techniques such as stereographic photography to establish an accurate picture of a site. However, these services are still expensive and furthermore don't provide much information on the electrical properties of building surfaces.

2.3 The Propagation Model

Once all the paths connecting the receiver (Rx) and the transmitter (Tx) have been found, the energy reaching the Rx from the Tx can be calculated from the (coherent) sum of the energy travelling along each ray.

In a ray tracing method the direction and weighting of each contribution to the overall electromagnetic field for a given position is directly available. This can be very useful for identifying any potential source of interference or to propose multipath countermeasures in radar or communications applications.

The contribution of each ray is handled as a complex voltage value according to the phase and magnitude of the signal coming from this ray. The total signal reaching the Rx is therefore equal to the complex sum of each ray's contribution.

Because of the uncertainty of the position of buildings in the database compared to the wavelength ($\lambda \approx 8$ mm for the model developed at RAL), and more generally because of the limited level of detail contained in the database compared to reality, the absolute phase of each ray component will not generally be predictable (although the relative phase of rays at two nearby locations will be). Therefore it is not realistic to expect the tool to give the exact value of received power at a given location. Of more importance, in practice, will be the prediction of quantities such as:

- the mean field in an area;
- the variability of the received signal with position;
- the range of delays of the multipath signals;
- the main directions of arrival of the signals;
- the transfer function over the channel bandwidth.

The ray parameters can be used to obtain these quantities. For example, the incoherent power is obtained by an RMS sum of the magnitudes of the contributions, while the RMS delay spread is obtained from calculating the second central moment of the power delay profile. These derived ray-quantities can then be used to assess the quality of a digital communication channel.

The rest of this section describes some of the propagation models used in the current tool. Both the millimetric and the VHF applications are covered. At present, diffraction is not included in the millimetric application since it generally has little practical effect; at VHF diffraction is accounted for by means of physical optics. Building penetration is also not included because of the absence of experimentally measured parameters.

A measurement programme is underway at RAL to get a better understanding of the mechanisms involved, and to provide driving parameters for the prediction tool. Some preliminary results from this measurement campaign are presented later in this paper.

2.3.1 The free space model.
The free space transmission model can be found in numerous references, and is usually given with transmit and receive antenna gain, as follows:

$$P_r = \frac{P_t G_t G_r \lambda^2}{16\pi^2 d^2}$$

where
P_r is the power at the terminal of the receive antenna;
P_t is the power delivered to the transmit antenna;
G_t and G_r are the gains of the transmit and receive antennas;
λ is the free space wavelength;
d is the path length.

2.3.2 The specular reflection model
Once again, the reflection model is a standard one, and can be found in [9]. The loss due to a reflection is modelled by:

$$R = R_0 . \rho$$

where
R_0 is the reflection coefficient on smooth surface.
ρ is the loss due to surface roughness.

The smooth surface reflection coefficient is given by Fresnel's formulae:

$$R_0 = \frac{\sin\varphi - \sqrt{C}}{\sin\varphi + \sqrt{C}}$$

Where φ is the grazing angle and

$C = \eta - \cos^2\varphi$ for horizontal polarisation

$C = (\eta - \cos^2\varphi)/\eta^2$ for vertical polarisation

with $\eta = \varepsilon_r(f) - j60\lambda\sigma(f)$

and
$\varepsilon_r(f)$ is the relative permittivity of the surface at frequency f;
$\sigma(f)$ is the conductivity (S/m) of the surface at frequency f.

At 38 GHz, the surface roughness of a building wall is very often significant compared to the wavelength ($\lambda \approx$ 7.9 mm) and loss due to surface roughness during a (specular) reflection has to be taken into account. Surface roughness is characterised by the Rayleigh roughness criterion:

$$g = 4\pi(S_h / \lambda)\sin\varphi$$

where
S_h is the standard deviation of the surface height about its mean value;
λ is the free space wavelength;
φ is the grazing angle.

The rough surface loss is then approximated by:

$$\rho = \frac{1}{\sqrt{3.2x - 2 + \sqrt{(3.2)^2 - 7x + 9}}} \quad \text{where } x = 0.5g^2$$

2.3.3 Building penetration model
A simple and very empirical model is proposed for building penetration:

$$\rho = Kd$$

where ρ is the normalised loss, K is a coefficient depending on the material used for the construction of the building and d is the length of the path within the building. A more accurate model that takes into account the wavelength, the dispersiveness of the medium and the incidence of the power on the building may turn out to be necessary, and a more thorough investigation of this mechanism is awaited.

2.3.4 Physical optics
At VHF frequencies, the wavelength is comparable to the size of buildings and diffracted signals can be significant; a pure GO/specular reflection model is inadequate in this case. Most existing models use GTD to represent the effects of diffraction. An alternative description of the interaction of electromagnetic waves by a finite sized reflector is provided by physical optics, and this is the approach used in this implementation.

The reflection/diffraction process is modelled in two stages:
1. the field induced on the reflecting surface by the incident wave is calculated. This depends on the electrical properties of the surface; for a smooth planar surface, and uniform illumination by the source, the field amplitude can be assumed to be constant;
2. the field at a receiver point is obtained by integrating the surface field over the "aperture" of the reflecting surface, just as if the reflecting surface were a radiating antenna.

In PO the reflecting surface can be thought of as a radiating source, and normal antenna concepts can be applied to the reflecting surface. Thus the surface has an "antenna pattern", and we can distinguish between the near and far fields of the surface. For even illumination of a rectangular surface, the field can be calculated analytically in terms of Fresnel integrals.

The PO approach can be considered a ray tracing approach with "non-specular" reflections and "reflection coefficients" given by the antenna pattern. A major computational advantage is that the "ray" trajectories simply interconnect the centres of the reflectors; this can be done once and for all when the reflector geometry is known, *and does not depend on the transmitter or receiver positions.*

In contrast the GO/GTD image approach requires the ray paths to be calculated afresh for each new transmitter position; for a moderately complex

environment, the ray path calculation is time consuming and dominates the overall run time. A PO approach will thus be more efficient than GTD for applications where both terminals are mobile. As in the GO approach, angles of arrival of the multipath rays are obtained from the path geometry of the (non-specular) rays.

The task of finding all the ray paths in GTD is also significantly more expensive than in PO as each surface requires two edge-diffracted rays in addition to the specularly reflected (GO) ray. As before, a connectivity matrix can be used to check for obstructions on the ray paths between transmitter and receiver.

Figure 3 shows an example of the ray paths in such a method for studying the effects of the reflection of VHF signals from hilly terrain surrounding the receiving point, Rx. The field contribution from every facet is added to obtain the overall effect on the transmission channel.

2.4 Features of the Computer Implementation

The computer model has been developed as a practical tool. With this model one may perform a point-to-point study as shown in Figure 4, or an area coverage study as shown in Figure 5.

With this simulation tool the user can easily modify:

- the properties of any facets of the database;
- the position, orientation and polarisation of transmitter and receiver;
- the antenna pattern of either receive or transmit antenna by using one of the built in standard models or by loading the user's own model.

Furthermore, functions specific to digital communications (such as delay spread analysis) are also available to the user—Figure 6.

This model is intended to be used with a communication system development tool such as SPW [10]. Figures 7 to 9 give an idea of how the propagation model can be incorporated in the simulation of a mobile communications system. The standard statistical channel models (such as Rayleigh fading, or the Jakes model) is replaced by a site-specific channel model generated by the propagation tool. This for instance would give a better estimate of the effects of changing the location of transmitter and receiver or of changing the type of antenna.

3 VALIDATION OF THE MODEL

In order to validate the model predictions a simulation was made for a group of buildings on the RAL site, for which a suitable building database was available.

Figure 10 shows the result of this simulation, in the form of the possible ray paths overlaid on a map of the buildings. The model predicts that the power arriving at the receiver should come from two different directions represented by the last part of rays 1 and 2.

A series of measurements were made, giving a qualitative comparison. The walls involved here were mostly made out of brick. The transmitter and receiver were both at 1.5 m above ground and the antennas used for transmit and receive were standard gain horns with an azimuthal beam width of 17°. Figure 11 shows the signal actually measured at Rx, as a function of receive angle, for the configuration in Figure 10.

One can see some agreement between the angle of arrival of ray 2 of Figure 10 and the peak measured in zone B on Figure 11. However the absence of any incoming field at 0° due to the ray path 1 in Figure 10 seems to suggest that specular reflections by a brick wall of third order or higher are not significant. Furthermore, peaks were seen (at A and C in Figure11) which were not predicted by the ray trace model. These are believed to be due to first order non-specular reflections.

4 CONCLUDING REMARKS

The paper has described a programme of work to develop a practical propagation tool that can be used from VHF to millimetric frequencies. This has necessitated a re-evaluation of the practical significance of the various propagation mechanisms in different frequency bands, and a flexible approach to the modelling.

In addition to the modelling effort, RAL are making measurements to characterise the effects of buildings and vegetation on propagation at millimetre waves (see, for example [6]). There is a rapidly growing civilian use of frequencies around 38–42 and 60 GHz for digital point-to-point links and point-to-multipoint distribution systems. The results of these measurements will also improve the parameterisation of the models used in the propagation tool, which in turn will be used in the design of the new generation of digital systems.

5 ACKNOWLEDGEMENT

This work has been funded as part of the National Radio Propagation Programme by the Radiocommunications Agency of the Department of Trade and Industry in the UK.

6 REFERENCES

[1] Craig K.H. and Levy M.F., "Parabolic equation modelling of the effects of multipath and ducting on radar systems", *IEE Proc. F* 138, No.2, 153–162, 1991.

[2] Levy M.F., "Parabolic equation modelling of propagation over irregular terrain", *Electron. Lett.* 26, No. 14, 1153–1155, 1990.

[3] Levy M.F., "Diffraction studies for microcellular applications", Eighth Int. Conf. on Antennas and Propagation (ICAP 93), *IEE Conf. Pub.* 370, Pt. 1, 76–79, 1993.

[4] Craig K.H. and Levy M.F, "Field strength forecasting with the parabolic equation: wideband applications", Sixth Int. Conf. on Antennas and Propagation (ICAP 89), *IEE Conf. Pub.* 301, Pt. 2, 461–465, 1989.

[5] Anderson H, "A Ray -Tracing Propagation Model for Digital Broadcast System in Urban Areas", *IEE Transactions on Broadcasting*, Vol. 39, Number 3 , September 1993.

[6] Seville A, Yilmaz U., Charriere P.G.V., Powell N. and Craig K.H., "Building scatter and vegetation attenuation measurements at 38 GHz", Ninth Int. Conf. on Antennas and Propagation (ICAP 95), *IEE Conf. Pub.* 407, Pt. 2, 46–50, 1995.

[7] McArthur R.J. and Craig K.H., "Propagation in the urban environment using building cover databases", URSI Comm. F Open Symposium, UK, 8.1.1–5, 1992.

[8] Maver T.W., in "Computers in Art, Design and Animation", Springer Verlag, 274–282, 1989.

[9] CCIR, "Reflection from the surface of the Earth", Reports and Recommendations of the CCIR, Report 1008-1, 1990.

[10] Cadence Design Systems Inc., "Signal Processing WorkSystem (SPW)", Alta Group, California, USA, 1990.

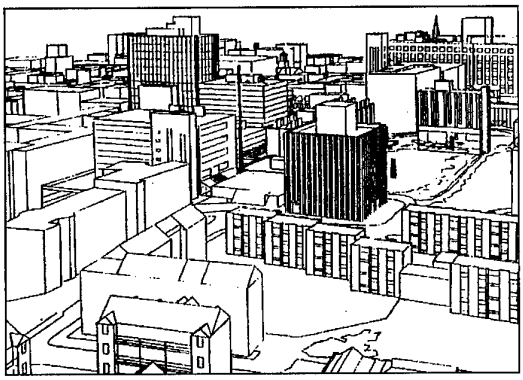


Figure 1: Illustration of the detail available with architectural databases.

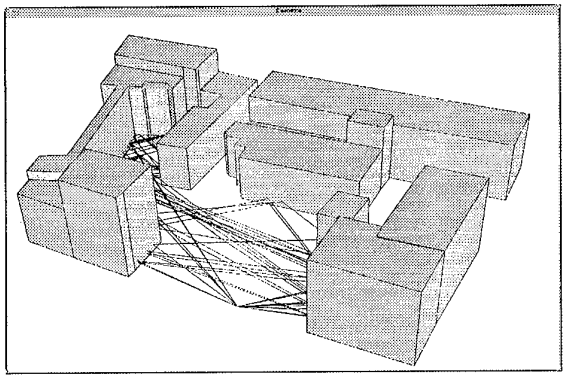


Figure 2: Illustration of the degree of detail of a 3D database.

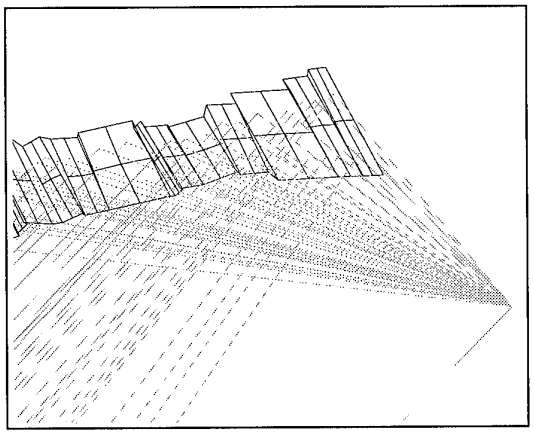


Figure 3: Illustration of a study at VHF frequency using PO method.

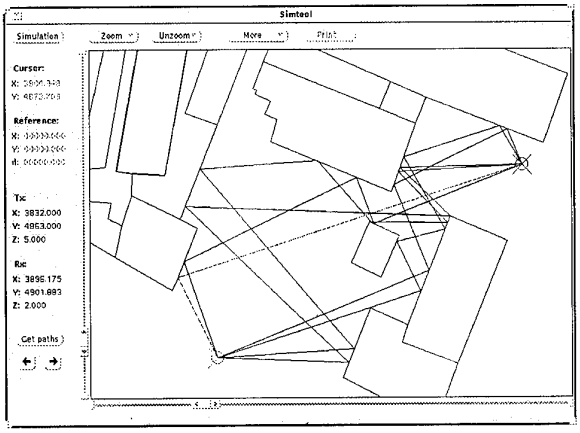


Figure 4: Example of point to point study.

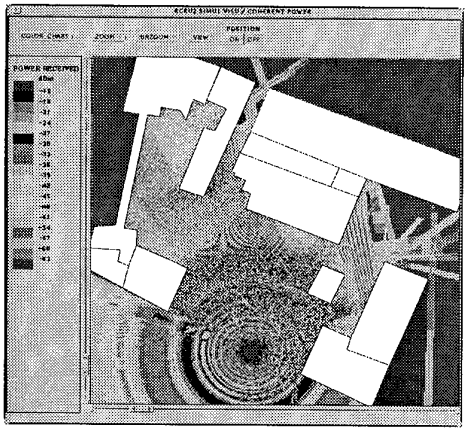


Figure 5: Example of coverage diagram obtained

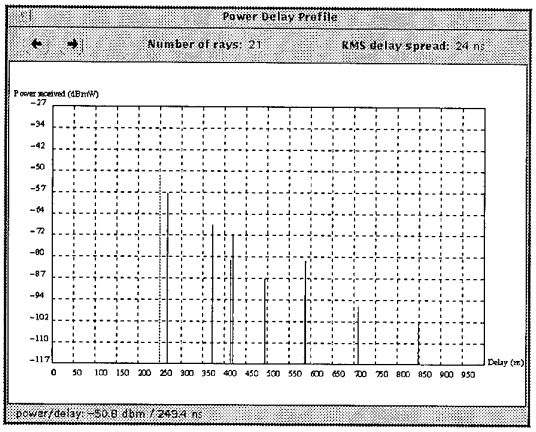


Figure 6:Power delay profile obtained.

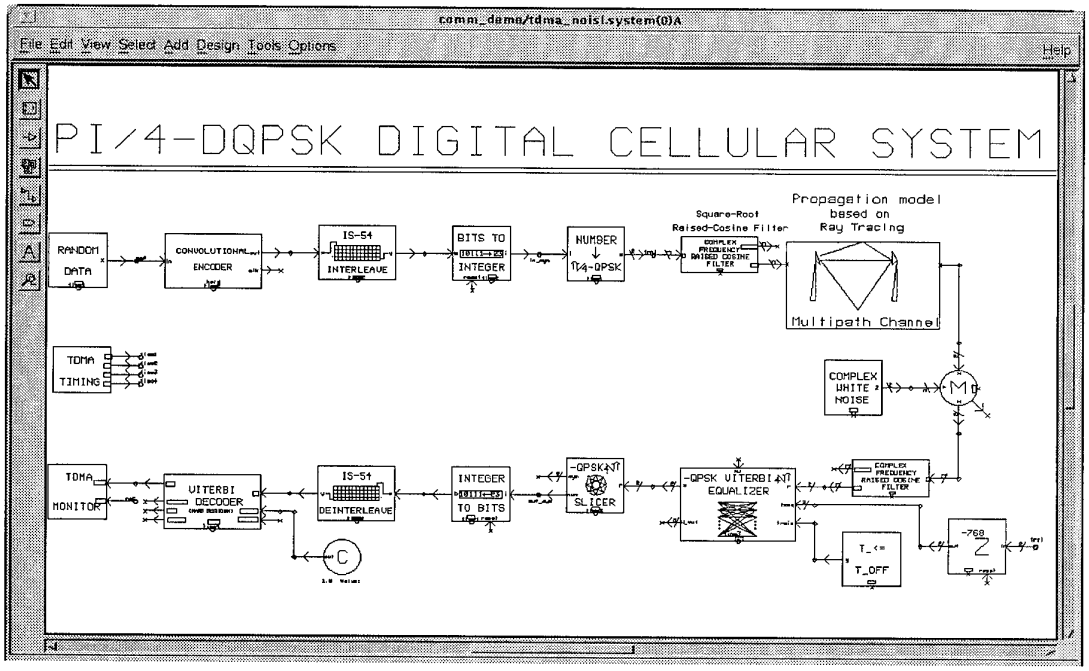
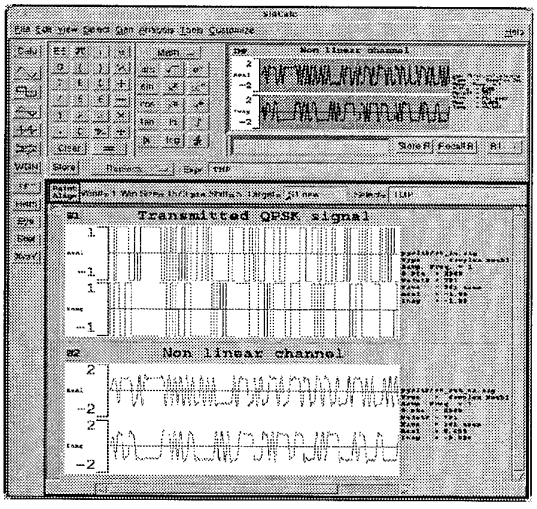


figure 7: SPW block diagram editor describing a digital communications channel.



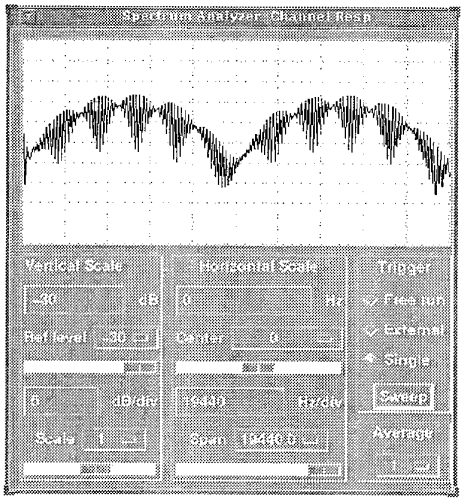


Figure 8 and 9: Example of analysis carried out on the system defined in figure 7.

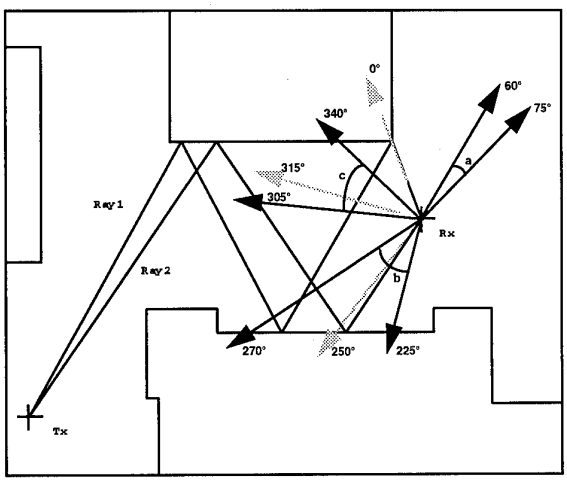


Figure 10: Prediction software ray-trace diagram.

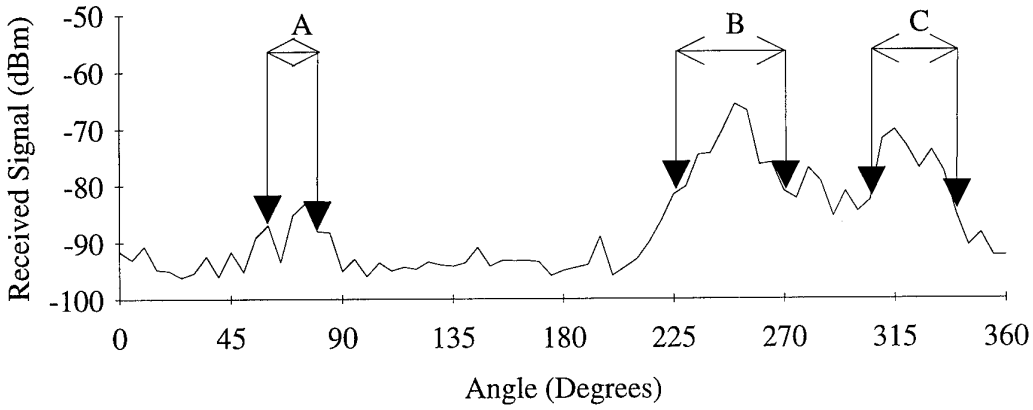


Figure 11: Measured signal around a 360° scan.

DISCUSSION

Discussor's name: J. F. Harvey

Comment/Question:

1. Comment:

I believe it is very important to understand the details of propagation in complex terrain in order to design the US Army's digital battlefield systems. This kind of model is important to advance this understanding.

2. Question:

Does your model simulate effets of complex terrain on the polarization of the EM waves, particularly for VHF.

3. Question:

Are your confident that the diffuse reflection contribution, which you neglect, is less than the specular reflection contribution for high order reflections?

Author/Presenter's reply:

2. No.

3. We have not yet dealt with a situation requiring more than first order reflections.

Discussor's name: A. D. Papatsoris

Comment/question:

Have you considered excluding phase information from your model? If yes, how does the coherent signal compare to the incoherent?

Author/Presenter's reply:

Yes. The tool can display other quantities than coherent fields. An incoherent sum gives a value of the signal including slow fading effects but excluding the effects of fast fading due to constructive/destructive effects of phase differences.

DISCUSSION

Discussor's name: F. Davarian

Comment/Question:

Please comment on the effect of room and office furniture on your prediction model.

Author/Presenter's reply:

We currently have only applied the model to outdoor coverage. Penetration through buildings at millimetre wavelengths is very sensitive to small scale features such as window locations (and quite possibly furniture). We are making measurements to try to characterise this. For now we have proposed a very simplified attenuation model.

Discussor's name: Not given

Comment/Question:

Question 1:

Have you any plans to perform experiments for comparison purposes?

Question 2:

Given the uncertainty in the various parameters needed and the effort needed to set up your simulation model, what is the advantage of simulated results as opposed to moving the transmitter and receiver positions in an experiment to determine the required information?

Author/Presenter's reply:

Question 1:

Yes a measurement programme is currently undertaken at RAL. Its purpose is the validation of the current model and its improvement, the investigation of the effects of vegetation and the study of through-building propagation at 38 Ghz.

Question 2:

The prediction tool can give an estimate of the quality of the received signal with a certain degree of confidence in many cases. Furthermore, it can be the only way to assess the radio coverage of a system in cases such as hostile environment in enemy territory.

DISCUSSION

Discussor's name: L. Bertel

Comment/Question:

Comment prenez-vous en compte les effets d'antennes dans votre modèle? Utilisez-vous les réponses complexes des antennes (amplitude et phase) ou seulement les diagrammes de directivité?

Translation

How do you take into account the effects of antennas in your model? Do you use the complex responses (in amplitude and phase) of the antennas or only their directivity patterns.

Author/Presenter's reply:

The simulation tool takes into account the antenna gains using digitised antenna patterns. The phase change due to antenna is not taken into account.

Although we calculate the coherent field, the phase at any particular receiver will be indeterminate because of the uncertainty of building positions compared to a wavelength. However, the relative phase from point to point will be correct, and this gives fade rates etc.

As long as antenna phase varies slowly with angle, these quantities will be unaffected by ignoring antenna phase.

UHF and Microwave Propagation Prediction in an Urban Environment

T. M. Willis, B. J. Guarino and J. D. Moore
AT&T Bell Laboratories
101 Crawfords Corner Road
Holmdel, NJ 07733-3030 USA

Raymond Luebbers and Joseph Schuster
Electrical Engineering Department
The Pennsylvania State University
University Park, PA 16802 USA

Henry L. Bertoni, George Liang and Nai Wo
Polytechnic University
Six MetroTech Center
Brooklyn, NY 11201 USA

Theodore S. Rappaport and Prabhakar M. Koushik
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061 USA

1. SUMMARY

This paper discusses ongoing work by AT&T Bell Laboratories, Pennsylvania State University, Polytechnic University and Virginia Polytechnic University in developing site specific urban propagation computer models. Both two and three dimensional ray tracing algorithms have been developed by the different organizations and applied in the Rosslyn VA area at 900 and 1900 MHz. After the models have predicted the propagation loss from specific transmitter sites to receiver locations, measurements have been taken to determine the actual values. This paper presents some comparisons of the different models and measurements. Where the computer models have worked well, they are typically within about 7 dB of measurements. In some areas of Rosslyn there are known problems with the building data base where predictions differ from measurements, however work is continuing to improve the models in other problem areas. This will be accomplished by comparing the different models to one another and the measurements and then refining the models.

2. INTRODUCTION

Before deploying a cellular or micro-cellular type microwave communication system, it is usually necessary to understand the propagation conditions in the intended service area in order to select antenna sites and set system design parameters to obtain the required coverage. Often design tools using statistical propagation models are used for this purpose. However, these tools are usually not accurate enough for urban environments and extensive on-site measurements are required before the design can be finalized. These measurements add to the time and expense of the system design and may not be practical in many scenarios.

An alternate approach is to use site-specific tools that model microwave propagation by including actual local environmental elements such as buildings, other objects, vegetation and terrain. These tools can be expected to provide better accuracy than can be obtained with statistical models and with less time, expense and intrusion than required for measurements.

Many such models and tools have been proposed in the scientific and engineering literature. To evaluate some of the more promising of these and encourage their further development into a complete urban propagation tool incorporating the best features of each, the United States government is sponsoring a project involving AT&T Bell Laboratories, Pennsylvania State University, Polytechnic University and Virginia Polytechnic University. This project includes a series of propagation predictions from each of the four groups which are then compared with actual measurements made independently. These comparisons are used to evaluate the accuracy of the predictions, identify the relative strengths of each model and refine the modeling techniques in areas where the measurements and predictions differ significantly.

The site selected for the study was about one square km area of Rosslyn, Virginia, USA. Rosslyn is an urban area lying across the Potomac River from Washington, DC which offers an irregular street plan with a variety of building types, building heights and street widths. There are residential areas with light vegetation in the western area of the study with high-rise, commercial buildings with an average height of about 10 stories in the east. The commercial area also includes some open areas and many elevated pedestrian cross-walks over the streets. In addition, terrain elevation varied from 15 to 55 meters over the

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

study area. These factors allowed a number of environmental variables to be addressed in the study.

3. BUILDING DATA BASE

Another major factor in selecting the Rosslyn site for this study was the existence of extensive photographic data obtained during aerial overflights. From the photographs, the major buildings and other structures were identified and the external surfaces of the buildings were digitally encoded into a set of three dimensional, convex polyhedra using a wire-grid structure. Additional post-processing of this data was performed to remove some of the detail which could greatly lengthen the calculation time without contributing significantly to the accuracy of the predictions. The final data base was generated in the standard DXF format used by many architectural and other CAD programs. The use of this standard format for interfacing all of the propagation prediction tools allowed easy interchange of building data between the four study groups and future flexibility with other data bases and tools. Figure 1 is a perspective view of this data base for a part of the city.

4. MODEL DESCRIPTIONS

4.1 AT&T Model

The computer tool developed by AT&T Bell Laboratories to predict outdoor electro-magnetic propagation currently supports two different ray tracing methodologies. The first is a two dimensional ray tracing technique while the second is a full three dimensional model. This computer tool was developed from an earlier tool designed to predict propagation inside of buildings that is discussed in papers by R. A. Valenzuela et. al.[1] [2]. The two dimensional technique has the advantage of being much faster to calculate than the three dimensional method, however it is only valid when the transmitter and receiver locations are significantly below the tops of the buildings in the area. This technique is sometimes referred to as the infinite canyon model. Both ray tracing techniques determine specular reflections off building surfaces and diffractions at vertical edges of buildings. The three dimensional model also uses specular reflections from horizontal building roofs and the ground and diffractions from horizontal edges.

The propagation tool requires as inputs the locations, orientations and antenna patterns of all transmitter sites and locations of all receivers. It also requires the transmitted power levels of all the transmitters. Finally a description of all the planar surfaces of the buildings is required as input. This surface description file must indicate the position and shape of every building surface as well as what type of surface it is. The user of the tool may select the building wall type from a list of typical building materials such as concrete, glass, wood, sheetrock, etc.

Once the tool has read the input data, it attempts to find all ray paths from a transmitter to a receiver location. These rays may find their way to the receive location by reflecting off building surfaces or diffracting around building corners formed by two building surfaces. The user of the tool may specify the maximum number of reflections and diffractions to consider in finding these ray paths. Once all the rays desired have been found, the electric field each ray provides at the receive location is calculated. The predicted power at the receive location is then calculated from these fields and may be performed coherently adding fields or incoherently adding power contributions.

The field from each ray is calculated by determining the strength of the radiated ray and the loss that ray undergoes. The radiated power is calculated from the transmit power and antenna pattern. The loss has two components, a free space loss and an additional scattering loss. The free space loss is simply the amount of loss due to the radiation of energy out into open space and is inversely proportional to the distance traveled by the ray squared. In addition to this loss, some energy is lost from the ray during each reflection or diffraction it undergoes. To compute the amount of energy lost in a reflection, the building surfaces are modeled as multilayered planar homogeneous lossy dielectric slabs with a plane wave impinging upon them. Good choices for the number, thickness, permittivities and conductivities of these dielectric layers is still an area of research. To date, values have been chosen to try and represent actual building construction and materials. Good agreements with measurements for the outside two dimensional model have been reported by Erceg, Rustako and Roman[3] using material parameters for concrete. To compute the amount of energy lost in a diffraction, the Geometric Theory of Diffraction (GTD) for an infinite lossy dielectric wedge, as reported by Luebbers[4], has been used.

4.2 Penn State Model

The Penn State propagation model is a hybrid of Shooting and Bouncing Rays (SBR)[5] and the Geometrical Theory of Diffraction (GTD)[6] [4] [7]. Both two-dimensional (2D) and three-dimensional (3D) versions have been developed. The 2D version assumes that the transmitter and receiver are located below the building roof level. The 3D model assumes that either the transmitter or receiver (or both) are located at a height above some of the buildings so that propagation over buildings must be included. The 2D version is therefore a "canyon" model. For the 2D model the ground is approximated as a flat surface which can be slanted. The 3D version is fully three-dimensional, with full polarization dependence

included in the reflection and diffraction coefficients. For the 3D version the building surfaces can be convex flat polygons of arbitrary shape, including slanted edges. The 3D version ground is modeled as connected polygonal plates, so it can be irregular. Diffractions from the ground plate edges are not included in these calculations, but could be added.

When making propagation calculations the SBR method is applied twice. For the first application many rays are shot from the transmitter location, reflecting off buildings and ground. For the 2D model these are disks, but for the 3D model the rays shoot out through a spherical area. The purpose of this "shooting" is to find the diffracting edges for a given transmitter location (and ray direction for the 3D model). This "shooting" is done regardless of the receiver locations. As the rays are "bouncing" off the buildings, the rays closest to building edges are determined. These rays are sorted, and each diffracting edge is identified. If double diffraction is to be applied, additional cones of rays (disks for 2D) are "shot" from the diffracting edges and bounced through the buildings. Once this is completed all the diffracting edges are known. For 3D the diffraction points on each diffracting edge for each ray incident on the edge are known.

Now the rays are all shot again. Each receiver point has a collection aperture surrounding it, typically several meters on a side. For each ray shot from the transmitter, including any subsequent diffracted rays, all rays collected in a receiving aperture are saved. For each ray shot from the transmitter all receiver points are considered. Then the next ray is shot and all receivers are again considered. As the rays are being shot, the rays received by each receiver aperture are sorted and one ray for each unique path is selected. For multiple rays following the same path the ray closest to the center of the receiver aperture is kept. The ray amplitudes are evaluated using reflection coefficients for a plane wave incident on a dielectric half-space. Thus the reflection coefficients are a function of incidence angle. The diffracted rays are evaluated using UTD wedge diffraction. Diffracted rays are combined coherently with the corresponding reflection and shadow boundary rays for each building surface. Rays from different building surfaces are combined incoherently. The ground reflected rays are combined coherently with the corresponding non-ground-reflected ray for the 2D model. In the 3D model the ground reflections are included but are not combined coherently with the corresponding non-reflected ray.

4.3 Polytechnic Model

Instead of a full three dimensional ray trace, the Polytechnic approach uses several simpler two dimensional ray traces to approximate the actual three dimensional rays that give the primary contributions to the received signal. To this end, we identify a hierarchy of ray classes that can be systematically searched by the program to find the primary contributors in each class. For each class, the program consists of a sub program that interrogates the building data base for relevant building information. A second sub program then calculates the ray contributions to the total signal. These individual programs are modular, and reused at later stages in the hierarchy.

In order to describe this hierarchy, we define for each receiver point a Vertical Planes (VP) [8], and a Slant Planes (SP) as shown in Figure 1. The VP's are defined as containing: 1) the receiver and the transmitter; 2) the receiver and an equivalent source point such as an image of the transmitter in a nearby building, or a diffracting edge; 3) an equivalent receiver point, such as the image of the receiver in a nearby building or an edge near it, and the transmitter; or 4) an equivalent receiver point and equivalent transmitter point. For each VP, there is an SP [9], which is perpendicular to the VP and containing the actual or equivalent transmitter and receiver points.

Within the VP or SP there are classes of rays that must be considered in order to account for the significant paths between the transmitter and receiver. In the VP the ray paths that must be considered include a direct path between the transmitter and receiver via diffraction over the tops of the buildings [10] and reflection near the transmitter prior and/or near the receiver subsequent to the diffraction over the top of the buildings. In the SP the significant ray paths can be considered to be three sub-classes [11] [12] and includes: those rays that reach the receiver by multiple reflections at the sides of the buildings, rays that are multiply reflected at the building sides and undergo one diffraction at a vertical corner and rays that involves two diffractions at vertical building corners and multiple reflections before, between and after the two diffractions. Finally the LOS ray, if it exist will be determined in both the VP and SP, therefore it should be included in the SP and neglected in the VP.

4.4 Virginia Tech Model

A brute force recursive technique is used to trace rays launched from the transmitter in three dimensions. Geometrical optics is used to trace the propagation of direct, reflected, and scattered fields. The transmitter is modeled as a point source generating rays uniformly in all directions. Constant angular separation between the rays is achieved by launching the rays through the vertices of an icosahedron inscribed in an unit sphere, with each of its triangular faces subdivided into smaller triangles. This method of launching rays provides wavefronts of equal shape and area that can

be easily subdivided. While this method is more computationally intensive than those based on image theory, it is much more flexible since different propagation models can be easily incorporated. The computational demands are drastically reduced by using bounding volume hierarchies for buildings and exploiting the parallelism of a network of workstations, the computation time can be further reduced.

The energy at a receiver location is determined by performing a non-coherent superposition of the contributions due to line of sight, reflected, and scattered fields. The contribution from the LOS path is calculated using the two ray model. The reflected ray follows a $1/D^2$ dependence according to the Friis free space formula where D is the total ray path length. The reflection coefficients are varied based on the angle of incidence and this appears to give better predictions over using constant values. A reception sphere is used to determine the reception of a specular reflected ray at a receiver location. Rough surface scattering is taken into account by subdividing the surface into several small facets (so the receiver is in the far field), applying the bistatic RCS to each, and then doing a non-coherent summation of the contribution from each facet. In heavily shadowed regions, diffraction is dominant and is taken into account with wedge diffraction.

5. MEASUREMENTS

5.1 Equipment Description

AT&T owns and operates two test vans equipped with electronic equipment for collecting propagation data in any frequency range up through the microwave region. One van is used for transmitting, while the other for receiving RF energy. The propagation loss measurements are performed by locating a transmitter at the top of a mast on the street level parked transmit van or on a building roof at specific sites in Rosslyn. Then measurements are collected by the receiving van driving down a particular street or parked at a specific site.

The transmit van has various synthesized signal generators, 4 independent "pods" for leveling capability to keep the antennas vertical, and a 10 meter pneumatic mast upon which the transmitting antennas are placed. The receive van houses a Data Acquisition System (DAS) for collecting, storing and analyzing receive signals, a Positional Acquisition System (PAS) to determine the van's location, and other miscellaneous equipment. Both vans operate off of 10 kilowatt gas generators. A block diagram of the receive van is shown in Figure 3, and a brief description of the major components is discussed below.

The DAS consists of 4 spectrum analyzers, signal generators for calibrations, antennas, LNAs, cabling, and a Data Acquisition Computer (DAC). There are two operating modes available from the AT&T DAS: high and low resolution. During high resolution data collecting, the DAC samples the receive power form the spectrum analyzers at 10,000 times per second and stores this data on a Bernoulli floppy disk in the DAC. During low resolution data collecting, the receive power is sampled at 10,000 samples per second, then power averaged and the average recorded. While the van is moving, the data is averaged over each 1 meter of travel. When the van is moving less than 1 meter per second, averaging is done over a 1 second interval. The DAS appends a time stamp on each data record it stores on the Bernoulli floppy disk in the DAC.

The PAS consists of a Trimble Differential GPS system, a Loran-C and a dead reckoning system. If there is no differential solution to the van's position, standard GPS is used. If the GPS signal is insufficient, then either Loran-C or the dead reckoning system is used. The dead reckoning system estimates the position of the van by a magnetic compass, a speed sensor on the van's drive shaft, and knowledge of the van's last position. About every 5 seconds, the location of the van is placed into a file with a time stamp on the Bernoulli floppy disk in the DAC.

5.2 The Measurement Program

In our measurement program there were two types of antenna heights and two types of measurements made. Ground level (from 2 to 10 meters) and rooftop level antenna heights were selected to test the two and three dimensional models' prediction capability, while van receive measurements were performed as either a continuous drive (excluding traffic conditions), or a stationary measurement. Prior to data collection, FCC permissions were obtained, and area surveys were completed to ascertain the best locations for transmitting antennas and optimum drive routes for collecting data. Optimum antenna positions were decided as a compromise between the available locations, limitations on where the vans could be placed, and what immediate environment would best exercise the reflection, diffraction and refraction capabilities of the researcher's modeling programs.

Once specific locations for the transmit and receive (in the case of stationary measurements) antennas, were established, position measurements referenced to the data base were made to place the antennas at the intended locations to within about 1 meter. The procedure for data acquisition was simply to place the transmitting antenna at the intended location, perform calibrations to assure full dynamic range of the DAS, and collect data by either driving the receive van along the streets of Rosslyn, or placing the van at the specified locations for the stationary measurements. Unfortunately for the data collected while driving, the

GPS positioning system did not provide consistently accurate locations for the van in the downtown Rosslyn area. To overcome this difficulty, the position of the van was determined with the dead reckoning system using a known starting point. Once a data set was collected, it was plotted and analyzed to verify its validity. After data collecting, the DAS would be checked for any calibration drift. If any significant drift occurred, or the data appeared to be invalid, the data set would be repeated.

6. COMPARISON OF MEASUREMENTS AND PREDICTIONS

Figures 4-8 display a comparison of a small sample of the model predictions and measured data for a transmitter site and receiver locations down a particular street in Rosslyn. These plots show the measurement data as small dots. The solid line depicts a 5 meter local average of the measurements. The various predictions are displayed as marks on the plots; AT&T as a cross, Penn State as a plus sign, Polytechnic as a diamond and Virginia Tech as a circle. Figure 4 shows a comparison of the two dimensional models for transmitter 2b down Moore St. Transmitter 2b used omni-directional antennas at an elevation of 10 meters two blocks east of Moore St. Except at the intersection of 19th St. near the north end of Moore, it is shadowed from this transmitter site by buildings.

Figures 5-8 display comparisons of the three dimensional models and measurements. Figure 5 shows Moore St. predictions and measurements for transmitter 5. Transmitter 5 used directional antennas pointing down Moore St. from the roof of a 40 meter high building. This transmitter had a clear line of sight to all of Moore St. Figure 6 shows Lynn St. for transmitter 5. Lynn is one block to the east of Moore and is mostly blocked by buildings from transmitter 5. Figure 7 depicts transmitter 6 for Nash and 19th St. Transmitter 6 used directional antennas aimed to the north-east on top of a 45 meter high building. Nash is to the west of transmitter 6 while 19th St. is to the north. These streets have both line of sight and blocked regions to transmitter 6. Figure 8 shows Colonial Terrace for transmitter 7. Colonial Terrace loops through a condominium complex of two to three story brick buildings surrounded by many trees. Transmitter 7 used omni-directional antennas at a 10 meter elevation on the south side of the complex. Colonial Terrace has a clear line of sight back to transmitter 7 at the beginning, end and peak in the middle of the plot. Other areas are blocked by the condominium buildings.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the help provided by the City of Rosslyn, Virginia, and members of the Arlington County Police Department in facilitating the data collection reported in this paper. In addition, AT&T employees Don Jacobs, Gene McArthur, Tedd Rinaldi, Jim Johnson and Bill Witkowski were responsible for installing and maintaining the test equipment, planning the data collection runs and collecting the signal strength data.

REFERENCES

1. R. A. Valenzuela, "A Ray Tracing Approach to Predicting Wireless Transmission", *1993 43rd IEEE VTC Conference Proceedings,* IEEE catalog #93CH3305-0, ISBN 0-7803-1266-X, pg. 214-218

2. S. J. Fortune, D. M. Gay, B. W. Kernighan, O. Landron, R. A. Valenzuela and M. H. Wright, "WISE Design of Indoor Wireless Systems: Practical Computation and Optimization", *IEEE Computational Science & Engineering;* Vol. 2 no. 1, Spring 1995, pp. 58-68

3. V. Erceg, A. J. Rustako and R. S. Roman, "Diffraction Around Corners and Its Effects on the Microcell Coverage Area in Urban and Suburban Environments at 900 MHz, 2 GHz and 6 GHz", *IEEE Trans. Vehicular Tech.;* Vol. 43 no. 3, Aug. 1994, pg. 762-766

4. R. J. Luebbers, "Finite Conductivity Uniform GTD Versus Knife Edge Diffraction in Prediction of Propagation Path Loss", *IEEE Trans. Antennas and Propagation;* Vol. AP-32 no. 1, Jan 1984, pg. 70-76

5. H. Ling, R. Chou, and S. Lee, "Shooting and Bouncing Rays: Calculating the RCS of an Arbitrarily Shaped Cavity," *IEEE Transactions on Antennas and Propagation,* Vol. 37, No. 2, pp 194-205, February 1989.

6. R. G. Kouyoumjian and P. H. Pathak, "A Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," *Proc. IEEE,* Vol. 62, No. 11, pp. 1448-1461, November 1974.

7. R. Luebbers, "Propagation Prediction for Hilly Terrain Using GTD Wedge Diffraction", *IEEE Transactions on Antennas and Propagation,* Vol. 32, no 9, pp. 951-955, September 1984.

8. H.L. Bertoni, W. Honcharenko, L.R. Maciel, and H.H. Xia, "UHF Propagation Prediction for Wireless Personal Communications", *Proceedings of the IEEE.* Vol. 82, No. 9, pp 1333-1359, Sept 1994.

9. T. Kurner, D.J. Cichon and W. Wiesbeck, "Concepts and Results for 3D Digital Terrain-Based Wave Propagation Models: An Overview", *IEEE Jnl. on Selected Areas in Comm.* SAC-11, pp. 1002-1012, 1993

10. L.E. Vogler. "An Attenuation Function for Multiple Knife-Edge Diffraction." *Radio Science.* 17. pp 1541-1546. 1982.

11. S.T.S Chia, R. Steele, E. Green and A. Baran, "Propagation and Bit Error Ratio Measurements for a Microcellular System," *J. IRE,* vol. 57, pp S255-S266, 1987.

12. F. Niu and H.L. Bertoni, "Path Loss and Cell Coverage of Urban Microcells in High-Rise Building Environments," *Proc. of the IEEE* GLOBECOM'93, Houston, TX, 1993.

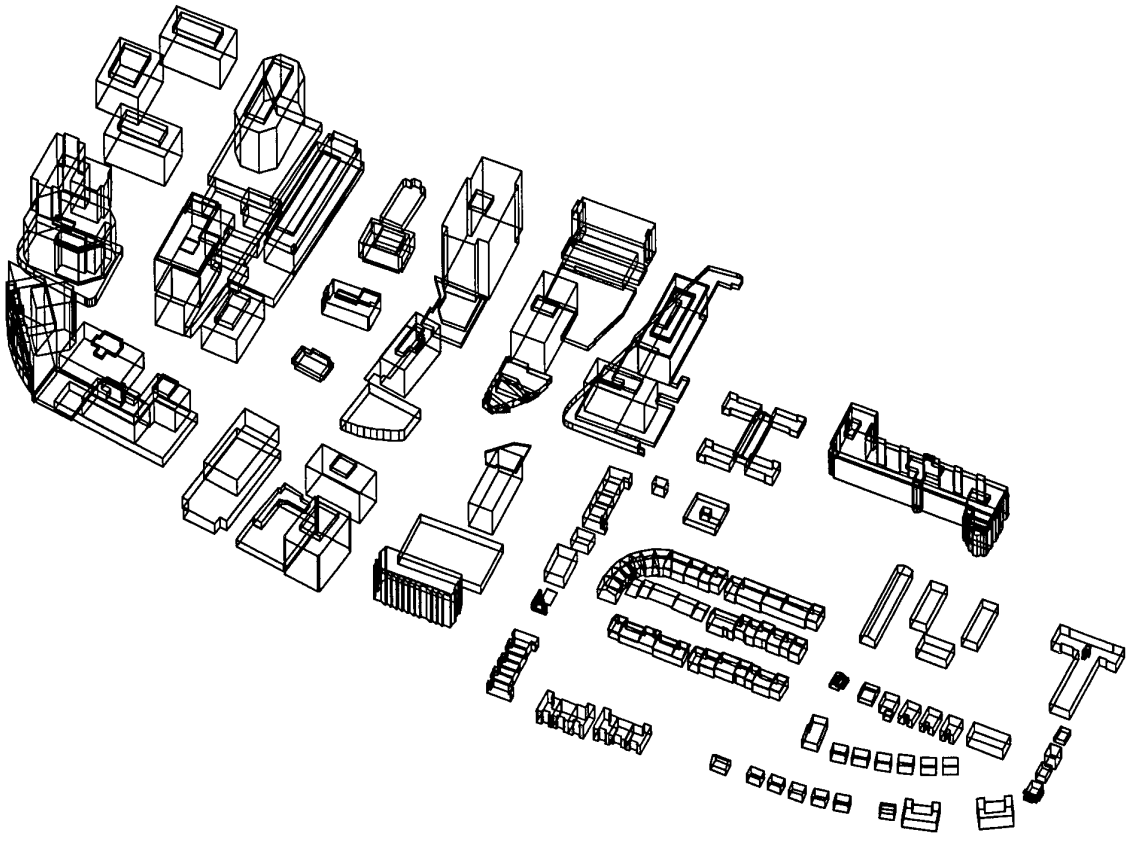


Figure 1 - Rosslyn Building Data

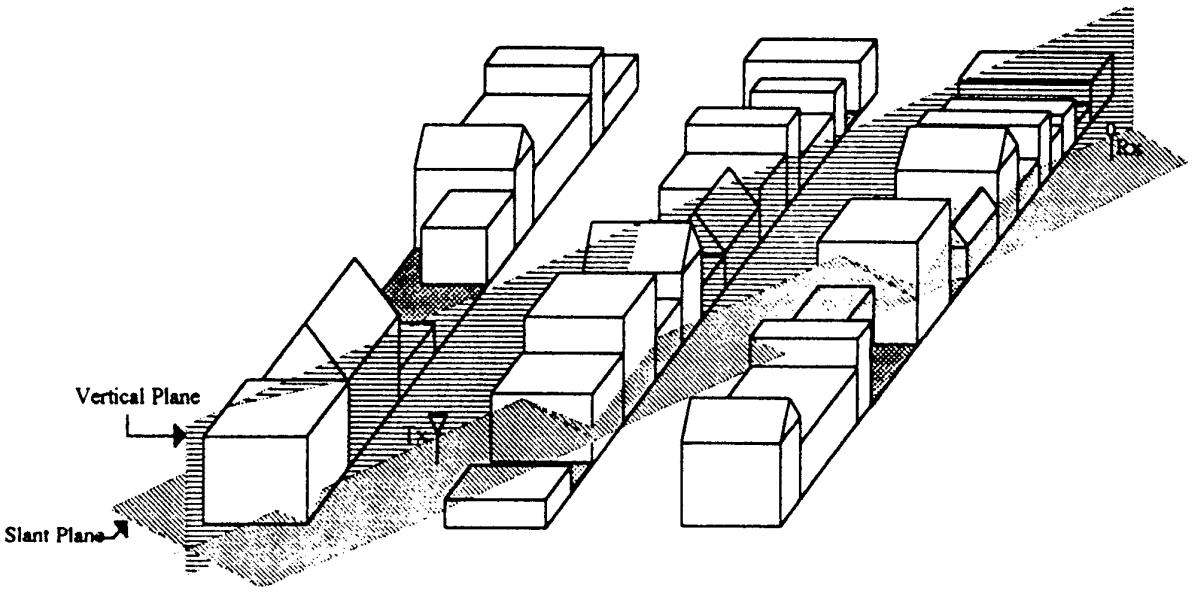


Figure 2 - Vertical and Slant Planes as They Intersect the Building Database

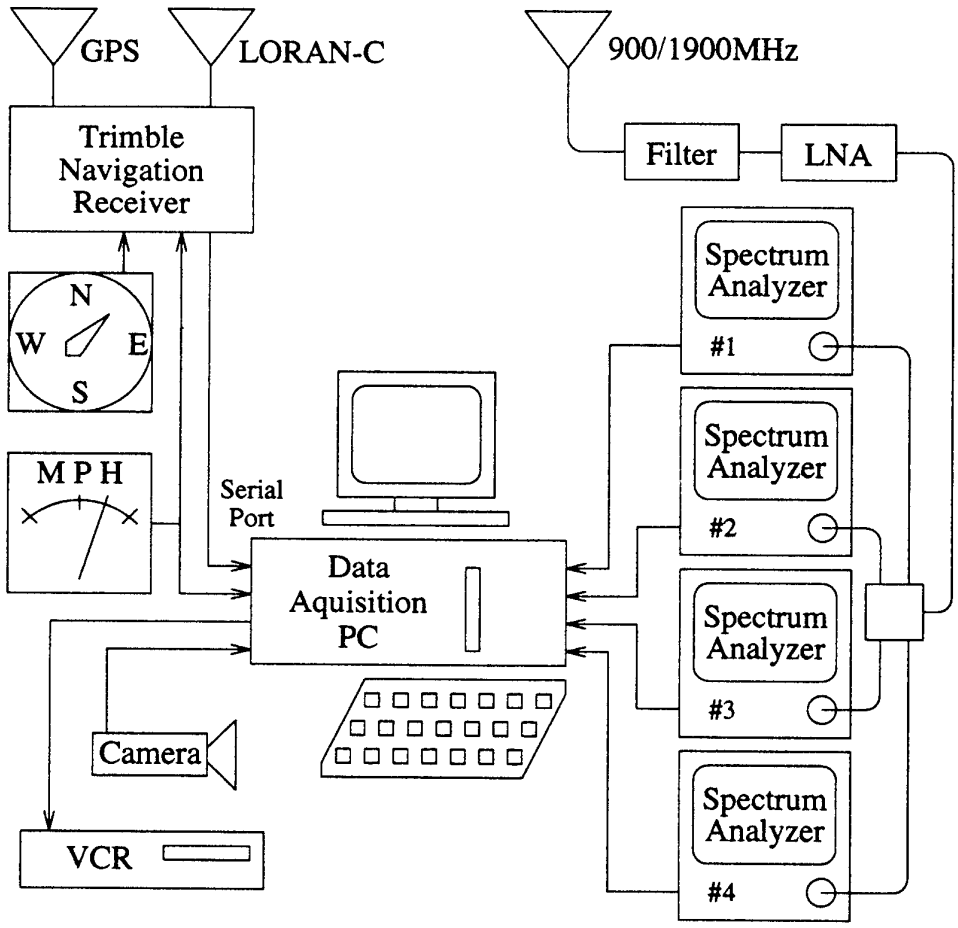


Figure 3 - Receive Van Measurement Equipment

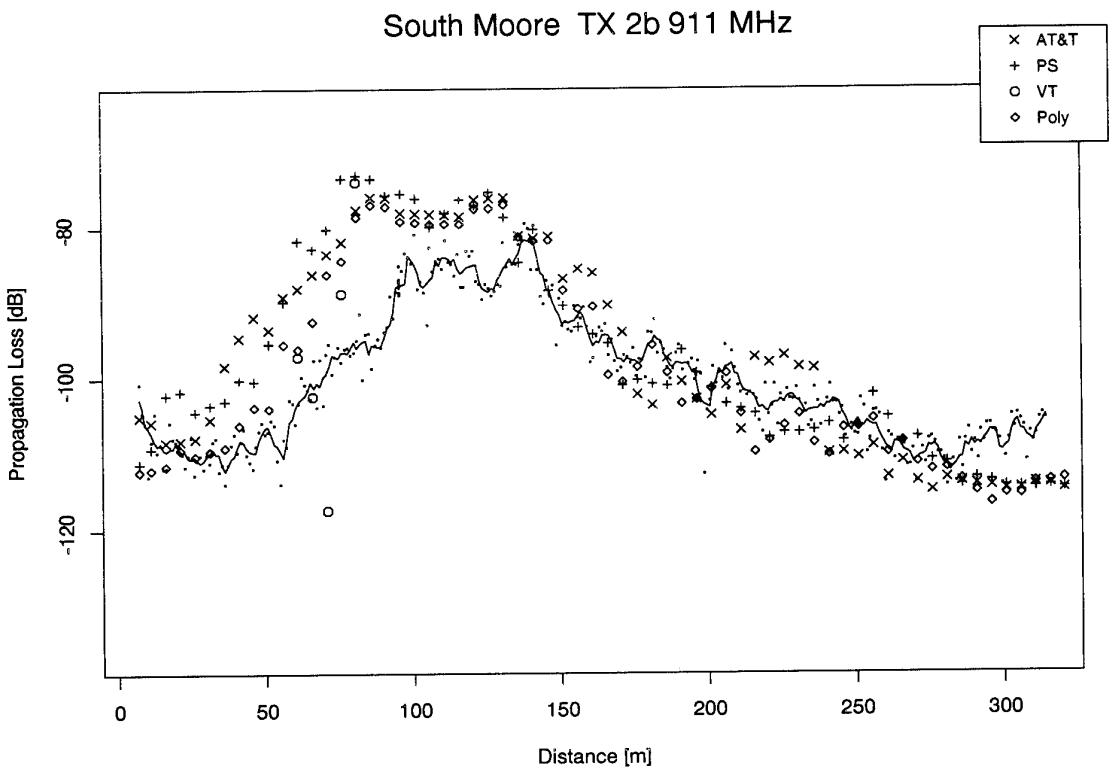


Figure 4 - Predicted and Measured Propagation Loss

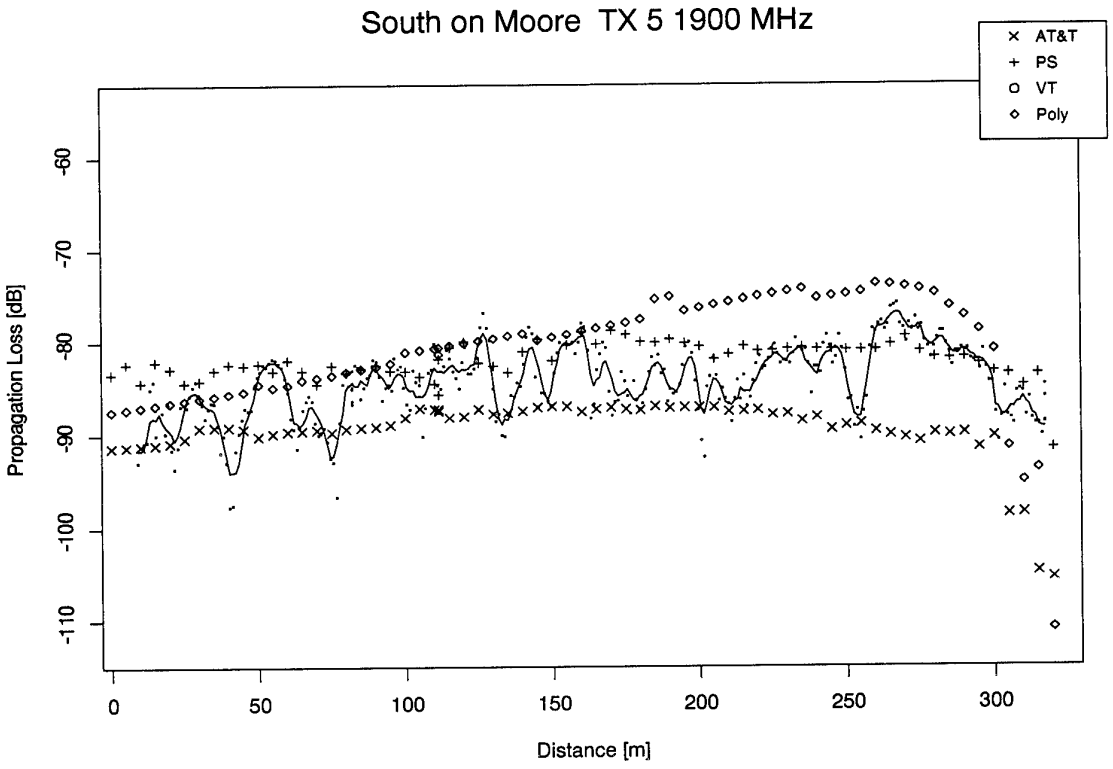


Figure 5 - Predicted and Measured Propagation Loss

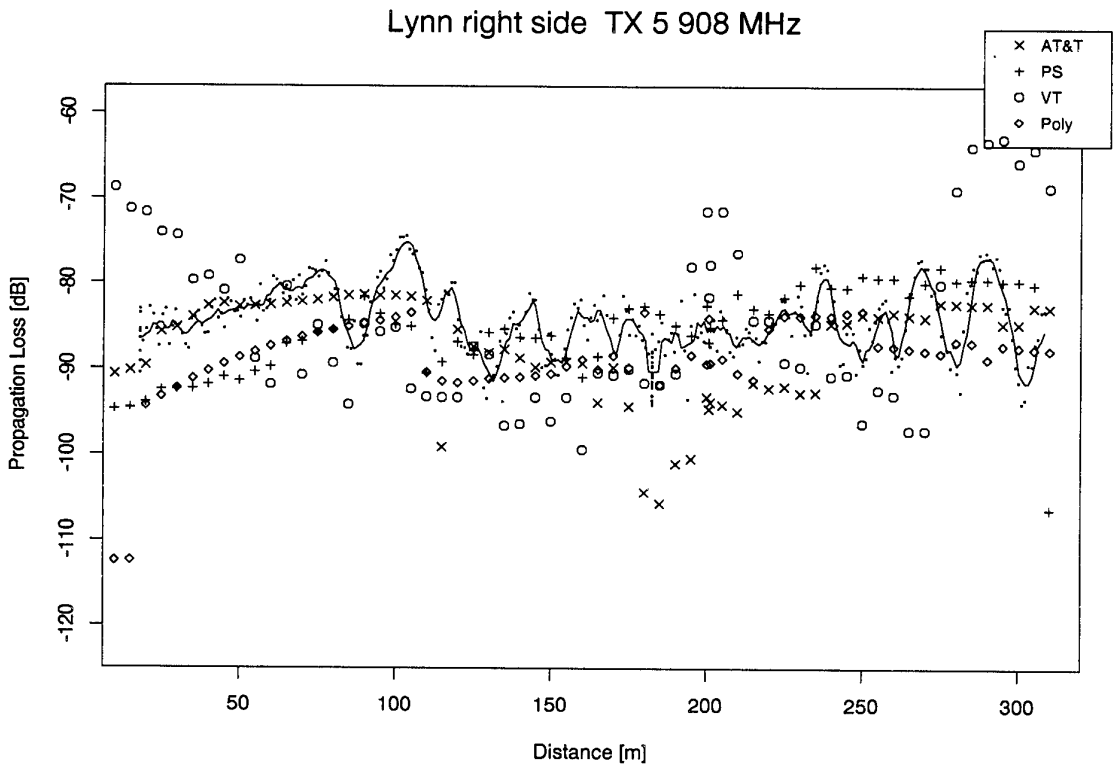


Figure 6 - Predicted and Measured Propagation Loss

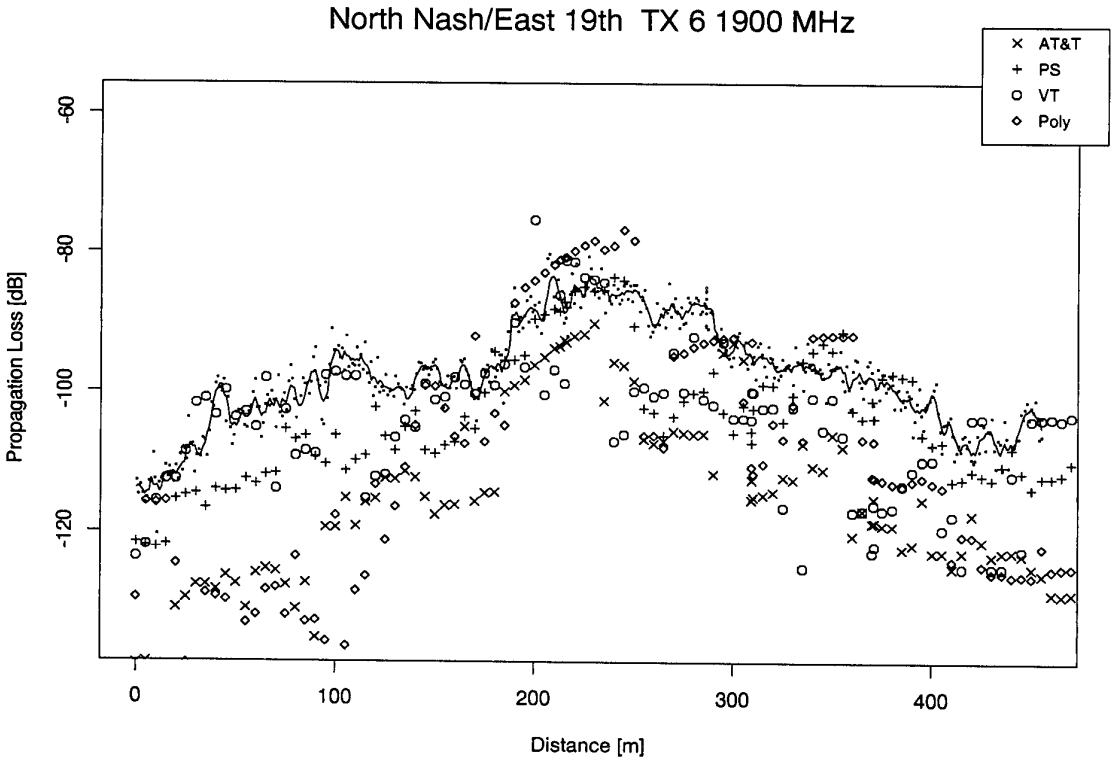


Figure 7 - Predicted and Measured Propagation Loss

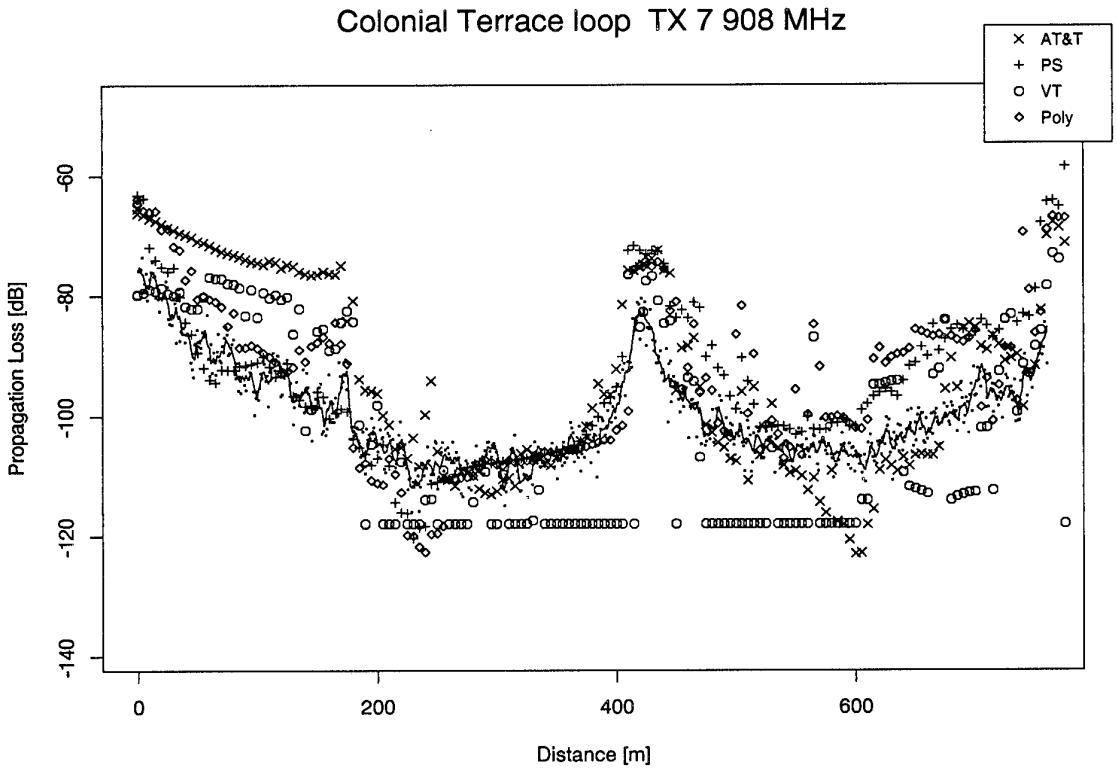


Figure 8 - Predicted and Measured Propagation Loss

DISCUSSION

Discussor's name: S. Karp

Comment/Question:

Was delay spread measured?

Author/Presenter's reply:

No.

Discussor's name: A. Altintas

Comment/Question:

1. Did you compare 2-D vs. 3-D models for the same transmitter position? If so, do you think that a 3-D model is necessary in general?

2. Did you compare your results with the Okumura-Hata model?

Author/Presenter's reply:

1. In our experience, a 3-D model is necessary for roof-top antennas, but not necessary if both antennas are lower than the roof tops.

2. We have not done so, but believe we are more accurate.

Discussor's name: C. Rigal

Comment/Question:

What would happen to the model for a transmitter placed on a satellite?

Author/Presenter's reply:

For transmitters placed on top of buildings, few rays are involved but they change rapidly when the receiver is moving. For that reason the computation may be more complicated.

DISCUSSION

Discussor's name: F. Davarian

Comment/Question:

Does the 7 dB model agreement indicate RMS eror, absolute error, or some other kind of error?

Author/Presenter's reply:

This is the standard deviation of the difference in dB between each measured and predicted value of path loss.

Discussor's Name: G. S. Brown

Comment/Question:

Did you make any phase measurements and/or phase prediction using the models?

Do you plan to do such measurements and/or model predictions in the future?

Author/Presenter's reply:

We have not made phase measurements. We have no plans to make such measurements.

Military Applications of Site-Specific Radio Propagation Modeling and Simulation

Ronald M Bauman
ARPA ASTO
3701 N Fairfax Drive
Arlington VA 22203-1714
USA

1. Summary

New electromagnetic propagation code suitable for use in complex urban environments has recently been developed and demonstrated by a collaborative group of researchers [1]. These methods use a combination of two- or three-dimensional ray tracing reflection, diffraction, diffuse scattering, and transmission rather than electromagnetic field theory to solve the path loss problem [2].

Although still under development, initial versions of the code predict propagation in complex urban environments with useful accuracy. Looking beyond this research, this paper addresses potential military use of site-specific propagation models and simulations, focussing on its use in distributed interactive simulations (DIS).

2. Introduction

"Uncertainty will be the norm as the Army moves into the 21st century . . . the Army must be structured, trained, equipped, and prepared for maximum flexibility . . . the challenge today is to determine what array of capabilities may be needed to perform a broader range of requirements . . . "[3] The implications of this uncertainty for communication and intelligence and electronic warfare (IEW) planners are that they need to prepare for operations in a broader range of environments and with greater constraints than classical open field warfare.

"We are entering an era of smaller, mainly unconventional . . . conflicts, waged for the most part inside rather than across . . . national boundaries."[4] Recent military missions support this notion in that they have occurred in rolling open desert and in urban areas in the Gulf war, mountainous and urban environments in former Yugoslavia, small villages in Somalia, and in the jungles of South America.

Short range radio propagation models, such as Longley-Rice, TIREM or GELTI[5], work well in rolling, outdoor, uncluttered electromagnetic (EM) environments in supporting military communication and IEW warfare planning for ground forces.

The empirically-derived Hata model [6] is often used to estimate path loss in urban environments. Neither the Hata nor other previously mentioned models were designed to support communication and IEW planning in complex cluttered EM environments where site-specific communication performance may be essential to mission success. Indeed, until recently, a usefully accurate site-specific model seemed a hopelessly complex notion.

Lacking site-specific models, commercial cellular, PCS and other mobile communication network planners collect site-specific data empirically. This is costly and time - consuming and constrains planning flexibility. Military planners do not often have the option of using empirical site-specific methods.

Besides the military impracticality of empirical techniques, a confluence of other factors is stimulating the need for new, high fidelity, site

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

- specific propagation models and simulators that work in a wide range of cluttered EM environments. These include: new military roles and missions with emphasis on efficiency of resources and the need to extend communication connectivity to individual combatants [7], the attractiveness of applying cheap, commercial radio to military applications but with frequencies new to the military, loss of spectrum for military use, and the growing importance of (DIS) in supporting mission training, planning, rehearsal, and evaluation.

Given the need for site-specific models and simulations to support EM-dependent applications, there are significant technical challenges that must be addressed in developing this new capability. These include: development of robust and efficient EM code that can achieve timely solutions on workstation-level machines, rapid generation of accurate three-dimensional terrain and feature data bases to support propagation models, creation of standards to support transportability and interoperability of terrain and other data bases that interact with propagation models and simulations, validating and building confidence in the models and simulations, supporting real-time communication and electromagnetic sensor performance predictions within interactive, distributed simulations, creating a simple-to-use runtime architecture, and achieving accuracy over a wide range of frequencies of interest to IEW and communication applications.

These problems are becoming tractable because of recent advances in: electromagnetic modelling techniques which have recently and successfully been applied to predicting propagation in Rosslyn Virginia[8], workstation speed and data storage, interferometric synthetic aperture radar (IFSAR) for rapidly collecting and generating accurate digital terrain elevation data (DTED) bases, and DIS and terrain data base standards[9] which are based on open systems object- oriented software standards.

3. The DIS Latency Problem

DIS deals with two interrelated problems: the infrastructure that allows linking together of various types of simulations at multiple locations and creation of realistic, complex virtual worlds for simulating highly interactive activities.

Entities in the simulation, whether live or virtual. interact with each other by broadcasting protocol data units (PDU) as defined in IEEE 1278. PDUs describe an entity's state and entities can also cause events. Knowledge about events is broadcast through PDUs to all relevant participants. Tolerable latency for human interactions is 0.3 seconds. Thus, near real-time updates must be supported throughout the DIS network.

DIS is being evaluated for use in supporting mission planning and rehearsal for small units. "To be of value, communications-effects models and simulations must faithfully reproduce the radio connectivity and network effects that forces will encounter as they maneuver across the battlefield"[10]. Simulations for supporting small units demand high fidelity in terrain and feature data bases and in the physical models that affect communication. But demands for high fidelity and real time are contradictory.

High fidelity, i.e., site specific, models demand high-speed, high computing-power workstations. Models developed to date have focussed on prediction fidelity rather than runtime speed. As a point of reference, workstations require minutes to hours to solve small numbers of path predictions using models based on three dimensional shooting and bouncing rays (SBR) with multiple diffractions [11]. But DIS demands solutions within the 0.3 second human reaction time.

A potential way around the latency problem is to run site-specific models off-line for a large number of paths, build tables of path loss, and, where needed, multipath structure, for the areas of interest, and distribute tables to all exercise participants. During runtime, entities would look up path loss and other important parameters in their local tables, based on end points of the communication links of interest.

This technique may work when applied to constrained communication or IEW problems, e.g., cellular base station network design or transportable communication infrastructures or cases where at least one end of the radio path is fixed and known, and only one end of the path is mobile, or when communication antennas are below roof tops, thus avoiding the need to account for diffraction over a third dimension- the roof tops . Other constraints, special to a particular scenario, could be imposed to reduce table sizes, for example, quantizing areas of loss into a few groups. Without constraints, complex scenarios, involving mobile entities interacting with other mobile or fixed entities in the presence of large numbers of reflecting and diffracting surfaces, e.g., houses and other buildings, may require tables too large to be tractable.

Another way around the latency problem exploits the parallelism inherent in the SBR technique. A massively parallel processor would be used as a runtime server to solve path loss calculations in near real time as entities move about the scenario data base.

Given the fastest of the techniques used in the Rosslyn test requires on the order of 10s of minutes to solve several paths, the speed-up required to meet 0.3 second latency is on the order of several thousand. Using a massively parallel machine with the equivalent of 64,000 processors, speed-up of several thousand-fold may be realizable. Other super computer architectures might also be capable of a several thousand-fold speed-up.

4. Path Loss Runtime Architecture

Given that either a look-up table or super computer path loss server solves the DIS latency problem for site-specific predictions, we next address basic architecture issues involved in DIS use of site-specific prediction.

Most human-in-the-loop (HIL) DIS simulations assume perfect communication, which in turn, can lead to skewed results[12]. As to the value of adding communication detail to DIS, "The effects introduced into the C2 system by imperfect communications would have profound implications regarding human interaction with the system. The behavior of an operator may differ significantly when all information is received correctly and on time as opposed to cases where data is missing or corrupted".[13]

Early work in applying propagation predictions to DIS[14,15] used Longley-Rice[16] to calculate the path loss between tank simulators in a fully distributed architecture. The propagation model and terrain data base reside in each entity's host. Path loss calculations involve reading terrain elevation data between transmitter and receiver whenever the radio of interest moves more than 50 meters. Line of sight determination and a loss factor are calculated for each transmitter to receiver path.

Each entity broadcasts a transmitter, receiver, or signal PDU, which we denote TPDU, RPDU and SPDU respectively, to advise all others in the simulation of state changes in the transmitter, receiver or signal. The SPDU supports interactive digitized voice or data in packetized format.

This fully distributed architecture works well for small numbers of entities but does not scale to large numbers of entities nor where high power computing is needed. The limiting element is workstation processing power.

A basic scalable architecture for solving the site-specific path loss problem is shown in Figure 1. Terminal entities (TE) may be live or virtual, machine or human, stationary or mobile. Associated with each TE is a host computer and a terrain and feature data base. A path loss server (PLS) performs all processing-intensive calculations for a confederation of TEs. Terrain and feature data bases may be different for each type of entity and for different purposes, e.g., EM or visualization.

Humans may participate in DIS events as virtual TEs through an I-PORT[17] or as live TEs through a distributed, mobile embedded instrumentation network[18]. Machines controlled by humans may participate as virtual

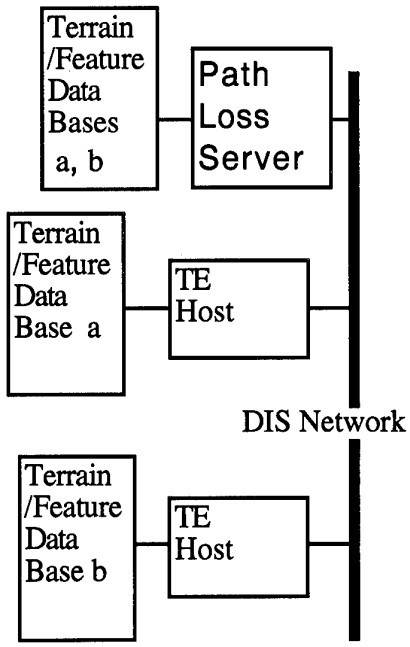


Figure 1. Path Loss Basic Runtime Architecture

TEs either through HIL simulators or as live TEs through a distributed, mobile instrumentation network. Computer generated human or machine constructive entities may also be TEs.

5. DIS Scenario

To illustrate how site-specific EM interaction might take place within DIS we construct a simple scenario. Assume an I-PORT virtual TE wishes to communicate with a live tank TE as an event within a larger DIS scenario. To stay within limits of current DIS technology, we assume that both the tank TE and dismounted infantryman I-PORT TE are in a city but not within visual range of one another. To simplify the example, we ignore details about communication protocols and signal processing details and focus on propagation-related events.

I-PORT initiates a voice transmission TPDU. At the same time, I-PORT also issues an entity state PDU (ESPDU)[19]. The local I-PORT host looks at the TPDU and creates a multicast group for appropriate TEs, based on I-PORT's specific transmitting EIRP as a function of I-PORT's posture, overall state (e.g., alive, with ample energy, located at X, Y, Z, moving easterly at 1kilometer per hour), antenna pattern, range of other TEs, types of other TEs, most recent other TE RPDUs (are other TEs receivers tuned to the proper channel?) and ESPDUs, and other filtering criteria aimed at minimizing network and other TE host loading.

If at least one potential TE is in the multicast group, the TPDU is sent through the DIS network to the PLS and all other relevant TEs in the multicast group.

TEs that receive the TPDU from I-PORT begin their calculations to determine how much energy and energy spread (if time and frequency dispersions are important and accounted for) they are receiving. In parallel, the PLS begins the SBR and refraction calculations for each I-PORT-TE path in the multicast group, based on TE position and motion vectors derived from ESPDUs as related to the terrain and feature data base.

Upon reaching a solution, the PLS transmits path loss (and possibly energy spread) and angle-of-arrival data to each TE in the original multicast group in the form of a path loss PDU (PLPDU).

Upon receipt of the PLPDU, each relevant TE completes its calculations to determine what action to take as a result of the I-PORT TPDU and the PLS PLPDU. The receiving TE may account for its antenna gain pattern relative to its local orientation and angle-of-arrival data. A quantized look-up table may be used to determine voice quality. Moreover, since this is a voice communication scenario, the receiving TE knows to expect additional TPDUs from I-PORT and to use that knowledge to either buffer or send data to an output device for the human tank operators.

PLS also knows this is a voice communication application and can anticipate, based on dead reckoning of the position and motion vectors contained in the ESPDUs of I-PORT and the other TEs in the relevant multicast group, what

paths in the near future are likely to be required for path loss calculations. Thus, the PLS can anticipate which areas of the terrain and feature data base are, or will shortly be, active and begin calculations in these areas.

In general, knowledge about the application, in this case voice communication, the motion and position of TEs within the terrain and feature data base, and the constraints associated with a particular type of TE (e.g., a tank TE is more likely to be in streets or at least at the terrain level of an urban data base than on tops of buildings), will provide the PLS with additional means to use its processing power efficiently.

We can create an EW scenario by substituting a live jammer for I-PORT. But, instead of launching real EM power, the jammer transmits PDUs. Simulated EM power would, in general, be much greater for a jammer, and, therefore, interference with a larger number of TE receivers is more likely, so the multicast group in this scenario would tend to be larger than in the communication case. Details about effects of received TPDUs would be different, but the process would be essentially the same as for the communication scenario.

Effects of jamming could be simulated in either real or virtual communications by creating bit error patterns in the victim TE data output stream or by other local means. Alternatively, the jammer signal could be constructed to form a special class of corrupted communication signal for processing by the victim TE receiver. In any case, EW and victim TE could be any combination of live and virtual.

Similar strategies may be applicable to signal interception and direction finding (DF) missions . In this case, the intercept receiver TE would receive simulated signals via PDUs in the same way as in the communication scenario. Control panel settings on the real or virtual intercept receiver, in combination with look-up tables, would determine output signal or angle of arrival information. Some missions may require local complex algorithmic processing instead of or in combination with table look-up. For this special case, the intercept host may need to be unique but otherwise conform to DIS standards for interoperability.

6. Terrain and Feature Data Base Acquisition

Host and PLS processors are challenged by the combination of DIS and site-specific prediction processing load. Therefore, minimizing data base complexity is worthwhile. One way to do this is to require no more detail than is essential for a given TE's data base. Different versions of the data base may be resident in a TE's host, denoted in Figure 1 by small case letters. A data base designed to support a TE operating at 10GHz may contain more detail than is needed for operation at 10MHz.

More often, too little rather than too much terrain and feature data is the problem. Highly accurate three dimensional terrain and feature data bases are not generally available[20]. Although photogrammetric techniques have been successfully applied to site-specific propagation problems they are manually intensive and slow. However, the Advanced Research Projects Agency (ARPA) is developing techniques to rapidly and automatically produce digital terrain elevation data bases [21].

High fidelity photogrammetric data bases are obtained by aircraft overflight of the area of interest. As a photographic technique, this process requires good lighting, good weather, and low overflight altitude (a few thousand feet). These requirements may be difficult to realize in military operations, especially where quick reaction is needed.

Interferometric Synthetic Aperture Radar (IFSAR) is an emerging technology that has potential to produce high fidelity digital terrain elevation data bases[22]. IFSAR overflights can be much higher and can tolerate greater standoff distances than photogrammetric missions. Moreover, X- band radar used in the JPL IFSAR system is robust to mild weather and can be operated in darkness.

Accuracy of the resulting digital elevation map of the Fort Irwin area is on the order of 1 to 3 meters and can be improved.TOPSAR has also been used to generate DEM of urban areas at nominally 100 square kilometers per hour[23].

Because of its robustness, accuracy and acquisition rate, IFSAR holds promise as a means to rapidly generate digital elevation maps (DEM) of areas of interest.

Whether by photgrammetric, IFSAR or other means, acquiring DEM is the first step in a process to prepare a data base for use by the propagation code. Figure 2. depicts a simplified flow diagram of a process to produce a terrain and feature data base ready for interaction with propagation code.

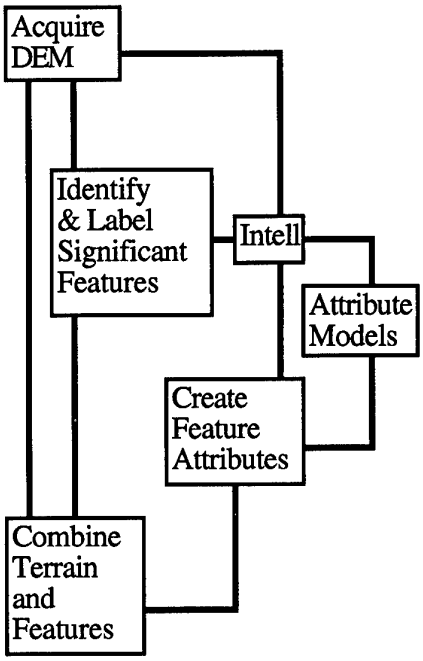


Figure 2. Terrain and Feature Data Base Process

The process begins with an assignment of a photogrammetric or IFSAR mission to acquire a DEM of given resolution and accuracy for an area of interest. DEM includes the terrain and features on the terrain.

Next, features of importance to EM propagation, e.g., buildings, streets and trees, must be identified, and feature facets labeled. EM parameters, e.g., reflection coefficient, must be assigned to each facet.(This corresponds to surface texturing in the visualization data base process and might thereby be called radio or EM texturing). Attribute assignment may require input from separate data bases and may require human judgement.

For example, where regions of the DEM include forest or mountains, other modular models specifically designed to treat these specific types of features [24], [25],may be introduced into the process. Correlation of EM attributes with intelligence surface feature data bases may be helpful in speeding up or automating the process.

Geophysical and meteorological effects[26] may also have to be accounted for by other modular models. Although these effects are not attributes of the urban features, they may affect long ray paths. Bodies of water and littoral boundaries may also have to be accounted for with special modules.

Once relevant facets and their attributes are labeled, and other relevant propagation-related effects are accounted for, they can be combined with the terrain data base, which might be in the form of triangulated irregular networks[27] often used to represent surfaces for visualization. A combined terrain and radio-textured feature data base is needed to interact with the propagation code.

7. Military Communication Models

Military planners are usually interested in communication performance rather than path loss or signal dispersion. Depending on its use, the communication model or simulation would take different forms.

For realtime DIS events, path loss and dispersion are appropriate. As addressed earlier in this paper, the nature of received signals is

determined locally from information derived from the DIS network. Realtime DIS events include mission rehearsals, mission evaluation and joint training events.

For mission planning, communication performance in specific geographic areas must be determined for a set of specific communication assets e.g., nomenclatured radios and antennas. In this case realtime processing is not necessarily a requirement and, therefore, longer solution times may be tolerable.

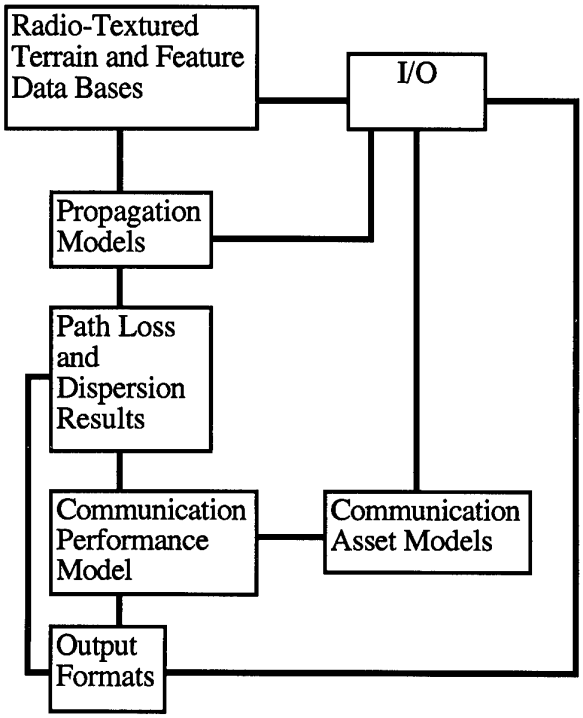


Figure 3. Communication Model

The mission planning model might take the form of Figure 3. In running the model, the mission planning operator determines the geographic area of interest, the communication assets which either are to be used or evaluated and the desired output format.

The output format may take the form of a signal quality map of the area of interest, or it might highlight the differences in performance between alternative radio assets, or for a given radio asset the differences in performance as a function of frequency, or optimum locations for a set of transmitter sites.

The output display or record may indicate a quantization of the communication performance, e.g., bit error rate or signal-to-noise ratio, or "go, no-go" areas.

8. Military Missions

Military planners, in deciding where to locate bases, use a combination of elevation maps, physical and political attributes of perspective sites. Site-specific propagation models would be an additional tool to help decide base locations and in predicting areas of communication coverage or outage. Being able to communicate reliably with each element of a task force is vital to situational awareness.

In Somalia, lack of situational awareness was often attributed to lack of communications. In this peacekeeping mission, accurate knowledge of the location of poor communications would have been valuable to planners and to the tactical forces.

In areas like the former Yugoslavia, knowledge of the border location in this area is vital to the peacekeeping mission and knowledge of the location of peacekeepers is also vital. As borders are generally unmarked in this region, a system called "Soldier 911" was developed to keep track of peacekeepers patrolling the border and their position relative to the border. Through use of Global Positioning Satellite (GPS) receiver and a radio to report-back position, the system warns the soldier as he or she approaches the border and also provides voice connection with headquarters.

Unplanned-for radio or GPS outages could lead to border breeches. Thus, the ability to accurately predict radio and GPS signal quality is important and of current interest in planning military missions.

Also of current interest is the ability to predict radio navigation performance in urban environments. Tests in New York city revealed

that the LORAN magnetic field is generally more available than GPS[28]. A combination of the two radionavigation systems, an accurate clock, and an inertial system can significantly increase the availability of good navigation fixes throughout the city [29].

A model built by the US Army Electronic Proving Grounds[30] predicts the number of paths from an urban location to GPS satellites. This model also requires detailed urban, suburban and mountainous terrain and feature data bases to support site-specific navigation performance.

9. Future Developments

Current work in site specific prediction[31,32] addresses nominally 800 MHz to 2000MHz. Military communications must be able to compatibly operate within world-wide frequency allotments which, in some areas of the world are established, but in others are not[33].

Tactical military communications extends from low VHF to microwave frequencies near the infrared. Moreover, radio navigation systems of potential interest to urban tactical missions work at frequencies near 200kHz. EW sensors and transmitters operate over a similar frequency range Thus there is need to extend the current work to cover the broad range of frequencies of tactical interest to the military. As a minimum, it is important to understand and characterize the frequency limits of current site-specific models.

Spectrum for military use is dwindling and therefore efficient use of remaining spectrum is of growing importance. Solving military cosite interference problems is often at odds with the notion of efficient spectrum management, although much effort is devoted to balancing the two problems[34].

Cosite interference is often addressed by avoiding use of precious but otherwise useful spectrum. In urban environments, classical cosite avoidance techniques would not make efficient use of spectrum because uncertainties about signal strength are large. Site-specific prediction tools would be helpful in determining frequency reuse in urban areas by identifying, with some degree of precision, signal strengths over the area of interests. With this knowledge, interference can be avoided by integrating site-specific signal strength information with classical frequency planning codes.

High altitude platforms, e.g., satellites, unmanned air vehicle (UAV) or other aircraft, can be used as relays or base stations to support range extension of tactical communications or connectivity within an area of interest. Site-specific propagation code should be extended to evaluate communication and EW performance between mobile or fixed TEs in the urban environment and high altitude platforms.

Obtaining high fidelity urban DTED with correlative EM attributes is manually intensive, slow and expensive. Research in automating the process of feature extraction for visual use [35] promises to address these problems. Although the research focusses primarily on supporting advanced DIS for human visualization, it appears that it might also be extended to support EM simulations in urban environments.

Although this paper addresses only urban outdoors EM modeling applications, clearly there is also a need to extend this work to address urban indoors, and indoors-outdoors models. The techniques already established for the outdoors case may be more than adequate to handle the indoors case. However, obtaining high fidelity building-specific EM data bases may present the military a formidable problem.

10. Summary and Conclusions

This paper addresses potential uses of site-specific propagation code to support tactical military planning, training and operations. Non-real time applications of site-specific code, such as mission planning, communication planning, communication, intelligence or EW system evaluation, can be supported with workstation- class processors. The several

hours of processing required to plan missions in complex urban areas may be acceptable in these cases.

However, for real time applications, e.g., human-in-the-loop DIS training or mission rehearsal, or even real-time communication support to real missions, significant speed-up in run time code and faster processing is required. The use of a massively parallel or other super computer as a real-time server is proposed as a DIS network element for solving DIS latency requirements. A notional runtime architecture is developed for supporting real - time DIS-based simulations.

Live training, test and evaluation or mission planning events, especially in the EW case, gain an added degree of flexibility by simulating rather than emitting real EM energy. Site-specific DIS-based simulations can support realistic training of EW or signals intelligence personnel without the need for radiating either high power or classified signals.

Realistic training and rehearsal can be supported for individual combatants through the use of I-PORTs equipped with DIS-based communication. The vision for such capability is contained in General (Retired) Paul F. Gorman's paper [36]. Soldiers or Marines in I-Ports would enter and move through a virtual representation of a real city and rehearse their mission with realistic communications, ranging from perfect, to slow fading, to periodic outages and reconnections, simulated by site-specific communication models running in real time.

Thus, trainees would experience realistic communications, and designers, working with the trainees, would learn where improvements are required.

11. Acknowledgement

The author wishes to thank Lt. Col. Robert Kocher, USA, for providing insights on the importance of maintaining command and control with the individual combatant in supporting missions in Somalia and the former Yugoslavia, and for information on operational use of "Soldier 911".

12. References

1.Willis, T.M. *et al.*, "UHF and Microwave Propagation in an Urban Environment", AGARD SPP Symposium, September 1995, Paper 8.

2. Rappaport, T., *et al.*, "Radio Propagation Prediction Techniques and Computer-Aided Channel Modeling for Embedded Wireless Microsystems", Quarterly Report, Virginia Polytechnic Institute and State University, MPRG-TR-92-10, June 1, 1992.

3. Sullivan, General Gordon R. and Coroalles, Lieutenant Colonel Anthony M., "The Army in the Information Age", Strategic Studies Institute, Carlisle Barracks, PA., March 31 1995.

4. Record, Jeffery, "Ready for What and Modernization Against Whom? A Strategic Perspective on Readiness and Modernization", Strategic Studies Institute, Carlisle, PA, April 10 1995.

5."Electromagnetic Environmental Test Facility (EMETF) Propagation Path Loss Modeling Capability Requirements Analysis Report" TECOM Project No. Publication No. EMETF-91-04-003, 26 May 1991.

6. Hata, M., "Empirical Formula for Propagation Loss in Land Mobile Radio Services," IEEE Trans. on Veh. Tech., VT-29, No. 3, August 1980. pp.317-325.

7. "SOLDIER 911 SPEEDS HELP TO THE FIELD", Federal Computer Week, P. 18, July 17, 1995.

8.Willis, T.M., *Ibid.*

9. Birkel, Paul, *et al.*, "Synthetic Environment Data Representation and Interchange Specifications", 12th Workshop on Standards for the Interoperability of Distributed

Simulations, Institute for Simulation and Training, March 1995.

10. Brockel, K.H., *et al.*, "Communication Realism for Digitized Battlefield Simulations Involving C2 On-the-Move" MILCOM '94 PREPRINT.

11. Luebbers, R.L. and Schuster, J., "Site-Specific Radio Propagation Prediction Methods for Urban Environments" , AGARD SPP Symposium, September 1995, Paper 41.

12. Siliato, J.M, *et al.*, "Modeling and Simulation to Support the Digitized Battlefield", MILCOM '94 Preprint.

13. Siliato, J.M., *Ibid.*

14. Gonzales, J.J., "Electromagnetic Propagation Modeling for Distributed Simulation", BBN, Cambridge MA 02138, Presented at the Interservice/Industry Training Conference, Orlando FL, 1991.

15. Pope, A.R., *et al.*, "SIMNET Simulation of Radio Communication: A Testbed for Investigating Combat Vehicle C^{3}I Technology", BBN Advanced Simulation, Report No. 7352, July 1990.

16. National Technical Information Agency. "A Guide to the Use of the ITS Irregular Terrain Model in the Area Prediction Mode", NTIA Report 82-100.

17. Gorman, P.F., "Supertroop via I-PORT: Distributed Simulation Technology for Combat Development and Training Development", Institute for Defense Analysis, Paper P-2374, August 1990.

18. Goblick, T.J., *et al.*, "Technical Requirements and Feasibility Analysis of 21st Century Live Play", MIT Lincoln Laboratory, Draft Report , 27 April 1995.

19. P1278.2 C/DIS) "Standard for Distributed Interactive Simulation: Communication Services and Profiles", The Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ 08855-1331.

20. Rappaport, T., *Ibid.*

21.Lukes, George E., "Cartographic Support for Advanced Distributed Simulation", Proc.SPIE, Vol. 2486, Society of Photo-Optical Instrumentation Engineers, April 1995.

22. Madsen, S.N., *et al.*, "Analysis and Evaluation of the NASA/JPL TOPSAR Across-Track Interferometric SAR System", IEEE Transactions on Geoscience and Remote Sensing, VOL.33, No.2, pp 383-391, March 1995.

23. Madsen, S.N., *Ibid.*

24. Welch, C. *et al.*, "A Model for Estimating Electromagnetic Wave Attenuation in a Forest (EWQAF) Environment", AGARD SPP Symposium, September 1995, Paper 11.

25. Kirlin, R.L., *et al.*, "Measured 900MHz Complex Impulse Responses in Mountainous Terrain:Relationship to Topographical Map Data", Proceedings,Wireless Personal Communications, June 9-11, 1993 Virginia Polytechnic Institute.

26. Brockel, K.H., *et al.*, "Tactical Line-of-Sight Path Reliability: Propagation Climate Factors", U.S. Army CECOM, Ft. Monmouth , N.J., Technical Report CECOM-TR-92-9, October 1992.

27.Rappaport, T., *et al.*, "Radio Propagation Prediction Techniques and Computer-Aided Channel Modeling for Embedded Wireless Microsystems", Virginia Polytechnic Institute, DARPA Quarterly Report MPRG-TR-92-18, September 1, 1992.

28. Peterson, B.B., and Miller, L.C., "Evaluation of Radionavigation Systems in an Urban Environment", U.S. Coast Guard Academy, New London, CT, Report 14-94, December 1994.

29. Peterson, B.B. and Miller, L.C., *Ibid.*

30. Wheeler, Raquel, *et al.*, "Analysis Software Environment User's Manual and Description", US Army Electromagnetic Environmental Test Facility (EMETF), 31 Dec 94.

31. Willis, T.M., *et al., Ibid.*

32. Bertoni, H.L., et al., "UHF Propagation Modeling for Wireless Personal Communications", Polytechnic University, CATT 93-69, December 1993.

33.Major, P.A. and Bohn, F., "Frequency Allotment Capability Concept Development Study", MILCOM '94 Preprint.

34. Brockel, K.H., *et al.*, "CECOM Application of E3 A Process Focused on Owning the Spectrum", MILCOM '94 Preprint.

35. Polis, Michael F., *et al.*, "Automating the Construction of Large-Scale Virtual Worlds", Computer, IEEE Computer Society, Vol.28, No.7, July 1995.

36.Gorman, Paul F., General USA (Retired), "Vision", Presentation to the Dismounted Battle Space Battle Laboratory, Fort Benning, GA, 15 February 1994.

DISCUSSION

Discussor's name: J. Harvey

Comment:

For the US Army the problem of modeling propagation in a complex terrain environment is further complicated by the requirement for a distributed communication net. Empirically "running the terrain" will work for a fixed base station system, but for many moving stations distributed throughout the terrain it is necessary to model and understand the details of the propagation.

Discussor's name: F. Davarian

Comment/Question:

For satellite application what frequency will the model be extended for?

Author/Presenter's reply:

Frequencies in the military communication band UHF, C Band, K band, and commercial frequencies of interest to the military.

DISCUSSION

Discussor's name: D. Yavuz

Comment/Question:

This question is for the presenters of papers 6 and as well.

Are there any chaotic/fractal effects that are anticipated since there are highly non-linear relationships involved in the processes modeled. In other words are you worried about "butterfly effects".

Author/Presenter's reply:

P.G.V. Charriere & K. Craig: (Paper 6)

You can predict fast fading effects around a given position of a receiver (within a radius of a few wavelengths at most). This method assumes that in this area that the paths connecting receiver and transmitter vary smoothly. It is certainly true that approximations are made (e.g. ignoring diffraction at millimetre waves) which can produce discontinuities in the result as initial conditions are changed (eg: change of transmitter position). However, this is not true chaotic behaviour, and we believe that the errors are bounded and quantifiable.

R. Luebbers: Paper (8)

We are not attempting to predict fast fading. We have not observed that small errors in the building geometries can produce large errors in the predictions.

R. Baumann: (Paper 9)

We have therefore not undertaken a formal non-linear analysis of the errors we have measured. That is not to say we may not do so, if we observe say evidence of chaos as the work progresses.

A Model for Estimating Electromagnetic Wave Attenuation in a Forest (EWAF) Environment

Conrad Welch
Catherine Lemak
Computer Sciences Corporation, Systems Engineering Division
3160 Fairview Park Drive, Falls Church, VA 22042
United States of America

Lee Corrington
The Electronic Proving Ground
Fort Huachuca, AZ 85613-7110
United States of America

SUMMARY

For decades, communicators have been concerned with radio frequency (RF) propagation path loss through foliage. This paper describes the development and implementation of the Electromagnetic Wave Attenuation in a Forest (EWAF) propagation path loss model, which uses conventional inputs to estimate plausible foliage losses. The EWAF is an elementary, heuristically based foliage propagation path loss model which represents a forest as a dissipative dielectric slab lying in a more lossy half-space representative of the ground. Modeling the foliage by a dielectric slab will provide an estimate of foliage path loss in excess of free space (EPL) expected in situations where meager information is available regarding the wood density and other electrical parameters of the forest. The EWAF model uses empirical foliage path loss information as a basis and is therefore closely coupled to actual path losses encountered in real-world situations. The EWAF model is not a rigorous attempt to solve the electromagnetic wave equations at the boundaries of the different representative dielectric slabs. Rather, EWAF is essentially a family of curve-fitting algorithms based on previously measured data for sparse and dense forests in the United States and Germany, as well as very dense forests in India, where the mean annual rainfall is about 3000 millimeters.

The EWAF model includes several features: the effects of antennas within the forest, outside the forest, and above the forest; wave polarization; forest density; canopy, trunk, and undergrowth losses; antenna beamwidth; wet foliage; lush foliage; and other physical/forest conditions. It is presently resident on a Sun SPARC network and provides a good estimator for EPL due to foliage anywhere in the world.

1. INTRODUCTION

Over the past 30 years, measurements and analyses have shown that actual losses encountered within forests are considerably less than had been estimated for propagation paths straight through foliage. Pounds and LaGrone[1] hypothesized in 1963 that propagation loss through foliage could be represented by propagation through a dielectric slab, where the antennas are submerged in the dielectric media.

Tamir[2,3] expanded this concept by developing the mathematics describing the dielectric slab propagation phenomenon. The literature contains much measured data for propagation losses through, around, and in the vicinity of foliage. Although mathematical models have been prepared that describe this propagation, very little has been done to provide communicators with a computerized tool for predicting foliage losses.

This paper describes a first step in the development of the EWAF computer model, which uses plausible inputs to estimate foliage losses. The model can be used by communications engineers to obtain an estimate of propagation losses through a forest or other foliage situations expected for a given set of operating conditions. The estimates will be useful for communications planners in situations where transmitter and receiver antennas (or perhaps only one) will be immersed in foliage. Representing the foliage by a dielectric slab will provide an estimate of foliage losses which could be expected in situations where meager information is available regarding the wood density.

The information available on propagation loss through foliage sometimes refers to the area in square feet of wood per acre of forest and assumes a uniform density throughout the forested region. In most instances, it is not possible to measure the wood density per acre. Furthermore, for large forested areas, the forest density is quite variable.

The EWAF model is not a rigorous mathematical treatment of a forest represented by a dielectric slab. It is actually based upon previously measured data for sparse and dense forests in the United States as well as very dense forests in India, as described by Tewari et al.[4], where the average annual rainfall ranges to 3000 millimeters. The model was implemented using Ada programming language and is part of a larger graphic user interface interactive engineering tool, the Analysis Software Environment. The model is resident on a Sun network at the Electronic Proving Ground, Fort Huachuca, Arizona, of the White Sands Missile Range.

2. PROPAGATION THROUGH A DIELECTRIC SLAB

Radio (or electromagnetic) waves passing through foliage represented by a lossy dielectric slab commonly incur path loss rates of perhaps 0.25 decibels per meter (dB/m) in excess of free space [referred to as excess path loss (EPL)]. For path lengths of 1 kilometer, for example, the attenuation of a wave due to foliage would be 250 dB. Communication over radio links having such losses would usually be impossible or very difficult. However, experimental results indicate that the EPL due to foliage is actually much less, more like 40 or 50 dB at 400 megahertz (MHz).

These differences between expected and actual results led

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

Pounds, LaGrone, and Tamir to the dielectric slab theory, where the wave travels up through the medium (incurring high losses) and then travels laterally along the boundary region at a much lower loss rate. Figure 1 depicts the propagation path geometry. This propagation in the boundary region is called the lateral wave and allows communication over much greater distances than would be possible otherwise.

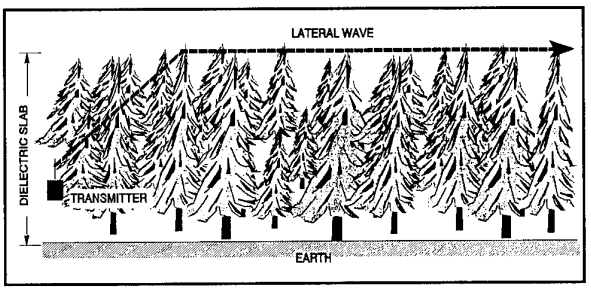


Figure 1. Propagation path along a dielectric slab

When the RF propagation is compared to optics and the refraction encountered when a wave passes from a more dense medium to a less dense medium, the geometry becomes equivalent to Figure 2. (The more dense medium is the forest, the less dense medium is the air.) A large portion of the wave escapes through the boundary into space for high angles of incidence with the boundary region. As the angle of incidence decreases, the refracted wave at the surface becomes more nearly parallel to the boundary. At a critical angle of incidence, the wave travels parallel to the air-forest boundary and forms the lateral wave, which incurs fairly low loss and permits communication over extended distances. Associated with the lateral wave is a lossy wave, called the leakage field, that travels back down into the dielectric slab. The leakage field provides the route down to the receiving antenna, which may be submerged in the forest at some distance from the transmitting antenna. The leakage field is mostly attenuated by the forest and the underlying ground and, for practical purposes, does not reflect back out of the forest. The receiving antenna may be within the forest, above the forest, or beyond the forest. The model handles these cases when the distance from the forest is not more than two or three forest heights (see para H). Below the critical angle of refraction, the wave is reflected back into the lossy dielectric and is highly attenuated by the foliage and earth beneath.

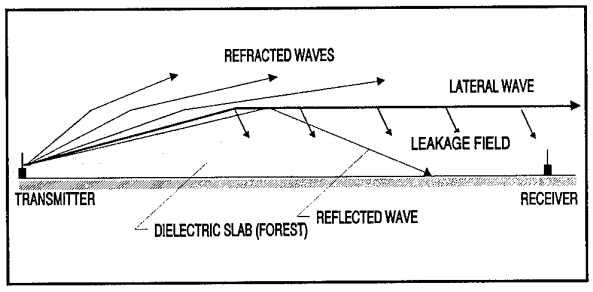


Figure 2. Refraction, reflection, and lateral waves and leakage field due to differences in medium density

3. ESTABLISHING FOLIAGE LOSS RATE

The relationship between the conductivity of the forest slab and the through-foliage loss for various frequencies is represented by the equation:

$$L_b = -20 \log e^{-2\sqrt{\sigma f}} \quad (1)$$

where

L_b = loss rate in dB/m (in the forest)
e = 2.71828
σ = conductivity in siemens
f = frequency in MHz.

This is the theoretical rate of attenuation of radio waves passing through a thin screen of trees as derived from Tamir's work by F.A. Losee[5] and further evaluated by Kivett and Diederichs[6]. This represents the propagation losses experienced by a direct wave through foliage and has close agreement with empirically collected data. The family of curves for various conductivities representing this loss is shown in Figure 3. The conductivities are directly related to the densities of the forested region. The top curve represents a very dense forest and the bottom curve represents sparse forest conditions. In the EWAF model, the engineer can select the forest density (represented by the conductivity) of the area for which estimates are desired.

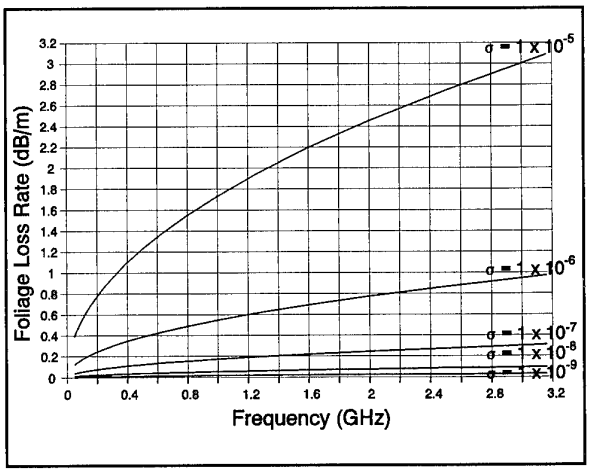


Figure 3. Frequency versus loss rate for several conductivities

Data are not available for all effects at all frequencies. Most effects are known at certain frequencies; for example, recent rain on a certain forest at 437 MHz was known to cause 3-dB increase in path loss over a known path. This can be equated to a specific increase in the conductivity of the forested region. Solving Equation 1 for conductivity yields:

$$\sigma = \frac{.0033136 L_b^2}{f} \quad (2)$$

where

L_b = loss rate in dB/m
σ = conductivity in siemens
f = frequency in MHz.

Equations 1 and 2 can be used to calculate conductivity at a certain frequency from a known path loss under particular foliage conditions and, from this value, determine the path

loss at other frequencies for these same conditions. The EWAF model makes extensive use of this technique to evaluate the effect of the density of the undergrowth, leaf moisture content, rain-wet foliage, and other conditions of the forested region which change the conductivity of the dielectric slab. The lateral wave loss rate, typically one-tenth to one-fifteenth of the foliage loss rate, is similarly adjusted.

While Equation 1 fits most of the empirical data, some empirical data may show greater or lesser changes with respect to frequency. Some are not hard data and may have been taken at a limited number of frequencies (as little as two). Such a small sample may not be representative of the true variation of foliage loss rate with change in frequency. Often, an author will have a limited amount of data and will draw a foliage loss curve through the data without considering the normal signal strength variations. Such a curve cannot be satisfactorily extrapolated to other frequencies.

4. EFFECT OF WAVE POLARIZATION

Most of the available empirical data show an increase in foliage loss for vertically polarized electromagnetic waves. The Consultive Committee on International Radio (CCIR) has issued a recommendation[7] which includes the added losses encountered when using vertical polarization. The recommendation covers the propagation losses for "an approximate average for all types of woodland." The CCIR data are plotted in Figure 4.

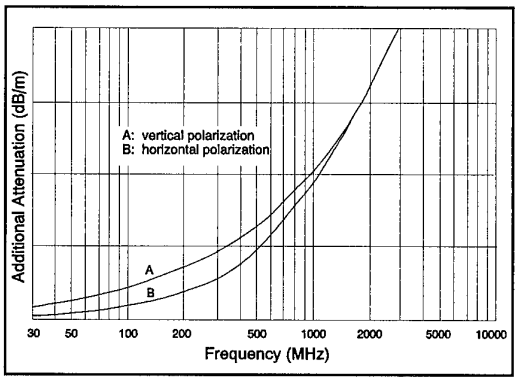


Figure 4. Specific attenuation of woodland (CCIR 1992)

The CCIR recommendation regarding polarization was incorporated into the foliage loss model for frequencies below 1.3 gigahertz (GHz). The EWAF model improves somewhat on the CCIR recommendation by allowing for differences in vegetation through changes in conductivity of the dielectric forest slab.

5. ANTENNA BEAMWIDTH EFFECTS

The beamwidths of the transmitting and receiving antennas play a significant role in the additional losses experienced due to propagation through foliage. This is particularly true for microwave links or tropospheric scattering links. When microwave or tropospheric scattering links are established within a forest, planners must take into account that the maximum range will be considerably reduced from operation not in a forest. The increased path attenuation calculation in the EWAF model is based upon the antenna response characteristics off-boresight for narrow beamwidth antennas.

6. FOLIAGE WATER CONTENT

Most researchers who collect empirical data in the field conclude that the attenuation in a forested region is primarily due to the amount of moisture on or within the vegetation. To quantify this effect, most data collectors accompany the data with comments regarding the lushness of the vegetation and the presence or absence of moisture on the foliage. The preparers of the EWAF model gave special attention to these parameters. Within the model, the conductivity of the dielectric slab is adjusted to reflect comments regarding moisture on the foliage, lushness of the canopy, and lushness of the undergrowth to accurately reflect the effect of the estimated overall water content.

7. EFFECT OF FOREST CANOPY, EXPOSED TREE TRUNKS, AND UNDERGROWTH

The density of the forest canopy, the volume of the wooded trunks, and the density of the undergrowth all affect the propagation path losses within a forest. Forests with relatively light canopies will transmit sunlight to the trunks and to the ground. The additional sunlight penetration into the forest tends to increase the foliage on the lower parts of the trees and increase the undergrowth. For example, Germany's Black Forest has a dense canopy, trunks that are not foliated, and no undergrowth. The result is a dark cavern carpeted with pine needles fallen from the high canopy. Other forests, like those in the panhandle of Florida, have such lush undergrowth that it is difficult to simply walk through.

The EWAF model allows the engineer to specify the density of the canopy, the tree trunk exposure, and the height and density of the undergrowth. The treatment of these layers within the dielectric slab, while not rigorous, does allow some consideration of these factors. In the future, the EWAF model will allow a rigorous treatment of these forest layers.

8. PROPAGATION OUTSIDE THE FOREST

The EWAF model is intended primarily to provide an estimate of the electromagnetic propagation path losses through a forested region when both the transmitting and receiving antennas lie within the forest. Occasionally, however, either one or the other of the antennas is above the forest or beyond the forest. Empirical data are available to assist in the estimation of path losses extending out of the forest. In the case of antennas above the tops of the trees, data are available relating the slope angle of the path from the top of the trees to the antenna as related to the lateral wave loss rate. These slope angles are very small in most cases and involve propagation in the lateral wave mode for an appreciable portion of path.

Transmission beyond the edge of the forest, on the other hand, requires a transition from the lateral-wave mode to a free-space mode which resembles wave diffraction over an obstacle. Transmission beyond the edge of the forest is highly dependent upon the diffraction angle from the lateral wave (parallel to the ground) to the receiving antenna near the ground. Figure 5 shows the transition from the lateral-wave mode to the free-space mode at a diffraction angle.

9. MODEL INPUT VARIABLES

The model accepts a wide variety of input parameters which describe the physical conditions under analysis. These

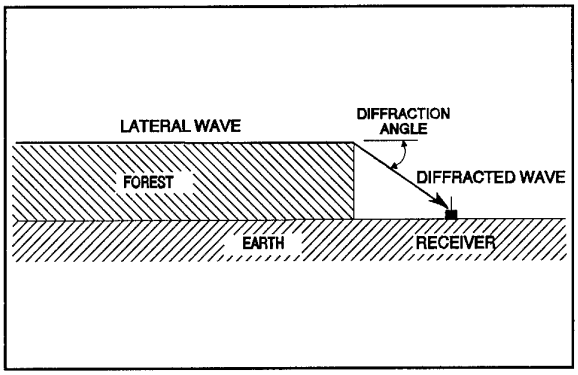


Figure 5. Lateral-wave transition to free-space propagation

parameters allow the model to ascribe physical constraints to the simulation and provide estimates of propagation loss through foliage commensurate with the scenario. These parameters are:

- Carrier Frequency
- Forest Density
- Type of Foliage (conifer/deciduous)
- Height of Forest
- Height of Transmitting Antenna
- Height of Receiving Antenna
- Height of Exposed Trunks
- Receiver Distance to Forest (or in forest)
- Transmitter Distance to Forest
- Presence of Moisture on Foliage
- Density of Undergrowth
- Polarization of Electromagnetic Wave
- Lushness of Forest
- Height of Undergrowth
- Receiving Antenna Beamwidth
- Transmitting Antenna Beamwidth

The term "forest density" is used in the EWAF model as the input to establish propagation parameters. This parameter is stated in terms of visual observation of the forest as light, medium, dense, very dense, jungle, and heavy jungle. These values are then equated to the conductivity of the forest.

10. CONCLUDING REMARKS

The EWAF model provides the design engineer or communicator with a tool which can be used to estimate probable propagation path losses through vegetation. The model is not mathematically rigorous but instead is based upon empirical data which produce results surprisingly similar to losses which are experienced in the field. It requires input parameters which are available or can be obtained without field measurement. Constants and factors are contained in tables easily changeable for fine tuning the model as more empirical data are analyzed. The prediction of electromagnetic wave propagation path losses through layers of the forest could be improved with additional analyses.

11. ACKNOWLEDGEMENTS

The authors wish to thank Dr. Joseph Whalen at Computer Sciences Corporation for his support and many informative discussions on electromagnetic propagation; Dr. Stephen Fast of the University of Texas for his help, support, and interest in the project; Mr. Daniel Searls of the Electronic Proving Ground for his help, encouragement, and support; and Mr. Jerry Burchard and Ms. Lora Langworthy for their diligent efforts in preparation of the manuscript.

REFERENCES

1. D.J. Pounds and A.H. LaGrone, "Considering Forest Vegetation as an Imperfect Dielectric Slab," Electric Engineering Research Laboratory, University of Texas, Austin, Report 6-53, May 1963.

2. Theodor Tamir, "On Radio-Wave Propagation in Forest Environments," IEEE Transactions on Antennas and Propagation, Vol. AP-15, No. 6, November 1967.

3. Theodor Tamir, "Radio Wave Propagation Along Mixed Paths in Forest Environments," IEEE Transactions on Antennas and Propagation, Vol. AP-25, No. 4, July 1977.

4. R.K. Tewari, S. Swarup, and M.N. Roy, "Radio Wave Propagation Through Rain Forests of India," IEEE Transactions on Antennas and Propagation, Vol. 38, No. 4, April 1990.

5. F.A. Losee, "The Influence of Trees on PLRS Communication Capability," Hughes Aircraft Company, July 1978.

6. J.A. Kivett and P.J. Diederichs, "PLRS Ground-to-Ground Propagation Test Technical Report (Draft)," Hughes Aircraft Company, January 1980.

7. Consultive Committee on International Radio, Recommendation 833, "Attenuation in Vegetation" (Question 9/5), International Telecommunications Union, Geneva, Switzerland, 1992.

DISCUSSION

Discussor's name: K. H. Craig

Comment/Question:

We have made measurements of attenuation rates in foliage at millimetric frequencies (38 GHz). Of course, at these frequencies multiple scatter gives the dominant signal rather than lateral wave effects. However, for our modelling, we need the electrical constants of the foliage. Would your equation (1) linking conductivity and attenuation rates apply at these frequencies?

Author/Presenter's reply:

The EWAF model was developed using a large set of measured data. There is nothing in the measured data to indicate a change in the conductivity of a forest medium with a change in frequency. The main problem is that at 300 MHz, the wavelength is fairly long (1 meter) and the forest is somewhat homogenous. For the model development, the highest frequency data immediately available was 3000 MHz (wavelength = 10 cm). The homogeneity of a forest at 3000 MHz is quite marginal. Extrapolating the data to 38 GHz may not be meaningful since the medium is probably not homogenous. AT 38 GHz the forest would effectively be "clumps" of high conductivity material separated by large areas of near zero conductivity. It is clear from the data, however, that any foliage would present a high loss rate at 38 GHz.

Discussor's name: G. S. Brown

Comment/Question:

Does your model distinguish between the coherent and incoherent components? This is important because Tamir's analysis applies only to the coherent field.

Author/Presenter's reply:

The EWAF model was prepared by using measured data and developing a set of curve fitting algorithms. The energy arriving at the receiving antenna is effectively the sum of all the components of the wave. No attempt was made to analyze or separate the wave into its components. The output of the model is simply an estimate of the total power arriving at the receiving antenna, as estimated for measured data.

DISCUSSION

Discussor's name: U. Lammers

Comment/Question:

Treetops should make a very rough dielectric boundary, even at a frequency as low as 400 MHz. Is it possible that scattering or diffraction between treetops are responsible for increased signal levels, rather than wave guidance along the dielectric interface?

Author/Presenter's reply:

There is no doubt that as the wavelength gets smaller in comparison to the size of an individual tree-top, the dielectric slab theory becomes questionable. Tamir himself questions the slab theory at higher frequencies. Could the mechanism be scattering of diffraction? Perhaps, but multiple diffraction and scattering occurrences do not seem to provide enough energy to explain the lower loss rates indicated by the data.

Atmospheric path loss and scintillations affecting communications in the EHF frequency range

H.H. Fuchs
R. Makaruschka
Forschungsinstitut für Hochfrequenzphysik der FGAN e.V.
Neuenahrerstr.20, D-53343 Wachtberg-Werthhoven, Germany

ABSTRACT

With the introduction of millimeter wave bands for satellite communications attenuation effects in the troposphere become adversely important. To investigate these propagation phenomena especially due to clouds an experiment has been set up on a 9420 m slant line-of-sight path in the German Alps between Garmisch and the top of the Zugspitze. The measurements were made at the frequencies of 35 GHz and 94 GHz. Hourly observations of the German Weather Service at both locations gave estimates of the height, thickness and type of the clouds. The measured data were evaluated in a statistical manner. Comparing the results with calculations based on standard drop size distributions for various cloud types an increase in the attenuation levels was observed. This difference can be explained by the presence of super-large drops in clouds as recently measured and reported. Though their contribution to the total liquid water content is small, the effect on the attenuation is enhanced. A second propagation experiment was performed on a 1340 m long, terrestrial line-of-sight path to evaluate the scintillation effects at 94 GHz. It was found that the power spectrum of the measured log-amplitudes follows the -8/3 slope of the theoretical pattern based on Kolmogorov's refractive index spectrum. The refractive index structure constant C_n^2 calculated from variances of the measured log-amplitude agree well with those deduced from meteorological measurements after applying a semi-empirical relation for height scaling the various sensors. This result establishes the applicability of theoretical relations for scintillation effects at mm-waves due to Tatarski and Kolmogorov.

1. INTRODUCTION

Communication systems operating in the frequency bands above 30 GHz exhibit several advantages in comparison to those used at present. This is mainly caused by the higher usable bandwidth, which allows higher traffic capacities. Additionally, the smaller beamwidths of conventionally sized antennas result in a minor degradation of system performance caused by interference. If, in spite of the lower probability, interference occurs, the system is able to cope better with these disturbances due to the higher efficiency of spread-spectrum techniques.

Before these EHF systems become operational it is required to examine the environmental conditions imposed on the propagation path. Usually, this task is performed using statistical methods based on long-year averaged values of atmospheric propagation effects for the site, path characteristics, and frequency under consideration, or the attenuation and scintillation amplitudes are calculated employing physical methods and meteorological parameters known as long-year average values. In practice these two methods are often mixed as in the frequency-scaling calculations for path losses recommended by the CCIR.

Our paper emphasizes the physical dependencies of the atmospheric propagation effects on meteorological parameters, especially for clouds and atmospheric turbulence at frequencies up to 94 GHz using theoretical and experimental techniques.

The propagation path losses are estimated using two different approaches: measurements on a slant path at 35 GHz and 94 GHz as well as on a short terrestrial path, and calculation from measured meteorological parameters. This consideration leads to an estimation of the numerical constants used in the CCIR semi-empirical attenuation calculation procedures.

The measured attenuation effects at 35 GHz and 94 GHz due to clouds are compared with calculations using standard dropsize distribution for various cloud types and those including super-large drops [1,2].

The scintillation effects are estimated on the basis of the refractive index structure function, which depends on turbulent fluctuations of temperature and water vapour assuming a Kolmogorov refractive index spectrum. Simultaneous amplitude measurements of mm-wave and optical scintillation in combination with the observation of the relevant meteorological parameters along the propagation path allow then to test the underlying physical models and to compare these results with CCIR recommended estimation methods.

On the basis of these results, the applicability of the prediction methods to propagation parameter calculations for higher frequency regions is assessed.

2. THE EXPERIMENTAL SET-UP

2.1 Slant Path Measurements

The experiment have been carried out in the Alps region of southern Germany in summer 1991. Two propagation links at 35 GHz and 94 GHz have been established along

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

a common slant path from the top of the Zugspitze to the southern outskirts of Garmisch. The geometric characteristics of the link are:
- altitude of both transmitters: 2966 m
- altitude of the receivers: 727 m
- path length: 9420 m
- elevation angle: 13.7°.

Table 1 summarizes the main system parameters.

Transmitter:		
	(GUNN injection-locked IMPATT-Oscillator)	
35 GHz	Output power	80 mW CW
	Antenna	2`- Cassegrain
	Gain	44 dB
	3-dB-Beamwidth	0.95°
	Polarization	linear
94 GHz	Output power	200 mW CW
	Antenna	2`- Cassegrain
	Gain	52 dB
	3-dB-Beamwidth	0.35°
	Polarization	linear
Receiver	(**35 GHz and 94 Ghz**, superheterodyne)	
	Sensitivity	- 70 dBm
	Bandwidth	30 MHz
	Dynamic range	60 dB
	Antenna	2′-Cassegrain
	Polarization	linear, hor. and vert.

Table 1: Performance data of the 35 GHz and 94 GHz system

At the receiver site a meteorological weather station was installed to collect the standard meteorological parameters including temperature, humidity, pressure, solar flux and wind data. A complete set of all data were digitized and recorded on tape by a DEC LSI 11 at a data rate of 10 s.
Further information about atmospheric characteristics was provided by the German Weather Service (DWD) which operates measurement stations near the receiver site at Garmisch and on top of the Zugspitze. Estimates of height, thickness and type of clouds within the propagation path together with additional temperature and humidity data at both stations were collected every hour in tabular form.

2.2 Scintillation Measurements

A second terrestrial propagation link at 94 GHz was setup to study the effects of atmospheric turbulence on the propagation of millimeter waves. The experiment was performed on a 1320 m path at an average height of 7 m over farmland. The obstacle free height along the path was larger than the first Fresnel zone size of 2 m. The 94 GHz system consisted of a Gunn-oscillator with an output power of 20 mW, all other performance data corresponded to those of Table 1. To characterize the turbulent fluctuations of the atmosphere within the propagation path and to compare these with the path-averaged scintillations at 94 GHz point measurements were carried out with a fast response temperature sensor (sonic probe) and a fast response humidity sensor which evaluates the strong absorption of the Lyman-α emission line of hydrogen. Both sensors were mounted 2 m above ground together with a sensitive propeller vane. An additional weather station gave averaged values of standard meteorological parameters for calibration purposes and subsequent analysis.
The millimeter wave and sensor signals were sampled at a rate of 125 Hz and digitised by a 12-bit ADC. To avoid aliasing errors the analog signals were first low-pass filtered with a cut-off frequency of 20 Hz. Finally, complete data sets of 20 minutes time interval were recorded on a storage medium by a PC.
For the data evaluation, variances of the log-amplitude fluctuations of the millimeter wave signal were computed and stored by taking consecutive intervals of 1 min length. In order to determine the power density spectra of the scintillation signal each data file was divided in one-third overlapping parts. After removing the mean value and linear trend and computing the power density spectrum of each part the spectrum of the complete time interval was obtained by averaging all individual spectra.

3. THEORY

3.1 Cloud Attenuation

Attenuation of millimeter waves by clouds is dependent on the droplet size distribution, liquid water content (LWC), temperature, wavelength and the physical composition and extent of clouds. Absorption by the cloud particles are the main contribution, where as the scattering losses are very small and can be neglected. If cloud droplet sizes is assumed to be less than 0.1 mm in diameter the Rayleigh approximation can be applied throughout the millimeter wavelength range. Then the absorption cross-section is a function of the volume of the droplets or the cloud water content (LWC). Knowing the droplet size distribution of clouds and their vertical development CCIR [3] calculation procedures may be applied to estimate the attenuation losses. For small drops, 1 μm ≤ r ≤ 50 μm, the average size distribution for any type of cloud can be described by a gamma distribution [4,5]

$$n(r) = \frac{N_0 r^{\alpha}}{\Gamma(\alpha+1)\beta^{\alpha+1}} \cdot e^{-\frac{r}{\beta}} \quad , \quad (1)$$

where n(r) is the cloud droplet distribution as a function of drop radii, α, β, β_S distribution parameters describing different cloud types, Γ the gamma function and N_o, N_S drop concentration in cm^{-1} for different cloud types.
Observations of particle size spectra have shown that also super-large drops are present in clouds which have a vertical extent of more than 300 m. The size distribution for droplet radii between 100 μm ≤ r ≤ 1.5 mm follows an exponential relation [2,4]

$$n(r) = \frac{N_S(1-\beta_S)}{85} \cdot \left(\frac{85}{r}\right)^{\beta_S} \quad . \quad (2)$$

Cloud Type	LWC [g/m³]			35 GHz A [db/km]			94 GHz A [db/km]		
	small drops	+ large drops	%	small drops	+ large drops	%	small drops	+ large drops	%
Stratocumulus	0.214	0.0116	5.1	0.221	0.352	59.3	1.013	2.113	108.6
Stratus	0.208	0.0047	2.2	0.215	0.253	17.7	0.987	1.28	29.7
Nimbusstratus	0.268	0.0094	3.4	0.276	0.825	198.9	1.268	3.199	152.3
Cumulus mediocris	0.996	0.0015	0.15	1.027	1.041	1.4	4.717	4.82	2.2
Cumulus congestus	1.735	0.0038	0.22	1.784	2.12	18.8	8.196	9.503	15.9
Cumulus congestus (min)	2.407	0.0144	0.6	2.475	2.582	4.3	11.37	11.68	2.7
Cumulus congestus (max)	2.407	0.0948	3.8	2.475	3.703	49.6	11.371	14.28	25.6
Cumulonimbus (min)	2.965	0.0545	1.8	3.044	3.525	15.8	13.992	16.665	19.1
Cumulonimbus (max)	2.965	0.726	19.6	3.051	7.501	145.9	14.02	41.74	197.7

Table 1: Liquid water content (LWC) of various cloud types and the effect of large drops on cloud attenuation at 35 GHz and 94 GHz [1, 2]

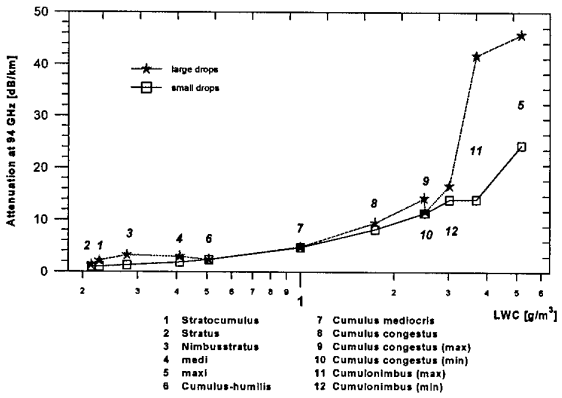


Fig.1: Effect of only small and additional large size drop size distribution on attenuation at 94 GHz for various cloud types as a function of liquid water content

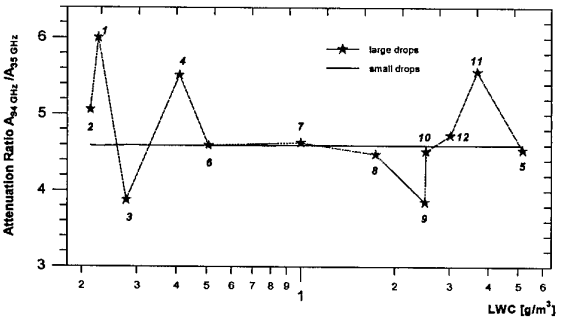


Fig 2: Effect of large drops on the attenuation ratio 94 GHz / 35 GHz as a function of liquid water content for different drop size distributions of various cloud types.

Though these large drops increase the total water content in the clouds only by a small amount, normally < 5 % , the effect on the attenuation of millimeter waves is significantly enhanced depending on the type of cloud. Taking Ajvazyan's distribution parameters and estimated LWC values [1], and Papatsoris' calculations of attenuation at 35 GHz and 94 GHz [2] Table 1 illustrates the changes in LWC and the specific attenuation at both wavelengths for various cloud types if the presence of large drops is considered.

Fig.1 shows the effect on attenuation at 94 GHz as a function of the total liquid water content for the cloud types listed in Table 1. As can be seen the implication on the propagation through clouds is not so serious because of the low absolute specific attenuation except for cumulonimbus cloud types. Especially those cloud types contain drops of large sizes which were held in balance by thermal convection induced upwinds and grow until they are too heavy and fall down as strong precipitation.

Fig. 2 shows the attenuation ratio 35 GHz / 95 GHz as a function of total liquid water content for the same cloud types of Fig.1. The straight line represents the calculation for only small drops based on the Rayleigh approximation yielding a size distribution independent constant ratio. Considering large drops with diameters comparable to millimeter wavelength, Papatsoris had to apply Mie calculations for his estimation of the millimeter wave attenuation. The data shows a different size distribution dependency for both represented frequencies.

3.2 Scintillation

Scintillation effects are induced by atmospheric turbulences. They occur in the troposphere in the boundary layer near the earth surface by air mixing mechanism and vertical transfer processes, in clouds due to air entrainment processes and, to a minor degree, during rain events. Time and space dependent small-scale variations and inhomogeneities of the atmospheric refraction caused by turbulent humidity and temperature

fluctuations generate rapid fluctuations of amplitude, phase and angle of arrival of a received signal. The intensity of the turbulence can be related to the structure function D_n and the structure parameter C_n^2 of the index of refraction. The theoretical treatment is based on Kolmogrov's theory of turbulence [5] and Tatarski's expressions for the scintillations [6]. The following assumptions are made:

- the von Kármán representation of the Kolmogorov refractive index spectrum [6]. The turbulent eddies sizes are assumed within the inertial subrange between the outer scale L_o and the inner scale l_o of the turbulence.
- Taylor's frozen-in hypothesis which assumes an unchanged shape and size of the eddies within a time required for the eddies to move a distance l [6]. This relates the space and time variations to the medium wind speed.
- an incident plane wave
- weak scintillations, the log-amplitude variances $\sigma_\chi^2 \ll 1$,

if the perturbation method is applied to determine the scintillation effects.

Using Kolmogorov`s model the three-dimensional refractive index spectrum $\Phi_n(\kappa)$ is expressed by (Tatarski)

$$\Phi_n(\kappa) = 0.033\, C_n^2\, \kappa^{-11/3}, \qquad (3)$$

where the wavenumber $\kappa = \frac{2\pi}{\lambda}$ is in the limits

$$\frac{2\pi}{L_o} < \kappa < \frac{2\pi}{l_o}. \qquad (4)$$

If the outer scale length is less than the first Fresnel zone size ($L_o < \sqrt{\lambda L}$) the variance of the log-amplitude [7] is

$$\sigma_\chi^2 = 23.39\, C_n^2 \left(\frac{2\pi}{\lambda}\right)^{7/6} L^{11/6}, \qquad (5)$$

where L is the path length. Theoretically, the power density spectrum of the log-amplitudes has a $f^{-8/3}$ slope. This shape may be used as a test criterion for the applicability of the Kolmogorov model.

To validate the magnitude of scintillation fading on millimeter wave radio links two deterministic physical models have been used for comparison of the structure parameters C_n^2. $C_{n,\,94GHz}^2$ was calculated from the variance of the log-amplitude σ_χ^2. $C_{n,\,in\,situ}^2$ was estimated from measured temperature and humidity fluctuations using a time-lag method [8, 9]. If the structure parameters were measured at different heights, h_{94Ghz} and $h_{in\ situ}$, a semi-empirical relation based on the Monin-Obukhov similarity theory is used [10]

$$\frac{C_{n,\,94GHz}^2}{C_{n,\,in\,situ}^2} = \left(\frac{h_{94GHz}}{h_{in\ situ}}\right)^{-\alpha}, \qquad (6)$$

where $\alpha = 4/3$ is valid for a moderately to high unstable atmosphere with low wind speeds.

4. RESULTS

4.1 Slant Path Measurements

Distributions of measured attenuation through clouds and clear air at the frequencies 35 GHz and 94 GHz on the 9420 m slant path during time periods with observed cloud layers within the propagation path are shown in Fig.3. The time period covered about 25 % of the total measurement period of nearly two months. At 35 GHz the maximum of the distribution corresponds to an attenuation value of 2.4 dB and at 94 GHz of 5.8 dB, respectively. Compared to the clear air propagation losses the additional attenuation due to clouds is on the average about 1 dB higher.

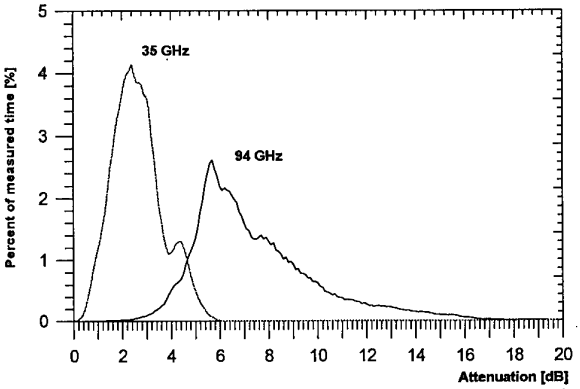


Fig.3: Distribution of measured total attenuation through clouds and clear air at 35 GHz and 94 GHz. Observation time: 210 hours.

Fig. 5 and 7 give examples of cloud attenuation over the 9420 m path length at 35 GHz and 94 GHz over a time period of 2 and 7 hours, respectively. Fig.4 and 6 give the corresponding time series of estimated height level of the clouds as they have been observed by the German Weather Service. Due to the method of visual determination the accuracy of the data is limited to ± 50 m.

The specific attenuation at 35 GHz and 94 GHz in stratocumulus clouds (Fig.5) is comparable over the time period yielding a constant attenuation ratio of about 2.5 to 3.5 depending on the temperature of the cloud droplets and the variation of the attenuation due to the atmospheric gases especially due to the water vapor concentration along the propagation path. Though the presence of large drops in this type of cloud increases the attenuation at 35 GHz and 95 GHz by a relatively high amount, 60% and 109%, respectively (Tab.1), the implication on the propagation is only limited because of the low total number of droplets.

Large attenuation discrepancies, however, can be observed with cumulonimbus cloud types (Fig.7). The increased cloud attenuation at 94 GHz in cumulonimbus clouds is caused by the higher drop concentration in this type of convective vertical developped clouds containing a substantial number of large droplets. Total attenuation values of up to 18 dB have been measured due to clouds

and clear air over the path. The attenuation level at 35 GHz over same time period reaches values of only 6 dB.

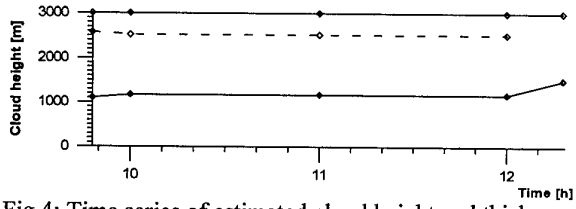


Fig.4: Time series of estimated cloud height and thickness up to the transmitter height. Dashed line indicating two separate cloud layers. Cloud type: stratocumulus. Date: July 1, 1990.

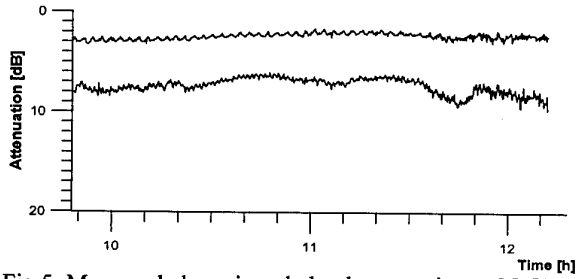


Fig.5: Measured clear air and cloud attenuation at 35 GHz (upper curve) and 94 GHz, path length: 9420 m, elevation angle: 13.7°, cloud type: stratocumulus. Date: July 1, 1990.

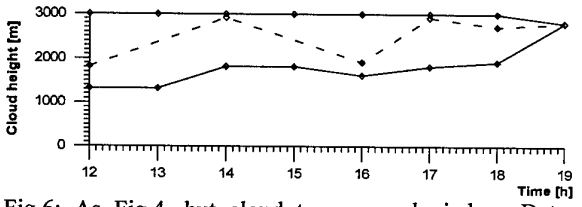


Fig.6: As Fig.4, but cloud type: cumulonimbus. Date: June 26, 1990.

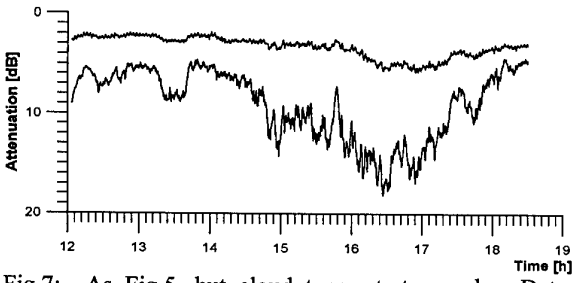


Fig.7: As Fig.5, but cloud type: stratocumulus. Date: June 26, 1990.

4.1 Scintillation measurements

Fig. 8 shows the averaged power density spectrum of amplitude at 94 GHz for a 20 min time period of a 5 hour lasting scintillation event. The straight line represents the $f^{-8/3}$ slope of the theoretical spectrum based on the Kolmogorov refractive index spectrum. The dashed line indicates the estimated outer scale length $L_o \cong 7.9$ m for a cross-wind component of $v_\perp \cong 1.5$ m/s. This value was originally determined from the one-dimensional temperature spectrum for the same time period yielding the same starting point for the inertial subrange [11]. The slope of the calculated spectrum agrees well with the theoretical one confirming the theoretical assumptions.

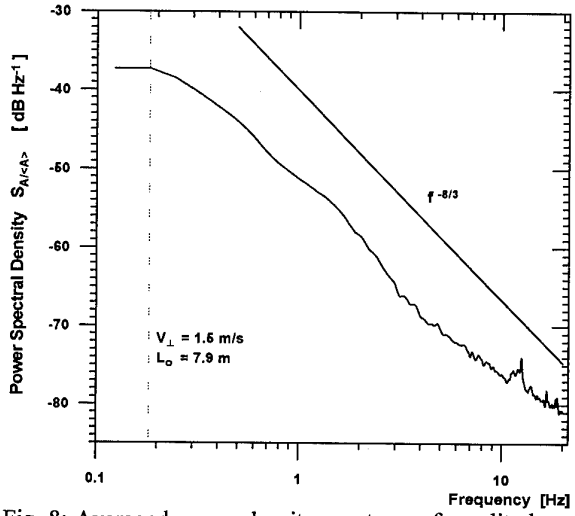


Fig. 8: Averaged power density spectrum of amplitude at 94 GHz.

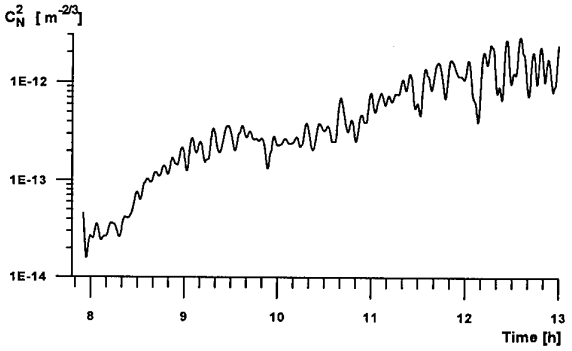


Fig. 9: Refractive index structure parameter C_n^2 calculated from *in situ* observed temperature and humidity fluctuations. Measurement height: 2 m above ground.

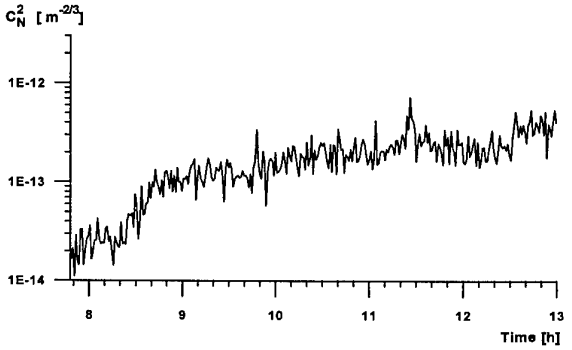


Fig.10: Refractive index structure parameter C_n^2 calculated from measured variances of log-amplitude fluctuations at 94 GHz. Measurement height: 7 m above ground.

Fig. 9 and 10 show the development of the refractive index structure parameters C_n^2 in this 5 hour time series for a typical sunny day with nearly no cloud cover exhibiting the influence of the growing atmospheric turbulences during the morning. Both data curves are calculated from two independent measurements, a point measurement and a path averaged one. The point measurement was performed at a height of 2 m with the fast response temperature (sonic) and humidity sensor (Lyman-α-hygrometer). The evaluation is based on the above mentioned time-lag method using time lags up to 0.5 s which correspond to spatial lags up to several meters depending on the average wind speed. The path averaged C_n^2 is derived according to equ. (5) from the variances of the log-amplitudes of the 94 GHz signal measured at a height of 7 m over a propagation path of 1340 m.
The results of both independent measurements are compared in the scatter diagram of Fig.11. The data given in this plot are 1-min averages. The regression line represents the semi-empirical relation for scaling the different heights of the sensors using the exponent $\alpha = -4/3$ suggested by Wyngaard et al. The scatter of the data can be explained by differences in the point and path-average measurement method. For this low height differences the correlation fits well with the theoretical height scaling exponent.

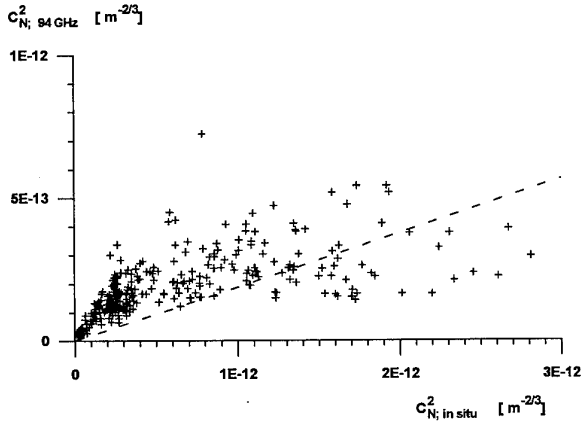


Fig.11: Scatter of the refractive index structure parameter C_n^2 calculated from scintillation at 94 GHz and C_n^2 calculated from measured temperature and humidity fluctuations.

5. CONCLUSION

Cloud attenuation at millimeter wavelengths can be affected by the presence of large droplets. Though their contribution to the total liquid water content in clouds is usually not more than 5 % they can enhance the specific attenuation by 100 % and more depending on cloud type and drop size distribution. At 35 GHz the increase on attenuation ranges from 10 % to 100 %, but is without significant implication on the link parameters because the absolute attenuation levels are still low. At 94 GHz, however, a similar increase results in a much more serious effect, especially concerning cumulus cloud types. The presented cloud attenuation measurements confirm this result for individual events. The cloud attenuation statistics for a nearly two month time period, however, shows only a moderate increase in the attenuation levels compared to clear air propagation losses on that link.
The measured scintillations at 94 GHz establish the applicability of the Kolmogorov refractive index spectrum and the turbulence theory of Tatarski. The refractive index structure parameter C_n^2 calculated from measured 94 GHz scintillations agree well with those calculated from temperature and humidity fluctuations if the height scaling suggested by Wyngaard et al. is applied. The results show that the applied theoretical relations describe the turbulence induced fluctuations at 94 GHz and may be used for modelling scintillation effects at millimeter waves for other links and for the prediction on the performance of communication systems.

REFERENCES

1. Ajvazyan, H.M., "Extinction and Radar Reflection Coefficients for mm and sub-mm Waves in Clouds", Int.J.Infrared and Millimeter Waves, *12*, (2), 1991, pp.177-190.
2. Papatsoris, A.D., "Implication of Super-large Drops in Millimetre-wave Radar Observations of Water Clouds", Electronic Letters, *30*, 21, 1994, pp.1799-1800.
3. CCIR Report 721-2: "Attenuation by Hydrometeors, in particular Precipitation, and other Atmospheric Particles." Geneva, April 1988.
4. Feigelson, H.M., "Radiation in Cloudy Atmosphere", Atmospheric Sciences Library, D.Reidel Publishing Company.
5. Kolmogorov, A.N., "The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds' Numbers," Doklady Akad. Nauk. SSSR, *30*, 1941, p.301.
6. Tatarski, V.I., "Wave Propagation in Turbulent Medium," McGraw-Hill, New York, 1961.
7. CCIR Report 718-3: "Effects of Refraction on Radio-Wave Propagation." Geneva, April 1990.
8. Kohsiek, W., "A Comparison between Line-Averaged Observation of C_n^2 from Scintillation of a CO_2 Laser Beam and Time-Averaged *in situ* Observations", J. Clim. Appl. Meteorol., *24*, 1985, pp. 1099-1103.
9. Hill, R.J., Bohlander, R.A., Clifford, S.F., McMillan, R.W., Priestley, J.T., Schoenfeld,W.P., "Turbulence-Induced Millimeter-Wave Scintillation Compared with Micrometeorological Measurements", IEEE Trans.Geosci. Remote Sensing, *26*, 3, May 1988.
10. Wyngaard, J.C., LeMone, M.A., "Behaviour of Refractive-Index-Structure Parameter in the Entraining Convective Boundary Layer", J. Atmos. Sci., *37*, 7, 1980, pp. 1573-1585.
11. Cole, R.S., Ho, K.L., Mavrokoukoulakis, N.D., "The effect of the outer scale of turbulence and wavelength on scintillation fading at millimeter wavelengths", IEEE Trans. Antennas Propag., AP-26(5), 1978, pp.712-715.

DISCUSSION

Discussor's name: A.D. Papatsoris

Comment/Question:

Sometime ago, I had theoretically treated the effect of super-large drops in clouds on attenuation and reflectivity at 35 and 94 GHz and I am happy to see those predictions verified by your experimental data. So, I was wondering whether you could share your data for further comparison and analysis?

Author/Presenter's reply:

We have made measurements over a period of nearly two months. This data is available and it would also be very interesting for us to compare your analysis with our attenuation measurements through clouds.

Characterisation of the transfer function of the clear atmosphere at millimetre wavelengths

A D Papatsoris
University of York
Department of Electronics
York, YO1 5DD
UNITED KINGDOM

1. SUMMARY

A model for the calculation of the absorption and dispersion spectra of the most important and millimetre-wave active atmospheric gases for arbitrary atmospheric geometries is presented. The model adopts a line-by-line summation technique to characterise the resultant resonant absorption and dispersion spectra of various trace constituents and major contributors H_2O and O_2. A spherically layered atmosphere is assumed with each layer taken to be in local thermodynamic equilibrium (LTE) with exponential profiles of density and pressure between layer boundaries. The relative air mass integral and the lineshape functions are also appropriately evaluated according to atmospheric conditions. Various model atmospheres may be selected or specified by the user.

2. INTRODUCTION

Close to the Earth's surface millimetre-wave absorption and dispersion by atmospheric gases is dominated by water vapour and oxygen. Nevertheless, other atmospheric gases including trace constituents, put their spectral signatures in the millimetre frequency spectrum. Although such contributions to attenuation and phase delay are negligible for communication systems operating close to the Earth's surface, they may become important for applications in remote sensing, weapons guidance, navigation and communications systems operating at higher altitudes. In this paper a modelling approach which takes into account contributions from trace gases is described and example calculations for selected atmospheric geometries are given.

3. THEORY OF GASEOUS ABSORPTION

3.1 Interaction of Radiowaves with Gases

Neglecting interactions between different modes of energy, the energy of an isolated gaseous molecule is given by:

$$E = E_r + E_v + E_e + E_t \quad cm^{-1}, \tag{1}$$

where E_e, E_v, E_r are the electronic, vibrational and rotational energies, respectively and E_t is the translational energy. The first three terms in equation (1) are quantized, and take discrete values only, these values being specified by one or more quantum numbers. Any combination of quantum numbers defines an energy state or quantum state, or term.

For absorption or emission to take place matter must interact with the incident field of electromagnetic radiation which in practice must involve an electric or magnetic dipole or quadrupole moment. In other words, for a molecule to be able to interact with an electromagnetic field and absorb or create a photon of frequency v, it must possess at least transiently, a dipole oscillating at that frequency. For emission or absorption this transient dipole is expressed in terms of quantum mechanics by the transition dipole moment.

Interactions resulting in transitions of the dipole moment can differ widely in strength. Electric dipole interactions are greater by a factor of order 10^5 than magnetic dipole interactions; electric dipole interactions are of the order of 10^8 times stronger than electric quadrupole interactions. Electric dipole transitions are therefore responsible for the strongest spectral lines, and are called permitted transitions. Other transitions are loosely named forbidden. The nature of a transition can be specified in terms of the quantum numbers of the upper and lower states [1, 2]. Such a relationship is known as a selection rule. When a transition takes place from one energy state to another, the frequency v of the absorbed or emitted quantum is given by Planck's relation $\Delta E = hv$, where h is Planck's constant and ΔE the energy gap between the energy states involved in the transition.

The most general transition involves simultaneous changes of electronic, vibrational and rotational energy. The minimum energy jumps commonly observed vary considerably between the three forms, and provide a convenient preliminary method of distinguishing between them. Minimum changes in rotational energy are typically of the order of 1 cm^{-1}. Rotation lines, that is lines due to rotational energy changes only, therefore form part of the microwave or the far infrared spectrums. Vibrational energy changes are rarely less than 600 cm^{-1}. This is so much larger than the minimum for rotational energy change that vibrational transitions never occur alone, but with many simultaneous rotational changes giving a group of lines which constitute a vibration-rotation band, usually in the intermediate infrared spectrum. An electronic transition typically involves a few electron volts of energy, and the resulting absorption or emission usually occurs in the visible or ultraviolet spectrum. At millimetre and sub-millimetre wavelengths the majority of transitions are purely rotational (with a few exceptions) and the modelling of interaction between radiowaves and gaseous molecules is generally simple.

The interaction of electromagnetic radiation with the absorbing matter within the vicinity of a single spectral line is described by the volume absorption coefficient k_v,

$$k_v(x) = S(x)f(v - v_0)\rho(x) \quad cm^{-1}, \tag{2}$$

where x denotes position, $f(v - v_0, x)$ is the lineshape factor, $S(x)$ is the line strength, v_0 is the centre of the absorption line and $\rho(x)$ the density of the absorbing gas. Generally, the volume absorption coefficient k_v is a function of the density of the absorbing substance, and the pressure and temperature of the atmosphere, all of which can vary along the path of observation. The volume absorption coefficient, $k_v(x_1, x_2)$, between two points x_1 and x_2 along a path of observation resulting from a single spectral line is given by,

$$k_v(x_1, x_2) = \int_{x_1}^{x_2} S(x)f(v - v_0, x)\rho(x)dx \tag{3}$$

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

and for a band of spectral lines by,

$$K_v(x_1, x_2) = \sum_{ij} \int_{x_1}^{x_2} S_{ij}(x) f_{ij}(v - v_{ij}, x) \rho_j(x) dx, \quad (4)$$

where $S_{ij}(x)$ is the strength of the i'th line of the j'th absorbing gas in the path from position x_1 to x_2, $f_{ij}(v - v_{ij}, x)$ is the line shape factor, v_{ij} is the position of the line centre, and $\rho_j(x)$ the density of the absorbing gas. Here we assume that when lines overlap their individual contributions to the optical depth simply add together. This is an essential simplification, adequately justified in terms of the relatively sparsely spaced spectral lines of various gases in the mm to sub-mm frequency spectrum, and the values of pressure and temperature in the atmosphere. Line strengths and positions may be calculated using quantum mechanics and are dependent on temperature, geometrical configuration of the molecule and the population of the lower energy state of the transition. The line strength of a spectral line involving a transition from state l to state u is given by,

$$S_{lu} = \frac{8\pi^3 v_{lu}}{3hc} \left[1 - \exp\left(\frac{-c_2 v_{lu}}{T}\right)\right] \frac{g_l I_a}{Q(T)} \exp\left(\frac{-c_2 E_l}{T}\right) R_{lu} \quad (5)$$

where h is Planck's constant, v_{lu} is the resonant frequency of the line, E_l is the energy of the lower energy state of the transition, g_l is the nuclear spin degeneracy of the lower level, $Q(T)$ is the total partition function, I_a is the natural isotopic abundance, c_2 is the second radiation constant and R_{lu} is the Einstein transition probability. The total partition function Q for temperatures encountered in the atmosphere is given by the following approximate expressions,

$$Q = \begin{cases} \frac{kG}{\sigma hB} T & \textit{for linear molecules,} \\ \frac{G}{\sigma}\left[\frac{\pi}{ABC}\left(\frac{kT}{h}\right)^3\right]^{\frac{1}{2}} & \textit{for nonlinear molecules,} \end{cases} \quad (6)$$

where σ is the symmetry number depending on the group of which the molecule is a member, h is Planck's constant, k is Boltzmann's constant, G is the common factor of the molecule and A, B and C are its rotational constants. Lineshapes depend upon the mechanisms of molecular collisions and motions, and are, therefore temperature and pressure dependent.

3.2 The Lineshape Function

The physical mechanism that dominates spectral line broadening close to the earth's surface is collisional interactions between molecules. This is also known as pressure broadening and is successfully interpreted by the Van Vleck-Weisskopf (VVW) lineshape function,

$$f(v, v_0) = \frac{1}{\pi}\left[\frac{1 - i\delta}{(v - v_0) - i\gamma_L} + \frac{1 + i\delta}{(v + v_0) + i\gamma_L}\right] \quad cm, \quad (7)$$

where δ is the interference coefficient suggested by Rosenkranz [4] to describe pressure induced line interference for the oxygen 60 GHz band ($\delta = 0$ for all other gases), and γ_L is the collisional linewidth. The collisional linewidth is a function of both pressure p and temperature T and is described by,

$$\gamma_L = \gamma_{L_0} \frac{p}{p_0} \left(\frac{T}{T_0}\right)^{-x} \quad cm^{-1}, \quad (8)$$

where x varies usually between 0.5 and 1.0 and γ_{L_0} is the value of the linewidth at normal atmospheric conditions, $p = p_0$ and $T = T_0$. At sufficiently low pressures, collisional broadening is negligible compared to thermal broadening (also known as Doppler broadening). Because molecular velocities have a Gaussian distribution, the Doppler lineshape is also a Gaussian,

$$f(v, v_0) = \frac{1}{\sqrt{\pi}\gamma_D} \exp\left[-\left(\frac{v - v_0}{\gamma_D}\right)^2\right] \quad cm, \quad (9)$$

and the Doppler line width is given by,

$$\gamma_D = 3.58 \times 10^{-7} v \sqrt{\frac{T}{M}} \quad cm^{-1}, \quad (10)$$

where M is the molecular weight of the gas. When collisional and thermal broadening become comparable, the resonance frequency in the collision shape expression should be convolved with the Gaussian shift probability distribution over all possible Doppler shifts. This convolution is called the Voight profile,

$$f(v, v_0) = \frac{i}{\pi} \int_{-\infty}^{+\infty} \frac{\exp(-t^2)}{z - t} dt \quad cm, \quad (11)$$

where z is a complex argument related to the collisional γ_L and thermal γ_D linewidths (see Penner [7]) and $t = (v - v_0)/\gamma_D$.

4. MODELLING APPROACH

4.1 Source of Spectroscopic Information

The energy states of gaseous molecules are well defined and discrete and therefore the positions of the resultant spectral lines can be calculated. The U.S. Air Force Geophysics Laboratory (AFGL) high-resolution transmission molecular absorption database (known under the acronym HITRAN [3]), gives the exact positions and other characteristics of the spectral lines of gaseous molecules between 0 and 17900 cm^{-1}. The 1986 edition of the HITRAN database was used to derive all the necessary spectroscopic information. The database was selectively accessed to provide suitable line files for the atmospheric gases of interest.

4.2 Modelling the atmosphere

4.2.1 *The refractive atmosphere*

In order to calculate the attenuation and dispersion spectra of atmospheric gases, we need to know their concentration and distribution along the path of observation. Furthermore, since the line intensity and lineshape function are sensitive to pressure and temperature variations, an efficient way of representing the propagation medium is needed. An excellent approach is to slice the atmosphere in many spherical concentric layers and assume exponential profiles of density and pressure between layer boundaries.

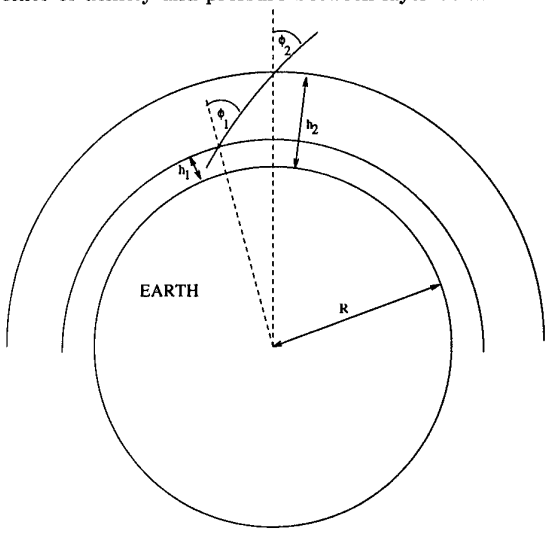


Fig. 1. *Relative air mass geometry.*

Fig. 1 shows a spherically symmetric atmosphere for which refraction effects are taken into account. The relative air

mass M between height points h_1 and h_2 is given by,

$$M(h_2,\phi_2) = \frac{\int_{h_1}^{h_2} \frac{n(h)(R+h)\rho(h)}{\sqrt{n^2(h)(R+h)^2 - n^2(h_1)(R+h_1)\sin^2\phi_1}} dh}{\int_{h_1}^{h_2} \rho(h) dh}, \tag{12}$$

where n is the index of refraction, dh is the height increment and ρ the gas density at point h along the ray path in the atmosphere. In this approach the relative air mass integral has been approximated by Uplinger's [5] simple closed form approximation which includes the effects of refraction and the variation of atmospheric density with altitude.

Comparison with ray traces in the U.S. Standard Atmosphere shows that Uplinger's approximation is accurate to within a few percent at an angle 90° to the zenith, with rapidly increasing accuracy for smaller to the zenith angles.

4.2.2 *Elimination of inhomogeneity of atmospheric layers*
Providing that each layer is found in LTE and each gas is uniformly mixed within it, an effective pressure, temperature and absorber amount may be specified for each layer. The effective values are those that make the absorption and/or emission of an equivalent homogeneous layer equal to that of the non-homogeneous layer. There is no single definition of the effective pressure or temperature because the effective values depend upon many factors and will in general be different for different absorbing gases. The most generally acceptable definition, however, is the density weighted average, also known as the Curtis-Godson approximation,

$$\bar{P} = \frac{\int P(h)\rho(h)dh}{\int \rho(h)dh}, \quad \bar{T} = \frac{\int T(h)\rho(h)dh}{\int \rho(h)dh}, \tag{13}$$

where P is the atmospheric pressure, T is the atmospheric temperature and ρ the total air density. Values of pressure, temperature and gas density at the layer boundaries are tabulated in various reference Atmospheres, ie [8]. Inside each layer the pressure and total air density are both assumed to follow an exponential distribution, whereas temperature is assumed to vary linearly with height.

4.3 Calculation of the Spectra of Atmospheric Gases

4.3.1 *Line-by-line summation technique*
The calculation of the resonant spectra of all atmospheric gases is calculated by considering individual contributions from all the spectral lines across the millimetre frequency spectrum. At some frequency separation, which depends upon constituent density and line intensity, contributions from distant lines can be neglected. Of course, the suitable lineshape function must be chosen according to the specific atmospheric conditions, pressure and temperature. The Van Vleck-Weisskopf, Voight and Gaussian lineshape functions are exclusively used for all atmospheric gases. For the evaluation of the complex Voight function Hui's [9] numerical algorithm has been adopted.

4.3.2 *Spectrum of O_2*
A modified VVW lineshape function must be used in the case of oxygen, as suggested by Rosenkranz [4] to describe pressure induced line interference between the closely spaced lines of the 60 GHz oxygen band. Line coupling occurs when collisions between a radiating molecule and broadening gas molecules cause the transfer of population between rotational- vibrational molecular states and result in a re-distribution of spectral intensity within a band. Rosenkranz presented a method in which weak coupling approximations and a first order expansion of the band's shape lead to an accurate description of the microwave spectrum of oxygen in the atmosphere. Interference coefficients have been given by Rosenkranz [6] and more recently by Liebe et al [10].

In addition to pressure induced line interference, Zeeman splitting of oxygen lines must be taken into account for altitudes above 30 km. Oxygen, being a paramagnetic molecule is subject to the Earth's magnetic field, which although weak, is capable of introducing a line splitting. This line splitting is also dependent on the polarisation of the electromagnetic radiation. A reasonable approximation valid for all microwave absorption lines according to Liebe [11], appears to be,

$$\gamma_Z = \sqrt{\gamma_L + 6.944 \times 10^{-13} B_0^2} \quad cm^{-1}, \tag{14}$$

where B_0 is the geomagnetic field strength ranging from 25 to 65 Tesla depending on geographic location and altitude. For a complete treatment of the Zeeman effect, including the change of polarisation of the radiowave, we refer to the comprehensive work of Liebe [12].

4.3.3 *Spectrum of trace gases*
The resonant spectra of trace gases are calculated employing the line by line summation technique. Each absorption line gives a contribution at the frequency of consideration depending upon its intensity and distance from the line centre. The lineshape function is as usually appropriately evaluated according to the atmospheric conditions.

4.3.4 *Spectrum of H_2O*
Although the line-by-line summation technique proves generally successful in predicting the absorption and dispersion from O_2 and other trace gases, it fails to interprete the open-field and laboratory measurements for water vapour. This seems not to be the case near the H_2O line centres, but in the window regions. This insufficiency has established the existence of excess water vapour absorption (EWA) to account for the discrepancy between theory and measurements. As a result an empirically derived term, known as the H_2O continuum, has to be added to the contributions of local lines below 1 THz to complete the characterization of the water vapour spectra at millimetre wavelengths.

The empirically derived H_2O continuum depends on the square of the water vapour density and shows a very strong negative temperature dependence. Various explanations have been proposed to bridge the gap between theory and measurements, including water dimers $(H_2O)_2$, submicron aerosol particles under conditions of high relative humidity and collision broadened far wings of infrared H_2O lines. Nevertheless, recently Ma and Tiping [14] presented a theory which attributes the continuum to the far wings of allowed rotational transitions of H_2O in the far-infrared portion of the spectrum.

In this model Liebe's [11] temperature and density dependence for the continuum absorption have been adopted. The water vapour continuum formula, however, differs from that of Liebe's because of the different number of resonant lines involved in this approach. For the dispersive continuum Hill's [13] expression, which calculates the contribution of all lines above 1 THz using the 1982 AFGL line-parameters compilation, has been adopted. So the water vapour continuum is given by,

$$k_a = 3.8 \times 10^{-6} v^2 \left[31.6 \left(\frac{300}{T}\right)^{7.5} e + p\right] e \left(\frac{300}{T}\right)^3, \tag{15.a}$$

$$k_d = 1.2566 \times 10^{-5} Q \sum_{j=1}^{4} A_j \left(\frac{296}{T}\right)^a_j \left(1 - B_j \frac{296}{T}\right) \left(\frac{v}{33}\right)^{2j}, \tag{15.b}$$

where p and e are the dry air and water vapour partial pressures in atmospheres, T is the absolute temperature in $^\circ K$, v the frequency in cm^{-1}, Q the absolute humidity in gm^{-3} and the coefficients A_j, a_j and B_j are given in Table I. k_a and k_d are given in units of cm^{-1} and $rad\ cm^{-1}$, respectively.

j	A_j	a_j	B_j
1	1.382	1.650	0.199
2	-0.214	0.162	3.353
3	-0.149	0.178	3.101
4	-0.109	0.192	3.005

Table I: Hill's dispersive continuum coefficients.

4.3.5 *Contributions from N_2 and CO_2*
Residuals from very strong CO_2 lines in the far-infrared and optical frequencies form a CO_2 continuum which, nevertheless, is too weak to add any contributions at millimetre wavelengths. Pressure induced nitrogen absorption becomes important at high pressures and for frequencies above 100 GHz. According to Stone et al [15] it can be expressed as,

$$k_{N_2} = 5.42\times10^{-10}v^2(1-1.972\times10^{-3}v^{1.5})p^2\left(\frac{300}{T}\right) \quad cm^{-1}. \tag{16}$$

Finally the term k_0 is used to describe the non-dispersive radio refractivity,

$$k_0 = 2\pi v10^{-6}[N_{dry}(p,T) + N_{wet}(e,T)] \quad cm^{-1}, \tag{17}$$

with the dry and wet components of refractivity given by,

$$N_{dry} = 78.6687 \times 10^3 \frac{p}{T} \quad ppm, \tag{18.a}$$

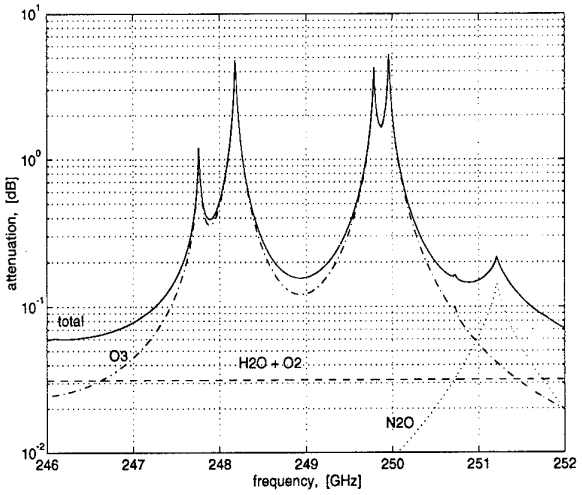
$$N_{wet} = 3.8 \times 10^8 \frac{e}{T^2} - 5.67 \times 10^3 \frac{e}{T} \quad ppm, \tag{18.b}$$

where p and e are the barometric and water vapour partial pressures in atmospheres, respectively, and T the temperature in $^\circ K$.

5. SPECTRA OF TRACE GASES

The effect of trace gases on radiowave attenuation and phase dispersion has always been regarded as insignificant at millimetre wavelengths. This is indeed true for applications operating close to the Earth's surface where attenuation and dispersion are dominated by O_2 and H_2O. At higher altitudes and for selected frequency intervals, however, trace gases are the principal contributors.

This is evident in Fig. 2 where characterisation of the transfer function of the atmospheric window centered at 250 GHz for an aircraft-satellite path is illustrated. The aircraft is assumed to fly at 10 km altitude looking to the satellite at an angle 80° to the zenith. Contributions from various trace gases are shown with O_3 being the main contributor. The frequency ranges from 246 GHz to 252 GHz while the frequency step is 1 MHz.



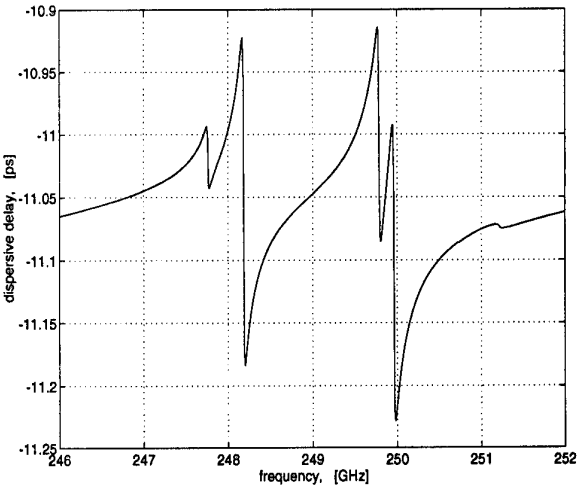


Fig. 2. *Calculation of the transfer function of the atmosphere along a 10-120 km slant path at 80° to the zenith.*
a. Comparison to the total attenuation predicted from the model and attenuation predicted only from O_2 and H_2O. Also shown are contributions from O_3 and N_2O.
b. Dispersive delay.

6. APPLICATIONS

The model described here, -also known as the Atmospheric Propagation Model [16]-, has been used to assess the potential of the 60 GHz absorption band for topside air traffic control and broadband transmission purposes [17].

To illustrate the typical losses encountered in the aircraft to satellite (AS) and aircraft to ground (AG) paths, calculations have been done for an example ascent path. An aircraft which takes off from New York, flies over the Atlantic Ocean and finally lands at London Heathrow and a geostationary satellite at 50° West, define the geometry of the application which is shown in Fig. 3. The elevation angle from the aircraft to the satellite varies during the course of flight between 15° and 35°, while the corresponding angle to the zenith varies between 6.86° to 8.36°.

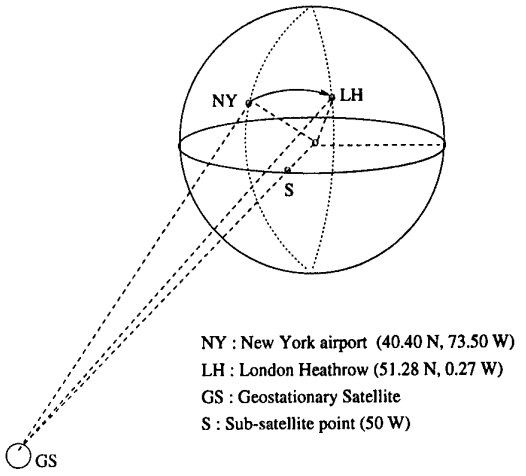


Fig. 3. *Aircraft-to-satellite path geometry.*

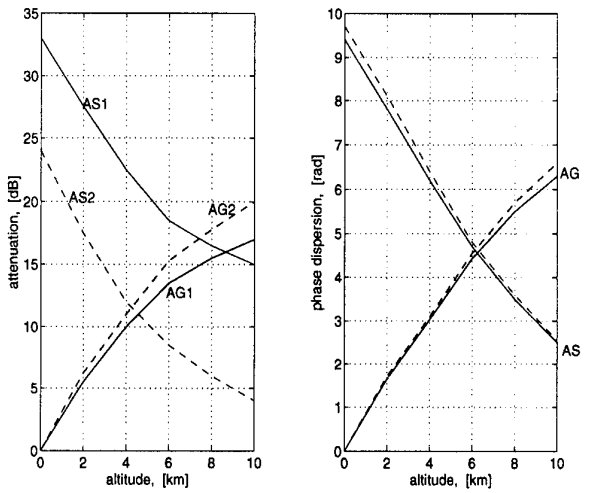


Fig. 4. *Path loss in dB and phase dispersion in rad as a function of altitude for a typical civilian aircraft ascent.* AS: aircraft-to-satellite path, AG: aircraft-to-ground path. — : f_1=54.671 GHz, -- : f_2=54.862 GHz.

The output from APM is shown in Fig. 4, where attenuation and dispersion profiles at selected frequencies in the 60 GHz O_2 absorption band both for AS and AG paths are plotted. It is then possible to evaluate the isolation between the AS and AG paths. At 4 km AS2=AG2=12dB, at 6 km AS2=8dB, AG2=16dB and the isolation is 8dB. As the aircraft advances to higher altitudes the isolation increases to 12dB at 8 km and 17dB at 10 km, while the AS path loss reduces attractively to 6.25 and 3.75dB, respectively.

7. CONCLUSIONS

A modelling approach for the calculation of the transfer function of the clear atmosphere has been presented. The model predicts the absorption and dispersion resulting from the principal contributors O_2 and H_2O and other trace gases such as O_3, N_2O, HNO_3, etc in the earth's atmosphere. Trace gases play an important role in the characterisation of atmospheric radiowave propagation when slant paths at high altitudes are considered. Their contribution to attenuation may not be negligible, indeed for selected 'window' frequency intervals trace gas contributions may be dominant.

References

[1] HERZBERG, G., *Molecular spectra and molecular structure I: Infrared and Raman spectra of polyatomic molecules*, Van Nostrand Reinhold Company, 1945.

[2] HERZBERG, G., *Molecular spectra and molecular structure II: Spectra of diatomic molecules*, Van Nostrand Reinhold Company, 1950.

[3] ROTHMAN, L.S., GAMACHE, H.M., GOLDMAN, A., BROWN, L.R., TOTH, R.A., PICKETT, H.M., POYNTER, R.L., FLAUD, J.M., CAMY-PEYRET, C., BARBE, A., HUSSON, N., RINSLAND, C.P., and SMITH, A.H., 'The HITRAN database: 1986 edition', *Applied Optics*, Vol. 26, No 19, pp. 4058-4097, 1987.

[4] ROSENKRANZ, P.W., 'Shape of the 5mm Oxygen Band in the Atmosphere', *IEEE Transanctions on Antennas and Propagation*, Vol. AP-23, No 4, pp. 498-506, 1975.

[5] UPLINGER, W.G., 'A simple model for relative air mass', 4th conference on Atmospheric Radiation, Toronto, Canada (American Meteorology Society), June 1981.

[6] ROSENKRANZ, P.W., 'Interference coefficients for overlapping oxygen lines in air', *J. Quant. Spectrosc. Radiat. Transf.*, Vol. 39, pp. 287-297, 1988.

[7] PENNER, S.S., *Quantitative Molecular Spectroscopy and Gas Emissivities*, Addison-Welsey, Reading, Mass., chapter 3, 1959.

[8] ANDERSON, G.P, CLOUGH, S.A., KNEIZYS, F.X., CHETWYND, J.H., SHETTLE, E.P., 'AFGL Atmospheric Constituent profiles (0-120 km)', AFGL-TR-86-0110, May 1986.

[9] HUI, A.K., ARMSTRONG, B.H., and WRAY, A.A., 'Rapid computation of the Voight and complex error functions', *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 19, pp. 509-516, 1978.

[10] LIEBE, H.J., ROSENKRANZ, P.W., and HUFFORD, G.A., 'Atmospheric 60-GHz Oxygen Spectrum: New Measurements and Line Parameters', *J. Quant. Spectr. Radiat. Tr.*, Vol. 48, pp. 629-643, 1992.

[11] LIEBE, H.J., 'MPM - An atmospheric millimetre-wave propagation model', *Int. J. Infrared & Millim. Waves*, Vol. 10, No. 6, pp. 631-650, 1989.

[12] LIEBE, H.J., and HUFFORD, G.A., 'Modelling Millimetre-wave Propagation Effects in the Atmosphere', AGARD *CP-454*, Paper No. 18, October 1989.

[13] HILL, R.J., 'Dispersion by atmospheric water vapor at frequencies less than 1 THz', *IEEE Trans. Ant. Propagat.*, Vol. 36, No. 3, pp. 423-430, 1988.

[14] MA, Q., and TIPPING, R., 'Water vapor continuum in the millimeter region', *J. Chem. Phys.*, Vol. 93, No. 9, pp. 6127-6139, Nov. 1990.

[15] STONE, N.W., READ, L.A., ANDERSON, A., DAGG, I.R., and SMITH, W., 'Temperature dependent collision induced absorption in nitrogen', *Canadian Journal of Physics*, Vol. 62, No. 184, pp. 338-347, 1984.

[16] PAPATSORIS, A.D., and WATSON, P.A., (1993), 'Calculation of the absorption and dispersion spectra of atmospheric gases at millimetre- wavelengths', *IEE PROCEEDINGS-H*, Vol. 40, No. 6, pp. 461-468.

[17] PAPATSORIS, A.D., and WATSON, P.A., (1993), 'Potential use of oxygen absorption bands for aeronautical communications', *Electronics Letters*, Vol. 29, No 22, pp. 1987-1988.

MILSATCOM - Where Are We Now?

J.B.Cairns & J.R.Watson
Matra Marconi Space (UK)
Gunnels Wood Road
Stevenage
Hertfordshire
SG1 2AS, England

1. INTRODUCTION

The military use of space for communications has seen dramatic changes since its first use in the 1960's. This paper presents some of the advances made, explores some of the current technical trends and proposes some key areas which need to be addressed in the future.

The use of 3-axis stabilised rather than spin stabilised satellites has enabled a significant increase in the DC power which can be generated using large solar arrays and has provided the platform necessary to mount sophisticated electronics and large, complex antenna systems.

The advantages of satellite communication over terrestrial systems in terms of coverage, availability and ruggedness are now widely accepted. The modern demand for highly flexible and rapid force deployment by NATO and Allied powers has further strengthened the case for satellite communications. Survivable communications is a vital component in the C^3I systems necessary for efficient military deployments during stress.

The Gulf Conflict demonstrated for the first time the full potential of space systems in military operations. Satellite communications in particular played an essential role in support of all operations both land based and shipborne. Satellite communications continue to provide vital support to NATO activities in Bosnia.

At a time when budget pressures are leading to critical reviews of expenditures amongst allied nations the need for the retention of a highly resilient space resource remains paramount. The end of the Cold War has not fundamentally reduced the potential threats to the Allied nations. Satellites remain a key resource to maintain both national security and for humanitarian and peace keeping roles worldwide.

2. TRENDS IN MILITARY SATCOMS

2.1 Mobility

There has always been a military requirement for reliable, light weight, communications terminals. Until the 1980's with the advent of new generation satellites high mobility was certainly not a feature of satellite communications. The TCS 500 terminal once deployed by the UK military exemplifies this : the terminal had a 13m antenna, a 10kW klystron, took 35 men to operate and a Belfast and 4 Hercules aircraft to deploy [6]. The terminal provided 1 secure speech and 12 telegraph channels. This was still hailed at the time as providing a significant improvement over the previous HF systems. The modern satellite terminal with equivalent performance to the TCS 500 system is now a man portable system typically weighing on the order of 10-15Kg deployed and operational within a few minutes by one person.

There remain practical limits to the miniaturisation of military systems in stressed situations however. This is governed by the anticipated threats to the network. Smaller terminals will always be more susceptible to uplink jamming despite the advances made to ECCM techniques. Although handheld systems of the types envisaged for next generation civil systems will have applications for routine communications, larger portable and mobile terminals will continue to be necessary to provide the EIRP necessary to guarantee at least low rate communications in times of stress.

2.2 Frequency Bands

Military satellite communications operate in UHF, X-band and EHF frequencies. Until the introduction of the FLTSATCOM EHF Package and MILSTAR systems, military traffic was exclusively carried at UHF and SHF.

There are several compelling reasons for the move to EHF in military satcoms. Wider available bandwidths provide more ECCM protection and higher frequencies, improves the LPE performance and increase the availability of circuits following exo-atmospheric nuclear events. EHF is less affected by black out and scintillation effects than SHF in these situations. There are however also disadvantages to the wide scale use of EHF namely system cost and circuit availability.

The main technique to improve the availability of a link due to rain-induced signal attenuation is to implement a so called rain margin. This margin however increases drastically with the desired link availability, particularly at high frequencies. At X-band margins of the order of 5 dB are normally employed to guarantee link quality at the typical 99.9% availability levels required.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

Corresponding margins at EHF can be of the order of 20 - 30 dB depending upon the climatic region. In the absence of stress SHF therefore provides a higher link availability. EHF could not, for example, meet the availability required of a civil ISDN network.

The major performance tradeoff in the choice of SHF or EHF is between the enhanced LPE and A/J performance at EHF and the enhanced overall availability at SHF. The replacement value of the currently installed infrastructure (mainly SHF) must also be considered. Since the added LPE and A/J protection is of greatest benefit to small disadvantaged terminals the general trend is to allocate EHF to serve this need in next generation systems. Large transportable and fixed strategic installations are not so disadvantaged against jammers in terms of EIRP. In the future it is probable therefore that they will continue to provide the high rate trunk backbone services for military traffic.

3. TECHNOLOGIES

The increasing trend in the mobility of military forces associated with the use of lightweight portable terminals will be a key driver of the design of MILSATCOMs.

The significant advances in the flexibility and mobility of military satellite communications has been primarily the result of advances made in satellite design. Key features include :

- the use of a range of ECCM techniques and on board processing in particular
- the use of A/J and high gain spot beam antennas which can be directed to any point on the earth
- intersatellite links
- life extending propellant systems

3.1 On Board Processing

On board processing (OBP) is becoming an essential feature of military satcom systems in order to obtain the full advantage of A/J protection on the critical uplink path against jamming attack. OBP will become a common feature on future MILSATCOM systems and is already used to various levels on the current SKYNET, DSCS and Syracuse systems particularly on the critical satellite command channel. MILSTAR and some FLTSATCOM satellites incorporate OBP on both the command and communications channels.

New generation military systems are employing frequency hopping spread spectrum (FHSS) processors. FHSS has several advantages over direct sequence spread spectrum (DSSS). In particular, significantly wider bandwidths can be practically implemented and the FHSS waveform is more robust against sophisticated jammer strategies. This has particular application to small terminals which require as much protection as possible against jammers.

It is probable that future military systems will benefit from the advances currently being made in the civil satellite sector. Although not specifically designed as anti-jam systems, the advantages of OBP in terms of improvements in link performance, channel capacity and data rate conversion [3] are of significant importance to the civil operators. The ability of OBP satellites to perform autonomous traffic routing and switching and the reduced dependence on the ground segment can also be important.

These advantages have encouraged civil systems to move towards OBP and their requirements in terms of processing capability are likely to be much greater than those of military systems. It is probable that in the future technology transfer will occur from civil to military systems.

In addition to its ECCM advantages for military systems OBP also enables the implementation of additional state of the art techniques for both the civil and military areas. Specific area that require or benefit from OBP capabilities are :

a) the use of multibeam antenna systems in which digital beam forming can allow the generation of large numbers of beams with independent characteristics thus saving spacecraft power.

b) highly complex channel to beam routing can be achieved compared with analogue implementation.

c) the introduction of inter-satellite links for integrating a constellation of satellites.

The first two can be associated either with regenerative or transparent architectures, while ISLs typically require regeneration of the user data.

3.2 Antenna Systems

Early spaceborne antenna systems of the type used on for example the SKYNET 2 satellite comprised simple earth cover horns and reflector system which had low gain (typically 17 dB). The resulting EIRP was consequently low and large aperture ground stations were therefore required to maintain the communications link. Such antenna systems also had no ECCM protection facilities to mitigate against the effects of uplink jamming.

Current and planned military antenna systems capable of providing jammer protection are now among the most complex elements of the communications payload. The systems not only contain the radiating structure but also the transmit/receive electronics and the control electronics required both to detect and null interfering sources. Two basic antenna types are generally used to provide the A/J function: phased arrays or multi-beam antennas. Each has advantages depending upon the specific requirement [4].

Fixed and mechanically steered high gain spot beams are also now used in modern military satellites such as SKYNET 4 to provide communications to small tactical

and man portable terminals.

Progress in the civilian field has also been significant. For example, the latest generation Inmarsat 3 payload incorporates a highly flexible antenna system the gain of which can be adapted electronically to cope with various traffic demands within the satellite field of view.

3.3 Intersatellite Links

Intersatellite links provide advantages in terms of eliminating the requirement for multihop through ground station relays. This has potential significance militarily since a nation can control complete global coverage of satellites from its own territory. Ground relays outside national jurisdiction could be more vulnerable in times of stress.

Intersatellite links at RF are chosen to lie in the oxygen absorption bands to eliminate the chance of interception from the ground. 60GHz links are currently employed on the MILSTAR system. Optical intersatellite links have also been proposed for future satellite systems although no operational optical system has yet been deployed in space. However optical technology has evolved to a point where the feasibility of such systems for space applications has been demonstrated.

3.4 Propellant Systems

Both military and civilian satellites are designed with a significant amount of redundancy and the use of life limited items is minimised wherever possible. The net result is that the primary end of life driver in satellites is often the amount of fuel carried ot maintain the satellite on station. New generation electrical/ion propulsion systems are now maturing which will enable satellites to be maintained on station for greater than the current levels of around 10 years. Ion propulsion can reduce the propellant requirement of geostationary satellites by some 90%.

Launch costs contribute some 30% of the total through life cost of a system and are normally proportional to satellite mass. Launch cost reduction combined with life extension in orbit can lead to significant savings on the total cost.

4. SYSTEM MANAGEMENT AND CONTROL

Future military satcoms will become increasingly complex to support both the greater traffic volumes, the greater diversity of users (particularly highly mobile netted users), improved spectrum monitoring and advanced ECCM techniques. Current command and control measures will be unable to cope with this increase in volume and sophistication. Coupled with this is the continual pressure to reduce the overall life cycle costs of systems which inevitably leads to requirements for reduced manning levels.

Given these conflicting constraints there will be an increasing requirement for higher levels of automation in system management and control. In the long term the use of AI techniques can be envisaged although this technology is still at an embryonic stage. Short to medium term solutions will include a high degree of information pre-processing, enhanced HCI displays, improved fault diagnostic techniques and the possible use of knowledge based systems to deal with satellite or network anomalies. Specific areas that could be addressed by these techniques are ECCM, satellite access control. link performance assessment and management.

The increasing sophistication of satellite on-board ECCM methods such as jammer detection and nulling will also require future spacecraft to have a greater degree of autonomy. Conventional TT&C operations which have hitherto been undertaken from the ground will not be feasible given the significant amount of data transfer required and the loop around time constraints which would introduce an unacceptable delay to the operation of, for example. antenna nulling algorithms. It will be essential that a high degree of processing is performed on board the satellite to enable only top level command and telemetry to be required.

5. THE MILITARY USE OF COMMERCIAL SATELLITE SYSTEMS

The basic technologies used for both civil and military satellites are essentially the same. There has therefore been pressure to determine if nations do indeed require a dedicated military satellite capability to meet the requirement.

Military systems are currently hardened against ECM, nuclear burst radiation effects, EMP and laser attack. The communications capabilities of military systems are however, being rapidly outpaced by commercial systems. The reasons for this are largely due to the military requirements for high survivability, security and reliability in a stressed environment. These requirements make military systems more expensive than the commercial counterparts, less flexible and technologically more conservative.

Present day commercial systems do not however, provide a universal answer to military communications. Although systems can and are used for routine military activities such as administration they are incapable of meeting fundamental military operational requirements due to the following constraints :

- They are highly vulnerability to traffic interception and exploitation.
- The jamming threat in many parts of the world is real. The technology to produce jammers at military and civil satellite frequencies is inexpensive, comparatively low tech, available and readily deployable.
- There is no direct national control over the space segment.
- Having total control of the system to modify

transmission, frequency allocations, bandwidth and beam coverage are key factors to the military user. A civil system cannot guarantee the same response time or traffic connectivity to an area of conflict.

- There are difficulties associated with international frequency co-ordination and related political and legal constraints.
- Commercial satellite communication system are not permitted by convention to carry operational war traffic, which at a time of crisis eliminates its use for military purposes. Ownership and control of the system is essential to ensure that rapid response is possible to re-allocate capacity, assign new priorities, re-point coverage zones or even charge the satellites orbital position.

Studies have now been conducted by many nations. In summary the basic conclusion is that civil systems are perfectly adequate in peace time scenarios but are highly susceptible to ECM and physical attack in times of stress.

Civil systems will continue to have a significant role in military communications despite these drawbacks. The US army is currently adding triband terminals to its communications infrastructure [1]. This will increase the flexibility and permit the use of commercial satellites, as directed by recent congressional action and Defence Department guidance. It is probable that similar techniques will be adopted by other allied nations.

6. CURRENT SATELLITE SYSTEMS

Although the following is by no means an exhaustive survey of operational satellites, these systems represent the current state of the art in satcom and military communications and include the main technical features discussed in this paper.

6.1 ACTS

The Advanced Communications Technology Satellite (ACTS) is a NASA experimental satellite launched in 1993. Its purpose is to investigate new technologies, such as on-board processing and utilisation of the Ka band.

Although the ACTS is essentially experimental and operates in the Ka band rather than the military EHF band, the system incorporates many of the features that will be utilised in future systems. ACTS key technologies include electronically hopping multiple spot beam antennas, on-board processing and switching and Ka band transmit and receive subsystems. Of particular relevance to EHF communications is the incorporation of power control in the ACTS ground stations and satellite to compensate for atmospheric fading due to precipitation.

6.2 FLTSATCOM

The FLTSATCOM system was the first operational system for mobile tactical users. Later flight versions (7 and 8) incorporate an experimental EHF package containing a spot beam and a global beam antenna system. The satellite also houses an autonomous resource controller which automatically sets up circuits at various data rates between the beams to support individual users. It has often been heralded as the first true "switchboard in the sky."

6.3 SKYNET System

The current 3 operational SKYNET 4 satellites and the proposed SKYNET 4 Stage 2 next generation systems provide the UK MoD with a rugged and flexible space resource into the next century. The system together with its predecessor SKYNET 2 has proven the benefits of satcoms to the UK in the Falklands, the Gulf and Bosnia.

The satellites contain SHF, UHF and experimental EHF payloads with protected broadcast and command channels and an on board nulling antenna system. Enhancements on the Stage 2 systems include steerable spot beams, more sophisticated nulling capabilities and commandable UHF channel frequencies. These facilities will provide a significant increase in capabilities in particular for the deployment of small mobile and man portable systems.

6.4 DSCS

The Defence Satellite Communication System (DSCS) provides the main milsatcom bearer system for the USA. The current generation DSCS III satellites contain advanced SHF multiple beam antennas that can be electronically switched to provide beam coverages ranging from earth cover to spot beam. The system also provides an advanced A/J nulling capability.

The DSCS satellite constellation has been designed for use into the next century. Follow on programmes may include further enhancements in OBP and A/J capabilities. Studies are on-going to determine if the new system will be an exclusive US development or form part of an international satellite programme with the UK and France.

6.5 Syracuse II

The French military space segment uses a multipurpose Telecom II satellite which operates in the C and Ku bands for civil applications and at X-band for military purposes.

Key features of the system include a degree of on-board autonomy and a flexible antenna system with fixed and steerable beams. The system operates at X-band and provides communications to fixed and mobile terminals including airborne and submarine platforms.

6.6 MILSTAR

The current state of the art in MILSATCOM is undeniably the MILSTAR system. This is a genuine highly survivable war fighting system. The two initial MILSTAR satellites provide interoperable voice and data

links up to 2.4kbps at EHF. Each satellite has a capacity of 192 uplink channels operated in one of nine satellite service beam areas. The next generation satellites, if approved, will include upgrades in communications capacity to 1.544Mbps and enhanced nulling antenna capability.

The system comprises highly directive antennas to reduce the probability of jamming or detection by enemy forces. A sophisticated on board processing system and communication waveform also provides a significant amount of A/J protection as well as making transmissions difficult to detect. The system is aimed primarily at support of tactical circuits to highly mobile units.

6. CONSIDERATIONS FOR THE FUTURE OF MILITARY SATELLITE COMMUNICATIONS

It is unlikely that fundamental changes will occur in the NATO and Western allied country infrastructures in the near future. The US, UK and France in particular have made significant investments in their respective satellite infrastructures and any new system will be required to be backwards compatible with these systems to maximise the investments made. This approach has the obvious cost advantages of using existing facilities. It can however, tend to proliferate the conservation design approach which is often associated with military systems.

Cost is and will remain a major issue in the procurement of future MILSATCOM systems. Western allied nations are currently actively considering the possibilities of collaborative military space programmes in the Inmilsat (US/UK/France), Bimilsatcom (UK/France) and Eumilsatcom (European) initiatives in a wide range of international study programmes. The various nations are taking advantage of the fact that their individual MILSATCOM space resources will be nearing their design lifetimes at approximately the same time. The aim will be to use this opportunity to cost share the resources although their remain a number of issues regarding control of the space segment resource and the optimum technology solution to meet the varying requirements which need to be resolved.

Advanced digital coding and data compression systems for both speech and video systems will enable more efficient use to be made of the available bandwidth. Future systems will also be anticipated to use more demand assigned multiple access (DAMA) techniques to provide more flexibility and increased capability. DAMA networks are already replacing dedicated single channel per carrier systems both in the military and civil (VSAT) areas [5]. INMARSAT currently operates commercial DAMA satcom networks for mobile users. DAMA systems are also operated in the military UHF band on the UHF - Follow on satellite, FLTSAT and LEASAT by the US DoD.

Military satellite communications are essential to provide the operational flexibility and reliability under a stressed environment. This can only reliably be provided with a dedicated resource, although civil systems have their place as part of the media mix.

It is debatable whether military satellite communications will wish to follow the trends into flexible satellite access and call management systems which are now becoming prevalent in the civil satellite sector. Military systems will always need to retain a strict command hierarchy.

The view of satellite communications merely providing a rear link back to HQ is also changing. Satellite terminals now also play a vital role in tactical battlefield operations. The new generation of highly mobile ground terminals operating at X-band and EHF are now key to military operations.

Military satellite communications has proved its worth in the Gulf which has been heralded as the first space conflict. Current and planned military satcom systems will continue to provide robust anti-jam communications so crucial to successful operations on the modern battlefield.

References

[1] Signal, July 1994

[2] Considerations for NATO Satellite Communications in the Post 2000 Era. AGARD-AG-330

[3] On-Board Processing, E W Brugel

[4] Considerations for NATO Satellite Communications in the Post 2000 Era, pp 100 - 101 AGARD-AG 330

[5] Satcom Giving You a Busy Signal? Dial-up DAMA

[6] What Future for UK Military Satcoms? Space Volume II, No. 2, pp 19 - 21

DISCUSSION

Discussor's name: L. Castanet

Comment/Question:

Concerning the use of EHF, you said that it permits fast recovery from scintillations.

The OLYMPUS campaign showed that scintillations are higher with frequency. Could you explain that point?

Author/Presenter's reply:

I was referring to scintillation effects after hostile nuclear events.

Discussor's name: G. S. Brown

Comment/Question:

What effect do you see the "assembly in space" approach having on the future of communication satellites?

Author/Presenter's reply:

Commercial systems insist on the ability to use alternative launch methods - which tends to favour conventional launch vehicles. In the medium term this is also sufficient to provide the mass/power/size needed for the missions. "Assembly in space" could be applied further in the future, but we are not so far seeing needs emerging. Military systems are likely to follow commercial practice for cost reasons.

Discussor's name: J. Aarons

Comment/Question:

What is the role of UHF transmissions in new military satellites?

Author/Presenter's reply:

UHF is very popular with military users because of the low cost and simplicity of the terminals, despite the vulnerability of UHF to jamming and exploitation. The simplicity and wide-beam width of the antennas also make UHF SATCOM easy to use. However, in addition to EW vulnerability it is very prone to interference.

Earth-Space Links and Fade-Duration Statistics

Faramaz Davarian
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109, U.S.A.

1. Abstract

In recent years, fade-duration statistics have been the subject of several experimental investigations. A good knowledge of the fade-duration distribution is important for the assessment of a satellite communication system's channel dynamics: What is a typical link outage duration? How often do link outages exceeding a given duration occur?

Unfortunately there is yet no model that can universally answer the above questions. The available field measurements mainly come from temperate climatic zones and only from a few sites. Furthermore, the available statistics are also limited in the choice of frequency and path elevation angle.

Yet, much can be learned from the available information. For example, we now know that the fade-duration distribution is approximately lognormal. Under certain conditions, we can even determine the median and other percentiles of the distribution. This paper reviews the available data obtained by several experimenters in different parts of the world. Areas of emphasis are mobile and fixed satellite links. Fades in mobile links are due to roadside-tree shadowing, whereas fades in fixed links are due to rain attenuation.

2. Introduction

Digital radio communication systems are influenced by the dynamics of fade events, that is, fade rate, fade duration, and inter-fade duration. In a digital receiver, signal decoding and synchronization (timing and phase) can be affected by the distribution of fade events. For long fades, a receiver could lose lock, and reacquisition may become necessary. Therefore, fades and their statistical distribution influence service availability, receiver loss of lock probability, and decoder performance. We will examine fade-duration statistics on Earth-space links.

On satellite links, fade events can result from atmospheric causes (rain) or, in the case of mobile satellite links, from tree shadowing and blockage. The analysis of measured data has shown that both types of fade give rise to fade-duration statistics that can be approximated with a lognormal distribution [1, 3,4].

The lognormal probability density function is given by

$$p_z(z) = \frac{1}{sz\sqrt{2\pi}} \exp[-\frac{(\ln(z)-m)^2}{2s^2}] \qquad (1)$$

where z denotes the random variable, and s and m are the standard deviation and the mean of ln(z). The cumulative distribution of z can be obtained by integrating the above function from Z to infinity resulting in the probability that z is larger or equal to Z:

$$P(z \geq Z) = \frac{1}{2} erfc[\frac{\ln(Z)-m}{s\sqrt{2}}] \qquad (2)$$

where *erfc* denotes the complementary error function. For convenience, Eq. (2) can be multiplied by 100 to express the probability as a percentage.

3. Fade-Duration Statistics in Mobile Satellite Links

Fade-duration statistics caused by tree shadowing were obtained from measurements made in south-eastern Australia at a frequency of 1.5 GHz and a path elevation angle of 51° [1]. The data were shown to fit the lognormal distribution with good accuracy when m = -1.514 and s = 1.215, for a 5-dB fade threshold. The distribution encompassed road types that exhibit moderate and extreme shadowing —see reference 1 for a description of shadowing types. This distribution is shown in Figure 1. Note that fade duration is given in meters rather than seconds. Distance durations may be converted to time durations by dividing the former by the speed of the vehicle, which was nominally 25 m/s in the Australian experiment.

For this example, 50 percent of all the fade events consisted of fades shorter than 0.22 m. At a speed of 25 m/s, this translates to less than 0.01 s. Ninety-nine percent of all fade events consist of fades with durations shorter than 4 m, which is equivalent to a time duration of 0.16 s. Therefore, it appears that, for this example, fade durations are generally short enough to avoid receiver synchronization problems (at least for most of the time). Other experimental data have shown that for tree shadowing, fade durations increase with decreased elevation angle. For instance, the fade duration for a 30° elevation angle has been observed to be twice that of a 60° elevation angle at the same probability level [1]. If

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

this observation is applied to the data presented in Figure 1, the fade-duration distribution for an elevation angle of about 25° can be estimated by adding the constant ln (2) to m. The result is shown by the dashed curve in Figure 1.

The lognormal probability density of Equation 1 is a good model for the duration of fades of moderate depth and length. Experiments conducted at 20-GHz in southern California, USA, indicate that when heavy shadowing causes deep fades with relatively long durations, an exponential model fits the data better [2].

4. Rain Attenuation

Fades due to rain attenuation have been investigated in several experimental campaigns. Signals propagating through the atmosphere experience both attenuation and scintillation effects. Since signal fluctuations due to scintillation contain higher frequency components than the ones caused by rain attenuation, low-pass filtering of the observed signal can remove the scintillation effect. To focus our discussion to fade events due to rain attenuation, only data with scintillation-induced fluctuations removed are considered.

Reference [3] reports on a four-year experiment in Texas, where the elevation angle is 6° and the signal frequency is 11 GHz. Fade-duration statistics for fade thresholds from 5 to 20 dB are shown to be approximately lognormally distributed with the 50% probability level corresponding to a fade duration of about 300 s. Figure 2 shows the lognormal fit to the data. In this example, 50% of all fades are shorter than 300 s, and 90% of all fades are shorter than 2500 s. For a threshold of 20 dB, measured fade durations are shorter than the lognormal model predictions at high duration values. The asterisk at the 2000 s duration value in Figure 2 shows the measured data for a threshold of 20 dB. This means that, for a fade threshold of 20 dB, fades longer than 2000 s are rare.

There is evidence that the lognormal model also applies to a mix of data taken at different sites and different frequencies. Figure 3, based on data from [4], shows an approximate lognormal fit to data obtained from three sites, one in the U.S. (Virginia) and two in Europe. The data include three frequencies of 12, 20, and 30 GHz, and fade thresholds in the range of 4 to 10 dB. The Virginia site[1] had the lowest path elevation angle of 14°. The 50 percentile value in this case is 200 s compared to 300 s in Figure 2. Another difference between these two cases is that the percentage of very long fades, that is, fades longer than 1000 s, is less in this case than in Figure 2. This may be attributed to the data in Figure 2 corresponding to the very low elevation angle of 6°.

Another subject of interest is the absolute number of fades at a given threshold. The absolute number of fades depends on the frequency of the signal. The number of fade events increases with increasing frequency. For the Virginia site, Reference [5] shows the number of fades as a function of fade duration for frequencies of 12, 20, and 30 GHz and an elevation angle of 14°. For example, at a fade threshold level of 5 dB the number of fade events with fades equal to or exceeding 1 minute is 108 at 12 GHz, 605 at 20 GHz, and 935 at 30 GHz. These values for a fade threshold of 15 dB are 6 at 12 GHz, 75 at 20 GHz, and 200 at 30 GHz. Clearly, the number of fade events decreases when the threshold level increases. Since system engineers use the fade threshold information to determine the required fade margin in a system, it is useful to further explore the relationship between the fade duration statistics and the fade threshold (attenuation level). The next section will address this topic.

5. Fade Duration and Attenuation Level

The probability of service availability is a very important parameter of link design in a satellite communication system. The availability of the service is directly related to the link margin. Therefore, it is important to examine the fade-duration distribution as a function of signal attenuation. Presenting fade duration in this fashion will enable system designers to conveniently plan the optimum value of the fade margin for a system. Note that too little margin will result in an unacceptable grade of service, whereas too much link margin will give rise to high system costs.

Reference [6] considers this issue by relating the occurrence of fades of a certain duration to attenuation level. Although the model presented in [6] is a preliminary one, it is noteworthy in that it addresses an important system need. In this approach, fade durations have been binned into predetermined intervals. For example the model for fade durations between 30–60 s is given by

$$FT_{12} = 8900\exp(-1.315A + 0.118A^2 - 0.004A^3) \quad (3)$$

where FT_{12} denotes the cumulative fade time in seconds for a frequency of 12 GHz and A is the attenuation level in dB ($3 \le A \le 16$). Figure 4 shows a plot of Eq. (3) for attenuation levels from 3 to 16 dB. For example, at a margin of 9 dB all fades lasting for 30–60 s add up to 500 s of link outage in one year of observation. The data used for this model were also obtained from the Virginia site mentioned earlier with a path elevation angle of 14° to the satellite.

Because the database used to develop Eq. (3) also included 20 and 30 GHz data, this equation was extended to a model that predicts fade time for three frequencies [6]. This model is given as

$$FT_f = 8900(-8.899 + 0.915f - 0.008f^2)\exp(-1.315A + 0.118A^2 - 0.004A^3) \quad (4)$$

where FT_f is the cumulative fade time in seconds for the 30–60 s fade duration interval at a given frequency, f is the frequency in GHz, and A is the attenuation in dB. The dashed curve in Figure 4 shows the above model for 20 GHz. An extension of the above model to other fade duration intervals, e.g., 60–120 s, is also given in [6].

Although Eq. (4) offers a powerful tool for determining fade-duration distribution as a function of attenuation, there is a caveat associated with this tool: The model was developed based on measurements at a single elevation angle and in a single rain climate zone. Therefore, its extension to other elevation angles and climate zones is not clear. The model may also be contaminated by statistical noise because it is based on only one year of observations. Measurements at different sites and for longer time durations will very likely contribute to the improvement of the model's accuracy and robustness.

[1] One full year of data was acquired in the 1990–92 period using the Olympus spacecraft.

6. Summary

Fade-duration statistics on satellite links can be approximated by the lognormal distribution. This distribution can be uniquely described by the mean and the standard deviation of the natural logarithm of the samples. This model applies to fades due to rain attenuation (fixed satellite) as well as fades due to tree shadowing (mobile satellite).

For shadowing-induced fades, fade-duration statistics seem to be insensitive to the amount of shadowing, i.e., light to heavy shadowing. However, a moderate dependence on elevation angle has been observed in field measurements. This dependence gives rise to longer fade durations with decreased elevation angle.

For rain-attenuation-induced fades, it appears that for temperate site locations, elevation angles above 14°, and frequencies between 12–30 GHz, the fade-duration statistics are not strongly dependent on site location, elevation angle, or frequency.

A simple model to relate fade time for a given fade duration to attenuation level is also discussed. This approach is useful for examining the impact of fades on the receiver, allowing for an estimation of total receiver downtime. Note that total downtime is the sum of the fade duration and system reacquisition time.

The models presented in this paper are based on a limited amount of experimental data. Further study is required to substantiate their application to specific slant path configurations.

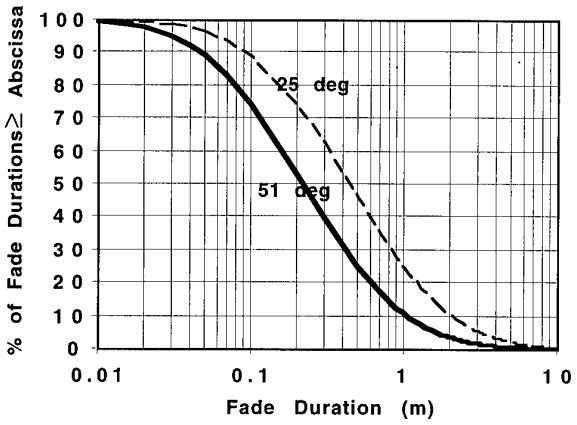


Figure 1. Fade Duration Distribution for Mobile Satellite Links (elevation angles 51° and 25°)

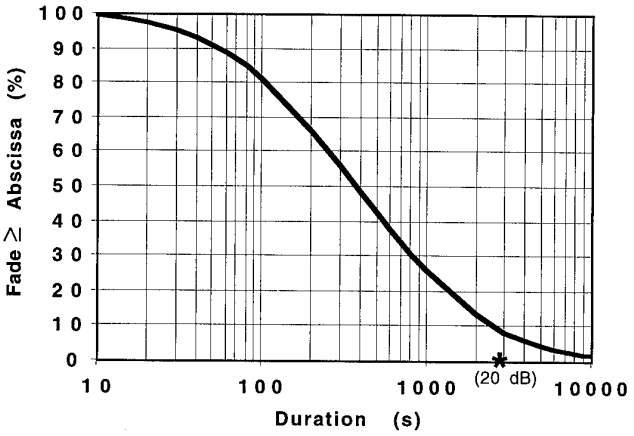


Figure 2. Texas 11-GHz Data

7. Acknowledgment

The work conducted for this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

8. References

1. J. Goldhirsh and W. Vogel, *Propagation Effects for Land Mobile Satellite Systems,* NASA Reference Publication 1274, Feb. 1992, pp. 41–45.
2. M. Rice, et. al., "K-Band Land-Mobile Satellite Channel Characterization using ACTS," *International Journal of Satellite Communications,* will be published in late 1995.
3. W. Vogel, et al, "Rain Fades on Low Elevation Angle Earth-Satellite Paths: Comparative Assessment of the Austin Texas, 11.2 GHz Experiment," *Proc. IEEE,* vol. 81, no. 6, Jun. 1993, pp. 885–896.
4. J. Baptista, et al, ed., *Reference Book on Attenuation Measurement and Prediction,* ESA WPP-083, Nov. 1994, pp. 101–107.
5. W. Stutzman, et al, "Results from the Virginia Tech Propagation Experiment Using the Olympus Satellite 12, 20, and 30 GHz Beacons," *IEEE Trans. Antennas and Propagation, vol 43, no 1, Jan. 1995, pp. 54–62*
6. H. Ajaz and A. Saffaai-Jazi, *Fade and Inter-Fade Durations in Ku- and Ka-Band Frequencies Measured From the Olympus Satellite Beacons,* Satcom Report 93-17, VPI&SU, Oct. 1993.

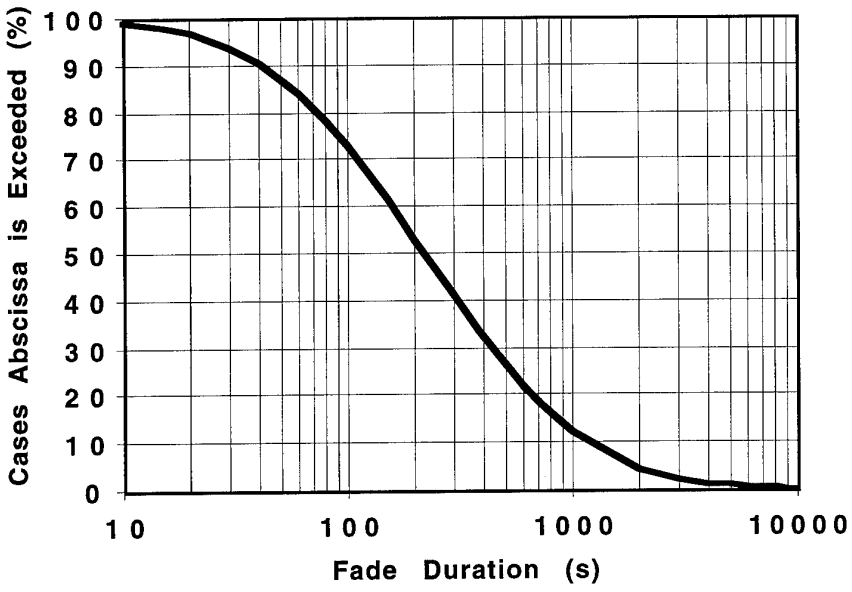


Figure 3. OPEX 12-, 20-, and 30-GHz Data

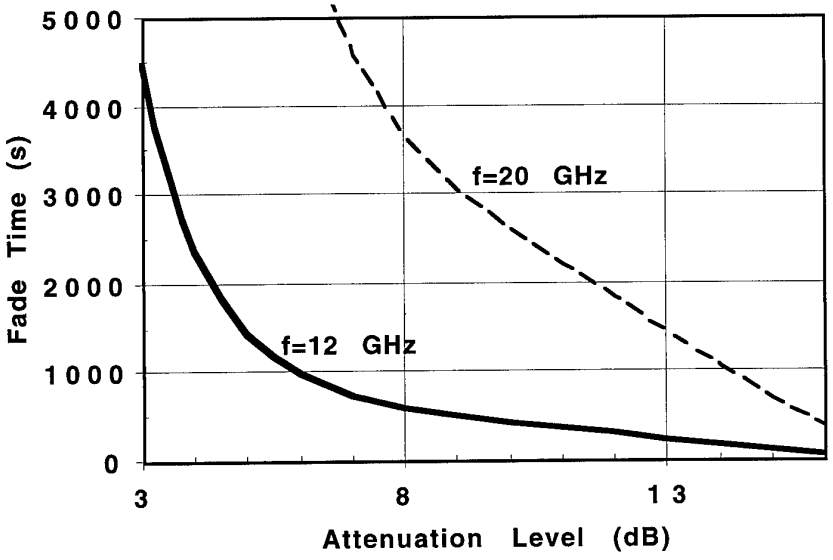


Figure 4. Annual Fade Time for 30–60 s Fade Durations

DISCUSSION

Discussor's name: W. Keydel

Comment/Question:

For spaceborne radar interferometry (two-pass) phase fluctuations become increasingly important. Did you ever look for phase distortions in frequency ranges relevant for spaceborne SAR (1 GHz - 10 GHz)

Comment: In SIR-C/X - SAR two pass interferometry together with your colleagues from JPL we have observed phase distortions probably caused by rain cells which simulated earth plate shifts up to dm. Additionally, modern modulation techniques like PCM also require phase stability.

Author/Presenter's reply:

Because our application was telecommunications, we did not investigate phase fluctuations.

Discussor's name: L. Castanet

Comment/Question:

I've seen in one of your previous communications that the ACTS campaign was open to the international community, as the OPEX. Could you give us some information on the data base. First of all, is it available to the scientific community? Finally, what kind of data is stored? I mean with or without processing?

Author/Presenter's reply:

The results are available for the scientific community, and new members are welcome to participate. Some results are beginning to be introduced into the data base. Both processed and unprocessed data will be available.

DISCUSSION

Discussor's name: L. Hoff

Comment/Question:

What is the sensitivity of the fade duration model to the signal level that causes the receiver to lose lock.

Author/Presenter's reply:

The model does not include the effects of the communication receiver losing lock or reacquiring lock. The measurement receiver used to measure fading lost lock only during very deep fades. Communication receiver fade durations will be longer than the fades shown because of losing lock during long and deep fades.

Discussor's name: N. Farsaris

Comment/Question:

What are the current requirements for link margins? (for a 20 GHz fixed link).

Author/Presenter's reply:

It depends on the application and system requirements. In my opinion, for a 20 GHz link 15dB of margin is likely to be adequate for most regions. This margin can be provided on an as needed basis to save cost.

Optogeometrical Characterisation of Erbium Doped Fibres

F. Ravet, B. Heens, E. Jaunart*, P. Mégret, , M. Blondel.

Service d'Electromagnétisme et de Télécommunications, Faculté Polytechnique de Mons, Boulevard Dolez, 31, B-7000 Mons, Belgium.
Tel.: +32 65 37 41 91, Fax.: + 32 65 37 41 99, E-mail: ravet@fpms.fpms.ac.be

*New Developments, BELGACOM, Bd E. Jacqmain, 177, B-1020 Brussels

Abstract

The reliability of the new photonic network components is becoming one of the challenges of the present years. The Erbium Doped Fibre (EDF) can be considered as one of these devices contributing to the optical communication systems upgrading. These are the reasons why easy-to-reach EDF optogeometrical properties seem to be more and more important. Within this context, an experimental set-up has been developed permitting the measurements of these properties (mode field diameter, core diameter, refractive index profile, Numerical Aperture...).

Our set-up is based on the Transmitted Near Field Technique (TNFT), combined with an infra red camera detecting the near field intensity magnified by an infra red aspherical lens. Such a method has been successfully demonstrated in the past for the optogeometrical characterization of single mode fibres. The combination of the TNFT with the infra red camera opens the possibility of yielding very quick measurements.

The importance of the camera lies in the two dimensional intensity detection: such a device records and instantaneously displays the intensity distribution. The NFI pattern is thus easy to get. The mode field diameter has been determined from the Near Field Intensity (NFI) pattern and we then calculated the refractive index profile difference. Finally, we determined the fibre core radius and numerical aperture. We then compared them with the Equivalent Step Index parameters we have calculated. We carried out all these measurements at 1310 nm and 1550 nm. So we have direct informations on the optogeometrical parameters in the second and third windows.

1. Introduction

Since the beginning of the eighties the Transmitted Near Field Technique (TNFT) has been succesfully used to characterize optical fibres. Authors have demonstrated that it is possible to derive optogeometrical parameters from the near field intensity patterns. Some carried out the measurements to get the refractive index profile [1]-[5],while others were more interested in determining the Mode Field Diameter (MFD) [1], [6], [7], the geometrical datas or the Equivalent Step Index (ESI) parameters [1], [8], [9].
In our work, we studied and we applied some of these results to the characterization of commercial Erbium Doped Fibres (EDF). Basically, we built our experimentation following the ITU recommendations G.651 and G.652 on TNFT and we treated the datas in agreement with theories from [1]-[5] on index profile and [1],[6],[7] on MFD. We then derived the core diameter and the maximum Numerical Aperture (NA_{max}) as defined in the G.651 recommendation. We calculated the ESI parameters [8],[9] and we compared them with our measurement results at 1310 nm and 1550 nm.

2. Theory

2. 1. The TNFT description

The TNFT is based on the Near Field Intensity pattern detection by the experimental set-up shown in figure 1 and described in the G651 and G652 recommendations. In fact, we detect the magnified fibre output intensity pattern with an infra red camera. Point 3 gives a more detailed description of the experimental set-up.

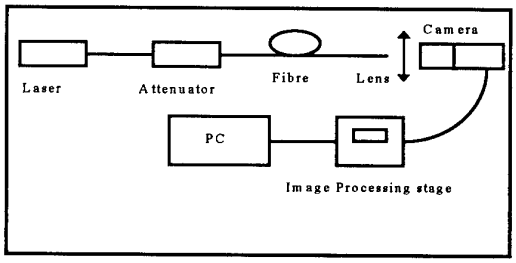


Figure1: experimental set-up.

2. 2. Refractive Index profile

It is possible to calculate directly the refractive index profile from the bidimensionsional intensity pattern [4]. Only one assumption was made: optical fibres are weakly guiding waveguides. It has been demonstrated that a stricter assumption could simplify the calculation [1]-[3], [5]: When optical fibres are axially symetrical, the useful formula could be derived from the scalar propagation equation. This formula is

$$n^2(r)=\frac{\beta^2}{k^2}+\frac{m^2}{k^2r^2}-\frac{1}{k^2I(r)}\times \left\{\frac{1}{2}\frac{d^2I(r)}{dr^2}-\frac{1}{4}\frac{1}{I(r)}\left(\frac{dI(r)}{dr}\right)^2+\frac{1}{2}\frac{dI(r)}{dr}\right\} \quad (1)$$

It gives the index profile as a function of the radial NFI $I(r)$. k is the wavenumber in the vacuum, β is the mode propagation constant, m is the azimutal mode number and r is the radial coordinate. In our case, we are only interested in the index profile difference

$$\Delta n^2(r)=-\frac{1}{k^2I(r)}\left\{\frac{1}{2}\frac{d^2I(r)}{dr^2}-\frac{1}{4}\frac{1}{I(r)}\left(\frac{dI(r)}{dr}\right)^2+\frac{1}{2}\frac{dI(r)}{dr}\right\} \quad (2)$$

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

assuming the single mode operation during the measurement. From (2) we are able to derive the fibre core radius and numerical aperture.
The interest of (1) and (2) lies in the fact that they establish a direct relationship between the experimental data *I(r)* and the investigated quantity

2. 3. Mode Field Diameter

Petermann has derived formulas to calculate the MFD. from the NFI pattern [1], [6]. We are interested in the definition related to lateral displacement losses in fiber interfaces as well as to dispersion optical waveguides.This definition is also referred in Recommendation G.652 for MFD calculation from NFI pattern. Hereafter, we give the expression of the definition for an axially symmetrical fibre, called the Petermann 2 MFD definition

$$MFD = 2w_{p2} = 2\left\{2\frac{\int_0^\infty rI(r)dr}{\int_0^\infty r(\frac{d\sqrt{I(r)}}{dr})dr}\right\}^{\frac{1}{2}} \quad (3)$$

w_{p2} is the mode field radius and is also known as the spot size. *I(r)* and *r* are respectively the radial NFI and the radial coordinate at the output of the fibre under test. Authors have also shown that is possible to calculate the MFD of axially non-symmetrical optical fibre based on a similar definition [7].

2. 4. Optogeometrical Characteristics

In agreement with the definitions from the G.651 and G.652 recommendations, we determined the core radius *a* and the maximum NA_{max} after having fixed a cladding reference level on the index profile. In our case, we considered the fixed cladding level as the mean value over the whole cladding. We obtained the square NA_{max} by substracting the maximum value in the core with that reference level.

$$NA_{max} = \sqrt{n_{max}^2(r) - n_{ref}^2} \quad (4)$$

We suppose that the core radius *a* is given by the intersection between the core profile and the cladding reference level.
We also calculated the square centre Numerical Aperture defined as follows

$$NA_{center} = \sqrt{n^2(0) - n_{ref}^2} \quad (5)$$

We should find the real NA between NA_{max} and NA_{centre}.

2. 5. ESI parameters

In this paper, we only considered the effective core radius a_{eff} and the effective Numerical Aperture NA_{eff}. These parameters can be calculated from the MFD value evaluated with the Petermann 2 definition [1],[8],[9]. In this way, we first had to calculate the ESI normalized frequency

$$V_{eff} = 2.405\frac{\lambda_{ce}}{\lambda} \quad (6)$$

where λ_{ce} is the measured cut-off wavelength and λ is the test wavelength. After having obtained a value for V_{eff}, we use the following formula [1],[8],[9] for the calculation of a_{eff}

$$a_{eff} = \frac{w_{p2}}{A + BV_{eff}^{-\frac{3}{2}} + CV_{eff}^{-6}} \quad (7)$$

Table 1 gives the value of the coefficients *A,B* and *C* following the definitions of [8] and [9].

References	*A*	*B*	*C*
[8],[1]	0.65	1.619	2.879
[9]	0.625	1.707	4.112

Table1: coefficients of (7)

We are then able to get the NA_{eff} from

$$NA_{eff} = 2.405\frac{\lambda_{ce}}{a_{eff}2\pi} \quad (8)$$

3. Experimental Set-Up and Measurements

We used two laser sources emitting respectively at 1550 nm (Δv<100kHz) and 1310 nm ($\Delta\lambda$<2.5nm).
An infra red aspherical lens ensured the magnification (M=182).
The camera was a vidicon. The scanned area was about of 9.5×12.7 mm^2. Each camera pixel had a horizontal length of 25 µm and a vertical length of 19.45 µm.
We tested two commercial EDF. The fibre 1 length was about 2 metre and the co-dopants are Ge and Al. Fibre 2 is a 40 metre long fibre co-doped with Al and P. Their cut-off wavelengths were calculated by the reference test method (G.652 recommendation). The fibres were measured without any particular preparation. During the measurement, the fibres under test remained on the bobbin.
We did not use any cladding mode stripper.
We had to attenuate the measuring signal to avoid any high power level on the camera surface because such a power level saturates the detector and can damage the camera.
After having detected the two dimensional NFI pattern, a computer samples one axis including the maximum intensity value. Then, we assumed that the fibre is axially symetrical. From that radial NFI pattern, we computed (2) and (3). All these calculations were performed numerically without using any fit function near the core region. Far from the core, we fittedpower function on the radial NFI to minimize any noise effect. We also smoothed $\Delta n^2(r)$ using FFT and filtering algorithms.

4. Results and Comments

Figure 2, 3 and 4 show the radial NFI pattern of fibre 2 while Figures 5, 6 and 7 present the $\Delta n^2(r)$ of the two fibres we measured. Table 2 contains the calculated optogeometrical characteristics and ESI parameters.

	Fibre 1 1310 nm	Fibre 1 1550 nm	Fibre 2 1310nm
w_{p2}	3.3	3.5	3.0
a (μm)	1.7	1.7	1.8
NA_{max}	0.18	0.22	0.18
NA_{center}	0.12	0.19	0.16
	A=0.625	B=1.707	C=4.112
a_{eff1} (μm)	2.0	1.5	2.0
NA_{eff1}	0.17	0.22	0.18
	A=0.65	B=1.619	C=2.879
a_{eff2} (μm)	2.1	1.7	2.1
NA_{eff2}	0.16	0.20	0.18

Table 2: optogeometrical and ESI parameters.

We observe that the measurement wavelength does not seem to influence the global shape and the core radius value. On the other hand, it clearly appears that the optical properties are wavelength dependent.
If we compare these results with the ESI parameters, we see that NA_{max} and NA_{eff1} are identical while each $a_{eff1,2}$ presents differences with the measured core. Moreover the effective core value vary with the wavelength. One of the reasons could be the origins of definition (7) and its valadity degree in our study case. We are currently studying the causes of these differences and are trying to understand the effect of the assumptions associated to these definitions.

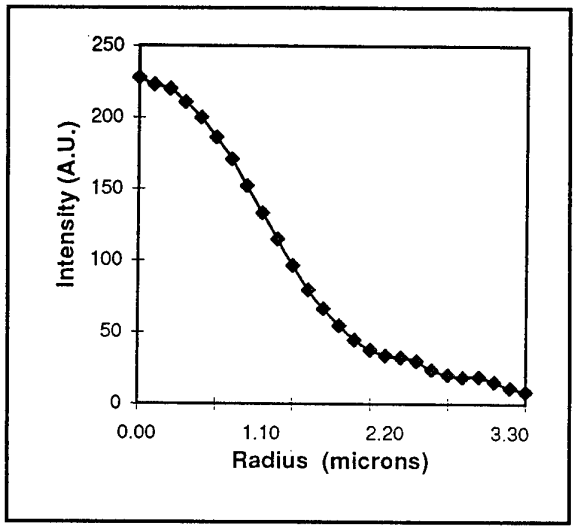


Figure 3: Fibre 1 NFI profile, measured at 1310 nm.

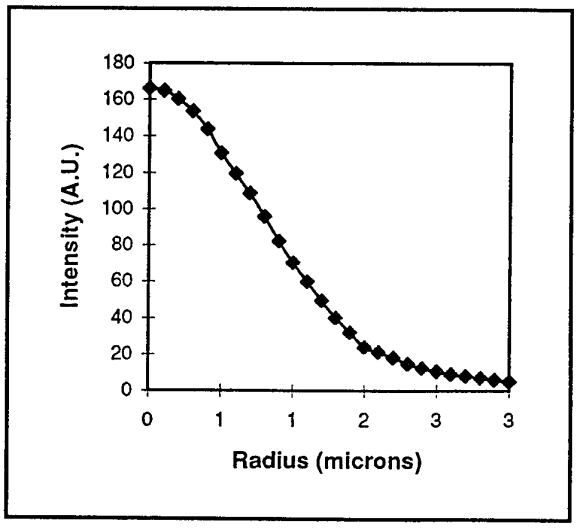


Figure 2: Fibre 2 NFI profile, measured at 1310 nm.

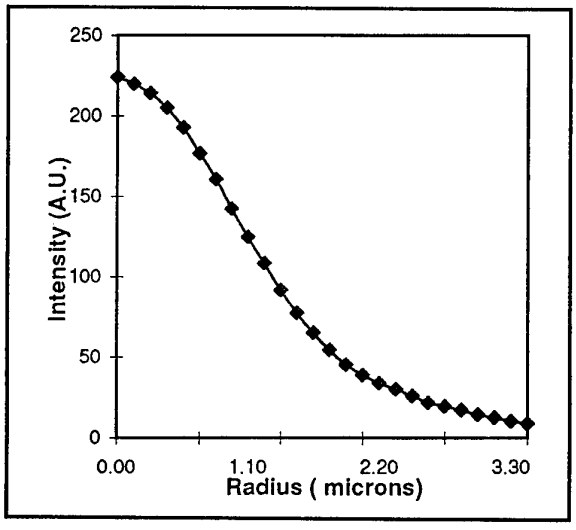


Figure 4: Fibre 1 NFI profile, measured at 1550 nm.

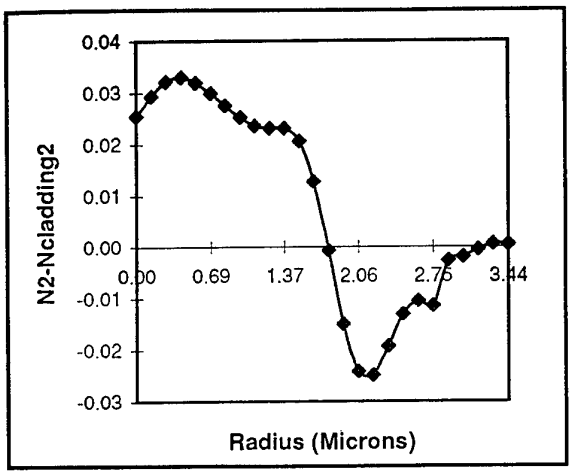


Figure 5: Fibre 2 refractive index profile, measured at 1310 nm.

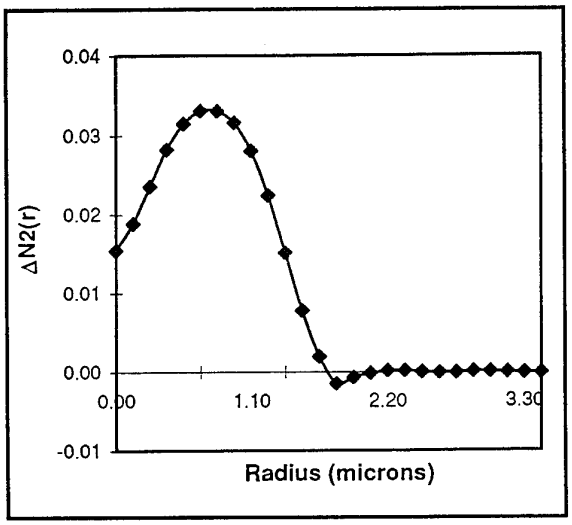


Figure 6: Fibre 1 refractive index profile, measured at 1310 nm.

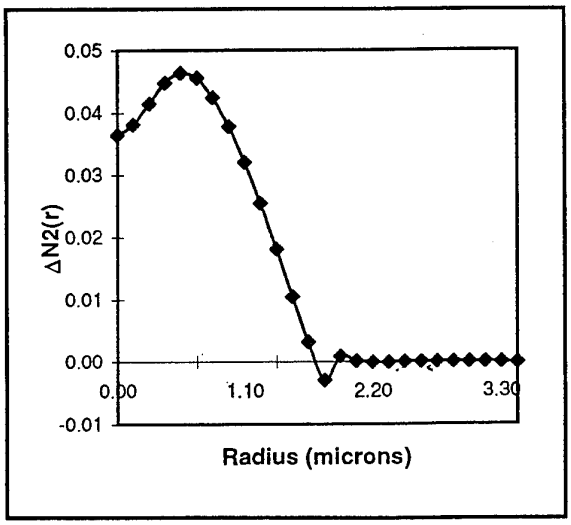


Figure 7: Fibre 1 refractive index profile, measured at 1550 nm.

5. Conclusions

In this work, we developped an experimental set-up based on the Transmitted Near Field Technique. The scanning device was a CCD camera that detected the Near Field Intensity magnified with an infra red aspherical lens. From the NFI, we derived the refractive index profile difference, the MFD, the core radius, the NA and the ESI parameters of two commercial EDF. All these measurements were carried out at 1550 nm and 1310 nm. The difference we obtained between the three core definitions does not exceed 0.4 μm. For the three *NA* definitions, the difference is less than 0.05.
Our set-up gives a good idea of the fibre optogeometrical parameters with not need of a particular preparation of the tested fibres. The measurement can be done directly on the bobbin.

6. Acknowledgement

We would like to thank the COST 241 Management Committee for having kindly loaned the second EDF.
The cut-off wavelength has been measured with an experimental set-up developed within the context of COST 241 S.G.1.2. (G.652 test method).
This work has been funded by FNRS, REGION WALLONNE and BELGACOM in the frame of RD/TR/UN/1 contract.

7. References

[1] S. Geckeler, *Optical Fiber Transmission Systems*, Artech House, 1987, pp. 116-145.
[2] K.Morishita, "Refractive-Index-Profile Determination of Single -Mode Optical Fibers by a Propagation-Mode Near-Field Scanning Technique", *JLT*, vol LT-1, n°.3, pp.445-449, Sept. 1983.
[3] K.Morishita, "Refractive-Index Profile Measurement of Single -Mode Optical Fibers by the Propagation-Mode Near-Field Method", *JLT*, vol LT-3, n°. 2, pp.244-247, Apr. 1985.
[4] K.Morishita, "Index Profiling of three dimensional Waveguides by the Propagation-Mode Near-Field Scanning Technique", *JLT*, vol LT-4, n°.8, pp.1120-1124, Aug. 1986.
[5] G. Coppa, P. Di Vita, and U. Rossi, "Characterisation of Single Mode Fibers by Near-Field Measurements", *Electron. Lett.*, vol. 19, n°. 8, pp. 293-294, Apr. 1983.
[6] K. Petermann,"Constrains for Fundamental Modal Spot Size for BroadBand Dispersion Compensated Single-Mode Fibres", *Electron. Lett.*, vol. 19, pp.712-714, 1983.
[7] F. Villuendas, F. Calvo, and J. B. Marqués, "Measurement of Mode Field Radius in Axially Nonsymetrical Single-Mode Fibers with Arbitrary Power Distribution", *Optics Letters*, vol. 12, n° 11, pp. 941-943, Nov. 1987.
[8] F. Martinez, C. D. Hussey, "(E)ESI Determination from Mode Field Diameter and Rfractive Index Profile Measurements on Singl-Mode Fibres", *IEE Proc.*, vol. 135, Pt. J. , n°. 3, June 1988.
[9] J.C. Augé, L.B. Jeunhomme, A. Kelekis, "The Prediction of Single-Mode Fibre Transmission Characteristics from Mode Field Diameter", IOOC-ECOC'85, pp. 333-335.

DISCUSSION

Discussor's name: A. Altintas

Comment/Question:

Did you test erbium-doped fibres with different doping levels, and in this case, did you observe different optogeometrical properties?

Author/Presenter's reply:

We tested two different commercial fibres. No fundamental behaviour was noticed connected to the doping level. Furthermore, the general conclusion of the paper (the necessity to optimize the A, B, C parameters for the effective parameters calculation (Mode Field Diameter, Numerical Aperture) was still valid.

MEASUREMENTS OF DOPPLER SPREAD ON HIGH LATITUDE HF PATHS

M J Angling, P S Cannon, N C Davies[1]
B Lundborg[2],
V Jodalen[3],
K W Moreland[4]

[1] Radio Science and Propagation Group
Defence Research Agency
Malvern, Worcs, WR14 3PS
United Kingdom

[2]Swedish Institute of Space Physics, Sweden.
[3]Norwegian Defence Research Establishment, Norway
[4]Communications Research Centre, Canada.

1. SUMMARY

The increased sophistication of the modulation schemes employed by high frequency (HF) data modems has necessitated a better understanding of the effects of the ionosphere on propagated signals. This paper describes an HF pulse compression radio sounder, that is being used to characterise high latitude HF paths. Results from operations of the sounder in Scandanavia are also presented.

2. INTRODUCTION

The performance of beyond line-of-sight (BLOS) HF communication systems is dependent on the match of system parameters to the propagation environment. Antenna gain and transmitter power are very important in this respect, though preliminary results indicate that the match between the signalling waveform and the time and frequency spread characteristics of the propagation path deserves equal or greater attention.

BLOS propagation of HF signals occurs via the ionosphere which introduces time dispersion of single propagation mode signals. It can also break the signal into multi-mode and multi-hop components all of which can arrive with different time delays. Furthermore, time dependant changes of the ionosphere can introduce both Doppler shifts and spreads to the transmitted signal. Doppler spread, which forms the basis of this paper may be due to both bulk motion of the ionosphere [e.g. *Cannon et al.*, 1991, 1992][1,2] and/or to the turbulent internal motion of plasma irregularities which are common in the auroral and polar cap ionospheres [e.g. *Tsunoda*, 1988; *Fejer and Kelley*, 1980][3,4]. The difficulties imposed by Doppler shifts and spreads can cause the break down of communications systems. Those which are designed to operate in multipath propagation conditions are especially vunerable. In conventional systems, multipath is overcome by operating at low data rates using simple modulation schemes (e.g. 75 baud frequency shift key (FSK) modulation). However, the low data rates and the consequentially low bandwidths, make the system more susceptible to frequency errors. These systems are no longer acceptable since they provide a less than optimum data rate when the environmental conditions are benign and yet the transmissions may still break down when the environmental conditions are stressed.

Whilst all of these effects occur at mid-latitudes they are more prevalent and dynamic in the high latitude and equatorial regions. If HF modems are to be designed to exhibit a robust performance under such punitive conditions, it is necessary to experimentally measure and quantify the HF channels in question. Many such measurements have been carried out, but these have generally used continuous wave (CW) and low data rate FSK waveforms [e.g. *Shepherd and Lomax*, 1967; *Warrington et al.*, 1994][5,6]. These experiments have yielded much valuable information, yet they are somewhat inappropriate to modern sophisticated (high data rate) modems. There have been, however, two studies of note which sought to characterise HF channels in more detail through the measurement of the channel scattering function [*Basler et al.*, 1988 and *Wagner et al.*, 1988][7,8] The scattering function describes the way that the signal energy is redistributed in both delay and frequency as a result of the transmission channel and as such it meets the requirements of the modem designer. The scattering function can be approximately determined by a suitable channel sounder which repeatedly, at an interval t_i, measures the complex valued impulse response as a function of path delay, $f(t_d)$. The samples at each delay value t_d are processed with discrete power spectrum estimation techniques to derive the Doppler spectrum.

This paper briefly describes a channel sounder known as DAMSON (Doppler And Multipath SOunding Network) which has been developed to characterise the disturbed narrow band channel (3 kHz) by measuring its scattering function. The real time nature of the DAMSON processing makes it unique in the frequency range, but in its approach to measuring the channel its operation is similar to that of *Wagner et al.*,[1988][8]. The DAMSON system is described in more detail in *Davies and Cannon,* [1993][9]

3. DESCRIPTION OF DAMSON

DAMSON operates from remote sites on pre-selected frequencies. The system is based on commercially available equipment (such as HF communication receivers and transmitters, PC computers etc.) and makes extensive use of digital signal processing (DSP) techniques. The Global Positioning Satellite (GPS) system provides the system timing ensuring that transmit and receive stations are accurately synchronised and also allowing absolute time of flight (TOF) measurements to be made.

DAMSON characterises the propagation path using a number of sounding waveforms which can be flexibly scheduled - of

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

these the delay-Doppler (DD) waveform has been specifically designed to accurately measure the channel scattering function as described earlier. For the measurements reported in this paper a Barker-13 code is used, modulated at 2400 baud onto a bi-phase PSK carrier. Each code lasts 5 ms with a pulse repetition interval of 12.5 ms. The codes are sent 128 times giving an integration time of 1.6 s and a frequency resolution of 0.62 Hz. The Doppler frequency range of the measurements is ± 40 Hz of the central frequency. The processing gain for the DD measurements can reach 32 dB for non-spread signals.

The other operating modes include a noise measurement period lasting 2.8 s and 4.3 s of CW transmissions. Although the CW measurement mode is a very sensitive detector of signals, due to the long integration time and low receive bandwidth, it is unable to unambiguously distinguish between different propagation modes either in the time or frequency domain. A time of flight (TOF) mode also exists which is similar to the DD mode but which is able to unambiguously measure the signal TOF and large multipath delays due to its low pulse repetition frequency (prf). Like the DD mode, the TOF mode also uses pulse compression waveforms but with a shorter integration time. Due to its low prf the TOF mode is unable to provide an accurate measure of the Doppler shift and spread.

4. DAMSON DELAY-DOPPLER DATA ANALYSIS

DAMSON transmissions occur on a number of frequencies which may, or may not, be matched to the expected or prevailing ionospheric conditions. On many occasions transmissions may be below the lowest usable frequency (LUF) or above the maximum usable frequency (MUF), resulting in no ionospheric support for signal propagation. On other occasions the reception of DAMSON signals may be severely hampered by co-channel and adjacent channel interference. Nevertheless DAMSON data files are always stored at the scheduled frequencies and it is necessary to subsequently extract those files where signals can be analysed and discard the rest. This is achieved by the DAMSON analysis software [*Angling*, 1995][10]. On replay the data is re-displayed and where possible a box is drawn around the data enclosing those elements of the spectrogram which lie above a noise dependant (via a sliding scale) threshold. Further analysis can then take place on the boxed data. If the box cannot be drawn the data is either categorised as too difficult to analyse at present (but there is hope in the future) or it is discarded. The Doppler spreads quoted in this paper are those corresponding to the central region within the box which contains 95% of the total boxed power.

5. PROPAGATION PATHS

This paper considers the Doppler spreads observed from the 19th to the 25th of April 1995 on a single path from Isfjord Radio on Svalbard (78.063 N, 13.627 E) to Tuentangen, near Oslo, Norway (59.937 N, 11.086 E), during which time the DAMSON system was operated almost continually.

The antenna used at Isfjord is a rhombic pointing due south fed by 250 watts. With a perfect ground the antenna is predicted to illuminate preferentially the higher elevation angles; at 5 and 10 MHz the peak gain lies at angles of 35° and 18° respectively. The antenna is consequently better matched to *F* region paths than it is to *E* region paths. At the receiver an 80 m sloping Vee antenna is used pointing due north.

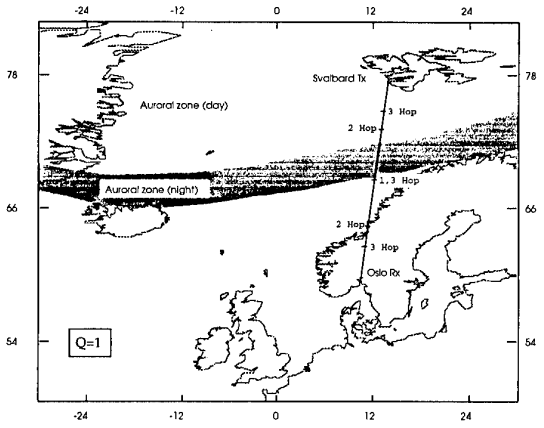


Figure 1: Map showing DAMSON deployment and the position of the auroral zone for $Q = 1$.

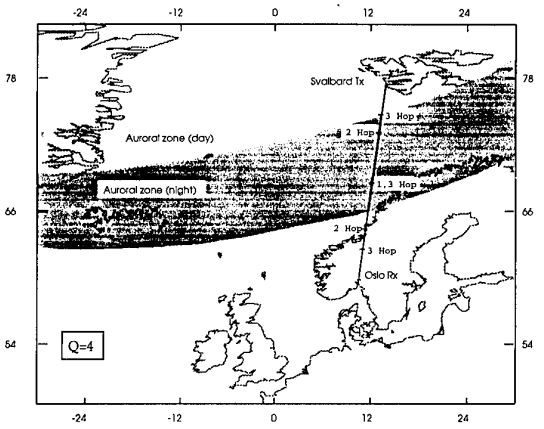


Figure 2: Map showing DAMSON deployment and the position of the auroral zone for $Q = 4$.

The geometry of the 2017 km path is shown in Figures 1 and 2, together with the night (~02 LT) and daytime (~12 LT) extreme positions of the auroral oval for $Q = 1$ and $Q = 4$ according to the model due to *Whalen*, [1970][11]. These values are representative of the mean Q values observed during the first and second halves of the data collection period. Also shown in the diagram are the geographic apogee positions of the one, two and three hop signals; the midpoint is situated at 69.00 N, 11.83 E.

6. RESULTS

Figure 3 shows a typical DAMSON output for a measurement at 0000 UT on the 20th April 1995 at a frequency of 9.04 MHz; at this time there was significant spectral spread. The top panel shows the low resolution TOF multipath profile with a range of 40 ms. The second panel shows a higher resolution version of this using the DD waveform.

The third panel shows a grey shaded representation of the multipath scattergram. The single mode spectral spread is very evident, as is a co-channel interferer running horizontally across the scattergram at -30 Hz. The aforementioned box has been drawn around the scattering function. The estimated two sided Doppler spread of 50 Hz and multipath spread are

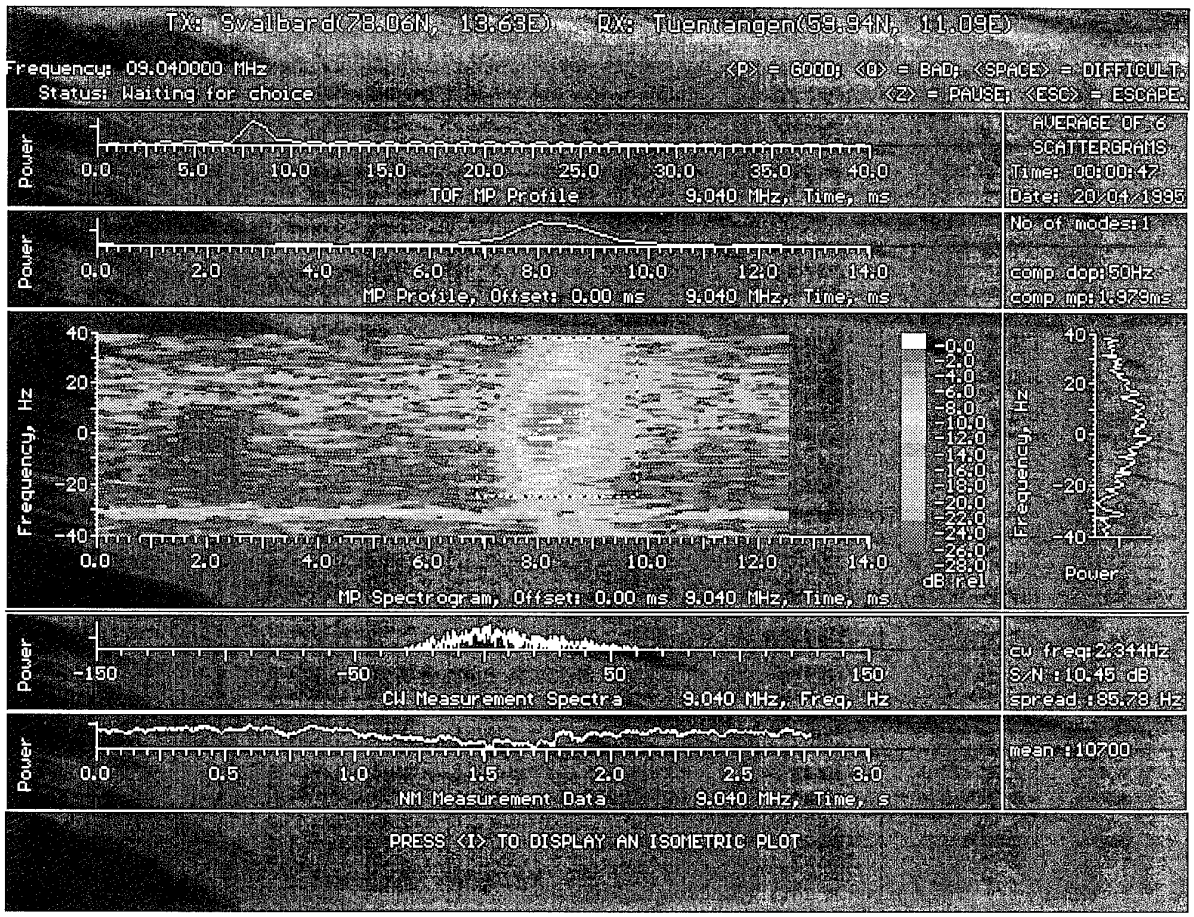


Figure 3: Example display from DAMSON analysis program.

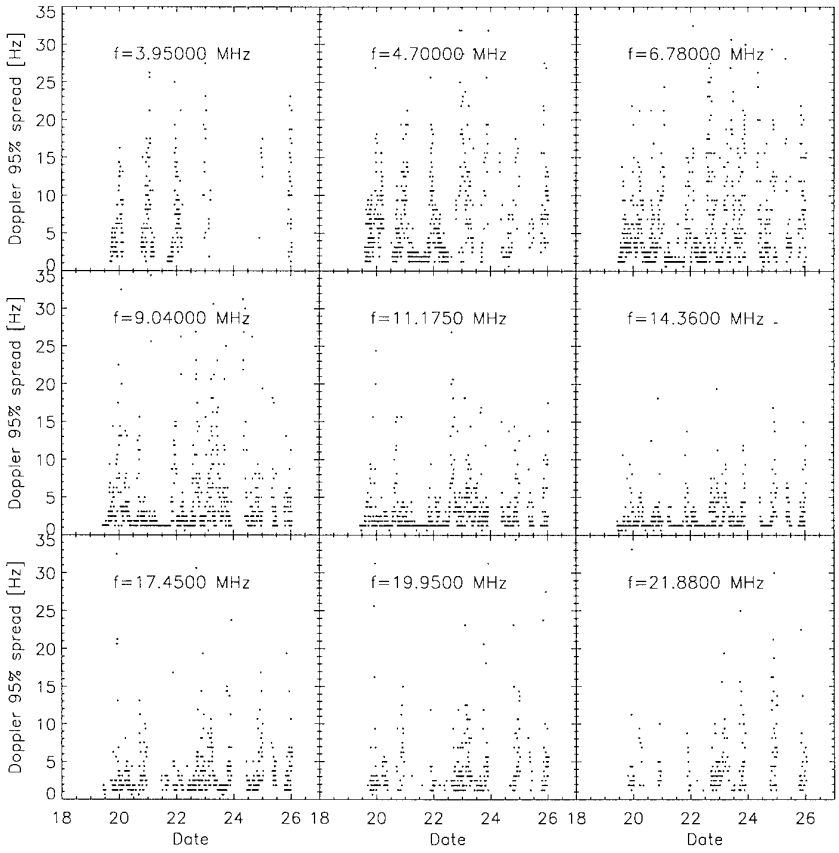


Figure 4: Doppler spread data plotted agaist time.

shown in a small panel on the right. The signal power at high spectral spreads is seen to be ~12 to 16 dB below the peak of the mode. *Wagner et al.*, [1995][12] suggest that such signals are due to a scattering rather than a reflection process.

The small Doppler shift associated with the majority of the power in the scattergram is in line with that expected for an oblique path as opposed to backscatter or back-reflection. In the latter cases the radar measures twice the radial ionospheric velocity component [e.g. *Cannon et al.*, 1991, 1992][1,2]. However, for forward *great circle path* reflection or scattering in a homogeneous ionosphere, with no vertical motion, the positive Doppler shift for the up-going wave is cancelled by an equal negative Doppler shift for the down going wave. Any residual Doppler shift is a result of asymmetry on the propagation path, off great circle propagation, or vertical motions.

In this scattergram the higher frequency Doppler components arrive up to a millisecond behind the zero Doppler shifted component. This may due to reflection or scattering from locations off the great circle path combined with the effects of an ionospheric flow component which is perpendicular to the path.

The small panel directly to the right of the scattergram displays the time (incoherently) integrated Doppler frequency spectra. Some asymmetry may be seen about the zero shifted component, though other more pronounced examples have been observed. There are a number of possible explanations for such asymmetries, which include preferential scattering from discrete areas each with a different line-of-sight (up and down leg) velocity. An inhomogeneous, time dependent ionosphere introduces other possibilities. Others are based on an asymmetric path with an ionospheric velocity component parallel to the direction of propagation.

In order to better understand the impact of Doppler spread all of the data (ie. all the modes for each analysable scattergram) have been amalgamated from the period in question and the 95% power percentiles are displayed as a time sequence in Figure 4. Diurnal variations in Doppler spread can be seen, with the maximum spread values occuring during the night - the period when the ionosphere is most disturbed.

The data can also be displayed in histogram form (Figure 5). This example illustrates the Doppler spread for 841 points, this being the number of analysable points, at a frequency of 9.04 MHz. Spreads of up to ~20 Hz are shown, though in general, the spreads are much less than this. They are, however, still significant for all but the most robust of modem designs.

In order to display the Doppler spread data in a more concise fashion the data are plotted as a function of the different sounding frequencies as shown in Figure 6. The symbols connected by straight lines repesent the median values of the spread distributions, while the horizontal bars represent the distribution's quartile values. The small dots are data points, which are regularly spaced due to their quantisation by the resolution of the instrument. It is now evident that the median spectral spreads are higher towards the extremes of the frequency range than at frequencies towards the centre of the range. It should be noted that the Doppler equation predicts that in the abscence of other effects, Doppler shift, and hence spread, should be directly proportional to frequency (see for example *Cannon et al.*, [1991][1]). While this seems to be true at the top of the frequency range (15 - 20 MHz), the Doppler spreading at the lower frequencies must be dominated by a different effect.

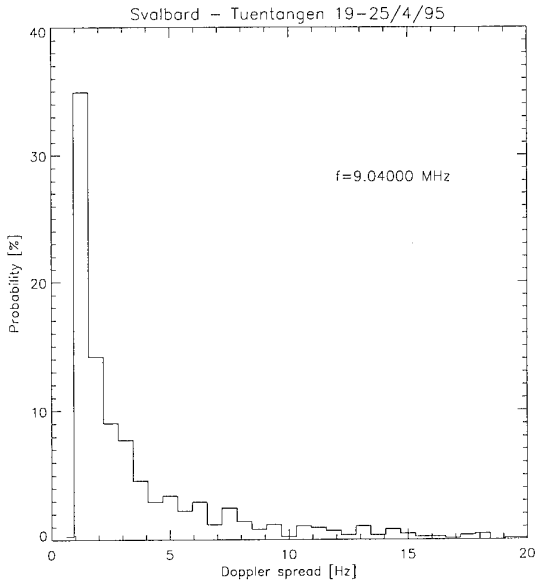


Figure 5: Example of histogram produced by DAMSON analysis program

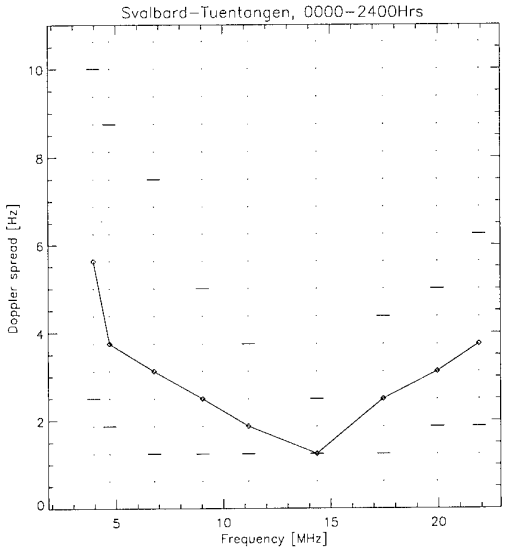


Figure 6: Doppler spread variation with frequency for 24 hour period

It was noted that during the period of the experiment two levels of geomagnetic activity were observed. From the 19th of April until midday of the 22nd, the 15 minute activity index Q exhibited a mean value of ~1, while between midday of the 22nd and the end of the experiment (2400 UT on the 26th of April) Q increased to a mean value of ~4 (Figure 7). One effect of this change in Q can be seen in Figure 4 - during the low Q period, many data points can be seen in the 3.95 MHz panel, but few can be seen in the 21.88 MHz panel. The

opposite behaviour is true during the high Q period. This is due to changes in the LUF and MUF, which is induced by the change in Q - when Q is high the additional precipitation causes absorption of low frequency signals, but the MUF is also increased by the associated extra ionisation. This is in broad agreement with the LUF/MUF changes predicted by the communications prediction program ICEPAC for the Svalbard - Tuentangen path during this period.

In order to investigate the impact of geomagnetic activity on Doppler spread, the low and high Q data periods were separately replotted against the sounding frequency. Furthermore, the day was sub-divided into 6 hour periods (Figures 8 to 11). These graphs reveal that during high Q the data exhibits consistently larger Doppler spreads than the low Q data. This is as expected, due to the increased ionospheric flows and turbulence when Q is high. Significant Doppler spreading can be seen in the pre and post-midnight sectors in the lower part of the frequency range - particularly for the high Q data, where median Doppler spreads of >10 Hz are seen. Trends similar to those seen in Figure 6 are also apparent.

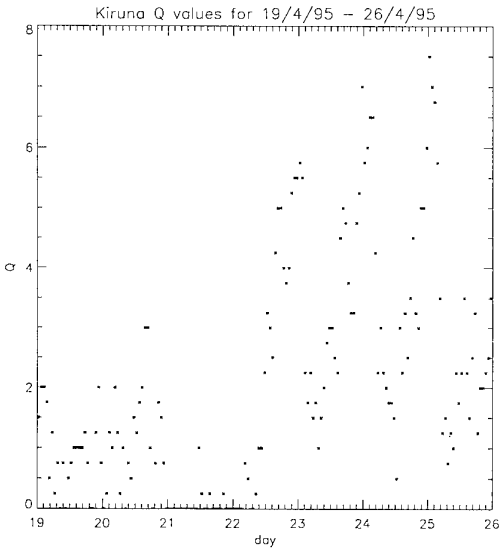


Figure 7: Q values for analysis period.

The trend of Doppler spread decreasing from 5 to 15 MHz is believed to be associated with a change from multi-hop to single hop propagation. During the high Q period (after midday of the 22nd April) multi-hop paths are thought to have control points located further into the active auroral ionosphere (Figure 2) than single hop paths. Therefore, the multi-hop signals exhibit higher Doppler spreads because of the more turbulent conditions encountered at the higher latitudes. This effect is most prevalent at night, when the auroral region has moved to its southernmost positon. During low Q conditions (before midday of the 22nd April), the control points for both single and multi-hop paths are southwards of the auroral zone in the daytime (Figure1), producing less Doppler spreading. At night, however, when multi-hop paths have their control points in, or to the north of, the auroral zone, Doppler spreading is apparent on the lower frequencies.

Above 15 MHz, Doppler spreading is seen to increase. This is in accordance with the Doppler equation, suggesting that the previously described mechanism is no longer in operation. It is thought that as the transmission frequency increases from 15 to 22 MHz, single hop propagation is predominant. Although the higher frequency rays may penetrate deeper into the ionosphere than the lower frequencies, the control points of the path remain in approximately the same geographic position. Thus, without the northward shift of the control points, the dependance of Doppler spread on frequency is revealed.

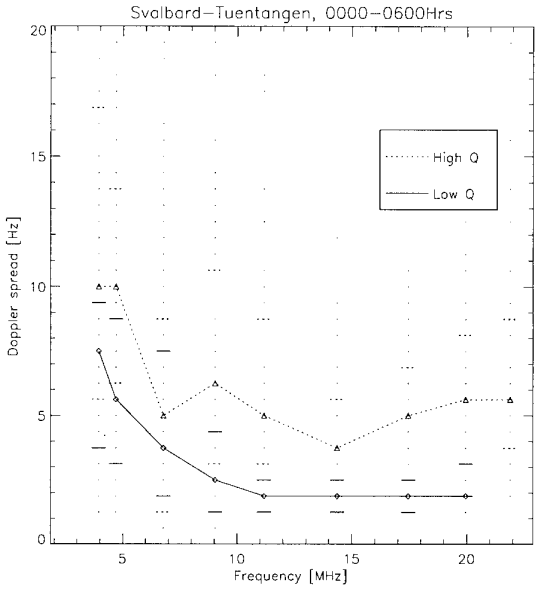


Figure 8.
Doppler spread variation with frequency; 00 - 06 Hrs

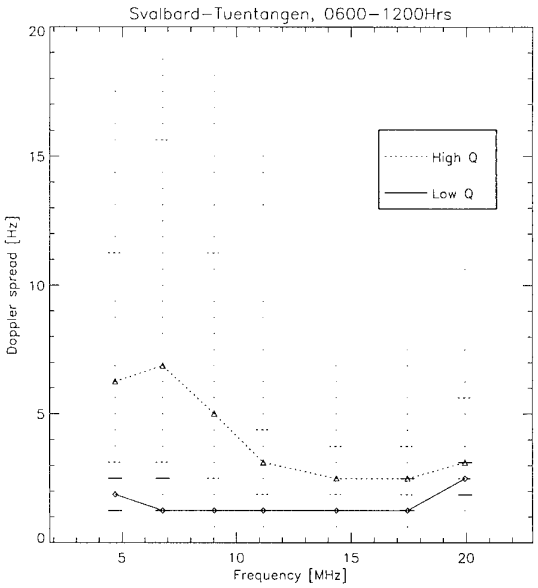


Figure 9.
Doppler spread variation with frequency; 06 - 12 Hrs

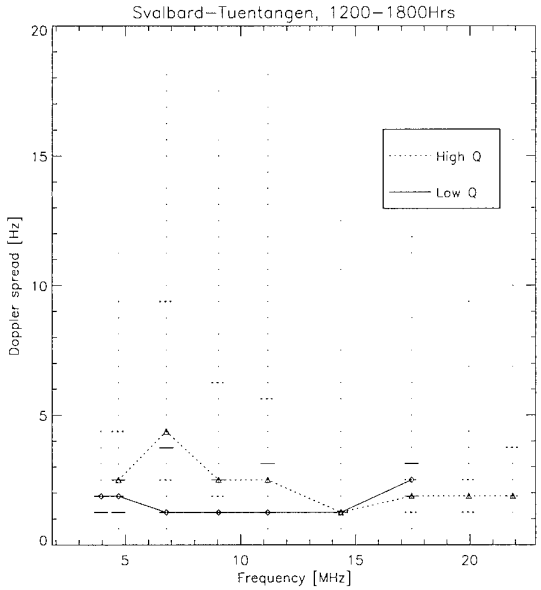


Figure 10.
Doppler spread variation with frequency; 12 - 18 Hrs

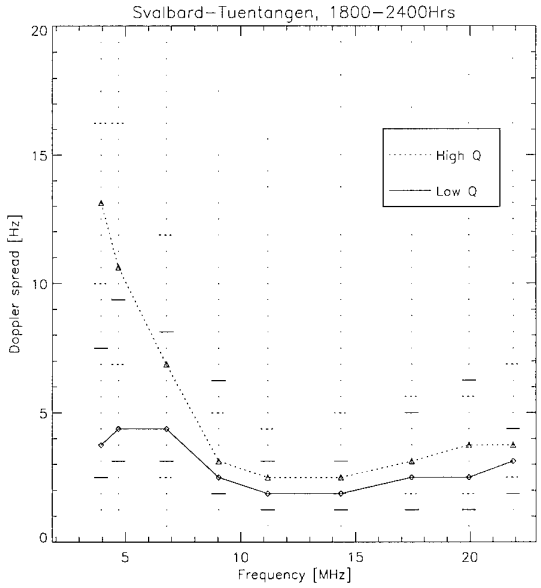


Figure 11:
Doppler spread variation with frequency; 18 - 24 Hrs.

The trend of Doppler spread decreasing with increasing frequency (up to ~15 MHz) is consistent with data collected in August 1994 [*Cannon et al.*, 1995][13]. Also, the proposed decrease in the number of hops as frequency increases is consistent with the propagation modes predicted by ICEPAC, though further analysis will be needed to confirm this as the primary spreading mechanism.

7. CONCLUSIONS

These results from DAMSON demonstrate its utility for making measurements pertinent to HF systems operating in the demanding enviroment of the high latitude region. A well defined trend of Doppler spread decreasing with increasing frequency (up to ~15 MHz) has been found that is believed to be driven by a change from multi to single hop propagation. A clear dependance of the median spreads on the geomagnetic index Q has also been found. These two results provide hope that it will be possible in the future to model and predict the Doppler characteristics of HF channels.

The analysed data have displayed median values of Doppler spread in excess of 10 Hz during high Q conditions in the pre-midnight sector, though individual measurements can show much greater spreads - results which are highly significant for the design of HF modems.

8. REFERENCES

1. Cannon, P. S., B. W. Reinisch, J. Buchau and T. W. Bullett, 1991, 'Response of the polar cap F-region convection direction to changes in the interplanetary magnetic field: Digisonde measurements in Northern Greenland', J. Geophys. Res., 96, (A2), 1239-1250.

2. Cannon, P. S., G. Crowley, B. W. Reinisch and J. Buchau, 1992, 'Digisonde measurements of polar cap convection for northward interplanetary magnetic field', J. Geophys. Res., 97, (A11), 16877-16885.

3. Tsunoda, R. T., 1988, 'High-latitude F region irregularities: A review and synthesis.', Rev. of Geophys., 26, (4), 719-760.

4. Fejer, B. G. and M. C. Kelley, 1980, 'Ionospheric irregularities', Rev. Geophy. and Space Phys., 18, (2), 401-454.

5. Shepherd, R. A. and J. B. Lomax, 1967, 'Frequency spread in ionospheric radio propagation', IEEE Trans. Comm. Tech., COM-15, (2), 268-275.

6. Warrington, M., B. S. Dhanda and T. B. Jones, 1994, 'Observations of Doppler spreading and FSK signalling errors on HF signals propagating over a high latitude path', Proceedings of IEE Conf on HF Radio Systems and Techniques, Conf Pub. 392, York, UK.

7. Basler, R. P., P. B. Bentley, R. T. Price, R. T. Tsunoda and T. L. Wong, 1988, 'Ionospheric distortion of HF signals', Rad. Sci., 23, (4), 569-579.

8. Wagner, L. S., J. A. Goldstein and W. D. Meyers, 1988, 'Wideband probing of the transauroral channel: Solar minimum', Radio Sci., 23, (4), 555-568.

9. Davies, N. C. and P. S. Cannon, 1993, 'DAMSON - A system to measure multipath dispersion, Doppler spread and Doppler shift on multi-mechanism communications channels', Proceedings of AGARD Symp on Multiple mechanism propagation paths (MMPPs): Their characterisation and influence on system design, CP-543, Rotterdam.

10. Angling, M J, 1995, 'Functional description and user guide for the DAMSON analysis software; version 5.0.', Report, DRA/CIS/CIS1/WA95026/1, Defence Research Agency, UK.

11. Whalen, J. A., 1970, 'Auroral oval plotter and nomograph for determining corrected geomagnetic local time,

latitude and longitude for high latitudes in the northern hemisphere.', Report, AFCL-70-0422, Air Force Cambridge Research Laboratories, USA.

12. Wagner, L. S., J. A. Goldstein, M. A. Rupar and E. J. Kennedy, 1995, 'Delay, Doppler and amplitude characteristics of HF signals received over a 1300 km transauroral skywave channel', Submitted to Rad. Sci.

13. Cannon, P. S., N. C. Davies, M. J. Angling, V. Jodalen, K. W. Moreland, B. Lundborg, 1995, 'Initial results from DAMSON - A system to measure multipath, Doppler spread and Doppler shift on disturbed HF channels', Proceedings of 9th International Conference on Antennas and Propagation, Conf. Pub, 407, Eindhoven, Netherlands.

© British Crown Copyright 1995 /DERA Published with the permission of the Controller of Her Britannic Majesty's Stationary Office

DISCUSSION

Discussor's name: G. S. Brown

Question/Comment:

On your multiple hop paths, have you observed any sensitivity to the roughness of the sea due to attenuation of the sea surface reflected component?

Author/Presenter's Reply:

This is not an effect that we have considered. However, we feel that, although the sea state may modify the characteristics of the received signal, the effects will be small and extremely difficult to separate from the actions of the ionosphere.

DISCUSSION

Discussor's name: C. Goutelard

Comment/Question:

1. Dans vos modèles de propagation, avez-vous pris en compte les propagations possibles hors du grand cercle? Nous avons dans une communication AGARD faite à Ankara vers 1986-87 signalé l'importance de ces effets qui s'accompagnaient de déviations doppler importantes.

2. Comptez-vous utiliser les résultats de mesure de SUPERDARN, expérience à laquelle votre pays participe?

Translation:

1. *In your propagation models did you take into account possible off great circle propagation? In an AGARD paper given in Ankara in 1986-87 we pointed out the importance of these effects which were accompanied by significant doppler deviations.*

2. *Do you intend using the results of the SUPERDARN project, in which your country participates?*

1. We have observed the effects which we believe to be associated with off great circle propagation. First, the existence of highly delayed modes observed on relatively short paths. Secondly, modes whose spread components are delayed in relation to their zero doppler shifted component. Future work will address these two observations.

2. We wish to use data from other instruments to increase our understanding of the mechanisms effecting the DAMSON transmissions. We would therefore be interested in SUPERDARN data.

Observations of Doppler spreading and FSK signalling errors on HF signals propagating over high latitude paths

B.S. Dhanda, E.M. Warrington and T.B. Jones
University of Leicester
Leicester, LE1 7RH
UK

SUMMARY

The auroral zone ionosphere forms a highly dynamic and disturbed medium which can severely degrade the performance of HF radio links when the signals propagate through these regions. In order to investigate this effect, a series of experiments have been undertaken over a variety of high latitude paths for which the Doppler spreading imposed on the signal by the ionospheric reflection processes and FSK data signalling error rates and distributions were measured.

The Doppler spreads observed over two of these high latitude propagation paths are examined in this paper for two experimental campaigns, one conducted during the summer months and the other during winter months. One of these paths was contained entirely within the polar cap and another of the paths was sometimes entirely within the polar cap and at other times crossed the auroral oval. Times of increased Doppler spreading on the second path were found to be well correlated with periods for which one of the ionospheric reflection points was within, or close to, the boundaries of the statistical position of the auroral oval.

The bit error rate of the FSK signals was not influenced by the high level of Doppler spreading observed on the path crossing the auroral oval. However, the distribution of bit errors was markedly affected by changes in the Doppler spreads. At times of high Doppler spreading, the length of the error bursts was significantly less than for the cases where the Doppler spreading was small.

1 INTRODUCTION

The high latitude ionosphere within the auroral zone forms a highly dynamic and disturbed medium which can severely degrade the performance of HF radio links when the signals propagate through these regions. Characteristics of the signal which may be affected include the signal strength, the fading rate and the frequency range over which signals may propagate (i.e. an increase in the LUF and / or a decrease in the MUF). Furthermore, large Doppler frequency shifts and spreads may be imposed on the signal due to movements in the ionosphere near to the reflection points. In order to examine these effects, signals from a computer controlled transmitter operating on a predetermined frequency and modulation schedule, located at Clyde River on Baffin Island in the Canadian North West Territories (70.5° N, 68.5° W) were received at several locations so as to give paths both within the polar cap and crossing the auroral oval. Measurements were made during two, one month duration experimental campaigns undertaken during July / August 1988 and during January / February 1989. Measurements are presented here for two representative paths, one to Alert (1340 km) which is contained entirely within the polar cap and the second to Prudhoe Bay (2960 km) which is sometimes within the polar cap and at other times trans-auroral oval.

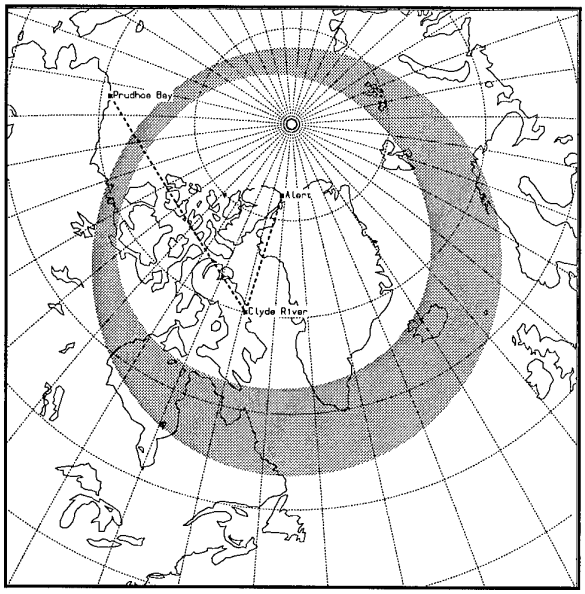


Figure 1. Map showing the locations of the transmitter and receiver sites. Also shown is the average position of the auroral oval at 0000 UT for moderate geomagnetic activity.

Large Doppler shifts and spreads associated with high latitude propagation are known to degrade the performance of data communications systems, the precise nature of the degradation being dependent upon the modulation format. In addition to the CW signals for the measurement of the Doppler characteristics, the Clyde River transmitter also radiated an 850 Hz shift FSK pseudo-random data signal. It was anticipated that at times of large Doppler spreading, significant changes in the bit error rate and error distributions of this signal

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

would occur. These signalling error parameters and their relationship with Doppler spreading for the path to Prudhoe Bay are also examined in this paper.

2 EXPERIMENTAL ARRANGEMENT

Illustrated in Figure 1 are the positions of the transmitter and receiver sites. The average position and extent of the auroral oval, which is well correlated with to the level of geomagnetic activity, is also illustrated in this figure for a time of 0000 UT and moderate (Kp=3) geomagnetic activity (see Feldstein and Starkov [1]).

The transmitter system was based around an amateur radio Icom 735 transceiver and 2KL power amplifier modified to include frequency stable reference oscillators to allow accurate measurements of the Doppler spread imposed on the signal by the reflection processes in the ionosphere. The receiver system was based around a Racal RA6790/GM remotely programmable HF receiver provided with quadrature outputs at the modulation baseband. These baseband signals were simultaneously sampled and recorded by a microcomputer system when measurements of signal strength and Doppler spread were being made. For the measurement of FSK signalling errors, the baseband signals were input to two tone detection filters, the outputs of which were sampled and recorded by the microcomputer system. The data were stored on magnetic tape for subsequent processing and analysis.

The transmission schedule comprised two parts:

(a) A series of two minute sequences transmitted once per hour on each of 14 frequencies in the range 3 - 23 MHz. Each sequence included a 30 second period of continuous carrier during which the Doppler spectrum of the received signal was measured followed by the callsign transmitted in Morse code for recognition purposes.

(b) A series of two minute sequences transmitted on six frequencies once per hour. Each sequence contained a pseudo-random binary sequence which was transmitted using 850 Hz shift FSK for 30 second periods at speeds of 50, 75 and 150 baud. The data sequence comprised continuously repeated 168 bit blocks each containing three 13-bit Barker sequences for synchronisation purposes followed by 129 bits of a pseudo-random sequence.

Details of the frequency schedules employed are given in Table 1. It should be noted that the format of the signals was not ideally suited for the studies reported here since they were developed as diagnostic aids to other types of measurement. Better statistics concerning the signal parameters would have been obtained with long duration transmissions rather than the short signals radiated for these tests. Despite this limitation, useful results have been obtained.

Table 1. Frequency schedules employed during the two experimental campaigns.

Minutes into the hour	Frequency (MHz) Summer	Frequency (MHz) Winter	Signal type
0	6.800	6.800	FSK
2	9.941	9.941	FSK
4	13.886	13.886	FSK
13	3.185	3.185	CW
15	4.900	4.900	CW
17	6.800	6.800	CW
19	9.941	9.941	CW
21	13.886	13.886	CW
23	18.204	18.204	CW
25	20.900	20.900	CW
30	6.905	6.905	FSK
32	10.195	10.195	FSK
34	14.373	14.373	FSK
38	3.230	4.455	CW
40	5.200	6.905	CW
42	6.905	10.195	CW
44	10.195	14.373	CW
46	14.373	17.515	CW
48	17.515	20.300	CW
50	20.300	23.169	CW

3 DOPPLER SPREAD

3.1 Signal recognition

Special care was taken to recognise the times when the measurements were contaminated by interfering signals. This was achieved in these experiments by applying the following tests to the received signals:

(a) The frequency distribution of the spectral components was tested. A dominant peak well away from the expected frequency would, for example, indicate reception of a co-channel signal.

(b) The presence of and correct position of the modulation sidebands occurring during the Morse code portion of the signal was tested.

Any measurements for which these tests indicated the presence of a signal other than our own were rejected from further analysis.

3.2 Measurement method

Several methods of quantifying the Doppler spread imposed on a signal by the ionospheric reflection process have been employed by various workers. These often involve fitting an idealised curve to the measured signal. The present observations indicate that the form of the spectrum varies considerably between measurements and consequently parameters of fitted curves would not be a reliable measure of the Doppler spread. In view of this, a simple technique was developed in which the area under the normalised signal spectrum, corrected for the base-line noise level was measured. This area was multiplied by 20 and the resulting figure (referred to here as the spread index) employed as a measure of the Doppler spread. For example, for a simple triangular shaped amplitude spectrum of base width 10 Hz, a spread index of 100 Hz is assigned. This simple method of quantifying Doppler spread is found to provide a good indication of the spectral broadening on a scale 0 to 150 Hz (approximately)

independently of the received signal strength and the precise shape of the spectral peak.

3.3 Observations

Considerable variation was observed in the Doppler spread of the received signals. At times, the spectra were very narrow, similar to those observed for mid-latitude paths; at other times large spreads, with a spread index often in excess of 100 Hz, were observed. As expected, at times of high Doppler spread the amplitude fading was fast and irregular, whereas at times of low Doppler spread the amplitude fading was slow and much more regular. Details of the measurements made at each receiving site are considered in the following sections.

In order to obtain better statistics, measurements for closely spaced frequencies have been combined. Only times for which the received signal passed both recognition tests have been included. Furthermore, in the graphs of the diurnal variation of the average spread indices and the corresponding standard deviation (Figures 2, 3, 5 and 6), points have only been plotted for times at which there are 5 or more measurements passing both recognition tests.

3.4 Alert

The variation in the Doppler spread index, averaged over the entire campaign, of the received signals for all frequency pairs as a function of time of day are presented in Figures 2 and 3 for the summer and winter campaigns respectively. The vertical bars indicate the standard deviation of the measurements for each hour of the day.

Summer

At the lowest frequencies radiated by the Clyde River transmitter (3.185 and 3.230 MHz), the Doppler spreads of the received signals are relatively low, with spread indices of around 20 Hz for the period 0100 until 0700 UT (i.e. centred around the path mid-point midnight). For the remainder of the day, the spread indices are very low, around 5 Hz. The diurnal variation in Doppler spread is more marked at the next higher frequencies (4.900 and 5.200 MHz) with a comparatively high Doppler spread centred on path mid-point midnight and low values centred around mid-point noon. In this case the duration of the low spread index period is shorter than at the lowest frequencies.

A diurnal variation of the form observed at the two lower frequency pairs is not evident at 6.800 / 6.905 MHz and 9.941 / 10.195 MHz. On these frequencies the spread index is around 25 Hz for most of the day with some evedence of a peak at around 1200 UT. At 13.886 / 14.373 MHz and above, the signal frequency is above the MUF around local midnight, and propagation occurs only during the day. On these higher frequencies, there is evidence of a peak in the Doppler spread occurring shortly after the signal begins to propagate (i.e. at around 1100 UT).

The frequency of occurrence of Doppler spreads with spread indices in the ranges 0-30, 30-100 and >100 Hz observed on this path, expressed as a percentage of all signals passing both recognition tests, are given in Table 2(a) for each frequency pair. From these tabulated data, and the graphs of Figure 2, it is clear that the Doppler spread for this path in summer is less than an spread index of 30 Hz for the majority of occasions and very rarely does the spread index exceed 100 Hz.

Winter

During the winter campaign (see Figure 3), the signal only propagated occasionally at and above the 13.886 / 14.373 MHz frequency pair. At the lower frequencies, the spread index is much larger than the values observed during the summer period. A diurnal variation, with two minima, one at around 0300 UT and the other at around 1700 UT, is apparent.

The frequency of occurrence of various Doppler spreadvalues over the range of propagation frequencies are given in Table 2. In contrast to the summer period, the Doppler spread is much higher in winter, the spread indices being in the range 30-100 Hz for most of the time and exceeding 100 Hz for a significant portion of the time.

Variation with geomagnetic activity

During the summer period the level of geomagnetic activity was low: Ap had a mean value of 9 and a peak of 12 during the 25 day measurement period. There is a tendency for the more active (Ap>10) days to be associated with slightly higher Doppler spreads, although the opposite effect is evident at 9.941 / 10.195 MHz.

The winter period was geomagnetically more active, with a mean Ap of 21 and a peak of 45. As during the summer period, there was a tendency for the more active (Ap>20) days to have slightly higher Doppler spreads.

Table 2. Occurrence of Doppler spreads observed on the Clyde River to Alert path expressed as a percentage of all signals recognised.

(a) Summer

Frequency (MHz)	SI<30 Hz	30-100 Hz	SI>100 Hz	Recogn rate
3.185/3.230	92%	7%	1%	29%
4.900/5.200	74%	25%	1%	58%
6.800/6.905	72%	27%	1%	67%
-/10.195	80%	18%	2%	67%
13.886/14.373	97%	3%	0%	30%
18.204/17.515	83%	17%	0%	15%
20.900/20.300	85%	14%	1%	12%

(b) Winter

Frequency (MHz)	SI<30 Hz	30-100 Hz	SI>100 Hz	Recogn rate
3.185/-	13%	81%	6%	28%
4.455/4.455	31%	64%	5%	49%
6.800/6.905	22%	68%	10%	40%
9.941/10.195	24%	61%	15%	28%
13.886/14.373	0%	65%	35%	1%
18.204/17.515	6%	83%	11%	4%
20.900/20.300	-	-	-	<0.1%

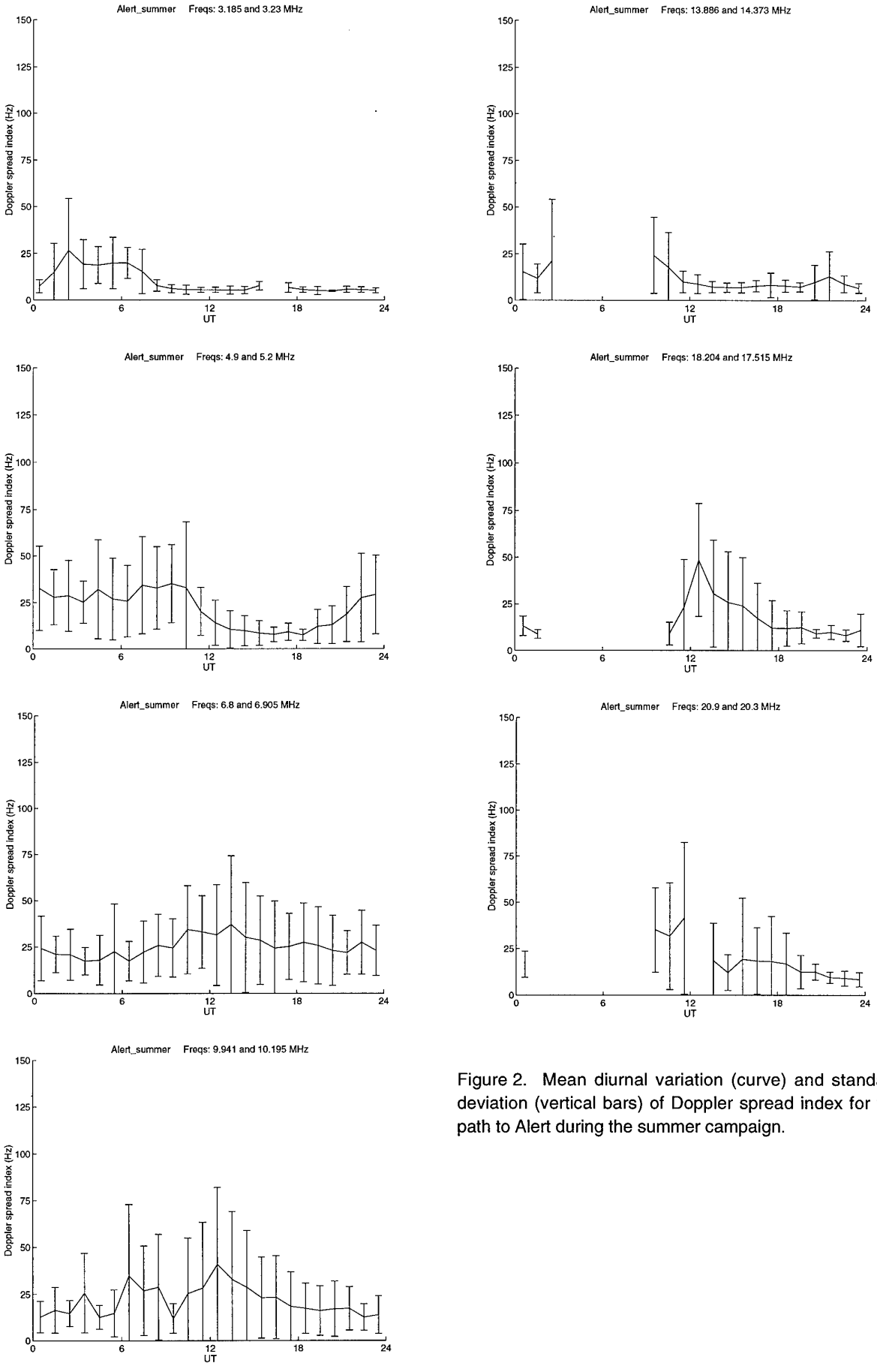
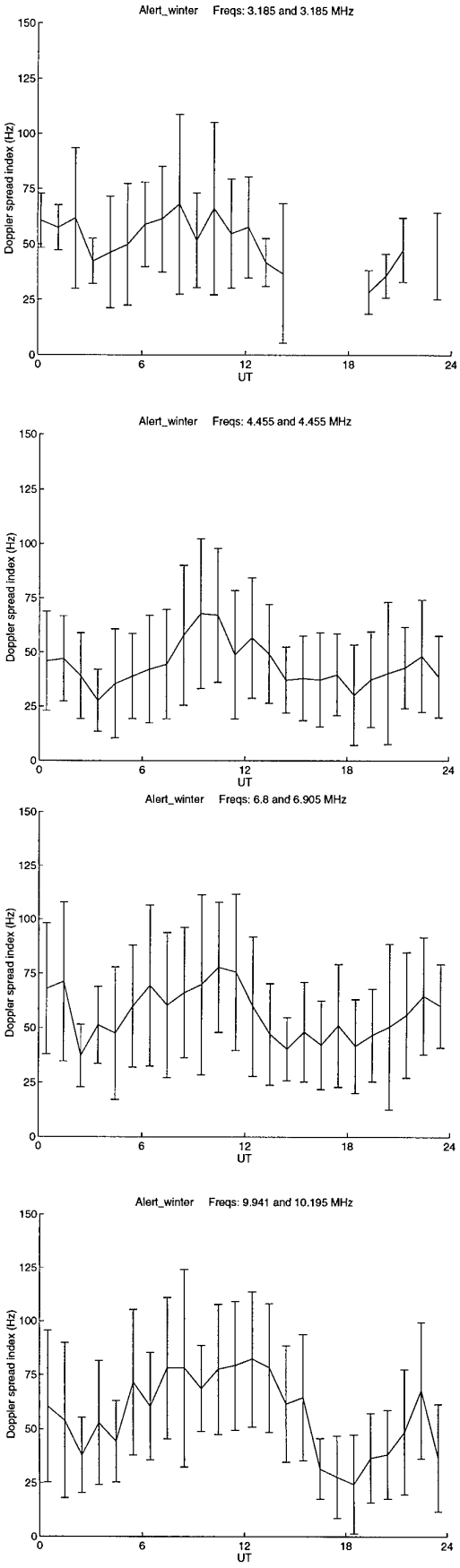


Figure 2. Mean diurnal variation (curve) and standard deviation (vertical bars) of Doppler spread index for the path to Alert during the summer campaign.



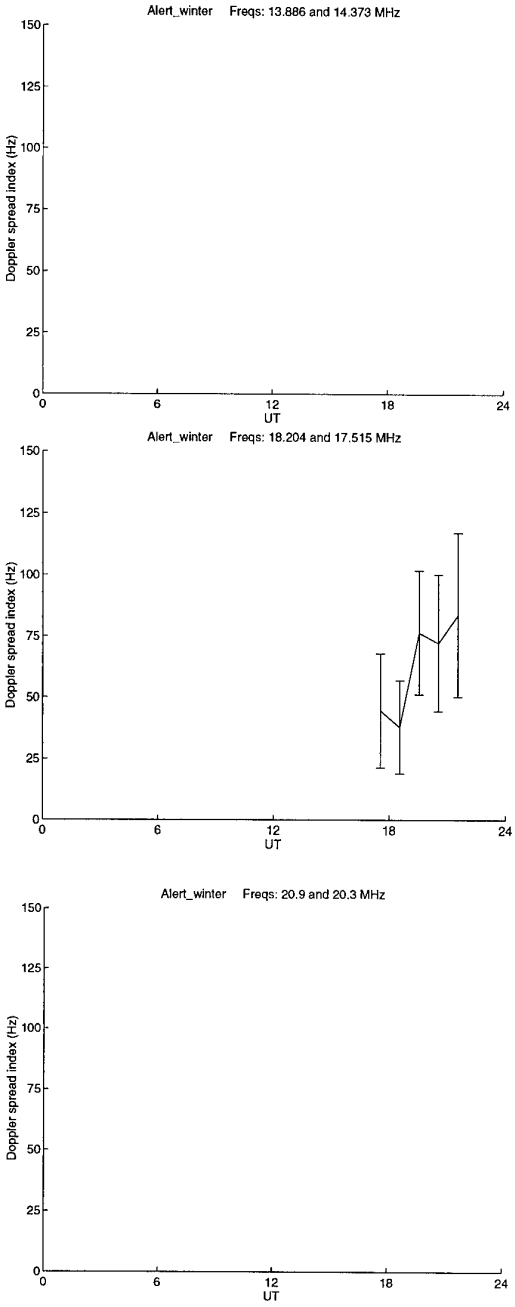


Figure 3. Mean diurnal variation (curve) and standard deviation (vertical bars) of Doppler spread index for the path to Alert during the winter campaign.

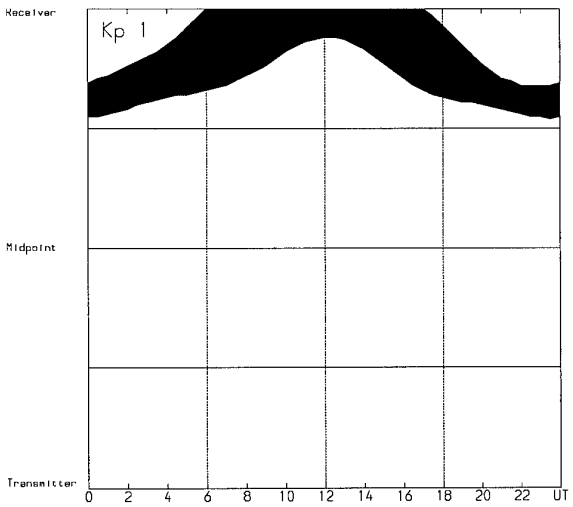


Figure 4. Diurnal variation of the mean position of the auroral oval relative to the propagation path from Clyde River to Prudhoe Bay (transmitter at the bottom and the receiver at the top).

3.5 Prudhoe Bay

The path from Clyde River to Prudhoe Bay alternates between being contained entirely within the polar-cap and being trans-auroral oval. This is illustrated in Figure 4 in which the statistical position of the auroral oval for moderate geomagnetic conditions is indicated with respect to the path reflection points. Although the path mid-point is always located within the polar cap, the second 2-hop reflection point is close to the boundaries of the mean auroral oval between approximately 1600 UT and 0500 UT.

Summer

There is little evidence of propagation at the two lower frequency pairs (see Figure 5). At 6.800 / 6.905 MHz during the period of propagation, the mean spread index is around 60 Hz with no apparent diurnal trend. At all higher propagating frequencies there is a clear diurnal variation in Doppler spread with a minimum around 1200 UT. This variation is well correlated with the average position of the auroral oval with respect to the propagation path: the path is almost entirely within the polar cap at around 1200 UT and low Doppler spreading is observed. In contrast, large spreads are observed when the second 2-hop reflection point is close to or within the active auroral zone, at around 0000 UT,.

Winter

For the winter period (see Figure 6), the Doppler spreads are significantly larger than those for the summer period, typical spread indices being in the order of 90 Hz. Unlike the summer period, there is little or no discernible diurnal variation in the Doppler spreads.

Variation with geomagnetic activity

During the summer period, there is little difference in Doppler spreads observed between the geomagnetically more active days (Ap≥10) and for the geomagnetically less active (Ap≤10.).

Table 3. Occurrence of Doppler spreads observed on the Clyde River to Prudhoe Bay path when the second 2-hop reflection point is located within and outside the expected (mean) boundaries of the auroral oval.

(a) Summer, well within the auroral oval (2000 - 0400 UT).

Frequency (MHz)	SI<30 Hz	30-100 Hz	SI>100 Hz	Recogn rate
3.185/3.230	-	-	-	0%
4.900/5.200	-	-	-	0%
6.800/6.905	12%	77%	11%	13%
-/10.195	20%	49%	31%	70%
13.886/14.373	8%	62%	30%	73%
18.204/17.515	13%	66%	21%	30%
20.900/20.300	33%	67%	0%	1%

(b) Summer, well outside the auroral oval (0900-1400 UT).

Frequency (MHz)	SI<30 Hz	30-100 Hz	SI>100 Hz	Recogn rate
3.185/3.230	-	-	-	0%
4.900/5.200	11%	79%	10%	7%
6.800/6.905	13%	77%	10%	55%
-/10.195	44%	47%	9%	65%
13.886/14.373	71%	27%	2%	71%
18.204/17.515	72%	26%	2%	17%
20.900/20.300	13%	75%	12%	3%

(c) Summer, all times.

Frequency (MHz)	SI<30 Hz	30-100 Hz	SI>100 Hz	Recogn rate
3.185/3.230	-	-	-	0%
4.900/5.200	11%	80%	9%	4%
6.800/6.905	9%	81%	10%	33%
-/10.195	25%	54%	21%	67%
13.886/14.373	33%	53%	14%	72%
18.204/17.515	35%	52%	13%	30%
20.900/20.300	16%	51%	33%	3%

(d) Winter, all times. Note that the occurrence statistics are very similar for times when the path is within or crossing the auroral oval; hence only the overall values are given.

Frequency (MHz)	SI<30 Hz	30-100 Hz	SI>100 Hz	Recogn rate
3.185/-	-	-	-	<1%
4.455/4.455	2%	92%	6%	7%
6.800/6.905	0%	45%	55%	53%
9.941/10.195	1%	62%	37%	39%
13.886/14.373	2%	60%	38%	37%
18.204/17.515	4%	59%	37%	33%
20.900/20.300	4%	63%	33%	6%

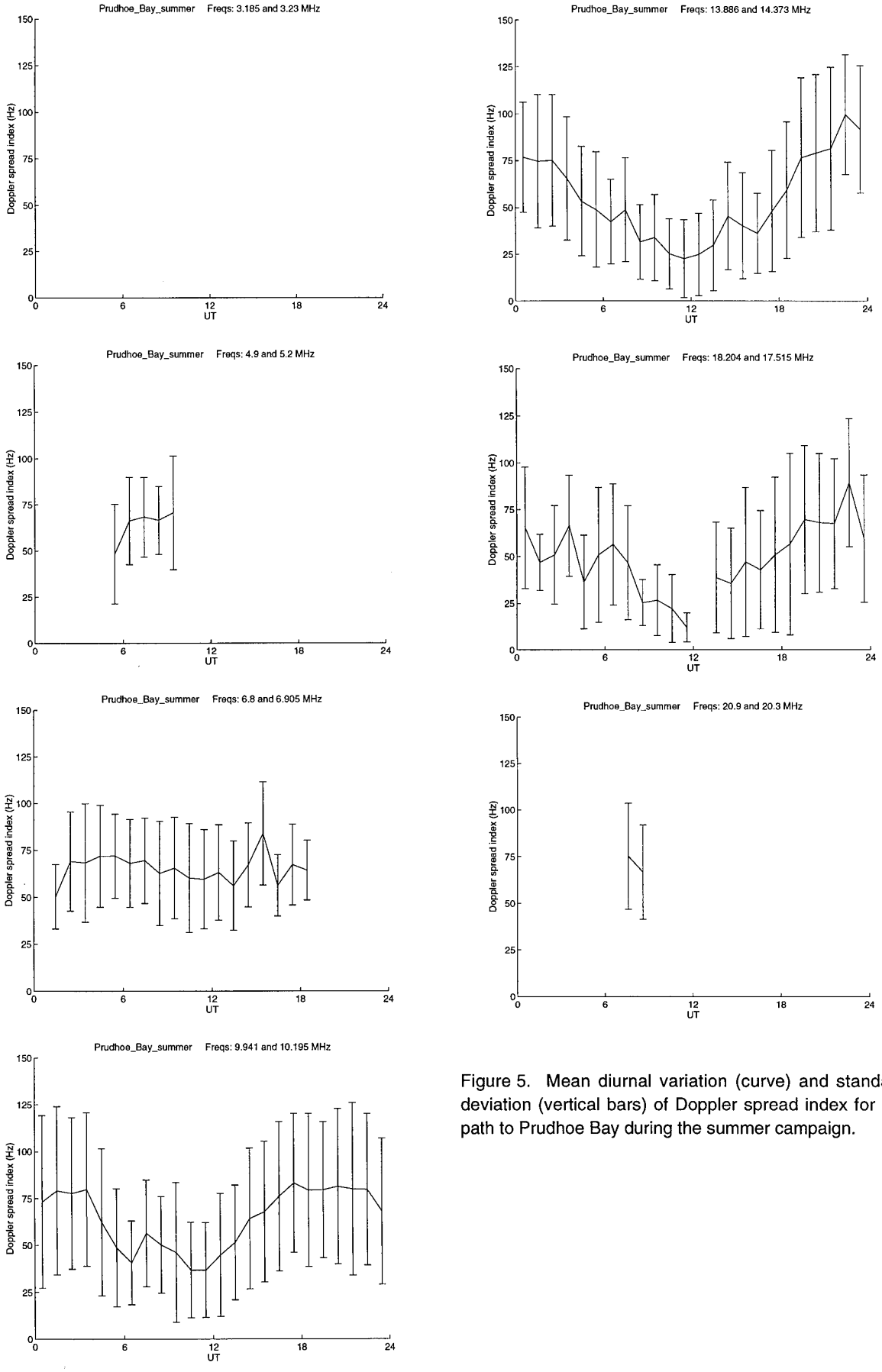
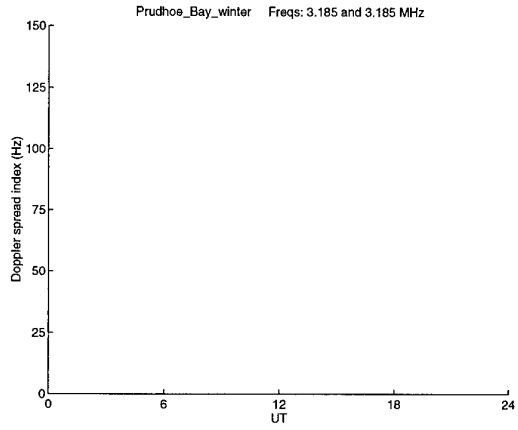
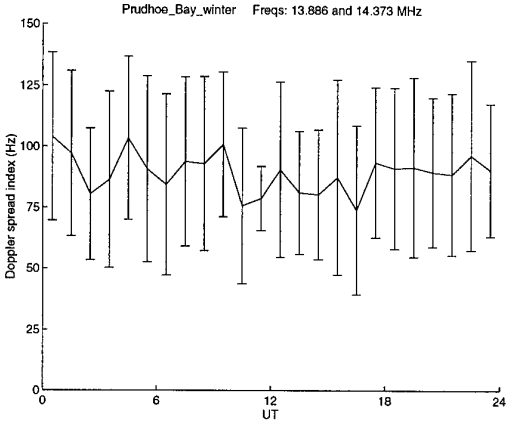
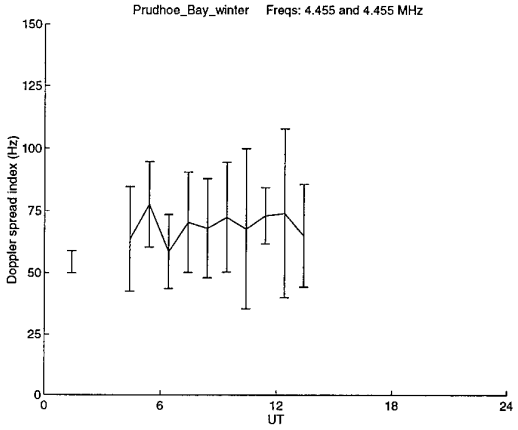
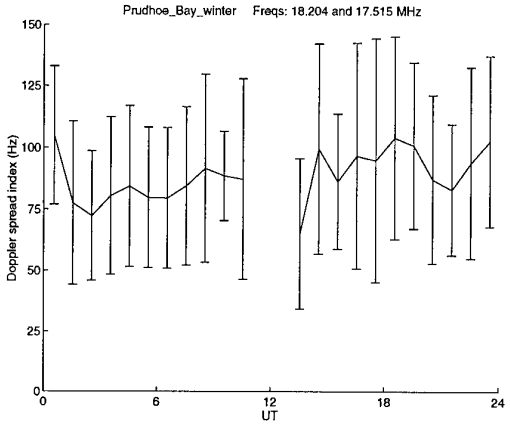


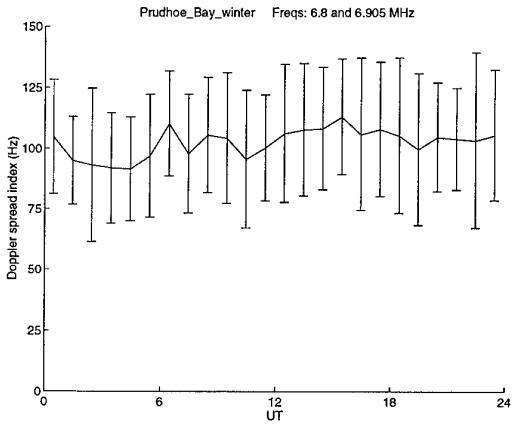
Figure 5. Mean diurnal variation (curve) and standard deviation (vertical bars) of Doppler spread index for the path to Prudhoe Bay during the summer campaign.

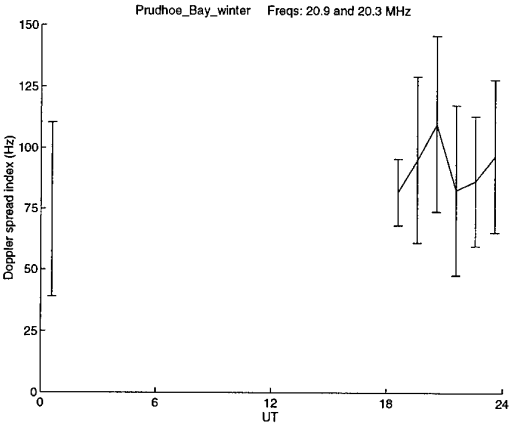












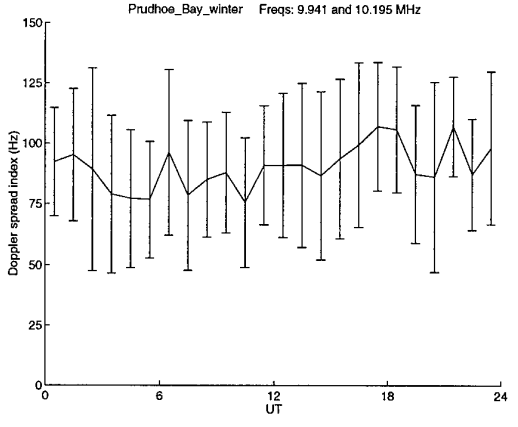


Figure 6. Mean diurnal variation (curve) and standard deviation (vertical bars) of Doppler spread index for the path to Prudhoe Bay during the winter campaign.

The most apparent feature is a change in the time of the minimum in the diurnal variation to around 1400 UT at 13.886 / 14.373 MHz. There is no apparent trend in the relationship between Ap and spread index during the winter period.

4 FSK SIGNALLING ERRORS

4.1 Measurement method

Synchronisation with the FSK data sequence was achieved by cross correlation of the received signal with the known 39-bit pseudo-Barker coded data sequence. After synchronisation, the received data were compared with the expected pseudo-random sequence and the position and distribution of the signalling errors determined.

It is interesting to note that errors rarely occurred in adjacent bits, although at times the errors did tend to cluster into groups, referred to here as error bursts (e.g. see Figure 7 for the Clyde River to Prudhoe Bay path). An error burst was defined as a group of bits starting and ending with bits received in error followed by at least 4 correctly received bits. The error free periods between error bursts are referred to as the guard space. These definitions are similar to those employed by Brayer [2].

4.2 Relationship to signal to noise ratio

The signal to noise ratio (SNR) was measured by comparing the signal power from the FSK tone detection filter in which the signal was expected (from the known data sequence) with the power from the other tone detection filter which was assumed to be a measure of the background noise alone.

The average diurnal variations of signal to noise ratio taken over the entire campaign for signal frequencies of 6.800 MHz, 9.941 MHz and 13.886 MHz are presented in Figure 8. Almost constant values of around 30 dB occur for the two higher frequency signals. A marked variation occurs for the 6.880 MHz signal with the highest SNR occurring at around path mid-point midnight when the D region absorption is least. It should, however, be noted that the SNR values indicated in this figure are averaged over the entire campaign; much lower values occurred at times, often as a result of destructive interference between the various propagation modes.

The probability of a bit being received in error was found to increase with decreasing signal to noise ratio (e.g. see Figure 9 for the 13.886 MHz 150 baud signal). The form of the observed variations are in good agreement with the expected (theoretical) distributions [3].

4.3 Error distribution measurements

The following analysis of the bit error distributions concentrates on the 13.886 MHz measurements recorded during the summer campaign on the path to Prudhoe Bay.

Averaged error distributions for this frequency at 50, 75 and 150 baud as a function of time of day are presented in Figure 10. There is evidence of an increase in error bursts of 2 or more bits around 1200 UT when the second 2-hop reflection point is expected to be furthest away from the auroral oval. The largest variation is observed at 50 baud where the probability of a burst of 2 or more errors increases from approximately 5% to 20%.

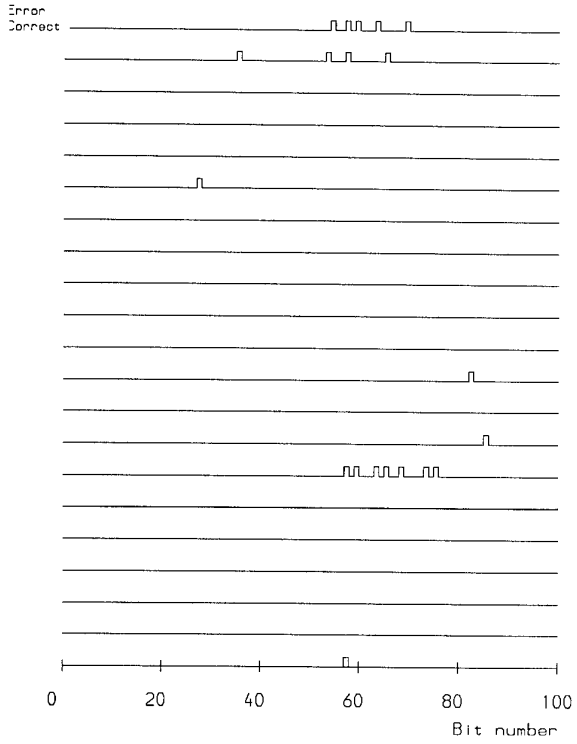


Figure 7. Example of the grouping of errors for the 13.886 MHz, 150 baud signal at 1300 UT on 25 July 1988. 2000 bits are shown, 100 on each row of the figure.

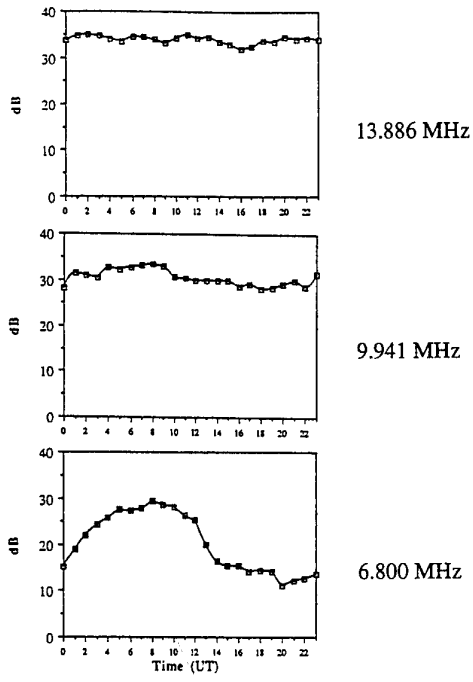


Figure 8. The average diurnal variation of signal to noise ratio taken over the entire summer campaign for signal frequencies of 6.800 MHz, 9.941 MHz and 13.886 MHz.

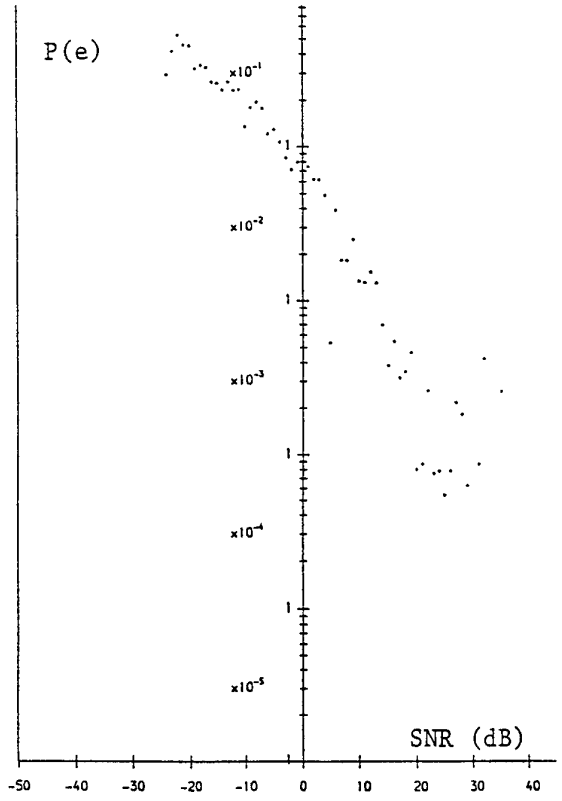


Figure 9. Variation of the probability of a bit being in error as a function of SNR for the 13.886 MHz, 150 baud signal. 25 July 1988.

The variation in the length of the guard space with time of day is presented in Figure 11. No correlation between the time of day and the length of the guard space is apparent.

4.4 Relationship of error distribution to Doppler spread

The distribution of bit errors changes markedly with changes in the Doppler spread. For periods with the lowest Doppler spread, the tendency for bit errors to occur in clusters is greatest. As the Doppler spread increases, the average length of the error bursts decrease. This is illustrated for the 13.886 MHz signal in Figure 12 in which the percentage of error bursts of a particular length is shown as a function of Spread Index for signalling rates of 50, 75 and 150 baud. It is also apparent from this figure that the error bursts are longer (in bits) at the higher data rates.

Little or no variation in the length of the guard space was observed at all baud rates. Almost all guard spaces were at least 10 bits in length with about 80% exceeding 50 bits and about 45% exceeding 100 bits (see Figure 13).

5 CONCLUDING REMARKS

Several interesting features of the propagation of HF signals over high latitude paths have been observed:

(a) Large Doppler spreads, with spread indices often in excess of 100 Hz, are frequently observed on all paths indicating the presence of a turbulent 'rough' ionospheric reflector which produces multiple sub-modes.

(b) The Doppler spreading of trans-auroral oval signals is well correlated with the average position of the auroral oval relative to the path reflection points. This region is a source of major disturbance.

(c) The bit error rate of low speed 850 Hz shift FSK signals is not influenced by the high level of Doppler spreading observed on trans-auroral oval paths. The error rate is well related to the received SNR and is in good agreement with the theoretical distribution.

(d) The distribution of bit errors is markedly affected by changes in the Doppler spreading of the received signals. At high Doppler spreads, the length of the error bursts is significantly less than for the cases where the Doppler spreading was small. However, no significant affect on the length of the guard space between error bursts is evident.

The large spread in the frequency spectrum of the received signals is a well known feature of high latitude propagation. However, the present studies show that the Doppler spread is sensitive to auroral activity and to the geometry of the propagation path in relation to the auroral oval. Furthermore, it has proved possible to relate the Doppler spreading imposed on the signal by the ionospheric reflection processes to the error distribution characteristics of low speed FSK signals.

6 REFERENCES

1. Feldstein, Y.I. and Starkov, G.V. Dynamics of auroral belt and polar geomagnetic disturbances. *Planetary and Space Science*, 1967, **15**, 209-229.

2. Brayer, K. Error patterns observed on trans-equatorial HF communication links. *IEEE Trans. on Communication Technology*, **COM-16**, 1968, 215-221.

3. Schwartz, M., Bennett, W.R. and Stein, S. *Communication Systems and Techniques*. McGraw-Hill, 1966.

◇ Individually occuring errors

○ Error bursts of 2 or more

△ Error bursts of 3 or more

⊞ Error bursts of 4 or more

◆ Error bursts of 5 or more

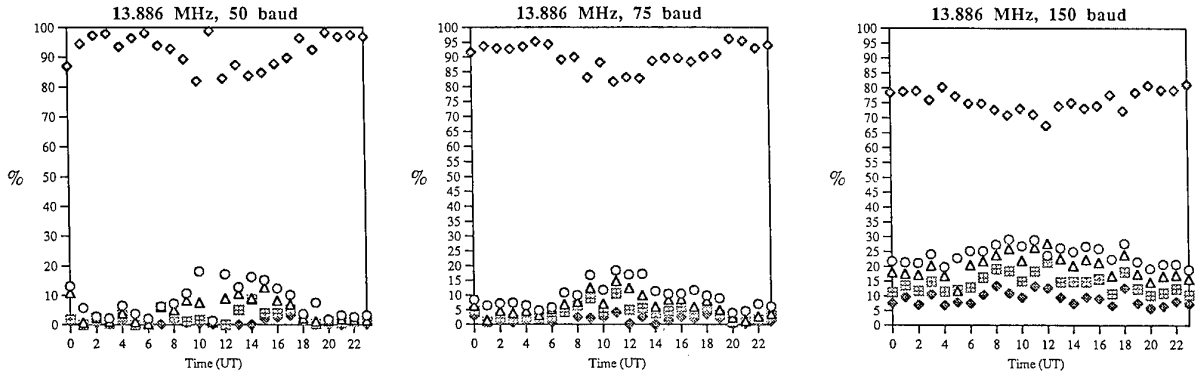


Figure 10. Average number of errors occurring individually or in bursts as a percentage of the total number of errors recorded on the 13.886 MHz FSK data transmission during the summer 1988 experimental campaign.

◇ Guard space of 10 or more

○ Guard space of 50 or more

△ Guard space of 100 or more

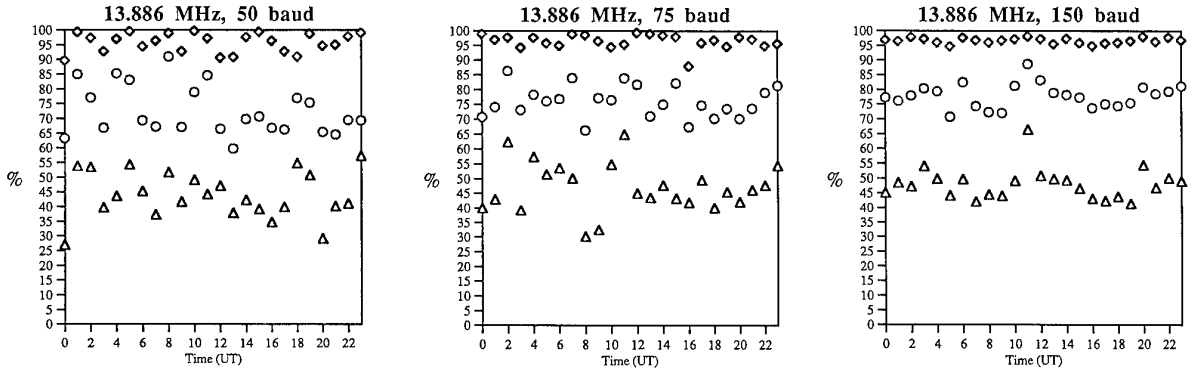


Figure 11. Average length of guard space as a percentage of the total number of guard spaces recorded on the 13.886 MHz FSK data transmission during the summer 1988 experimental campaign.

◇ Individually occuring errors

○ Error bursts of 2 or more

△ Error bursts of 3 or more

⊞ Error bursts of 4 or more

◆ Error bursts of 5 or more

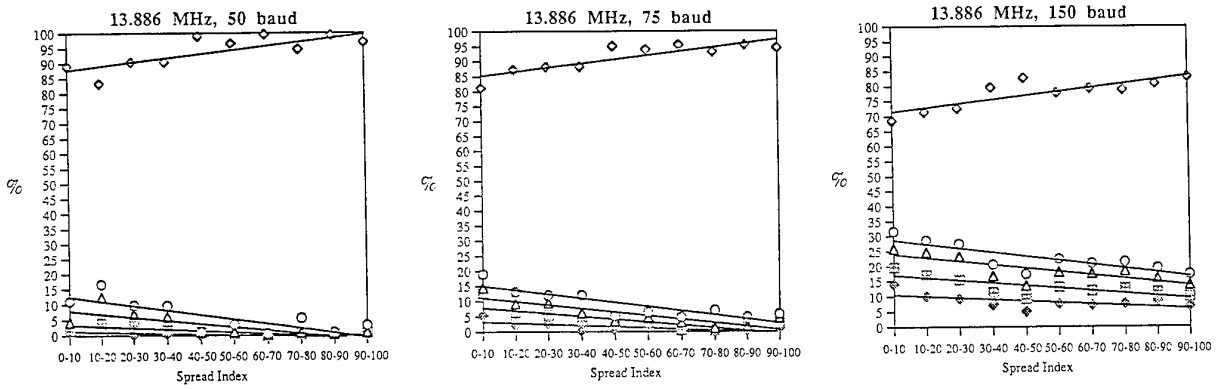


Figure 12. Average number of errors occurring individually or in bursts as a percentage of the total number of errors recorded on the 13.886 MHz FSK transmission during the summer 1988 experimental campaign plotted as a function of the corresponding spread index.

◇ Guard space of 10 or more

○ Guard space of 50 or more

△ Guard space of 100 or more

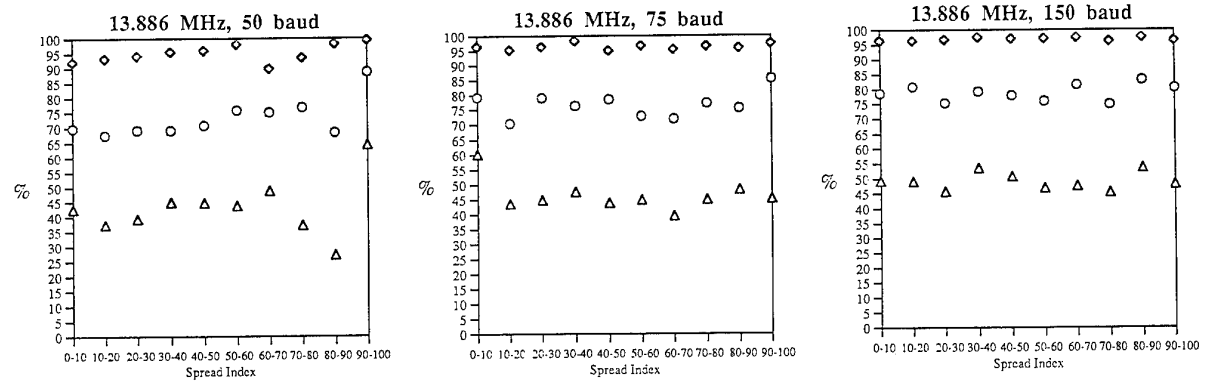


Figure 13. Average length of guard space as a percentage of the total number of guard spaces recorded on the 13.886 MHz FSK transmission during the summer 1988 experimental campaign plotted as a function of the corresponding spread index.

DISCUSSION

Discussor's name: P. Cannon

Comment/Question:

In respect to your error distribution on incoherent FSK signals, I wonder what your link margin uses? A large spread would place signal power outside of the tone filter but this might not matter in an incoherent system if there was sufficient power margin. This is, of course, providing that the spread power does not fall in the other tone filter. Could this explain why the FSK error rate does not increase with speed?

Author/Presenter's reply:

At the data rates employed in these experiments, the level of Doppler spreading does not exceed the tone detection filter bandwidth. Consequently, the detected SNR, and hence the BER, is not degraded. Problems would be expected to occur in systems employing narrow shifts at low band rates since power may be spread outside the detector filter bandwidths reducing the detected power and may even leak into the pass bands of other detection filters.

COMPARAISON ENTRE LES FORMES D'ONDE SERIE ET PARALLELE POUR LES TRANSMISSIONS HF

Didier Pirez, Dominique Merel
THOMSON-CSF RGS
66 rue du Fossé Blanc
92231 GENNEVILLIERS
FRANCE

RESUME :On compare des formes d'ondes capables toutes deux de démodulation cohérente, grâce à la présence de symboles de référence.

On cherche tout d'abord à comparer les performances intrinsèques des formes d'onde dans le cas de modèles de canaux de type CCIR (plusieurs trajets sur quelques millisecondes affectés d'évanouissements de Rayleigh indépendants). Sur ce type de canal la forme d'onde parallèle doit se comporter comme en présence d'un évanouissement plat. La forme d'onde série doit profiter d'un effet de diversité lié à l'indépendance des évanouissements, et donc présenter des courbes de taux d'erreurs de pente plus élevée, les performances intrinsèques en présence d'un seul trajet étant identiques.

Dans la pratique les performances de la forme d'onde série s'éloignent des performances intrinsèques. Ceci est dû en partie à la sous-optimalité des égaliseurs effectivement mis en œuvre. Ceci est dû également au fait que l'égalisation et le décodage correcteur d'erreurs ne sont pas optimisés conjointement. La forme d'onde parallèle peut quant-à-elle rester très proche de ses performances théoriques.

On donne également quelques indications sur les performances comparées des deux formes d'onde en fonction de l'étalement temporel, de l'étalement doppler ; on donne également quelques résultats de mesure sur le facteur de crête qui, moyennant un traitement préalable, n'est pas toujours défavorable au modem parallèle.

La forme d'onde série présente donc quelques avantages intrinsèques sur la forme d'onde parallèle mais ces avantages ne pourront être exploités qu'en se rapprochant des performances théoriques de l'égalisation (maximum de vraisemblance) et en optimisant conjointement l'égalisation et le décodage. Dans l'intervalle la forme d'onde parallèle présente une solution très performante, y compris jusqu'aux très hauts débits (9600 bits/s).

A. INTRODUCTION

Les formes d'onde parallèles sont historiquement les premières à avoir permis des débits importants sur le canal HF.

Dans les mises en œuvre qui ont été proposées, on reproduit dans le processus de démodulation un processus de démodulation bas-débit porteuse par porteuse : on retrouve des modulations FSK ou PSK différentielle à chaque fréquence.

Le codage par contre peut prendre en compte l'aspect haut-débit de la transmission. Ainsi dans l'ANDVT («Advanced Narrow-band Digital Voice Terminal»)a été proposé un codage de Golay portant sur l'ensemble des porteuses d'une trame.

Les modems série sont apparus ensuite, dès que les développements technologiques ont permis la mise en œuvre des algorithmes adaptatifs nécessités par l'égalisation.

On considère généralement que ces modems série ont des performances meilleures que celles des modems parallèles qui les ont précédés. Ils se sont imposés et continuent de s'imposer dans toutes les normes de transmission à moyen débit (entre 300 et 2400 bits/s).

Cependant, les algorithmes d'égalisation deviennent plus complexes lorsqu'on essaye d'atteindre des débits beaucoup plus élevés. Pour atteindre ces débits avec une complexité de mise en œuvre faible, nous avons développé à THOMSON-CSF RGS un nouveau type de modem parallèle, basé sur une démodulation et un décodage cohérent ([7]). Cette démodulation cohérente est rendue possible par l'adjonction dans la forme d'onde de porteuses de référence (connues du récepteur), qui servent à estimer la fonction de transfert du canal, ainsi que le niveau de bruit sur chaque fréquence.

Aux débits élevés, il est difficile de faire des comparaisons entre une forme d'onde parallèle de ce type et des formes d'onde série, encore peu disponibles pour les raisons de complexité de processus d'égalisation déjà mentionnées. Il est cependant possible d'effectuer cette comparaison à débit moyen.

Dans la suite de l'article on tente tout d'abord d'établir des performances théoriques des deux formes d'onde sur un modèle de canal à multi-trajets et évanouissements de Rayleigh. On discute ensuite les problèmes d'algorithmie et de mise en œuvre qui nous éloignent des performances théoriques. Enfin on envisage les axes de recherche qui pourraient faire évoluer ces résultats.

B. PERFORMANCES INTRINSEQUES

B.I. PRINCIPE DE LA COMPARAISON

B.I.1. Modèle de canal

Le modèle de canal est du type Watterson, c'est-à-dire que

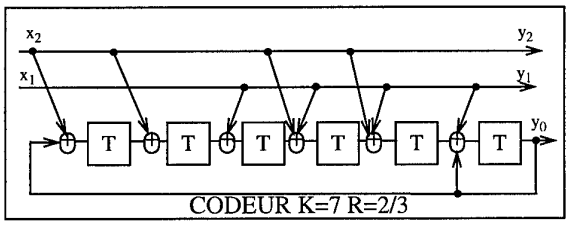
Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

l'on a des trajets multiples à évanouissements de Rayleigh indépendants([6]). Pour bien mettre en évidence les effets de diversité, on a pris ici trois trajets de même amplitude moyenne. De plus, pour faciliter les calculs dans le cas du modem série, on a choisi des écarts temporels multiples de la période symbole. Plus particulièrement, les retards sont 0, T et 3T, où T est la période symbole (du modem série). Il faut remarquer que pour une vitesse de modulation typique de 2400 bauds, ce canal n'est pas bien sévère, puisqu'il s'étale sur 1,25 ms.

Dans la suite on supposera que la réponse impulsionnelle du canal est en permanence parfaitement connue du récepteur.

B.I.2. Codage

On essaye ici d'évaluer les performances des formes d'onde sur une transmission codée. Pour ce faire on utilise une modulation codée qui passe 2 bits utiles par symbole en 8PSK([5]). Le codeur utilisé est un codeur convolutionnel systématique de longueur de contrainte 7, de rendement 2/3 (voir schéma ci-dessous). L'affectation des bits codés aux symboles 8PSK se fait selon la numérotation naturelle. On suppose que l'entrelacement est parfait, c'est-à-dire que les symboles successifs qui arrivent au décodeur sont passés par des canaux totalement indépendants. Avec ce code, on obtient alors une diversité de 4 sur le fading de Rayleigh.



B.I.3. Fonctionnement du modem parallèle

On suppose ici que les paramètres des trames modem, en particulier le temps de garde entre les trames, sont bien dimensionnés de telle manière que l'on soit dans l'hypothèse de travail habituelle du modem parallèle : il n'y a pas d'interférence intersymbole. Par ailleurs, on néglige les effets des non-stationnarités du canal pendant la durée de la trame (ces effets seront pris en compte dans un autre paragraphe).

Dans cette hypothèse, si un symbole 8PSK s_n est émis à un instant t_n sur une sous-porteuse de fréquence f_n, on peut considérer que l'on reçoit simplement un signal $z_n=A_n.s_n+b_n$, où A_n est une amplitude complexe, b_n est un échantillon de bruit gaussien.

Supposons que la réponse du canal s'écrive :

$$h(t,\tau) = \sum_i h_i(t)\,\delta(\tau-\tau_i)$$

c'est-à-dire que l'on a une réponse variable dans le temps, mais composée de trajets de retards τ_i fixes.

Il s'agit ici d'un modèle simplifié puisqu'on n'a pas pris en compte les réponses impulsionnelles des filtres radio de l'émetteur et du récepteur.

On peut alors réécrire l'amplitude complexe A_n sous la forme

$$A_n = \sum_i h_i(t_n)\, e^{-j2\pi f_n \tau_i}$$

:

On voit dans cette expression que la valeur de la fréquence de la sous-porteuse et la valeur du temps de retard du trajet interviennent uniquement dans des termes de déphasage par rapport aux amplitudes complexes des trajets. A_n s'écrit alors comme une somme de variables gaussiennes, c'est donc également une variable gaussienne.

Si maintenant on considère deux valeurs A_n et A_m prises à des instants suffisamment éloignés pour assurer la décorrélation des amplitudes complexes des trajets $h_i(t_n)$ et $h_i(t_m)$, on aura réalisé deux tirages indépendants de la variable gaussienne A_n.

Dans le modèle de canal considéré, les amplitudes des trajets sont également supposées décorrélées entre elles (modèle de Watterson). On peut donc calculer facilement la puissance de A_n :

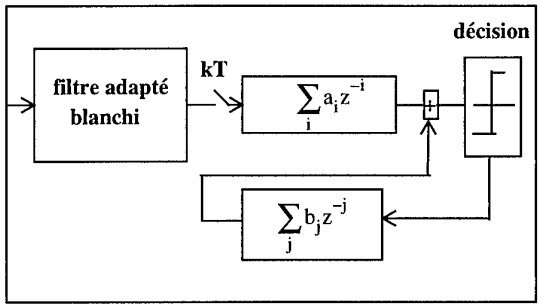
$$E\left(|A_n|^2\right) = \sum_i E\left(|h_i|^2\right)$$

On voit donc que le modèle théorique correspondant au modem parallèle pour le problème posé est celui d'un fading de Rayleigh parfaitement entrelacé, dont la puissance est la somme des puissances des trajets.

B.I.4. Fonctionnement du modem série

Démarche

On évalue ici le comportement d'un égaliseur à décision dans la boucle, qui est une structure communément utilisée aujourd'hui. La structure générale retenue est la suivante([3]) :



Dans la structure classique de l'égaliseur à décision dans la boucle, on trouve en tête un filtre adapté, et non pas un filtre adapté blanchi, suivi d'un échantillonnage au rythme symbole et d'un filtre transverse échantillonné. Cette structure est cependant tout-à-fait équivalente en théorie, et présente plusieurs avantages en pratique.

La démarche que nous allons suivre est un peu plus complexe que dans le cas du modem parallèle.

Il s'agit d'effectuer des tirages indépendants de la réponse impulsionnelle du canal décrite plus haut. Pour chaque tirage, on va calculer les expressions exactes des filtres constituant la structure de l'égaliseur. On pourra en déduire un rapport signal à bruit au niveau de l'organe de décision. On se ramène alors pour le signal égalisé au cas

$z_n= A_n.s_n+b_n$

où A_n est une amplitude complexe et b_n un échantillon de

bruit.

On fait de plus l'hypothèse qu'en sortie d'égaliseur, le bruit et le résidu d'interférence inter-symbole sont gaussiens, ce qui nous permet d'effectuer des tirages aléatoires de l'échantillon de bruit b_n

On est alors dans une situation proche de celle du modem parallèle. Pour ce dernier on fait un certain nombre de tirages de l'amplitude A_n et du bruit b_n, A_n suivant une loi de Rayleigh ; pour le modem série on fait des tirages de la réponse impulsionnelle du canal, et on en déduit une valeur de A_n compte-tenu du comportement modélisé de l'égaliseur.

Il faut remarquer que cette approche revient à considérer que les symboles passés décidés réinjectés dans l'égaliseur sont exacts. On traite plus loin le cas où ils sont effectivement décidés.

Calcul du filtre adapté blanchi

Le filtre adapté blanchi peut se mettre en théorie sous la forme de la concaténation d'un filtre adapté au canal, d'un échantillonage au rythme symbole, et d'un blanchisseur échantillonné.

Si H(f) est la fonction de transfert du canal, et $\gamma_b(f)$ la densité spectrale du bruit, le filtre adapté a pour fonction de transfert : $H^*(f)/\gamma_b(f)$.

Si on échantillonne derrière le filtre adapté, le canal équivalent a une fonction de transfert :

$$U(z) = u_0 \cdot \prod_i \left(1 - u_i z^{-1}\right) \cdot \left(1 - (1/u^*_i)\, z^{-1}\right)$$

Dans cette représentation, les zéros peuvent être rangés par paires, dans lesquelles on a deux inverses conjugués : on peut par exemple supposer que tous les u_i sont de module inférieur ou égal à 1. Le filtre blanchisseur consiste à éliminer tous les zéros situés à l'extérieur du cercle unité Il reste donc :

$$V(z) = u_0 \cdot \prod_i \left(1 - u_i z^{-1}\right)$$

Cette réponse est à phase minimale, c'est-à-dire en particulier que l'énergie est concentrée au maximum sur les premiers coefficients. Par ailleurs on sait que dans le cas de l'égaliseur à annulation de l'interférence inter-symbole («zero forcing DFE»), on a réalisé le filtre optimal et le filtre transverse qui suit est inutile([3]).

Pour le critère du minimum d'erreur quadratique qui nous occupe ici, le filtre transverse est encore nécessaire, mais le nombre de coefficients à utiliser peut être réduit si la réponse du canal est à phase minimale.

Dans le cas présent on a pris soin de choisir un modèle de canal où les trajets ont des temps de retard multiples de la période symbole. On peut alors faire tous les calculs au rythme symbole. Pour trouver la réponse du canal en sortie du filtre adapté blanchi, il suffit de remplacer dans la transformée en z de la réponse du canal tous les zéros à l'extérieur du cercle unité par leur inverse conjugué. Le filtre qui réaliserait cette opération est un filtre passe-tout.

Calcul du filtre transverse de l'égaliseur

On se fixe un nombre de coefficients du filtre transverse, qu'on prend grand devant la durée de la réponse impulsionnelle à traiter (ici on a pris 20 coefficients pour une réponse à phase minimale s'étalant sur 4 symboles).

Si on appelle h_i les coefficients de la réponse impulsionnelle à phase minimale, on peut écrire pour tout échantillon z_n :

$$z_n = \sum_{i=0}^{L-1} h_i \cdot s_{n-i} + b_n$$

Si on représente par un vecteur Z l'ensemble des échantillons présents dans le filtre transverse à un instant donné, et par S l'ensemble des symboles de modulation présent et futurs (la contribution des symboles passés est supposée complètement retranchée par la partie récursive de l'égaliseur), on peut écrire :

$$Z = \begin{bmatrix} z_n \\ z_{n+1} \\ \cdots \\ z_{n+N-1} \end{bmatrix}$$

$$Z = \begin{bmatrix} h_0 & 0 & \cdots & \cdots & 0 & 0 \\ h_1 & h_0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ h_{L-1} & \cdots & h_1 & h_0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & h_{L-1} & \cdots & h_1 & h_0 \end{bmatrix} \cdot \begin{bmatrix} s_n \\ s_{n+1} \\ \cdots \\ s_{n+N-1} \end{bmatrix} + \begin{bmatrix} b_n \\ b_{n+1} \\ \cdots \\ b_{n+N-1} \end{bmatrix}$$

$$Z = \Omega \cdot S$$

On cherche le filtre A qui minimise l'erreur quadratique moyenne : $E\left(\left|A^\dagger Z - s_n\right|^2\right)$

La solution est alors classique, en faisant intervenir les matrices d'autocorrélation de Z et d'intercorrélation avec la réplique s_n :

$$A = R_{ZZ}^{-1} \cdot R_{Zs_n^*}$$

ce qu'on peut ensuite exprimer en fonction des paramètres du problème :

$$A = \left(\Omega \cdot \Omega^\dagger + 2\sigma^2 \mathrm{Id}\right)^{-1} \cdot H$$

avec $H=(h_0,h_1,\ldots,h_{L-1},0,\ldots,0)^T$ et $2\sigma^2$ puissance du bruit.

On peut alors retrouver un modèle du type : $A^\dagger Z = \alpha_n \cdot s_n + \xi_n$ pour faire la comparaison avec le modem parallèle.

L'amplitude complexe α_n est assez facile à retrouver ; on a :

$$\alpha_n = A^\dagger H$$

Pour retrouver la puissance du bruit (en fait bruit plus interférence inter-symbole résiduelle) en sortie du filtre, on part de la puissance totale et on retranche la contribution de l'utile :

$$P_b = P_t - P_u$$

On a :

$$P_t = E\left(\left|A^\dagger \cdot Z\right|^2\right) = A^\dagger \cdot R_{ZZ} \cdot A = A^\dagger \cdot H$$

Compte-tenu de la valeur de α_n, on peut écrire :

$$P_b = A^\dagger \cdot H \cdot (1 - (A^\dagger \cdot H))$$

On peut en déduite le rapport signal à bruit :

$$\frac{S}{B} = \frac{A^\dagger \cdot H}{1 - (A^\dagger \cdot H)}$$

On peut également faire intervenir dans ce calcul la puissance d'erreur d'égalisation ; on peut montrer assez facilement que l'on a :

$$P_e = E\left(\left|A^\dagger Z - s_n\right|^2\right) = 1 - A^\dagger H$$

Le rapport signal à bruit peut donc encore s'écrire :

$$\frac{S}{B} = \frac{1 - P_e}{P_e}$$

On peut également retrouver la pondération optimale des symboles égalisés en vue d'un décodage correcteur d'erreurs : avec un modèle du type : $y_n = \alpha_n s_n + \xi_n$, ξ_n étant gaussien, on sait que la pondération optimale consiste à transmettre au décodeur : $u_n = \alpha_n^* y_n / E\left(\left|\xi_n\right|^2\right)$; dans le cas présent, ce calcul se réduit à :

$$u_n = \frac{y_n}{P_e}$$

La pondération optimale pour un égaliseur à décision dans la boucle consiste donc simplement à diviser les symboles égalisés par la puissance d'erreur d'égalisation.

On a donc retrouvé pour le modem série le modèle simple que l'on avait utilisé pour le modem parallèle. La différence essentielle étant qu'on ne connaît pas a priori la densité de probabilité suivie par le rapport signal à bruit (ou l'amplitude du canal) ; on l'obtient donc par simulation en calculant ce S/B à chaque tirage de la réponse impulsionnelle du canal.

B.II. RESULTATS

B.II.1. SUR SYMBOLES DECIDES EXACTS

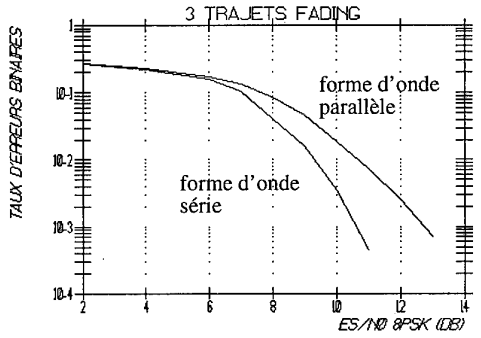
On suppose ici que l'on réinjecte dans la partie récursive de l'égaliseur les symboles exacts, et on applique le modèle ci-dessus, avec un bruit supposé gaussien et une amplitude calculée à partir du rapport signal à bruit dans le cas du modem série, et une amplitude suivant une loi de Rayleigh dans le cas du modem parallèle, mais avec un rapport signal à bruit moyen égal à la somme des rapport signal à bruit des trajets.

Le décodage de la modulation codée est effectué selon le maximum de vraisemblance avec la métrique :

$$\sum_n \mathrm{Re}\left(s_n^* \cdot \left(\frac{A_n^* y_n}{2\sigma_n^2}\right)\right)$$

où les s_n représentent la séquence de symboles à décider, et les y_n sont les échantillons reçus, qui se mettent sous la forme : $y_n = A_n s_n + b_n$, où A_n est une amplitude complexe et b_n un échantillon de bruit de puissance $2\sigma_n^2$

On obtient alors les courbes de taux d'erreurs ci-dessous, qui représentent le taux d'erreurs binaires en sortie de décodage, en fonction du rapport signal à bruit par symbole codé :



On constate qu'à faible rapport signal à bruit, les deux courbes se confondent, mais qu'à fort rapport signal à bruit, la pente de la courbe correspondant au modem série est plus forte, avec un avantage d'environ 2 db à un taux d'erreurs de 10^{-3}. Ceci correspond bien à un effet de diversité sur les évanouissements des trajets (diversité fréquentielle), effet dont profite le modem série mais pas le modem parallèle.

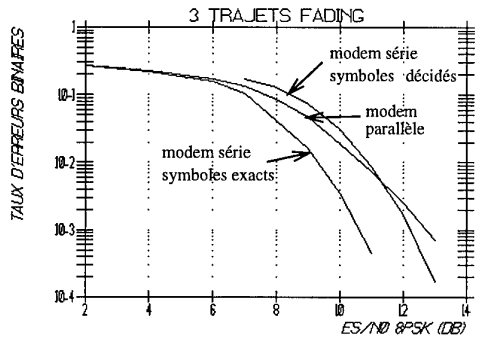
B.II.2. SUR SYMBOLES REELLEMENT DECIDES

Dans la structure de l'égaliseur, on prend des décisions sur les symboles 8PSK codés, ces décisions servant à éliminer l'influence des symboles passés dans l'interférence inter-symbole. L'analyse ci-dessus suppose que ces décisions sont bonnes, c'est-à-dire que l'influence des symboles passés est totalement retranchée. Cette hypothèse n'est évidemment pas réaliste, et l'effet des erreurs de décisions intervient dans les performances intrinsèques

de l'égaliseur.

Pour analyser cet effet, on ne peut plus se contenter du modèle simplifié précédent. On doit donc simuler le processus d'égalisation en tant que tel. Pour ce faire, on entrelace les symboles par bloc, de manière à garder l'hypothèse d'indépendance des tirages de la réponse impulsionnelle sur des symboles codés successifs. On simule donc pour chaque tirage le passage par le canal de transmission bruité correspondant, et l'égalisation par les filtres calculés comme précédemment.

Sur la figure ci-dessous, on a reporté les deux courbes déjà obtenues pour le modem parallèle et le modem série, ainsi que deux courbes obtenues par cette nouvelle simulation, l'une avec symboles exacts et l'autre avec symboles effectivement décidés.



On constate que la décision des symboles fait perdre largement au modem série l'avantage qu'il avait sur le modem parallèle.

Deux phénomènes sont en jeu :

-Le premier est que la décision porte sur les symboles codés, qui ont donc un taux d'erreurs beaucoup plus élevé que les bits décodés.

-Le deuxième phénomène est la propagation des erreurs. Les erreurs de décision provoquent une moins bonne égalisation, qui à son tour provoque des erreurs de décision. Dans les formes d'onde série normalisées, on insère à intervalles réguliers des séquences connues («mini-probes» en anglais), qui ont pour but premier de faciliter l'estimation du canal, mais qui servent également à casser les paquets d'erreurs de ce type. L'effet réel est donc sans-doute moins critique que ce que nous exposons ici, mais bien-sûr avec une perte de débit utile.

C. DEGRADATIONS LIEES A LA MISE EN ŒUVRE

C.I. PERTES LIEES A L'ESTIMATION DE CANAL

Les évaluations précédentes supposaient le canal parfaitement connu. Dans la pratique, l'algorithme d'estimation de canal est un paramètre crucial du fonctionnement d'un modem, qu'il soit série ou parallèle.

Dans le modem parallèle développé à THOMSON-CSF RGS, l'estimation de canal est réalisée par un échantillonnage direct de la fonction de transfert du canal par des porteuses de référence insérées dans certaines trames.

La qualité d'estimation est alors conditionnée par le taux de répétition de ces porteuses de référence en temps et en fréquence, par rapport à la vitesse de variation du canal (donnée par l'étalement doppler) d'une part, et par rapport à la durée de la réponse impulsionnelle (pour l'échantillonnage en fréquence) d'autre part. Evidemment, comme ces porteuses de référence prennent une part du débit disponible, on est conduit à réaliser des compromis entre la qualité d'estimation et le débit utile que l'on peut envisager.

Dans certaines formes d'onde série, on peut atteindre en principe un fonctionnement similaire, grâce à la présence de séquences connues longues répétées de manière périodique. C'est le cas du STANAG 4285 où la séquence de référence pourrait suffire à échantillonner le canal pour des étalements doppler modérés. Lorsque cette caractéristique n'est pas présente, on doit prendre en compte les symboles décidés dans le processus d'estimation, avec toutes les dégradations qu'apportent les erreurs de décision.

Dans le processus d'estimation de canal, on peut compter également l'ensemble des algorithmes liés au suivi de synchronisation, et au suivi de dérive de fréquence, dont la qualité a une importance considérable dans les performances.

C.II. PERTES LIEES AU DIMENSIONNEMENT DES FILTRES

Dans l'évaluation des performances intrinsèque de l'égaliseur, nous avons pris en compte une réponse impulsionnelle de faible longueur (4 symboles), que l'on supposait déjà à phase minimale, et un filtre transverse relativement long : 20 coefficients.

Pour le canal HF, on considère généralement que les réponses impulsionnelles peuvent atteindre une durée de 6 ms dans certains cas, soit un peu plus de 14 périodes symboles à 2400 bauds.

Par ailleurs, le nombre de coefficients des filtres est en général limité par la puissance de calcul nécessaire à leur adaptation.

On a donc là une source de dégradation supplémentaire.

C.III. PROBLEMES LIES AU FACTEUR DE CRÊTE

Le facteur de crête, c'est-à-dire le rapport entre la puissance crête de l'amplificateur nécessaire pour passer la forme d'onde, et la puissance moyenne émise, est souvent considéré comme le principal problème des formes d'onde parallèles.

En effet, on peut considérer que le facteur de crête d'une forme d'onde série de type STANAG 4285 est d'environ 4 à 5 db, alors que le facteur de crête d'une forme d'onde parallèle est en théorie égal au nombre de porteuses, mais se situe en pratique entre 10 et 12db dès que le nombre de porteuses est suffisant.

On peut cependant faire évoluer ce résultat en écrêtant le signal à l'émission. Cette opération amène des dégradations de performances à la réception, et il faut donc choisir le seuil d'écrêtage en fonction d'un compromis entre la puissance émise et ces dégradations.

Des essais menés à THOMSON-CSF RGS ont permis de mesurer des facteurs de crête de l'ordre de 4 db sur un émetteur réel,

pour un modem parallèle utilisant le code décrit dans cet article, avec des dégradations très faibles en rapport signal à bruit moyen.

On voit donc que l'on rejoint sur ce critère les performances du modem série.

Par ailleurs, les facteurs de crête mesurés sur les formes d'onde série dépendent de la constellation utilisée pour les symboles de modulation. Ils seront plus élevés pour des modulations de phase et d'amplitude que pour les modulations de phase utilisées couramment aujourd'hui. Dans le cas du modem parallèle, ce paramètre est indifférent.

C.IV. EXEMPLES DE PERFORMANCES

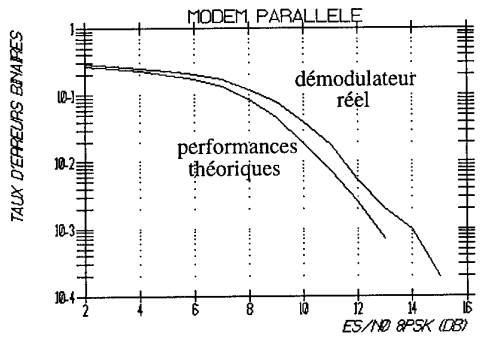
On a simulé une forme d'onde série et une forme d'onde parallèle sur le canal à trois trajets décrit ci-dessus, avec un étalement doppler de 1 Hz. Le code utilisé est le même que ci-dessus.

La forme d'onde série est celle du STANAG 4285, qui offre donc un débit utile de 1200 bauds, soit 2400 bits/s avec le code que nous avons choisi.

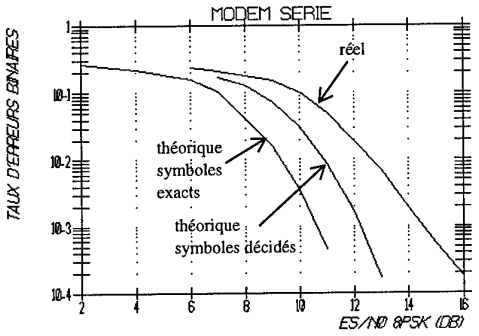
La forme d'onde parallèle est spécifique ; elle a 41 porteuses, une durée de trame de 22,5 ms, et présente un débit utile de 1355 bauds.

La simulation des démodulateurs est complète, c'est-à-dire qu'elle intègre l'estimation de canal, le suivi de synchronisation et de dérive de fréquence.

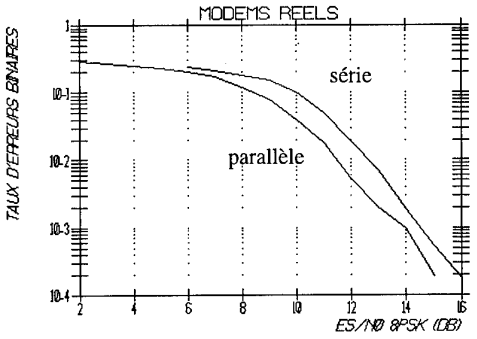
On donne ci-dessous les résultats pour chacun des modems puis on compare les réalisations des deux types de formes d'onde.



On voit que dans le cas du modem parallèle les pertes par rapport à la courbe théorique sont limitées à moins d'1 db.



Dans le cas du modem série, les pertes par rapport à la théorie sont plus importantes, même par rapport à la courbe «symboles décidés».



Lorsqu'on compare les deux réalisations, on constate que l'avantage potentiel initial du modem série a disparu complètement dans les exemples que l'on a pris ici : les performances du modem parallèle restent proches de la courbe théorique, alors que la complexité du modem série induit des pertes beaucoup plus importantes.

D. QUELQUES EVOLUTIONS POSSIBLES

Une des évolutions probables pour les formes d'onde série est le remplacement des égaliseurs à décision dans la boucle par des algorithmes de type «maximum de vraisemblance simplifié», dont le problème sera de trouver le bon compromis entre la complexité et la perte de performances par rapport au maximum de vraisemblance.

Un autre problème important sera l'intégration dans un même processus de décision de l'égalisation et du décodage correcteur d'erreurs, même lorsque la transmission est entrelacée. On a vu que la décision des symboles codés 8PSK au niveau de l'égaliseur amenait des dégradations de performances importantes dans notre exemple.

Enfin un point important sera l'évolution des algorithmes d'estimation de canal. Les exigences sur la qualité de l'estimation de canal seront d'autant plus fortes que l'on cherchera à augmenter le débit de la transmission.

Dans le cas du modem parallèle, on pourrait imaginer de retrouver une partie de la diversité sur les trajets par des concaténations de codes, un premier niveau exploitant la diversité fréquentielle, et un deuxième niveau la diversité temporelle. Il faut remarquer que dans ce cas il n'y aurait plus de différences fondamentales entre une forme d'onde série et une forme d'onde parallèle, le premier niveau de décodage s'apparentant à l'égalisation.

E. CONCLUSION

Nous avons tenté dans cet article de comparer les performan-

ces des formes d'ondes série et parallèle cohérente en séparant les effets intrinsèques à la forme d'onde des effets liés aux choix de mise en œuvre.

Il apparaît que pour un modèle de canal de type Watterson, avec plusieurs trajets indépendants, une forme d'onde série égalisée par un égaliseur à décision dans la boucle peut mieux profiter de la diversité fréquentielle liée à l'indépendance des trajets que la forme d'onde parallèle.

Malheureusement, cet avantage est quasiment entièrement annulé par la propagation des erreurs dans la boucle récursive sur les symboles décidés.

Par ailleurs, la simplicité de la démodulation de la forme d'onde parallèle fait que les performances des systèmes mis en œuvre peuvent rester très proches des performances théoriques, ce qui est plus difficile avec la forme d'onde série.

Il est cependant probable que dans un avenir plus ou moins proche, les modems série accomplissent des progrès significatifs. Dans l'intervalle, les formes d'onde parallèles constituent un choix alternatif de faible complexité de mise en œuvre pour les débits moyens (jusqu'à 2400 bits/s), et une solution performante pour explorer les débits supérieurs, probablement jusqu'à 9600 bits/s.

F. BIBLIOGRAPHIE

[1] FORNEY

«Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference»

IEEE Transactions on Information Theory, mai 1972 pp 363-378

[2] UNGERBOECK

«Adaptive Maximum Likelihood Receiver for Carrier-Modulated Data-Transmission System»

IEEE Transactions on Communications, mai 1974 pp 624-636

[3] BELFIORE-PARK

«Decision Feedback Equalization»

Proceedings of the IEEE, août 1979 pp 1143-1156

[4] QURESHI

«Adaptive Equalization»

Proceedings of the IEEE, septembre 1985, pp 1349-1387

[5] SCHLEGEL-COSTELLO

«Bandwidth efficient coding for fading channels : code construction and performance analysis»

IEEE Journal on selected areas in Communications, décembre 1989, pp 1356-1368

[6] CCIR

«Simulateurs de voies ionosphériques sur ondes décamétriques»

rapport 549-2

[7] PIREZ-GOMBAULT

«Modem parallèle cohérent pour transmissions de données haut débit en HF»

GRETSI 1991 pp 457-460

CHANNEL EVALUATION FROM PREDICTED ZERO-CROSSING ANALYSIS

P. W. Piggin and M. Gallagher [+]

Communication and Signal Processing Research Consortium
Department of Electronic and Electrical Engineering
The University of Leeds
Leeds
LS2 9JT
UK

1. SUMMARY

This paper presents an investigation into Real-time Channel Evaluation (RTCE) from predicted zero-crossing analysis of: (i) a pilot-tone, and, (ii) Frequency Shift Keying (FSK) modem tones within a HF (High Frequency) radio system. This work extends the initial investigations from the early 1970s which used an analogue implementation analysing phase perturbation by time-differential phase comparisons of a low-level pilot-tone inserted in-band. The results to be presented in this paper compare theoretical analysis and measured results in Gaussian and flat fading conditions, and hence identify the close relationship between phase threshold, zero-crossing error rate and SNR (Signal-to-Noise Ratio). It will conclude with recommendations for extending the applicability of the technique to FSK (Frequency Shift Keying) systems and outline methods for incorporation into frequency management systems for various different radio systems operating in the HF, VHF and UHF bands.

2. INTRODUCTION

The utilisation of a low-level, co-channel pilot-tone as a diagnostic probe to monitor current channel state, can provide an insight into communication system performance. Specifically, both the amplitude and phase of the tone is affected by underlying propagation conditions and co-channel interference, with phase sensitivity providing the best estimates of prevailing conditions [1]. Consequently, analysis of phase perturbation by time-differential phase-comparison can be used to estimate the prevailing SNR.

An original analogue implementation [2], required long observation times (e.g. >200 seconds) to obtain a channel estimate for high SNRs. Latterly, Grayson [3] by concentrating on reducing the observation time, identified that, by using a Digital Signal Processing (DSP) based implementation, complete channel observations could be completed in 2.5 seconds. Work here with an enhanced DSP implementation optimises observation time still further to produce meaningful channel quality estimations in the region of 750 ms.

Application of this research is focused on effective spectrum management, since efficient system operation depends upon the ability to react to a range of parameters that characterise both propagation and channel noise. The use of RTCE in an automated communication system can reduce the use of high radiated powers by selecting frequencies and transmission times for which the received SNR is maximised and interference avoided. By effectively incorporating appropriate RTCE techniques, significant improvement in system performance can be achieved, as, in many instances, interference caused by congestion is the limiting factor in a communications system performance, rather than the propagation conditions.

3. RTCE BY PREDICTED ZERO-CROSSING ANALYSIS FROM PILOT-TONES

Pilot-tone RTCE uses a low level tone placed in-band close to frequencies of interest. This pilot-tone is then analysed to consider aspects such as instantaneous phase, amplitude, and other properties that are affected by the noise and channel perturbations. In this application the pilot-tone is used to provide channel quality information to determine the required transmission SNR at a given instant in time. Although both the amplitude and phase of a pilot-tone are affected by propagation conditions and interference, it is the phase that most directly represents operational conditions [4]. Consequently, analysis of phase perturbation by time-differential phase-comparison is most beneficial for providing channel performance information.

In a digital system, the incoming baseband signal is sampled and filtered to extract the pilot-tone using a highly-tuned narrow-band FIR (Finite Impulse Response) filter. The occurrence of a zero-crossing is predicted by adding a constant value of ($n/2$) onto the previous detected crossing, (n being an integer and τ representing the period of the pilot-tone). The timing of the predicted crossing point is used to calculate the phase error of the received tone and, if this is greater than a chosen threshold, a failure is recorded. The rate of phase comparison failure is then used to infer channel quality. This system does not require any form of decoding or demodulation to determine the channel quality and in this respect is *non-parametric* system; in addition to this no synchronisation of the zero-crossing of the pilot-tone is required.

[+] Both authors are formerly from the Department of Electronic Engineering, University of Hull, Hull. UK.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

$$P_e|_{SS+N} = 1 - \int_{-\theta_{Th}}^{\theta_{Th}} \int_0^{\infty} \frac{1-k_0^2}{\pi} \cdot \frac{y}{(y^2 - 2k_0 y\cos\theta_\tau + 1)^2}$$
$$\left\{1 + \left\{\tfrac{1}{2}\left(\frac{A_0}{\sigma_N}\right)^2 \frac{(1-k_0)}{(1+k_0)} \cdot \frac{y^2 + 2y\cos\theta_\tau + 1}{y^2 + 2yk_0\cos\theta_\tau + 1}\right\}\right\} \qquad (1)$$
$$\exp\left[-\frac{1}{1+k_0}\left(\frac{A_0}{\sigma_N}\right)^2 \left\{1 - \tfrac{1}{2}(1-k_0)\frac{y^2 + 2y\cos\theta_\tau + 1}{y^2 + 2yk_0\cos\theta_\tau + 1}\right\}\right] dy d\theta_\tau$$

A pilot-tone subject to additive Gaussian noise was used in initial tests. The theoretical results were derived from equation (3) in [1], given in equation (1). $P_e|_{SS+N}$ is the probability of zero-crossing error for a signal in Gaussian-distributed noise, θ_{Th} is the pilot-tone phase threshold, σ_N RMS noise amplitude at the output of the RTCE filter (digital implementation), A_0 is the signal amplitude (normalised), and k_0 is the envelope of the autocovariance of the output noise of the RTCE filter. k_0 is given by

$$k_0 = (\rho_0^2 + \lambda_0^2)^{\frac{1}{2}} \qquad (2)$$

ρ_0 and λ_0 are given by

$$\rho_0 = \frac{1}{\sigma_N^2} \int_0^{\infty} N(f)\cos(\omega - \omega_0)\tau df \qquad (3)$$

and

$$\lambda_0 = -\frac{1}{\sigma_N^2} \int_0^{\infty} N(f)\sin(\omega - \omega_0)\tau df \qquad (4)$$

where τ is the comparison interval for zero-crossing (correlation interval), ω_0 is the angular centre frequency of the RTCE filter and $N(f)$ is the spectral density function of the noise at the output of the RTCE filter.

For a 200 Hz tone with period (t) of 5 ms the comparison period was 2.5 ms ($\frac{1}{2}$). The threshold setting θ_{Th} was set to multiples of the maximum resolution of $\pi/35$ i.e. $\pi/35$, $2\pi/35$ and $4\pi/35$. As expected, for a smaller phase threshold, the probability of error $P_e|_{SS+N}$ was higher than for a larger threshold, for any given SNR.

In calculating the theoretical probabilities of error, it can be seen from equations (1) to (4) that the phase error rate is not only dependent on the SNR, but also on the autocovariance of the noise k_0, which, in turn is dependent on the phase-comparison interval τ and the threshold setting θ_{Th}. The DSP based digital pilot-tone filter design produced an Equivalent Noise Bandwidth (ENB) of 174 Hz., larger than the original analogue implementation. The value of the ENB from [5] was used to determine σ_N with numerical integration of (5).

$$\sigma_N = \sqrt{\int_0^{\infty} |N(f)^2| df} \qquad (5)$$

Hence equation (1) was evaluated by substitution.

3.1. Concluding Remarks

Practical results were compared to the theoretical, as shown in Figure 1. As can be seen the results for both theory and experiment are in close agreement at all threshold values. It was noted that at high values of SNR, where the value of $P_e|_{SS+N}$ was very small, the measured results had a tendency to diverge from the theory. The probable reason for this phenomenon is that the number of comparisons possible at high SNRs is not statistically significant when compared to the probability of error expressed in the theory.

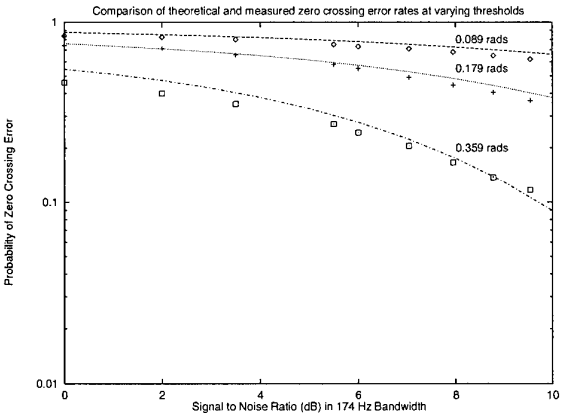


Figure 1 Comparison of theoretical and measured results for zero-crossing error rates for a pilot-tone at various threshold values under Gaussian conditions. (The lines are theoretical values and the points the measured results).

3.2. Optimisation of the Required Number of Phase Comparisons

Initially the number of phase comparisons carried out were grouped and averaged over a nominal large amount. These *groups* of phase comparisons were then averaged to give a measure of the probability pilot-tone phase error. The number of groups averaged was varied as a function of the SNR (averaging a high number of groups of comparisons in high SNR conditions), thus producing a statistically valid result at low probabilities of error. As Figure 2 shows, for a range of SNR values, the probability of pilot-tone phase error converges as the number of averaged groups of phase comparisons increases.

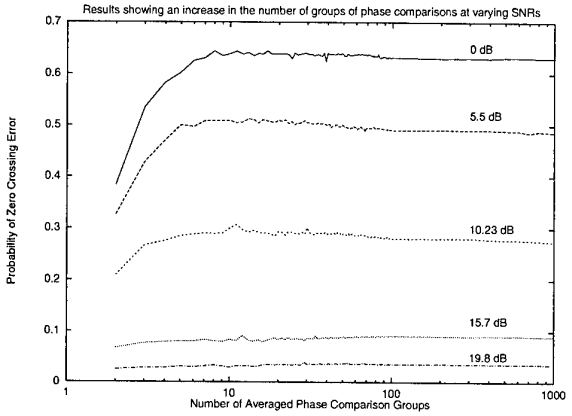


Figure 2 The probability of pilot-tone phase error becomes convergent with an increase in the number of averaged groups of comparisons increases. These results represent a phase threshold of 0.179 rads.

An investigation was carried into the actual number of phase comparisons required for optimal channel evaluation. In order to determine a reliable estimate of $P_e|_{SS+N}$ the results in Figure 3 were produced. Here a number of repeated trials at a constant SNR value are superimposed to show how when the averaged number of comparisons is increased the resultant value of $P_e|_{SS+N}$ becomes convergent to its final value. Results here clearly show that at low averaged comparison totals the estimate of $P_e|_{SS+N}$ can be very inaccurate as none or very little averaging takes place; however, by 300 phase comparisons $P_e|_{SS+N}$ is seen to be approaching its final value. This represents an observation time for RTCE in the order of 750 ms (300 phase comparisons at 2.5 ms duration each).

4. RTCE BY PREDICTED ZERO-CROSSING ANALYSIS FSK TONES FROM A HF MODEM IN A NON-FADING CHANNEL

Often a system requiring Frequency Management has no provision to use a pilot-tone RTCE system due to a fixed implementation. A limited power budget at mobile terminals

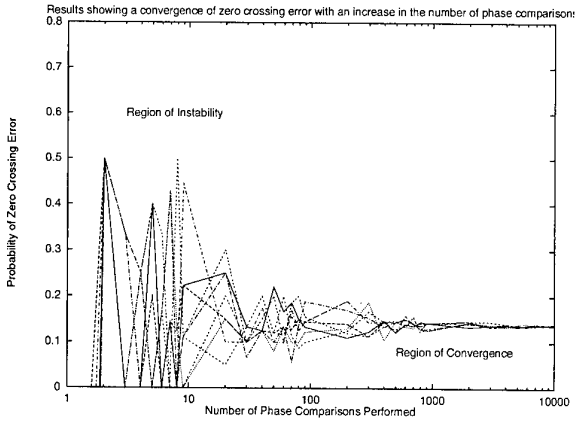


Figure 3 A family of results showing increasing number of comparisons, producing convergent results. These results represent a phase threshold of 0.179 rads at a SNR of 14.8 dB.

may preclude the expenditure of any additional RF power for such transmission. From a security aspect it would not be prudent to use a *fingerprint* transmission such as a pilot-tone to inadvertently indicate presence and location of transmissions. Thus the modulation tones used for data transmission should ideally replace the pilot-tone for RTCE purposes. Given below are the rules which govern the requirements for the modem tones to be successfully used in zero-crossing analysis [6].

A MFSK system with carefully chosen M tone frequencies and symbol rate B may allow the RTCE technique to be used without the need for a pilot-tone. Specifically, the modulated output must be amplitude continuous (symbol transitions occurring at zero-crossing points) and the M transmitted frequencies must satisfy the constraint;

$$f_i = mt_c \qquad i = 1,2,3,\ldots,M \tag{6}$$

where m is an integer and t_c is the time between comparisons. Similarly, the symbol rate, B, must satisfy the condition

$$B = pt_c \tag{7}$$

where p is also an integer. Finally, the pilot-tone digital filter must be designed such that it passes the MFSK tones. Instead of a phase threshold, the implementation would now accept a value of timing tolerance within which the zero-crossing should occur. This tolerance is required since other zero-crossings occur that are not used in the analysis. Clearly, it would be possible to derive RTCE information from the phase deviation of the information-bearing tones, making the pilot-tone redundant.

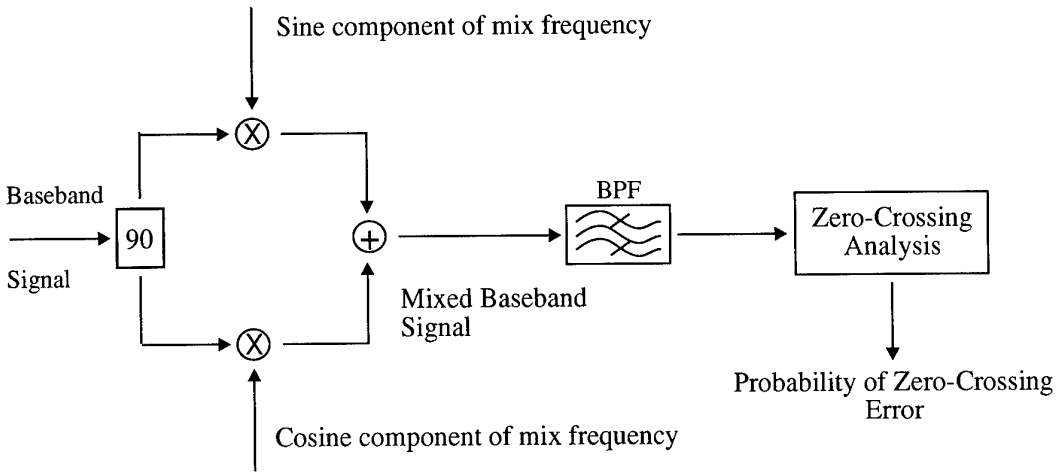


Figure 4 Block diagram of DSP software

Unfortunately, it is not always possible to choose the modem specification to be used, with this application an amount of signal processing was required prior to zero-crossing analysis being performed. This necessitated the use of digital filtering techniques. Here an image reject filters [7], producing a quadrature phase shift (by means of a 93-tap Hilbert filter) of the input signal and a subsequent mix with sine and cosine components at a chosen mix frequency, resulted after Bandpass Filtering in modem tones mixed to a frequency that allowed equations (6) and (7) to be satisfied. A block diagram is given in Figure 4, showing the extraction of tones from an FSK modem used in trials.

Theoretical results were developed from the modified original equation for the phase-error rate of a pilot-tone accompanied by Gaussian-distributed noise given by equation (1). In order to adapt the equations given to determine the zero-crossing error rate for FSK modem tones with respect to that of a pilot-tone, it was necessary to consider each tone separately, and in the limit as equi-probable. It was assumed that with a short correlation interval (τ), the FSK waveform was continuous over the period of zero-crossing analysis. Once the individual probabilities of error had been calculated, the mean of the two error rates gave the final value, such that:

$$P_e|_{SS+N} = \frac{P_e|_{SS+N_{tone1}} + P_e|_{SS+N_{tone2}}}{2} \qquad (8)$$

where

$P_e|_{SS+N_{tone1}}$ Probability of zero-crossing error for one of the modem frequency tones

$P_e|_{SS+N_{tone2}}$ Probability of zero-crossing error for the other modem frequency tone

The phase threshold or *window* (θ_{Th}) was based upon the number of sample points between zero-crossings of the sampled modem tones. With θ_{Th} being an angular representation, here assumed to be taken around the zero-crossing point; it is given in terms of n, an integer describing the digital representation of the threshold in terms of the number of samples. These two entities have the following relationship:

$$\theta_{Th} = \frac{(2n-1)\pi}{29} \qquad (9)$$

Results presented here use a value of n ranging from 1 to 3 giving values for θ_{Th} of $\pi/29$, $3\pi/29$, and $5\pi/29$ respectively for this example.

Expression (8) was evaluated numerically. Figure 5 compares theoretical values of probability of zero-crossing phase error for the modem tones with measured results from simulated trials. The results are independent of baud rate, as described by the theory, as τ (correlation interval) is constant throughout; this is shown in practice since varying the baud rate of the modem has no effect on the probability of zero-crossing error; however, very slow rate modems would produce some practical differences.

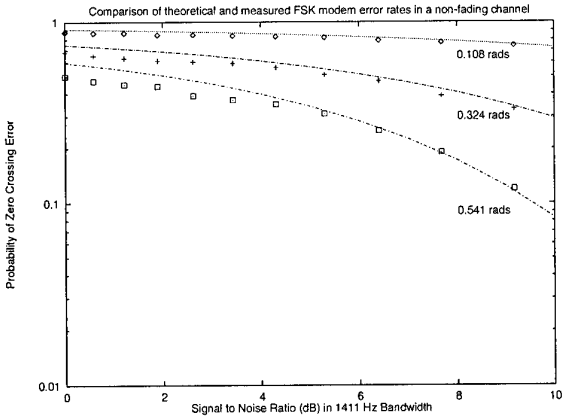


Figure 5 Comparison of theoretical and measured results for zero-crossing error rates for FSK modem tones at various threshold values under Gaussian conditions (The lines are theoretical values and the points the measured results).

4.1. Concluding Remarks

As can be seen from Figure 5 at larger window sizes and at lower SNR values the measured values of probability of zero-crossing error were in disagreement with that given by the theory. This is explained by the large valid window size (θ_{Th}) recording a lower error rate for the high noise content signal. The software implementation is not discriminating enough at such wide window values to determine exactly where subsequent valid zero-crossings should occur; hence, invalid crossings are considered as valid. In conclusion, it is clear from the graph that the small window size results are far better correlated with theory, and so should be used in any application.

5. RTCE BY PREDICTED ZERO-CROSSING ANALYSIS OF FSK TONES FROM A HF MODEM IN A RAYLEIGH FADING CHANNEL

Presented here are theoretical and measured results for zero-crossing analysis in a fading environment. The theoretical results are based upon equation (1): the expression for the probability of zero-crossing error for a steady non-fading channel ($P_e|_{SS+N}$).

For such a fading channel, the first order pdf (probability density function) of the Rayleigh fading signal is given by

$$\rho(A_0)_{RS} = \frac{A_0}{2\sigma_s^{\,2}} \exp\left(-\frac{A_0^{\,2}}{2\sigma_s^{\,2}}\right) \quad (10)$$

where σ_s is the RMS signal amplitude and A_0 is the instantaneous signal amplitude. The subscript *RS* signifies a Rayleigh fading signal.

For slow fading as defined above, the signal may be assumed constant for the duration of the sampling interval τ. The phase error probability for the fading signal plus noise is therefore given by:

$$P_e|_{RS+N} = \int_0^\infty \rho(A_0)_{RS} P_e|_{SS+N} dA_0 \quad (11)$$

i.e. the steady signal phase error for a given signal level weighted with the probability of that signal level occurring. From this, expression (12) can be derived from (11).

In the same way as described previously, equation (12) was adapted to give the probability of zero-crossing error for FSK modem tones with respect to that of a single pilot-tone.

Figure 6 gives a comparison of the theoretical and measured results. The measured results were produced over a simulated HF channel, using a DSP based HF simulator. The simulator can represent 3 independently fading skywave paths and one ground wave path, with up to 25 ms path delay on each skywave path. Additionally, the fade rate and path loss attenuation can be defined for each skywave. The simulator has a Rayleigh fading profile which is described by the Watterson fading model [8].

The simulator parameters stated above can be defined in such a way that a model of *good* and *poor* fading (as detailed by CCIR) can be produced. Using 2 skywave paths only, with a path loss attenuation of 3 dB, a *good* fading channel can be represented with a fade rate of 0.5 Hz and 0.5 Hz with a path delay of 3.5 ms and 3.0 ms respectively, for each skywave. A *poor* fading channel can be represented with a fade rate of 1.0 Hz and 1.0 Hz with a path delay of 5.0 ms and 3.0 ms respectively, for each skywave.

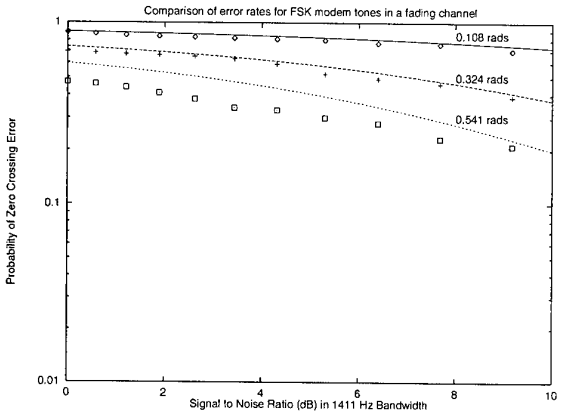


Figure 6 Comparison of measured and theoretical results for probability of zero-crossing analysis in a Rayleigh fading channel. (The lines are theoretical values and the points the measured results).

5.1. Concluding Remarks

It is again clear that correlation between the theoretical values and the measured values is stronger at smaller window sizes (θ_{Th}). As this value increases, it is noted that the correlation becomes more tenuous. The reason for this is the same as in the similar results for a steady signal environment, i.e. with a large valid window size the recorded error rate for the high noise content signal is much lower. The software implementation is not discriminating enough at such wide window values to determine exactly where subsequent valid zero-crossings should occur; hence, invalid crossings are considered as valid. Again, in conclusion it is clear from the graph that small window sizes result in far better correlation with theory, and so should be used in applications. This problem is inherent within the design and cannot readily be compensated.

6. APPLICATION OF ZERO-CROSSING RTCE WITHIN A FREQUENCY MANAGEMENT SYSTEM

With the technique described above, RTCE from predicted zero-crossing analysis has the ability to produce an SNR estimate on a given propagation path. This can be used in parallel with other frequency management techniques such as propagation prediction programs to enhance the predicted SNR estimate. RTCE from predicted zero-crossing analysis

$$P_e|_{RS+N} = \int_0^{\infty} \frac{A_0}{\sigma_s^{\,2}} \exp\left(-\frac{A_0^{\,2}}{2\sigma_s^{\,2}}\right) \left(\frac{1-k_0^{\,2}}{\pi} \int_{-\theta_{Th}}^{\theta_{Th}} \int_0^{\infty} \frac{y}{(y^2 - 2k_0 y\cos\theta_\tau + 1)^2} \cdot\left\{1+\left\{\tfrac{1}{2}\left(\frac{A_0}{\sigma_N}\right)^2 \frac{(1-k_0)}{(1+k_0)} \cdot \frac{y^2+2y\cos\theta_\tau+1}{y^2+2yk_0\cos\theta_\tau+1}\right\}\right\} \exp\left[-\frac{1}{1+k_0}\left(\frac{A_0}{\sigma_N}\right)^2\left\{1-\tfrac{1}{2}(1-k_0)\frac{y^2+2y\cos\theta_\tau+1}{y^2-2yk_0\cos\theta_\tau+1}\right\}\right] dy\,d\theta_\tau\right) dA_0 \quad (12)$$

has been studied extensively, and it is clear that a measure of the probability of zero-crossing error can infer channel quality and provide an SNR estimate for a given modulation scheme. Within a communications network, modem traffic can be used to yield such estimates, and to build an image of the SNR conditions within the bounds of the network. This can all be done in a non-parametric fashion with no demodulation required, thus allowing the frequency management algorithm to determine if communication is possible with the prevailing SNR conditions, and, if not to modify operation accordingly.

In addition to the specified network modem traffic, effective *passive assessment* of RTTY and FAX stations is possible using non-parametric differential phase comparisons. With a known modulation type, RTCE from predicted zero-crossing analysis can be undertaken to provide a more comprehensive SNR estimate in the area of interest.

6.1. Relationship between FSK phase analysis and data channel performance

In order to relate results shown above to channel performance metrics such as BER (Bit Error Rate) measurements, knowledge of the correspondence between probability of phase error and SNR is insufficient. What is required is a relationship between the phase error analysis and the error rate of the modulation scheme being used. This can be done for both fading and non-fading channel environments. Figure 7, below, suggests a method to relate the two entities with a conversion included. The relationship between SNR and E_b/N_0 (signal energy per element to noise per unit bandwidth ratio) is given in a form by expression (13) [9].

$$E_b/N_o\,(dB) = S/N\,(dB) + 10\log\frac{ENB}{f_d} \quad (13)$$

where ENB is the Equivalent Noise Bandwidth of the digital RTCE filter and f_d is the data transmission rate per second.

Considering Figure 7, analytically the following can be deduced:

$$\text{SNR} \cong c(P_x) \quad (14)$$

where $c(\,)$ is a function described by the left-hand error rate curve in Figure 7, and P_x is the probability of zero-crossing error. Hence

$$P_e = e\left(d\left(c(P_x)\right)\right) \quad (15)$$

where P_e is the estimated value of the probability of bit error (BER), $d(\,)$ is a function described by the right hand error rate curve in Figure 7 and $e(\,)$ represents a conversion from SNR to E_b/N_o. $c(\,)$ and $d(\,)$ can be represented by look-up-tables in a computer system.

The concept indicated in Figure 7 enables zero-crossing analysis to produce a BER estimate range for known traffic within the bounds of *good* and *poor* fading channels. This BER information can then be used as frequency management information to enhance the circuit performance.

7. CONCLUSIONS

This paper demonstrates that RTCE from predicted zero-crossing analysis can be performed without the need for an inserted pilot-tone co-located with the information-bearing modem tones. The advantages of this are numerous: a reduction in system complexity and reduced power consumption within the transmitting system. Applications using the zero-crossing analysis RTCE technique would be best served by using a narrow window size (θ_{Th}), especially at low SNR values. RTCE from predicted zero-crossing analysis is an important non-parametric channel quality and evaluation technique, since it is possible to infer system performance without demodulation, synchronisation, error control or decryption.

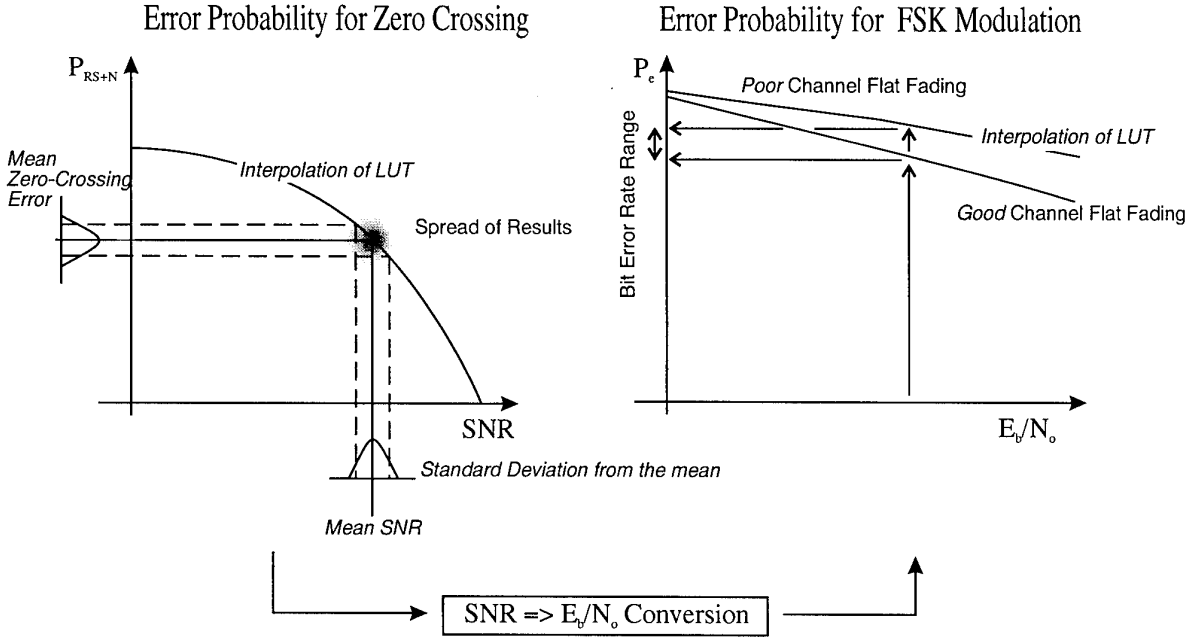


Figure 7 The generation of FSK probability of error values from the probability of zero-crossing error

The use of optimisation techniques to increase the speed of channel evaluations is over 50% faster than the previous DSP implementation. The assumption made here, with an implementation with a decreased observation time, is that the time to perform a channel evaluation is not significantly less than a fading interval. When considering HF system fading the principle periods of interest are from less than one to ten seconds. This deals with ionospheric motion, concerned with small-scale inhomogeneities associated with micromultipath and multihop/multilayer multipath [10]. Nulls due to such fading effects may effect channel evaluation. However, it is not believed fading of this nature would adversely limit system performance, as fading periods and the observation times for channel evaluations are comparable.

This research can be extended using in-band and out-of-band digital filters for constant envelope signalling methods e.g. MFSK. It is believed that this research has merits for a broad range of radio systems and forms an important element in the spectrum management process.

8. REFERENCES

[1] Betts, J.A., et al, "C.W. sounding and its use for control of HF (3-30 MHz) adaptive systems for data transmission", Proc. IEE, Vol. 117, 1970, pp 2209-2215.

[2] Betts, J.A., Darnell, M., "Real-time HF channel estimation by phase measurement on a low-level pilot-tones", AGARD CP 173, 'Radio systems and the Ionosphere', Paper 18, Athens, 1975.

[3] Grayson, M., "Real-time channel evaluation derived from predicted zero-crossing", Proc. of 3rd Bangor Symposium on Communications, May 1991, pp 186-189.

[4] David, F., Franco, A.G., Sherman, H., and Shucavage, L.B., "Correlation measurements on an H.F. transmission link", IEEE trans., 1969, COM-17, pp. 245-256.

[5] Stremler, F. G., "Introduction to Communication Systems", Third Edition, Addison Wesley 1990 (ISBN 0-201-51651-9), Chapter 4, pp 163-218.

[6] Grayson, M., "Improving Synchronisation Techniques for Time-Varying Dispersive Radio Channels", Chapter 4, Thesis Ph.D., University of Hull, UK., October 1993.

[7] Hickman, I., "Newnes Practical RF Handbook", Newnes 1993 (ISBN 0-7506-0871-4), Chapter 10, pp 160-176.

[8] Watterson, C. C., Juroshek, J. R., and Bensema, W. D., "Experimental Confirmation of an HF Channel Model" IEEE trans. Commun. Technol., vol. COM-18, pp. 792-803, Dec. 1970.

[9] Maslin, N., "HF Communications: A system Approach", Pitman 1987 (ISBN 0-273-02675-5), Chapter 9, pp 169-209.

[10] Goodman, J. M., "HF Communications: Science and Technology", Van Nostrand Reinhold 1991 (ISBN 0-442-00145-2), Chapter 4, pp 268-269.

9. ACKNOWLEDGMENTS

This work is jointly supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and EDS Defence Ltd, UK.

DISCUSSION

Discussor's name: F. Davarian

Question/Comment:

The theoretical treatment of the method is based on Gaussian noise interference, but the proposed use of the system for frequency management would imply an environment with interfering transmissions. Could you comment on this?

Author/Presenter's reply:

Narrow band transmissions could certainly be a problem with the pilot-tone scheme, since the transmission bandwidth is much wider than that of the pilot-tone. However, the pilot-tone analysis was an initial investigation into the zero-crossing method. The problem is much less severe for the FSK tones, since the analysis is performed for the whole bandwidth of the FSK transmissions. It must be admitted that the error analysis method would not be strictly applicable in the case of in-band interferers rather than Gaussian noise, but the error analysis would still show a high error rate and thus indicate an unsuitable transmission frequency.

New Technologies to Improve HF Modems Performance for Standardized Waveforms

Achim Brakemeier
Daimler-Benz AG, Research and Technology
P.O.Box 2360, D-89013 Ulm
Germany

Andre Kotlowski
Daimler-Benz Aerospace AG, Sensor Systems
Wörthstraße 85, D-89077 Ulm
Germany

1. Abstract

For military short wave communication systems two waveforms have been standardized -- STANAG 4285 and MIL-STD-188-110A. For these waveforms modems are available on the market.

In this paper design rules that have led to these waveforms are recalled briefly and compared with modern signal design techniques.

Knowing the shortcomings of standardized waveforms, the incorporated redundancy has been intensively used for advanced modem design. These new technologies have proved great performance enhancements in Daimler-Benz Aerospace new HF modem Echotel ETM1810/M for STANAG 4285 and MIL-STD-188-110A waveforms. Improvements have been achieved mainly in tolerance to Doppler and Delay Spread.

Results are presented in this paper

2. Standardized Waveforms

For digital short wave communication systems with high data rates the modulation has been chosen to be single carrier modulation based on phase shift keying. This choice is a good one (ref. 5) because of spectral efficiency (1...3 bit/Hz), power efficiency (relatively small crest factor) and complexity. The length of the channel impulse response is about 12 symbols, given a maximum delay spread of 5 ms and a symbol rate of 2400 symbols/s. Therefore the complexity is moderate and can be handled by standard signal processors.

The standardized waveforms that we will discuss have a common structure. Transmission begins with a preamble followed by a sequence of frames. Each frame consists of known sequences (probes) and unknown sequences (data) as shown in Fig. 1.

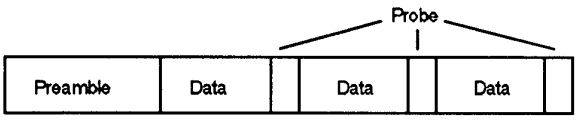


Fig. 1 frame structure

The length of the preamble, the length of the data sequences and the length of the probe sequences are specified in the standards.

In STANAG 4285 (abbr. STANAG, ref. 4) the preamble has a length of 80 symbols, the data sequences have a length of 32 and the probe sequences normally have a length of 16 (short probe). But every fourth probe has a length of 80 symbols (long probe) and is identical with the preamble.

In MIL-STD-188-110A (abbr. MILSTD, ref. 3) the preamble has a length of 1440 symbols (short preamble) or 11520 symbols (long preamble) and probe sequences and data sequences have a length of 20 or with high data rate (2400, 4800 bps) probe sequences have a length of 16 and data sequences have a length of 32 symbols.

Although both waveforms offer the same usable bit rates from 75 bps (coded) up to 3600 bps (STANAG uncoded) or 4800 bps (MILSTD uncoded) the fields of applications of these waveforms are quite different.

The STANAG is designed for robust transmissions having a short start up phase. During long data transmissions reentering is possible because the short preamble is periodically repeated (as every 4th probe). In addition frequency hopping is possible. The MILSTD is suitable only for long data transmissions, otherwise the very long preamble (compared with the STANAG preamble) reduces dramatically the usable data rate.

3. Channel Model

The short wave communication channel has a bandwidth of 3 kHz. It is a channel with time variant frequency selective fading. It can be modelled with the wide sense stationary uncorrelated scattering channel model (WSSUS, ref. 2). Channel simulators for short wave are normally of the tapped delay line type. Discrete signal pathes are modelled each having an individual delay spread and Doppler spread. A single path causes Rayleigh fading. For the results presented here we rely on four channel models:

- AWGN - channel: bit error rate (BER) for a single non-fading path versus gaussian noise. (STANAG Test 1)
- CCIR moderate: BER for a two path equal power Rayleigh fading channel versus gaussian noise. (Doppler spread 0.5 Hz, delay spread 1 ms, STANAG Test 4)
- Delay Channel: BER for a noiseless two path Rayleigh channel versus delay spread. (two sided Doppler spread 0.5 Hz, STANAG Test 7)
- Doppler Channel: BER for a noiseless two path equal power Rayleigh channel versus Doppler spread. (delay spread 1 ms, STANAG Test 6)

These channel models are described in the STANAG description, Appendix B, test configurations (ref. 4).

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

For this paper the STANAG Mode is tested without coding and without interleaving. The MILSTD Mode is tested with coding as described in the standards and short interleaving (0.6 s).

4. Classic Design Rules

In the past the waveforms were designed to make demodulation easy. The first task is preamble detection. To allow preamble detection in low noise environment the preamble should be as long as possible. Different steps in preamble detection were recognized: detect the signal, estimate the doppler offset, correct the doppler offset, do time synchronisation, do exact bit synchronisation, set up symbol rate sampler. With low memory and low signal processing capabilities these tasks could not be overlapped and therefore different segments in the preamble had to be used. This led to the very long preamble in MILSTD.

In STANAG it was recognized that a long preamble could degrade performance because the channel possibly had changed during preamble transmission. A shorter preamble was standardized just long enough to do one single task of the preamble detection, but now the signal was assumed to be stored in memory and all tasks could be done by simply reusing the stored signal.

After preamble detection the main difficulty was to cope with the time variant channel and with multipath spread. It was seen that the signal should be broken into independant signal elements (frames). This could be done by inserting probes having a length of the expected maximum length of the channel impulse response (CIR). The symbol rate was chosen to 2400 symbols/s to make a signal of 3 kHz bandwidth with a moderate roll off factor (0.25). Other systems at that time used e.g. 1800 symbols/s in a 2.7 kHz bandwidth making it even more harmless to do the spectral shaping. With a symbol rate of 2400 symbols/s the CIR had a length of 12 symbols and it seemed enough to insert probes of length 16 giving the modulation pulse a chance to decay properly.

Without adaptation the channel should not vary very much from probe to probe. Therefore the data sequences were chosen to have a length of 32 symbols. The hope was to restrict the length of an error burst. On the other hand with data directed equalisation the receiver complexity grew exponentially with the length of the data sequences and error propagation was more likely.

STANAG allows frequency hopping. If a frequency hop occurs the HF transmitter and the HF receiver have to switch the center frequency and a hop time is specified. It lies within the long probe, reducing the effective length of the long probe. A theoretical maximum hop rate of 9.375 hops/s can be handled because there are 9.375 long probes/s..

5. Modern Design Rules

The first task of the demodulation process is the preamble detection. Normally this is done with correlation techniques and requires a minimum total length of the preamble to achieve a given false alarm rate and rate of missing detection. But the channel varies and therefore the preamble must be segmented. The minimum segment length is of about twice the maximum length of the CIR. Better are three times the length of the CIR.

$$\text{LEN(PREAMBLE_SEGMENT)} > 2\ \text{LEN(CIR)} \qquad (1)$$

The preamble segments should not be much longer otherwise the correlation peak can decrease because of the frequency offset between transmitter and receiver center frequencies.

If the preamble length is too short for a given false alarm rate and rate of missing detection more than one preamble segment must be used. It is not necessary to have the preamble segments directly concatenated. It is possible to use preamble segments that are separated by data symbols. For example the STANAG 4.th probes form an appropiate set of preamble segments.

After preamble detection the demodulation process starts and it is known that an optimum demodulation process needs the knowledge of the CIR. This knowledge can be implicit (as in data directed equalisation) or explicit. A modern demodulation process uses explicit channel evaluation. When the CIR is known the equaliser type can be selected according to the measured properties of the channel.

It is evident that enough redundancy is built into the waveform to allow a near optimum channel estimation, although it is not strictly necessary. If the bit error rate on uncoded transmission is low enough it is possible to use the data symbols received in the past to do the channel evaluation. But this process can cause catastrophic error propagation.

A data independent demodulation process is possible if the probes have a minimum length of twice the length of the CIR:

$$\text{LEN(PROBE)} > 2\ \text{LEN(CIR)} \qquad (2)$$

The requirements (1) and (2) indicate that preamble and probes can be identical as it is for STANAG with every 4.th probe.

The requirement (2) is not fulfilled with STANAG and MILSTD and the short probes having length 16 respectively 20 symbols. These probes have not enough information to do a full data independent channel estimation. Theoretically only channels with a delay spread $T_D < 3$ ms can be handled independent from data. But in practice the restriction is on channels with delay spread $T_D < 2$ ms because otherwise the requirements for time synchronisation would become very hard and the sidelobes from the modulation pulse would be suppressed.

Generally the channel is not time invariant. Therefore channel evaluation and adaptation is an ongoing process during the whole data transmission. But the speed of channel variations is restricted. The channel models (that are verified by channel measurements) assume a Gaussian type of noise filter to produce the Rayleigh fading behaviour. These filters have an overall two sided bandwidth (3σ range) of about

$$F_{D_3\sigma} = 3\ (2\ \sigma) \qquad (3)$$

where $F_D = (2\ \sigma)$ is the two sided Doppler bandwidth given in the description of the channel models. From Shannon's sampling theorem it follows that only $F_{D_3\sigma}$ samples of the CIR per second are necessary. This gives the required number of probes per second:

$$\text{Number(PROBEs)/second} > F_{D_3\sigma} \qquad (4)$$

Therefore for a maximum Doppler spread of $2\ \sigma = 4$ Hz about 12 probes/s are needed.

STANAG has a probe rate of 37.5 probes/s (short probes), MILSTD has a probe rate of 50 (high datarate) or 60 probes/s. From a theoretical point of view these probes are too often and too short. It would have been a better design to use longer probes more seldom.

STANAG has 9.375 long probes per second, allowing a Doppler spread of ca 2 $\sigma \approx 3$ Hz without using the short probes.

6. Channel Evaluation

As stated above the most important property of a modem is the ability to evaluate and adapt the channel impulse response with high accuracy to achieve high quality. In principle only such probes are useful for channel evaluation that do not have spectral nulls or deep fades in their Fourier spectrums. If there are spectral nulls the spectrum is invertible and these parts of the CIR cannot be evaluated. Noise enhancement occurs if there are deep fades in the spectrum. Best performance for a given length of the CIR (measured in efficiency) requires a flat Fourier spectrum. This can be achieved if the probe is a CAZAC sequence (Constant Amplitude Zero Auto Correlation). CAZAC sequences of length L have normally symbols from the L-PSK symbol alphabet. Binary sequences (symbols from the 2-PSK symbol alphabet) with nearly as good properties as the CAZAC sequences are the PN sequences (Pseudo Noise) that have a flat spectrum with exception of the frequency line at f=0 Hz that has half the magnitude of the rest. If PN or CAZAC sequences are periodically repeated for at least one length of the CIR a simple correlation algorithm exists to evaluate the coefficients of the CIR.

In STANAG the long probes (and the preamble) consist of a periodically repeated PN-sequence of length 31. Therefore the CIR can be estimated with high accuracy. It proves to be a good choice for a probe sequence. With this sequence the CIR can be evaluated up to a length of 12.5 ms. If frequency hopping is used the effective length of the long probe is reduced. However, with a rest of 45 symbols there is enough information in the probe to do a data independant channel evaluation avoiding error propagation caused by wrong decisions.

The short probes in STANAG and in MILSTD only allow an data independant channel evaluation up to a length of the CIR of 3 ms. But the reduction in correlation length (by a factor of 2) gives a 3 dB loss in accuracy (error variance). Furthermore these short probes are not optimum. Partly they have deep fades (factor 10) giving an additional loss of 10 dB. In summary high noise enhancement occurs if these bad conditioned short probes are used for channel evaluation. MILSTD and STANAG do not have optimized short probes. Additionally the short probes vary from probe to probe. In STANAG there are 3 different short probes and a long probe, in MILSTD there are 4 or 10 different probes. This makes the demodulation process difficult without having a remarkable gain. The only effect of using different probes is a slightly better spectrum because the signal has weaker periodicies. But this better spectrum could have been achieved in the same way by choosing probes with good correlation properties.

7. New Technologies

Knowing the shortcomings of standardized waveforms, the incorporated redundancy has to be intensively used for advanced modem design. Some of the ideas are summarized below:

- Preamble Detection: Use of more than one long probe for preamble detection in STANAG.
- Channel Evaluation: In STANAG use the long probes for accurate channel estimations. Use the short probes for channels with short delay spread. In MILSTD avoid (do not use) bad conditioned short probes.
- Channel Adaptation: Average channel estimates from short probes give a better correlation gain and suppress noise. Interpolate the CIR estimates to avoid adaptation lag error.
- Equaliser: With a good CIR estimate all types of equaliser can easily be used. One good compromise to complexity is the decision feedback equaliser together with a channel matched filter. The equaliser coefficients are directly calculated from the CIR.
- Data Independance: Channel evaluation and therefore the adjustment of the equaliser coefficients are made independant on data symbol decisions for channels with small delay spread.
- Doppler Tracking: Based on CIR estimate.
- Sync Tracking: Based on CIR estimate.
- Modem Design: Demodulation algorithms are designed and tested via computer simulations.

A modern modem design is completely based on the CIR estimate. Therefore the CIR estimate should be as accurate as possible, having low noise and being valid at every time instant. Normally this requires adaptive channel evaluation. Setting up the equalizer's coefficients, tracking of frequency shifts and tracking of sampling rate can be done by adequately checking the CIR.

Once the CIR is known the choice of the type of the equalizer becomes secondary. Primarily the degree of intersymbol interference (ISI) has to be known that can be handled by a given type of equalizer. The length of an equaliser, i.e the maximum length of the CIR that can be handled, is responsible for the performance on the delay channel and it is the key factor for the complexity of the equaliser. A decision feedback equalizer nearly always is a good compromise of performance versus complexity.

8. Results

Basic tests for a communication system are bit error rates (BER) on the Additive White Gaussian Noise channel (AWGN) as shown in Fig. 2.

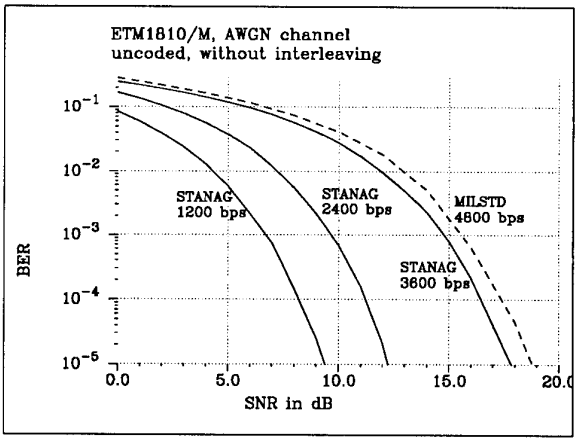


Fig. 2 SNR test, AWGN channel

These curves can be compared directly with the theoretically well known error curves for M-PSK transmission (ref. 1).

The tolerance to delay spread of the channel is tested by using the Delay Channel from chap. 3, i.e. a two path Rayleigh channel with increasing delay. The BER for the ETM1810/M modem was measured over a time of 56 min at 600 bps, 28 min at 1200 bps and 14 min at 2400 bps. The results are shown in Fig. 3 and Fig. 4

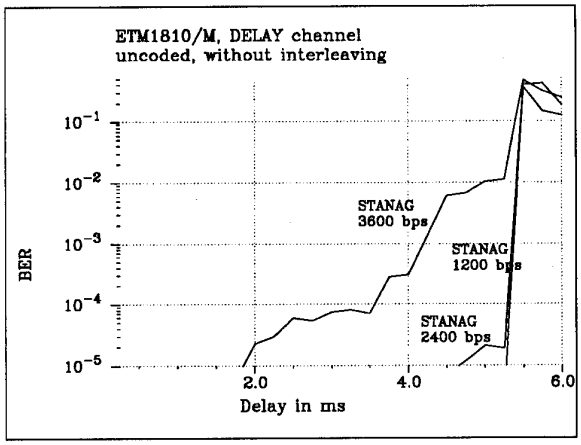


Fig. 3 Delay test, 2-path Rayleigh channel, STANAG

The design goal for the ETM1810/M modem was a maximum delay of 5 ms. It can be seen from the performance results that this design goal has been achieved very good for STANAG up to a data rate of 2400 bps with *uncoded* transmission. The uncoded 8-PSK transmission in STANAG provides with a BER < 10^{-4} up to a delay spread of 3.5 ms. For a 5 ms delay spread the BER is about 10^{-2}.

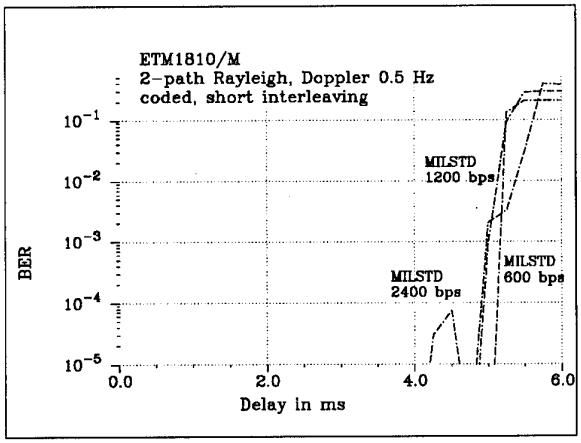


Fig. 4 Delay test, 2-path Rayleigh channel, MILSTD

The MILSTD uses coded transmission (rate 1/2 for 600 and 1200 bps and rate 2/3 for 2400 bps). The measured BER on the Delay channel is nearly zero within the delay range. Only for 2400 bps some errors occur at a delay spread of 4.25 ms and 4.5 ms. For MILSTD the design goal is nearly perfectly matched.

The 8-PSK transmission is less robust. The reason why STANAG and MILSTD do not show that good performance with 8-PSK are the waveforms themselves. The 8-PSK symbols have a shorter euklidean distance compared to 2- or 4-PSK. Therefore the channel impulse response estimate has to be better. However, the probes in the MILSTD 2400 bps mode are shorter (16 symbols) compared to the other modes (20 symbols) therefore the CIR estimate becomes worse. The short probes can only be used effectively if they are lenghtened by the surrounding data symbols. This introduces data dependance in the CIR estimate. In 8-PSK mode the data decisions are more uncertain and errors degrade the CIR estimate. In STANAG sometimes the short probes are used for channel estimation in situations when the channel varies very quick and the same effect occurs as in MILSTD..

The tolerance to Doppler spread is tested using the Doppler channel from chap. 3. The measured BER is shown in Fig. 5.

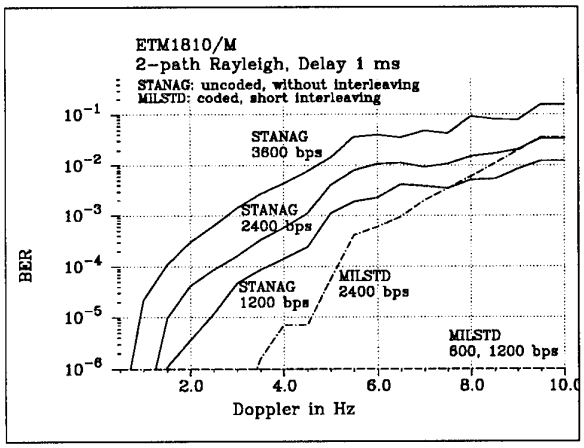


Fig. 5 Doppler test, 2-path Rayleigh channel

Fig. 5 contains performance results for the Doppler channel. In uncoded STANAG mode Doppler spreads up to 5 Hz with 1200 bps, 4.5 Hz with 2400 bps and 3 Hz with 3600 bps at a BER level of BER < 10^{-3} can be handled. This is about twice the Doppler spread that can be handled by conventional methods.

Coded transmissions have even higher tolerance to Doppler spread as shown with the MILSTD. The 2400 bps transmission with MILSTD can handle Doppler spreads up to 6.5 Hz at a BER level of BER < 10^{-3}. At this rate 8-PSK modulation is used together with a rate 2/3 coding. With 2-PSK or 4-PSK modulation and rate 1/2 coding the usable bitrate is 600 bps or 1200 bps. The tolerance to Doppler spread is much better than for 8-PSK. Now the MILSTD implemented in the Daimler-Benz Aerospace's ETM1810/M HF modem yields no errors during the measuring time (56 min at 600 bps) up to a Doppler spread of 10 Hz as shown in Fig. 5. Extended range measurements (not included in Fig. 5) have shown an errorfree behaviour up to 13 Hz Doppler spread for MILSTD 600 bps and MILSTD 1200 bps. For 16 Hz Doppler spread the measured BER was less than 10^{-3}.

In chapter 5. it was stated that the tolerable Doppler spread depends on the number of probes per second. In MILSTD 600 bps and 1200 bps are 60 probes per second and from (4) the theoretical maximum tolerable Doppler spread is about 20 Hz. With the 16 Hz tolerance to Doppler spread (at a 10^{-3} BER) the ETM1810/M modem is nearly optimum.

An additional effect influences the performace channels with high Dopper spread. The convolutional coding used in MILSTD is optimized for uniformly distributed errors. Normally on channels with flat fading characteristics -- and the two path Rayleigh fading channels belongs to this category -- the errors occur in bursts. An interleaver is used to spread the error bursts over time, however, it requires an interleaver length that is long compared to the mean time between error

bursts. This time is inverse proportional to the Doppler spread. The 0.6 s interleaver used in the MILSTD short interleaving is long enough for Doppler spreads above 5 Hz, but is too short for Doppler spreads below 1 Hz. On Doppler spread above 5 Hz the coding/interleaver combination works effectively and an uncoded BER below 10^{-2} is reduced below 10^{-6}. The STANAG has an uncoded BER < 10^{-2} for Doppler spread < 6 Hz (2400 bps) and < 9 Hz (1200 bps) as can be seen from Fig. 5. Within this range of Doppler spread the STANAG used with coding performs as well as MILSTD.

The combination of tolarance to delay spread and tolerance to Doppler spread is tested using the CCIR moderate channel. For a fixed delay spread of 1 ms and a fixed Doppler spread of 0.5 Hz the BER is measured versus SNR. The results for the ETM1810/M modem are shown in Fig. 6.

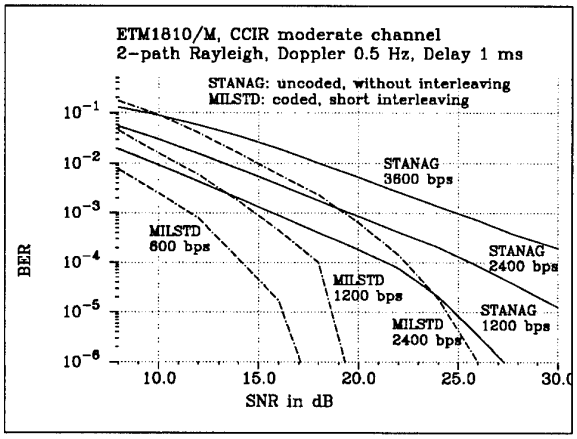


Fig. 6 SNR test, two path Rayleigh

Uncoded transmission with STANAG show BER curves that drop slowly towards zero. The curves behave like in a flat fading channel. The CCIR moderate channel is a relatively slow fading channel, error bursts (if they occur) are very long and in summary only a moderate average BER can be expected. The mean time between errors can be very long. With coded transmission (MILSTD) the BER curves drop faster.

9. Conclusion

In this paper the design rules have been recalled briefly that have led to the standards STANAG 4285 and MILST-188-110A. The focus was on the probes that are inserted in standardized waveforms to allow receiver set up.

The preamble in MILSTD is too long and inefficient, preamble detection can be made with very much shorter preambles as demonstrated by STANAG.

The probes in MILSTD and the short probes in STANAG are too short to do an data independant demodulation on channels with a long channel impulse response. However, for channels with a short CIR these probes can be used very effectively to build a modem with very high tolerance to Doppler spread. This was proved by measurements with the ETM1810/M modem.

The tolerance to delay spread mainly depends on the designed maximum length of the CIR and on the quality of an channel evaluation process. The ETM1810/M modem works nearly errorfree on two path Rayleigh channels up to the designed delay spread. Only for 8-PSK modulation the shortcomings of the short probes in the standardized waveforms cause some error bursts.

The performance on the additive white Gaussian noise channel is as expected from theory.

10. References

1. Proakis, J.G.: Digital Communications, 2.nd ed. McGraw-Hill series in electrical engineering, 1989.

2. Bello, P.A.: Characterization of Randomly Time Variant Linear Channels. IEEE Trans. Commun. Syst, Vol. CS-11, pp 360-93, Dec. 1963.

3. MIL-STD-188-110A. Military Standard, Interoperability and Performance Standards for Data Modems. Department of Defense, USA 1991.

4. STANAG 4285. Characteristics of 1200/2400/3600 Bits per Second Single Tone Modulators/Demodulators for HF Radio Links. North Atlantic Treaty Organization, Brussels 1989.

5. Brakemeier, A. Criterias to Select Proper Modulation Schemes. Nordic HF Conference HF95, Nordic Radio Society 1995.

DISCUSSION

Discussor's name: D. Yavuz

Question/Comment:

1. As you know narrow-band inteference excision is not an interoperability factor, therefore not in standards. However, excision is a very important performance factor. Since the preamble is used also for excision, the trade-offs are quite complex. Perhaps you would like to comment?

2. Does the ETM 1810 modem exist as a product, and how much does it cost?

Author/Presenter's Reply:

1. I agree. Narrow-band interference excision is an important factor, but it was not of interest in this paper. The reason is that the preamble can be used, but it is better to have an interference excision algorithm which protects the preamble instead of using it. The ETM 1810/M is able to suppress up to 4 interferers and it doesn't matter whether it is continuous wave or FSK interference.

2. The modem is available. However, as costs cannot be discussed here, I suggest you contact Daimler-Benz Aerospace, Ulm, Germany.

DISCUSSION

Discussor's name: C. Goutelard

Comment and Question:

On fait souvent des estimations avec des étalements doppler importants. Un canal à 2 ou 3 trajets et faible étalement se rencontre en pratique et est très pénalisant car il introduit des paquets d'erreurs très longs. Est-ce parce que ce cas paraît trop simpliste qu'il n'est pas souvent testé?

Translation:

Often estimations are made with wide doppler spread. A 2 or 3 path channel and narrow spread is usually encountered and this is a considerable handicap as it introduces very long error bursts. Is it because the case seems too simplistic that it is not very often tested?)

Author/presenter's reply:

Slow fading channels are very important in practice. A 2 or 3 path channel model is realistic and it is tested with channel simulators.

Coding and interleaving are used to cope with the problem of error bursts, but the length of the interleaver should be at least 3 times the length of the mean time between error bursts to be effective.

On very slow fading channels this requirement cannot be matched and then with the aid of an ARQ protocol the corrupted messages have to be repeated.

AIDE A LA DECISION POUR LA RECHERCHE DE CANAUX CLAIRS DANS LA BANDE DECAMETRIQUE

J. Wolf*, Y.M. Le Roux
LAB/PTI/GER
France Telecom — Centre national d'études des télécommunications
2, avenue P. Marzin, 22307 Lannion, France

Sommaire

Le Centre National d'Etudes des Télécommunications (CNET) dispose d'un analyseur de liaisons qui permet de mesurer, d'une part, les paramètres principaux de propagation du canal ionosphérique et, d'autre part, l'occupation spectrale de ce canal. On propose, dans cet article, une méthode permettant de discriminer et de quantifier simplement cette occupation spectrale afin de servir de base à des études ultérieures sur la recherche de canaux clairs dans un contexte opérationnel.

Introduction

Pour établir une communication par voie ionosphérique dans des conditions satisfaisantes, il faut tenir compte de divers phénomènes [1] :
- des évanouissements sélectifs en temps et en fréquence (fading) dus aux caractéristiques de propagation du canal ionosphérique, notamment celles liées aux réflexions multiples sur les couches E et F de l'ionosphère et à leurs évolutions au cours du temps
- des bruits et brouilleurs divers qui sont une cause de dégradation importante des communications.

On ne s'intéressera qu'à ces derniers dans le présent article.

On distingue généralement deux classes de bruits [2]:

- *les bruits de type A (ou brouilleurs)* :
Ce sont des bruits à bande étroite (vis-à-vis de la largeur de bande du signal utile), générés de façon artificielle, volontaire ou involontaire par des installations industrielles ou particulières (par ex : porteuses de transmission, systèmes d'allumage, lignes d'énergie,..)

- *les bruits du type B* :
Ce sont à l'opposé, des bruits à large bande, d'origine atmosphérique (décharges orageuses..) ou galactique (émission anisotrope provenant du gaz interstellaire de notre galaxie, comètes,..). Ces bruits étant la superposition d'impulsions électromagnétiques, leur modélisation par du bruit blanc gaussien, sur chacune des composantes en phase et en quadrature, peut être considérée acceptable dans de nombreuses applications, notamment dans le cas qui nous intéresse.

Pour mesurer l'encombrement spectral du canal ionosphérique, l'analyseur de liaisons du CNET [3] explore tout ou partie de la bande HF par tranches adjacentes de 10kHz. De ces spectres partiels, on peut déduire le spectre total dans toute la bande HF. La grande masse de données à traiter que cela concerne devient cependant rapidement une contrainte non négligeable pour le calcul, le stockage et l'interprétation des résultats.

L'objet de ce document est de présenter une procédure algorithmique permettant:
-de séparer le bruit des brouilleurs,
-de trouver les caractéristiques du bruit,
-de présenter un spectre "simplifié" permettant une économie de stockage et une lecture plus aisée.

I-Réarrangement non-décroissant d'une fonction.

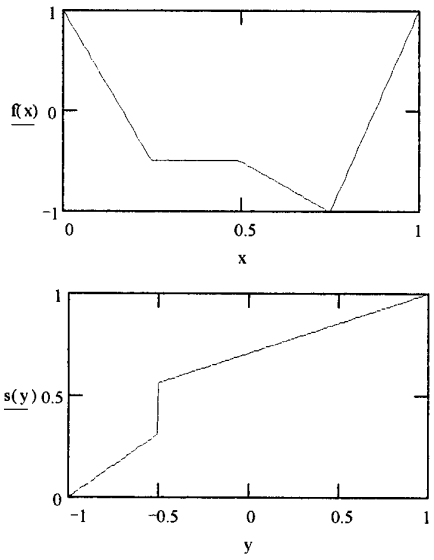
I-1.Définition.

Soit une fonction $f(x)$ définie en tous points de l'intervalle $D=(a,b) \subset R$.
Soit un ensemble A - une partie de D - et $mes(A)$ sa mesure. Se référant à [4], on appelle fonction-distribution de f:

$$s_f(y) = m\{x \in D: f(x) < y\} \quad y \in [\inf_{x \in D} f(x), \sup_{x \in D} f(x)]. \quad \text{(I-1)}$$

On peut noter que $s_f(y)$ est une fonction non-négative et non-décroissante.

Exemple I-1 :



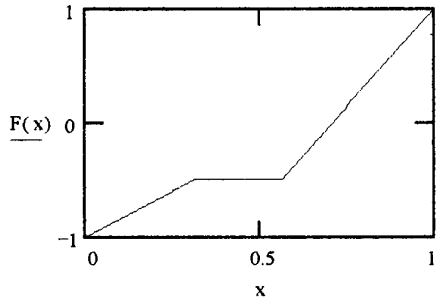
On remarque que, lorsque f possède un plateau, s possède une discontinuité et est continue à droite. On considère la transposée de f: (I-2)

$$f^*(x) = \sup\{y \in [\inf_{u \in D} f(u), \sup_{u \in D} f(u)]: \; s_f(y) \le x\} \quad x \in [0, m(D)].$$

$F(x)=f^*(x-a)$ est appelée **réarrangement non-décroissant de** $f(rend(f))$, sur l'intervalle (a,b) ou échantillon ordonné ou encore série variationnelle [5].
La figure suivante présente le réarrangement correspondant à l'exemple I-1.

*Professeur des Universités à l'Ecole Nationale Supérieure des Sciences Appliquées et de Technologie (Lannion)

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.



I-2.Propriétés :

Si $s(y)$ est continue et strictement croissante, $f^*(x)$ est la fonction inverse de $s(y)$.
On peut montrer que:

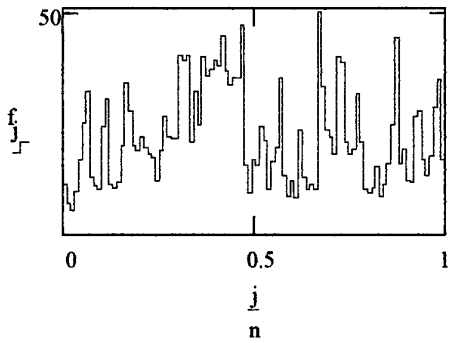
$$\int_a^b f(x)^p\,dx = \int_a^b F(u)^p\,du \quad \forall p \geq 0 \qquad \text{(I-3)}$$

I-3.Cas discret.

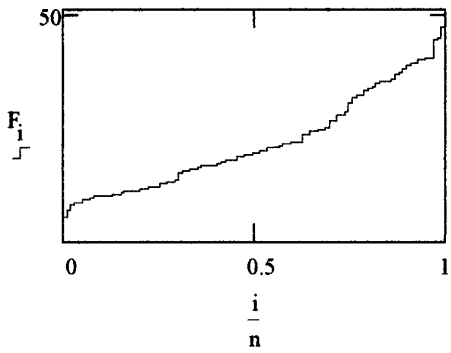
Prenons le cas où la "fonction" est définie par la suite de valeurs $\{y_i\ ,\ i=o,..,n\}$ à des "instants" supposés équidistants pour faciliter l'écriture : $x_i = x_0+i*h$, avec $h=(b-a)/n$ pas de mesure et $a=x_0-h/2$, $b=x_n+h/2$. On considère alors la fonction étagée:

$$f(x)=y_i \qquad x \in (x_i-h/2, x_i+h/2)\ i=0,...,n \qquad \text{(I-4)}$$

Exemple I-3: Signal discret contenant 100 valeurs,



et son réarrangement non-décroissant :



II - Processus aléatoire, stationnaire.

II-1-Bruit blanc.

On considère le processus aléatoire stationnaire défini par $X(t,\omega)$ où, à chaque instant $t \in [a,b]$, la variable aléatoire X_t est définie sur l'espace de probabilité $(\Omega, \mathcal{B}, P)$ avec $\Omega \subseteq R$.
A l'instant t, on note $P_t(z,X) = P(z,X)$ - l'indice t étant supprimé puisque le processus est stationnaire - la fonction de répartition de la variable aléatoire X_t. Les variables aléatoires X_t sont supposées stochastiquement indépendantes (bruit blanc strictement).

La variable aléatoire X_t, définie aux instants équidistants : $t_i=a+i(b-a)/n \quad i=0,..n$, prend les valeurs x_i.. A la suite des valeurs x_{i_5}, on associe la fonction étagée $x(t)$ telle que définie en (I-4). $x(t)$ est une trajectoire du processus. Pour évaluer le réarrangement non-décroissant de $x(t)$, on calcule $s_x(z,X)$, fonction-distribution de $x(t)$.

Si $P_n(z,X)$ est la fonction de répartition empirique de X, on a :

$$P_n(z,X) = \frac{\text{nombre de valeurs de } X \text{ telles que } (X < z)}{n+1}$$

$$P_n(z,X) = mes(t \in [a,b] \quad x(t)<z)/(b-a+h) = s_x(z,X)/(b-a+h)$$

D'après le théorème de Glivenko-Cantelli [5]

$$\text{Pour } n \to \infty \quad \sup_t \left| P_n(t,X) - P(t,X) \right| \underset{p.s}{\to} 0$$

$$\text{donc} \quad \lim_{n\to\infty} P_n(z,X) = P(z,X) = s(z,X)/(b-a) \qquad \text{(II-1)}$$

En résumé
Au processus aléatoire stationnaire (bruit blanc) $X(t,\omega)$ défini sur $[a,b]$, est associé un réarrangement non-décroissant $S(t,X)$ $(t \in [a,b])$. C'est la transposée de $s_X(z,X) = (b-a)P(z;X)$ où $P(z;X)$ est la fonction de répartition de la variable aléatoire X_t.

III-2-Réarrangement non-décroissant de $X(t,\omega)$ sur (a,b).

On considère un processus aléatoire stationnaire $X(t,\omega)$ sur l'intervalle $(0,1)$. La fonction de répartition de X_t à l'instant 0 est notée $P(z)=P_0(z)$.
On note $H(t)$ sa transposée, définie en (I-2). $H(t)$ est le réarrangement non-décroissant du processus aléatoire stationnaire $X(t,\omega)$ sur l'intervalle$(0,1)$.

$$H(t) = Sup\{v \in [\ \underset{u\in(0,1)}{Inf}\ x(u),\ \underset{u\in(0,1)}{Sup}\ x(u)]\colon P(v) \leq t\} \quad t \in [0,1].$$

On considère désormais le processus aléatoire stationnaire $Y(t,\omega) \equiv X(t,\omega)$ lorsque $t \in (a,b)$. Sa fonction-distribution est, d'après (II-1), $(b-a)P(z)$ et son *rend*:

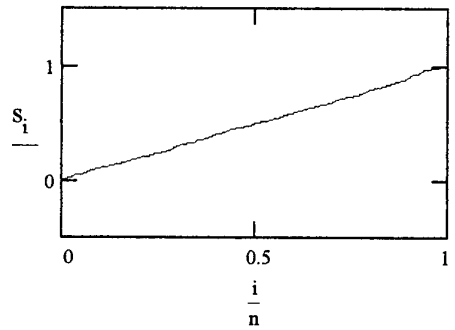
$$H_1(t) = Sup\{z \in [\ \underset{u\in(a,b)}{Inf}\ y(u),\ \underset{u\in(a,b)}{Sup}\ y(u)]\colon (b-a)P(z) \leq t\} \quad t \in [a,b].$$

On a donc

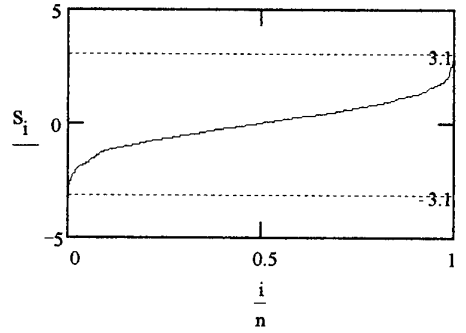
$$H_1(t) = H[(t-a)/(b-a)] \qquad \text{(II-2)}$$

II-3-Exemples de réarrangement de divers bruits blancs.

Exemple II-3-1: réarrangement non-décroissant d'un bruit blanc uniforme stationnaire sur (0,1) :



Exemple II-3-2 : réarrangement non-décroissant d'un bruit blanc gaussien N(0,2) stationnaire sur (0,1) :



III-Décompositon d'un signal.

III-1 Signal utile plus bruit.

On suppose disposer d'un signal "brut": somme d'un signal "utile" et d'un bruit:

$$s(t) = f(t)+b(t) \qquad t \in [0,1] \qquad \text{(III-1)}$$

Au bruit $b(t)$ on associe la variable aléatoire B dépendant du vecteur de paramètres Λ, admettant une densité de probabilité $d_B(t)$, et une fonction de répartition $P_B(t)$.
Exemple: dans le cas d'un bruit gaussien la variable aléatoire B dépend des deux paramètres: moyenne m et écart-type σ. On a donc :

$$\Lambda=\{m,\sigma\} \text{ et } d_B(t)= d_B(t;m,\sigma) \text{ et } P_B(t)= P_B(t;m,\sigma).$$

Au signal utile $f(t)$ on associe la variable aléatoire discrète F qui peut prendre les valeurs:
$\{f_0, f_1, \ldots, f_r\}$ supposées rangées dans l'ordre croissant, avec les probabilités respectives: $\{p_0, p_1, \ldots, p_r\}$.
En supposant qu'il existe des instants où, seul, le bruit est présent, on aura $f_0=0$.

Remarque : la justification de ce type de modèle se trouve dans le paragraphe V.

Connaissant les résultats sur la somme de deux variables aléatoires indépendantes, on peut immédiatement écrire la densité de la variable aléatoire S:

$$d_S(z) = \sum_0^r p_i \; d_B(z-f_i) \qquad \text{(III-2)}$$

III-2-Fonction de répartition

Dans les paragrahes précédents, on a vu qu'il est équivalent de parler de réarrangement non-décroissant ou de fonction-distribution. Comme cette dernière est aussi la fonction de répartition, on s'intéressera désormais à celle-ci comme transposée du réarrangement non-décroissant. En intégrant la relation (III-2), on a, aussi, immédiatement:

$$P_S(z) = \sum_0^r p_i \; P_B(z-f_i) \qquad \text{(III-3)}$$

III-3-Processus tronqués.

Soit la variable aléatoire continue X, dont l'ensemble des valeurs est $[a,b]$, et dont la fonction de répartition est $P(x)$. On définit deux nombres $k1$ et $k2$ qui sont tels que:

- si a est fini $k1=a$ sinon $k1=Max(k \in \mathbf{R})$ tel que $P(k)<10^{-n}$
- si b est fini $k2=b$ sinon $k2=Min(k \in \mathbf{R})$ tel que $P(k)>1-10^{-n}$

(pratiquement $n=3$).

On associe alors à la variable aléatoire X, une variable aléatoire "tronquée" X^*, dont la fonction de répartition est:

$$P^*(x) = \begin{cases} 0 & si \quad x \le k1 \\ \dfrac{P(x)-P(k1)}{P(k2)-P(k1)} & si \quad k1<x<k2 \\ 1 & si \quad k2 \le x \end{cases}$$

Exemples:
- cas de la loi normale $N(m,\sigma)$, si $n=3$, $k1=m-k\sigma$, $k2=m+k\sigma$ avec $k=3.1$.
- cas de la loi de Rayleigh de paramètre a, si $n=3$, $k1=0$, $k2=3.72/a$.

III-4-Interprétation. Cas divers.

On se propose, maintenant, d'interpréter la formulation (III-2) dans le cas d'un signal satisfaisant aux hypothèses du paragraphe III-1, pour un processus tronqué.

On considère l'expression (III-2) $d_S(z) = \sum_0^r p_i \; d_B(z-f_i)$.

On interprête $\{f_0, f_1, \ldots, f_r\}$ comme une population composée des $r+1$ individus $f_0, f_1, \ldots, f_r$. Dans le but ultérieur de pouvoir déterminer les différentes caractéristiques (densité ou bien fonction de répartition) du bruit, on considèrera cette population comme constituée de **deux** sous populations. La première Pop_1 contient l'élément unique f_0 et la seconde Pop_2 est constituée de tous les autres individus $f_1, f_2, \ldots, f_r$.

On remarque que:
. par construction de la variable aléatoire tronquée B^*, $d(z-f_0)=0$, lorsque $z \notin [f_0+k1, f_0+k2]$. L'influence de l'individu f_0 est limitée à cet intervalle. Puisque $f_0=0$, $d(z)$ contient toute l'information relative au bruit.
. de même l'influence de l'ensemble des individus $f_1, f_2, \ldots, f_r$ se limite, quant à elle, à l'intervalle $[f_1+k1, f_r+k2]$, ceci indépendamment des influences mutuelles des différents individus f_i $(i=1,r)$ entre eux.
On appelle *taux d'immersion* du signal utile dans le bruit, la quantité:

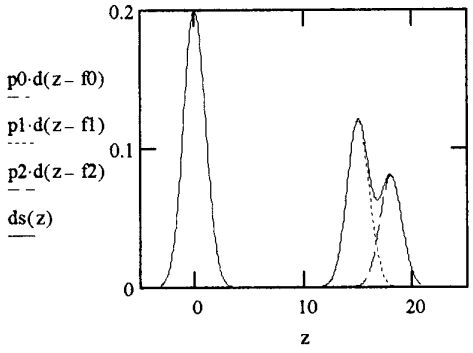
$$\omega = [(f_0+k2)-(f_1+k1)]/(k2-k1) = 1-(f_1-f_0)/(k2-k1)$$

Pour faciliter la compréhension, l'exemple suivant servira de support :

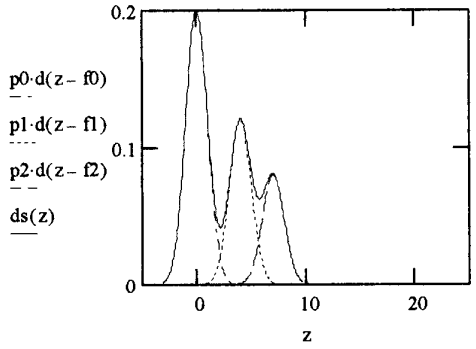
exemple III-4:
. ***B*** bruit gaussien *N(0,1)*, tronqué avec *k(=3.1)* constante de troncature.
. ***F*** peut prendre les valeurs *{0 , a ,a+3}* avec les probabilités *{0.5,0.3,0.2}*.

$d_S(z) = 0.5\, d(z) + 0.3\, d(z-a) + 0.2\, d(z-(a+3))$.

a) cas de deux sous-populations séparées :
Ceci signifie $(f_1+k1)-(f_0+k2)>0$ soit $\omega<0$.
Dans l'exemple $a=15$ ou $\omega = -1.4$.



b) cas de deux sous-populations mélangées :
Ceci signifie $(f_1+k1)-(f_0+k2) \leq 0$ et $f_1>f_0$ soit $0 \leq \omega < 1$
Dans l'exemple $a=4$ ou $\omega=0.35$.



La formule (III-3) aurait pu être utilisée et aurait donnée la même interprétation. La formule (III-2) est simplement plus lisible, évitant les cumuls des fonctions de répartition partielles. Dans le cas de deux sous-populations séparées, on constate que, sur l'intervalle $[f_0+k1, f_0+k2]$ la fonction de répartition de ***S*** est $P_S(z)=p_0\, P_B(z)$.

IV-Recherche des caractéristiques du bruit dans un signal brut.

Dans un premier temps, et dans le cas de deux sous-populations séparées, on se propose de déterminer les deux bornes *k1* et *k2* (puisque f_0 est supposé égal à *0*), ainsi que les autres paramètres, éventuels, attachés au bruit.

IV-1.Instruments mathématiques de mesure.

On considère l'intervalle de référence *[0,1]*, et l'intervalle $[u,v] \subseteq [0,1]$, $(v>u)$.

Soit, donc , le signal *s(t)* et sa fonction de répartition, *G(t)*, obtenue à partir de son réarrangement non-décroissant, définis sur *[0,1]*. Sur l'intervalle *[k1,k2]*, la fonction de répartition de la variable aléatoire tronquée représentant le bruit est $P^*(x,\Lambda)$.

Aussi, on considère la quantité:

$$Q_{u,v}(\Lambda) = K(u,v)\int_u^v \left[a + bP^*(t,\Lambda) - G(t)\right]^2 dt \qquad \text{(IV-1)}$$

où $K(u,v)$ est un coefficient de normalisation, b correspond à p_0 et a permettra de trouver d'éventuelles valeurs aberrantes.
Par exemple $K(u,v)=1/[(v-u)(G(v)-G(u))^2]$. La quantité $R=Q^{1/2}$, distance entre la fonction de répartition empirique et la fonction de répartition recherchée, est appelée **coefficient d'ajustement** (coefficient sans dimension pour le K choisi).

Puisque, par hypothèse, il existe un intervalle sur lequel la fonction de répartition du signal brut est égale à celle du bruit (à une translation et une homothétie, éventuelles, près), il reste à trouver cet intervalle, sur lequel la valeur du vecteur paramètre Λ est telle que $Q=0$.

IV-2.Relation fondamentale discrète.

On considère désormais le signal brut sous la forme d'une suite de données $\{s_i, i=0,..,n\}$. On a, comme indiqué en III-1:
$s_i = f_i + b_i \quad i=0,..,n$.
On se propose de décrire un algorithme permettant de trouver les caractéristiques du bruit $\{b_i \; i=0,..,n\}$, après avoir fait une hypothèse sur sa nature (gaussien par exemple) pour un taux d'immersion, du signal utile dans le bruit, nul.
Soit $P^*(x;\lambda_1,...,\lambda_\zeta)$ la fonction de répartition de la variable aléatoire représentative du bruit. On remarque que, lorsque $\zeta=2$, la donnée de *k1* et *k2* permet de déterminer λ_1 et λ_2.

D'une manière générale, soit la variable aléatoire ***X***. Au cours de $n+1$ épreuves, on obtient les réalisations x_i $(i=0,..,n)$ de ***X***. A partir de ces quantités x_i , par divers moyens, on est capable de trouver de bons estimateurs -voire des estimateurs optimaux- des différents paramètres de la loi de ***X***. La méthode du maximum de vraisemblance [6] sera utilisée la plupart du temps, parallèlement à celle dite de la plus courte distance.
Soit *[i/n, S_i , i=0,..,n]* le réarrangement non-décroissant de *[i/n, s_i , i=0,..,n]*. La fonction de répartition de la suite initiale est, alors, $\underline{S^{(n)}} = [S_i, i/n, i=0,..,n]$.

Dans le cas discret, la quantité Q définie par la relation (IV-1) devient la **relation fondamentale discrète**:

$$Q_{r,p}(\Lambda) = \frac{K(S_r, S_p)}{2} \sum_{i=r}^{p-1} (S_{i+1} - S_i)\{C_i + C_{i+1}\}$$

$$\text{où } C_i = \left[a + bP^*(S_i;\Lambda) - \frac{i}{n}\right]^2 \qquad \text{(IV - 2)}$$

Remarque: Pour intégrer (IV-1), la méthode des trapèzes a été utilisée. Celle des rectangles, un peu moins précise mais moins côuteuse en calculs aurait aussi pu convenir.

IV-3.Construction de l'algorithme d'ajustement optimal.

La résolution de (IV-2) est particulièrement délicate. On se propose donc de présenter une méthode donnant une solution approchée de (IV-2).

On pose $K(u,v)=1/[(v-u)(G(v)-G(u))^2]$.

Dans le cas des hypothèses III-4-a, on a $S_r=0$ (soit $u=0$). Alors, pour résoudre (IV-2), c'est à dire minimiser la quantité $Q_{r,p}$, on réalise les étapes suivantes :

- on a $S_r=0$ pour $r=0$ et on se fixe $p \in]0,n]$ (donc S_p).
- on pose: $a=G(S_r)=r/n=0 \quad b=G(S_p)-G(S_r)=(p-r)/n=p/n$.
- on calcule les estimateurs de $\lambda_1,...,\lambda_\zeta$, (par la méthode du maximum de vraisemblance).
- on fait, alors, varier p par valeurs décroissantes à partir de n.
- on calcule Q, aux valeurs correspondant aux paramètres estimés.

Quand, pour $p=p^*$, la quantité Q est nulle (lorsque en pratique, elle est minimum), on obtient la valeur des divers paramètres recherchés.

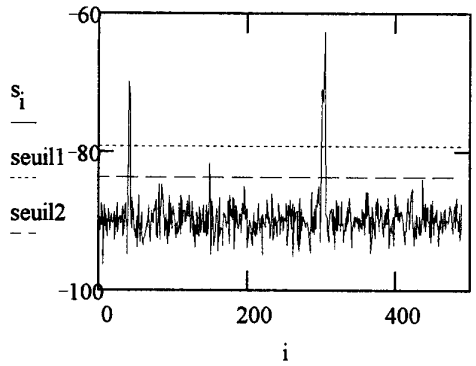
V-Rechercherche de canaux clairs dans la bande décamétrique.

Les bruits présents dans la bande décamétrique, qui sont une superposition d'ondes électromagnétiques d'origines diverses, peuvent, par exemple, être modélisés par un bruit blanc gaussien. Les brouilleurs, quant à eux, occupent une bande de fréquences pouvant ne correspondre qu'à une simple porteuse ou s'étendre sur quelques kHz.
C'est ainsi que la représentation de tout ou partie du spectre HF, par un modèle du type défini en III-1 apparaît comme raisonnable.

V-1.Validation du modèle.

Dans cette étude sont traités, essentiellement, des cas où le taux d'immersion est voisin de 0. D'autres travaux, en cours, laissent présager de bons résultats lorsque ce taux croît de manière significative.
On représente, pour chaque exemple, la fonction de répartition empirique du spectre étudié, ainsi que son ajustement optimal par une "gaussienne" sur l'intervalle ad hoc. Dans le cas gaussien les estimateurs utilisés sont les estimateurs traditionnels de la moyenne et de l'écart-type.

Exemple V-1-1: on commencera par un échantillon simulé de 500 points, composé d'un bruit gaussien $N(-90,2)$ et de 5 "brouilleurs" aux adresses 40 à 42, 150, 300 à 303, 304 et 306. Les niveaux respectifs exprimés en dBm sont : -71, -81, - 73, -65 et - 64. Le spectre synthétisé ainsi obtenu est le suivant :

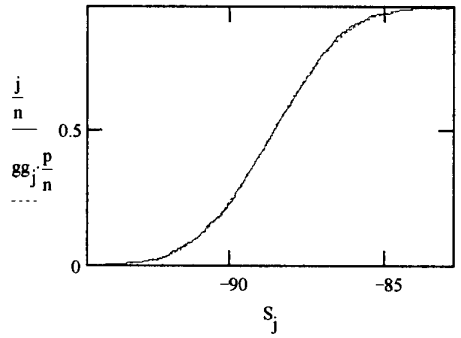


Les résultats fournis par l'algorithme d'ajustement optimal précédemment décrit sont dans ce cas :

moyenne calculée: - 90.01
écart-type calculé: 2.01
seuil (seuil1) par une méthode traditionnelle: - 79.21
seuil (seuil2) calculé: - 83.78

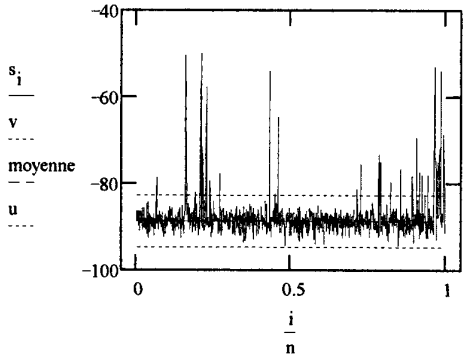
On voit tout l'intérêt de l'ajustement optimal, permettant en particulier de signaler le brouilleur ignoré par l'autre méthode citée.

Exemple V-1-2: (le spectre original se compose de 1000 points)
Cet exemple correspond à un enregistrement de l'occupation spectrale du canal ionosphérique, effectué le 6.03.92 à Lannion, au moyen de l'analyseur de liaisons du CNET. La portion de spectre présentée ici couvre la bande 5 - 6 Mhz et les niveaux sont exprimés en dBm.
La figure suivante présente les résultats de l'ajustement de la fonction de répartition empirique (___), par la fonction de répartition gaussienne (....), pour les 953 premiers points.



Valeur du coefficient d'ajustement: $R=0.0048$

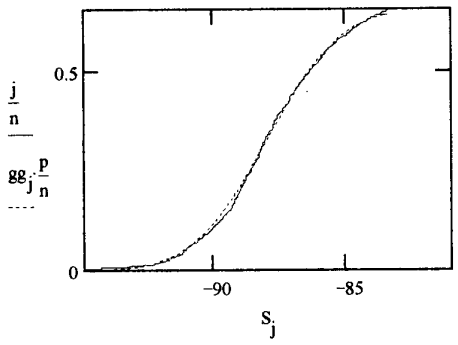
Représentation du spectre original :



caractéristiques du bruit: moyenne = - 88.71
écart-type = 1.91
valeurs extrêmes du bruit u = - 94.64
v = - 82.77

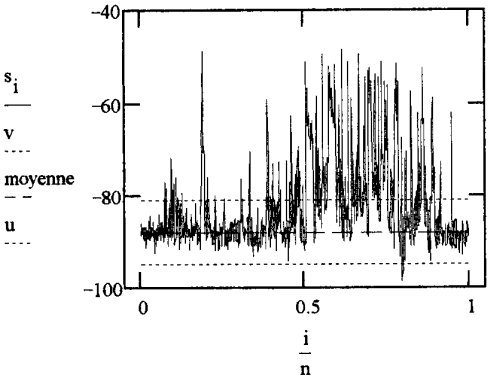
Exemple V-1-3: (le spectre original se compose de 1010 points)
Il s'agit ici d'un enregistrement obtenu le 13.03.92 à Lannion dans la bande 9 - 10 MHz.

La figure suivante présente les résultats de l'ajustement de la fonction de répartition empirique (___), par la fonction de répartition gaussienne (....), pour les 663 premiers points.



Valeur du coefficient d'ajustement: *R=0.0091*

Représentation du spectre original :



caractéristiques du bruit: moyenne = - 87.93
écart-type = 2.23
valeurs extrêmes du bruit: u = - 94.85
v = - 81.01

On en tire, pour chaque exemple présenté, un certain nombre d'informations :

- l'acceptation ou non de la représentation du bruit par un processus blanc gaussien.(on constate que, sur l'exemple V-1-2 l'ajustement est très.bon. Le coefficient *R* est deux fois moindre que dans V-1-3)
- les caractéristiques du bruit, c'est à dire sa moyenne *m* et son écart-type σ
- la zone $(m-3.1\sigma, m+3.1\sigma)$ où le bruit est, pratiquement, seul présent
- l'existence éventuelle de valeurs aberrantes (comme dans V-1-3-1)
- en conséquence, les zones où sévissent les brouilleurs

Par rapport à d'autres méthodes existantes [7], on constate les avantages suivants :

- possibilité de disposer d'un seuil de décision auto-ajusté aux conditions existantes dans le domaine des mesures :

méthodes traditionnelles : *seuil1* = - *72.80* V-1-2
seuil1 = - *52.54* V-1-3
ajustement optimal : *seuil2* = - *82.77* V-1-2
seuil2 = - *81.01* V-1-3

- une bonne séparation du bruit et des brouilleurs
- des résultats directement exploitables

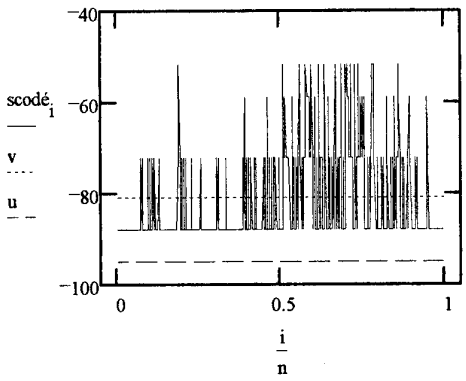
V-2.Représentation simplifiée de l'occupation spectrale.

Un récent article [8], montre que l'on peut passer d'un spectre codé sur *N* niveaux à un spectre codé sur *K* niveaux (*K*<<*N*) non uniformément répartis. Afin d'avoir une représentation simplifiée de l'occupation spectrale, on se propose ici :

- de mettre le bruit à sa valeur moyenne,
- d'indiquer ses valeurs extrêmes,
- de coder sur trois niveaux l'ensemble des intensités, hors des canaux où le bruit est seul présent.

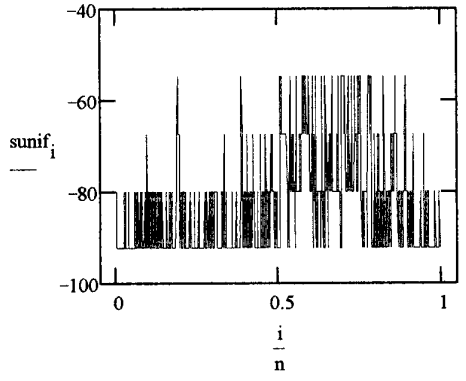
Exemple V-2.

On reprend l'exemple V-1-3 et les résultats du paragraphe précédent. Après ajustement optimal, les quatre niveaux, dont la moyenne, sont alors :



moyenne = - *87.93*
niveau1 = - *71.89*
niveau2 = - *58.73*
niveau3 = - *51.70*

Le spectre mesuré et son codage sur les 4 niveaux non uniformément répartis mettent nettement en évidence la très bonne préservation de l'information qu'autorise cette technique, en comparaison avec un découpage uniforme (ci-dessous).



Les niveaux sont, dans ce cas :

niveau0 = *-92.5*
niveau1 = *-79.9*
niveau2 = *-67.3*
niveau3 = *-54.7*

L'ajustement optimal et le codage associé sont particulièrement performants pour réduire la quantité d' information à stocker lors de campagnes de mesures de longue durée. Ce type de problème est d'actualité (cf. par exemple [9] et [10]).

Conclusion

Au cours de cette étude, tout en poursuivant l'objectif de la recherche des canaux clairs dans la bande décamétrique, plusieurs notions importantes ont été introduites :

- le réarrangement non-décroissant d'une fonction : il a l'avantage de donner très rapidement la fonction de répartition associée à cette dernière.
- le taux d'immersion : la liaison entre ce coefficient et le rapport signal à bruit reste cependant à parfaitement établir.

Ces instruments ont permis de définir une relation fondamentale, suffisamment générale, qui, dans cet exposé, a été utilisée pour trouver les caractéristiques d'un bruit gaussien. Cependant le cas d'autres types de distribution de bruits a déjà été testé, avec succès : Rayleigh, Rice, et plus généralement Nagakami... Les résultats obtenus, sur la recherche de canaux clairs, dans la bande décamétrique, sont encourageants.

Des prolongements de la méthode proposée ici, ainsi que d'autres applications sont d'ores et déjà envisagés :

- introduction de la dimension temporelle dans la recherche de canaux clairs, pour l'étude de leur cohérence
- prise en compte de taux d'immersion plus importants
- extension aux problèmes de détection radar par exemple.

Bibliographie

[1]**P. SCALART, Y.M. LE ROUX** Evaluation of a high speed digital communication system over ionospheric channels. *7th International Ionospheric Effects Symposium (IES '93).* pp 4A11.1-4A11.8 Washington DC, VA, U.S., mai 1993
[2]**A.D SPAULDING** Stochastic modelling of the electromagnetic environment, *IEE International Conference on communications (ICC '77)* pp114-123,Chicago, ILL, US. Juin 1977.
[3]**Y.M. LE ROUX, J.P. JOLIVET, J. MENARD** Analyse expérimentale, modélisation et simulation du canal de transmission ionosphérique, *Conférence SEE, Rennes septembre 1991.*
[4]**R.A. HUNT** On *L*(p,q) spaces. *L'enseignement mathématique,(1967) t.XII,fasc.4*
[5]**A. BOROVKOV** Stastistique mathématique. *Ed.MIR Moscou (1987)*
[6]**M. KENDALL, A. STUART** The advanced theory of statistics (1979) Vol.2 (4ème ed.)
[7]**G.F.GOTT, S. DUTTA, P. DOANY** Analysis of HF interference with application to digital communication, *IEEE Proceedings*, vol.130 n°5, august 1983.
[8]**PH. LE CLERC, Y.M. LE ROUX, J. LE SQUIN, J. WOLF** Application des principes de la classification optimale à l'étude de l'occupation spectrale de la bande HF. Introduction à la recherche de canaux clairs. *2èmes journées d'études SEE Propagation électromagnétique dans l'atmosphère, du décamétrique à l'angström.* Perros Guirec 15-17 mars 1994.
[9]**J.Y. LE SAOUT, F. GAUTHIER, R. FLEURY** Analyse.statistique de mesures d'occupation spectrale en bande HF. *2èmes journées d'études SEE Propagation électromagnétique dans l'atmosphère, du décamétrique à l'angström.* Perros Guirec 15-17 mars 1994.
[10]**T. CANAT, J. CARATORI, C. GOUTELARD** Modélisation spatio-temporelle des interférences électromagnétiques HF en Europe occidentale. Conférence AGARD, Proc. 486, Crète octobre 1990.

DISCUSSION

Discussor's name: C. Goutelard

Comment/Question:

Votre théorie est établie sur l'établissement de fonction de répétition ce qui nécessite un nombre de points minimum. Sur des bandes analysées de 1 MHz vous avez de toute évidence suffisamment de points. On ne souhaite pas toujours faire une analyse aussi vaste. Quelle est la bande minimale que vous jugez pouvoir analyser avec votre théorie.

(Translation:

Your theory is based on the establishment of a repetition function which requires a minimum number of points. On the 1 Mhz bands analysed you obviously have enough points. One would not always want to make such a broad analysis. What is the minimal band that you think you can analyse using your theory?)

Author/Presenter's reply:

Dans le cas de populations séparées, le nombre de points (de canaux) nécessaires est celui habituel - pour une loi Gaussienne - de l'ordre d'une cinquantaine dans les conditions d'expérimentation présentées.

Dans le cas de populations mélangées, (dans la mesure où le début de cette étude est récent) on ne peut déjà, donner une réponse chiffrée. Cependant, plus le taux d'immersion est grand, plus "l'échantillonage" ou la partie de la fonction de répartition à reconstituer, sera élevé.

(Translation:

In the case of separated populations the number of points (channels) necessary is as usual, for a Gaussian law, of the order of fifty under the experimental conditions presented.

In the case of mixed populations, (taking into consideration that this study was begun recently), it is not possible to give figures. However, the greater the level of immersion, the higher the "sampling" or the part of the distribution function to be reconstituted, will be.

COMPATIBILITE ELECTROMAGNETIQUE DANS LA BANDE DECAMETRIQUE : MODELISATION, PREVISION, EVALUATION PASSIVE DE L'IONOSPHERE

J. CARATORI, C. GOUTELARD
LETTI — Université Paris-Sud — Bâtiment 214
91405 ORSAY — France

Résumé

Le spectre de la gamme décamétrique révèle un nombre très important d'interférences dont la répartition dépend des utilisateurs et de l'état de l'ionosphère. Une modélisation des interférences peut être faite par l'observation des niveaux reçus sur l'étendue de la gamme. Il a été possible d'effectuer une telle classification par rapport au niveau du champ capté en un point de la surface terrestre.

Les études présentées ont été faites en Europe occidentale où la pollution électromagnétique est particulièrement importante. La caractérisation a conduit à définir des probabilités de clarté dont les variations journalières ont pu être représentées par des lois quantitatives, vérifiées avec un excellent accord, par les relevés expérimentaux.

Une modélisation théorique du spectre a été faite par un calcul numérique reposant sur la simulation de 2300 émetteurs répartis dans une zone de 7000 km de diamètre. Les résultats obtenus ont permis donner une représentation analytique de ces spectres et les confrontations avec les mesures effectuées ont été excellentes. A partir de ces premiers résultats, il a été envisagé d'utiliser les variations temporelles des spectres observés pour les corréler avec la fréquence critique de la région F2. Cette modélisation a conduit à définir un paramètre d'estimation de cette fréquence critique et les corrélations faites avec les mesures effectuées par le sondeur zénithal ont montré une très forte corrélation entre le foF2 réel et le paramètre introduit.

On a déduit de ces résultats une méthode de prévision totalement passive particulièrement bien adaptée aux systèmes de transmission discrets. Un dispositif a été introduit dans un système de télécommunication expérimental afin d'établir les liaisons sur les meilleures fréquences sans qu'il y ait d'échange prélable d'informations entre les stations. Les résultats encourageants obtenus ont conduit à implanter le système dans un récepteur qui permet la gestion automatique d'une liaison.

I. - INTRODUCTION -

L'encombrement spectral de la zone décamétrique est très important, ce qui prouve son exploitation par de nombreux utilisateurs. Le spectre, dont l'étendue est de 3 à 30MHz, révèle l'émission d'ondes de très forte puissance dues aux stations de radiodiffusion et la présence d'émissions professionnelles de puissances inférieures.

La qualité d'une liaison est caractérisée par son rapport signal/(bruit + interférences) ce qui a entraîné depuis une quinzaine d'années le développement d'études sur la caractérisation de ces interférences, voire leur prévision [1] [2] [3] [4].

Les caractéristiques variables de l'ionosphère liées à l'indice d'acvitité solaire ont incité les utilisateurs à définir des méthodes de prévision. Ces méthodes, bien connues, dont l'efficacité a été testée depuis de nombreuses années permettent de prévoir les caractéristiques de transmission des canaux de la gamme décamétrique avec une très bonne précision. Cependant, la qualité dépend tout autant du bruit et des interférences qui, souvent, sont les paramètres qui conditionnent réellement la qualité de la transmission. Les interférences qui apparaissent dans la gamme décamétrique sont extrêmement nombreuses et leur densité peut dépasser plusieurs centaines/kHz. Ces ondes se propagent à l'échelle mondiale ce qui entraîne des gênes extrêmement importantes. Il faut cependant noter que l'encombrement spectral n'est pas uniforme à la surface terrestre et que la pollution dans les zones Europe et Amérique paraît supérieure à celle que l'on peut observer dans d'autres régions du monde comme, par exemple, l'océan pacifique sud où la densité des interférences apparaît beaucoup plus faible [5].

Une vision plus large des prévisions de liaisons ionosphériques consiste donc à prévoir, non seulement les qualités de transmission des canaux, mais également l'intensité des interférences qui prédominent en général sur l'intensité du bruit. La prévision de l'activité des interférences a été développée par différents auteurs [3]

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

[4] [5] [6] mais leur modélisation est toujours apparue complexe.

L'étude présentée constitue une étape supplémentaire dans la prévision des interférences, dans leur modélisation, mais également dans leur utilisation pour prévoir les paramètres ionosphériques. Ce point est apparu par l'observation du spectre de la gamme décamétrique dont les évolutions sont journalières, saisonnières et liées aux variations undecennales des cycles solaires. Il est montré, dans cette étude, que la répartition des interférences dans le spectre HF est étroitement liée avec la fréquence critique de la région F2 qui est l'un des paramètres principaux des prévisions ionosphériques. L'exploitation de cette corrélation conduit à l'établissement d'une méthode de prévision passive qui a été expérimentée et les résultats obtenus sont discutés. Il apparaît, au terme de ces discussions, que l'exploitation du spectre décamétrique des interférences constitue une excellente méthode de prévision passive de la fréquence critique de la région F2 de l'ionosphère. Cette exploitation a conduit à la conception d'un système destiné aux opérationnels. Ce système, totalement passif, offre d'indéniables avantages pour des évaluations passives indétectables, particulièrement précieuses pour l'établissement fiable de liaisons discrètes.

II. - CARACTERISATION DES INTERFERENCES DANS LA GAMME DECAMETRIQUE -

L'observation des interférences dans la gamme décamétrique révèle une grande variabilité dont la figure 1 donne un exemple. Deux spectres analysés de 7 à 23MHz avec une résolution de 1kHz sont représentés. Le premier correspond à un spectre de jour, le second à un spectre de nuit enregistrés dans la zone europe occidentale. Ces spectres révèlent une grande dynamique des niveaux reçus qui s'étend de - 90dBm à - 15dBm. On peut noter l'existence de bandes prédominantes qui correspondent aux bandes de radiodiffusion dans lesquelles des émetteurs de très forte puissance, souvent de l'ordre de 1MW, sont opérationnels. Les points caractéristiques à relever sont :

- Une étendue du spectre plus grande le jour que la nuit. On peut noter qu'au-delà de 20MHz le niveau des interférences reçues décroît, ne subsistant finalement que les émissions locales. Cette disparition des interférences reçues dans la partie supérieure de la gamme, est due à l'abaissement de la fréquence critique foF2 durant la nuit.

- Un affaiblissement des fréquences inférieures durant le jour, créé par l'absorption de l'énergie par la région D.

On a cherché à caractériser ces interférences par l'établissement de lois traduisant leur variation temporelle. Il a été effectué une modélisation [5] qui a conduit à l'établissement de relations analytiques permettant de donner des lois de prévision sur les interférences à l'intérieur de chacune des 95 bandes allouées par l'UIT (Union Internationale des Télécommunications). Les résultats obtenus donnent la probabilité $P_{(S,J)}$ pour que, durant une heure donnée, un canal de largeur de bande 1kHz dont le niveau est inférieur à un seuil S pendant J-1 jours consécutifs le demeure pendant le jour J suivant. Cette probabilité dépend donc du nombre de jours observés J-1, et du seuil choisi S. Elle a été appelée probabilité de clarté. Les études de cette probabilité de clarté ont été menées avec un dispositif expérimental nommé SARABANDE (Système d'Analyse RApide de la BAnde DEcamétrique) qui a permis de traiter plusieurs centaines de millions de points de mesures pour vérifier la véracité des études théoriques menées. La figure 2 représente un exemple de mesure de la probabilité de clarté pour la bande [18 168kHz à 18 401kHz] réservée aux liaisons fixes pour des périodes d'observation de 17 jours. On peut noter sur les surfaces donnant ces probabilités de clarté qu'en-dessous d'un seuil minimum, la probabilité de clarté devient nulle, alors qu'au-dessus d'un seuil maximum, la probabilité de clarté est égale à 1. La zone intermédiaire fait apparaître des courbes qui sont assimilables à des exponentielles décroissantes déterminées par la théorie dans [5]. Ces courbes sont approchées avec des coefficients de corrélation supérieurs à 0,98 dans la majorité des cas, pour des représentations s'étendant jusqu'à une quinzaine de jours. Au-delà de cette période, la loi ne semble plus être valable pour des raisons qui ont été exposées dans les publications citées. Les résultats obtenus par ces études antérieures ont amené à l'étude présente par une modélisation plus systématique des interférences reçues en un point donné de la surface terrestre.

III. - MODELISATION DU SPECTRE DECAMETRIQUE -

On a cherché à modéliser le spectre décamétrique en s'appuyant sur des programmes de simulation de propagation afin de calculer les niveaux des champs reçus. Cette simulation a été faite en calculant le champ reçu moyen sur les 95 bandes allouées par l'UIT en distinguant les bandes de radiodiffusion des bandes professionnelles et amateurs dans lesquelles les puissances d'émission sont très différentes.

Pour réaliser cette simulation on a considéré une zone géographique de 7000 km de diamètre sur laquelle on été disposés 2300 émetteurs permettant de couvrir les 95 bandes allouées par l'UIT. Les études ont été menées pour :

- 3 valeurs de l'indice de l'activité solaire : 50, 100 et 150.
- 3 saisons : été, hiver, équinoxe.
- Pour chaque heure du jour : 0 à 24 heures.

Pour l'ensemble des émetteurs répartis sur la zone, on a calculé le champ reçu en un lieu donné (Paris : 48,8° nord/8,3° est) et on a, au lieu de réception, calculé la somme des puissances reçues de l'ensemble des émetteurs et traduit cette valeur en décibels par rapport au microvolt par mètre.

La figure 3 donne un exemple de modélisation du spectre reçu pour les bandes situées hors bandes de radiodiffusion. On peut noter que cette simulation fait apparaître les caractéristiques principales des spectres observés expérimentalement :

- De jour, le spectre est plus affaibli vers les basses fréquences et s'étend largement vers les hautes fréquences.
- De nuit, le spectre a une intensité plus grande vers les basses fréquences mais par contre s'affaiblit très rapidement vers les hautes fréquences.

Pour réaliser la modélisation du spectre on a recherché des lois représentant le plus fidèlement possible les résultats de la simulation. On a, pour cela, utilisé une première relation :

$$P = A - Bf - Cf^2 - Df^{-1} - Ef^{-2}$$

Dans cette expression *A, B, C, D, E* sont des paramètres qui dépendent de l'heure, du jour, de la saison et de l'indice de l'activité solaire. Cette représentation a donné d'excellents résultats, cependant elle est apparue trop compliquée pour être utilisée de façon opérationnelle. C'est la raison pour laquelle on a adopté une loi simplifiée donnée par la relation :

$$P = A' - B'f - C'f^{-1}$$

qui est déduite de la relation précédente par la suppression des termes en f^2 et f^{-2} dont on a pu constater qu'ils avaient une influence de second ordre.

Pour vérifier la validité de ces résultats on a effectué de nombreuses vérifications en installant un récepteur panoramique à proximité d'un sondeur zénithal. Le sondeur zénithal utilisé est celui de Poitiers situé en France et le récepteur panoramique a été installé à la station STUDIO2 du LETTI (STUDIO : **S**ystème de **T**raitement **U**niversel de **D**iagnostic **IO**nosphérique) distant d'environ 100 kms de Poitiers.

Pour valider la modélisation du spectre HF par la formule simplifiée on a effectué à STUDIO2 les relevés de spectres réels que l'on a ajustés par la loi simplifiée de P, de laquelle on a déduit la fréquence critique foF2 de la région F2. Cette fréquence critique, calculée à partir du spectre décamétrique, a été ensuite comparée à la mesure directe de foF2 effectuée à Poitiers.

Les résultats obtenus sont montrés sur la figure 4 qui représente les variations de foF2 pour les journées des 3 et 4 novembre 1990, mesurées directement à Poitiers par le sondeur zénithal et déduites du spectre décamétrique mesuré à la station STUDIO. On peut constater que les lois observées sont en excellent accord. Cependant, les fluctuations de foF2 déduit du spectre décamétrique apparaissent plus importantes que celles mesurées par le sondeur zénithal. Un lissage des mesures de la fréquence foF2 mesurée à partir du spectre décamétrique donne des résultats en excellent accord avec les mesures directes.

On a considéré que la modélisation qui a été effectuée était suffisamment bonne, et que les polynômes représentant le spectre étaient suffisamment fiables pour donner une représentation fidèle de ce spectre réel. Cependant, malgré ces bons résultats, cette méthode a été abandonnée compte tenu de sa trop grande complexité en vue de l'application dans les systèmes opérationnels. On s'est donc tourné vers une méthode plus pratique qui, sans atteindre la qualité des résultats de cette première méthode, conduit à des résultats acceptables pour des systèmes opérationnels.

IV. - MODELISATION OPERATIONNELLE -

Dans cette étape, on a cherché à définir une méthode de modélisation permettant de trouver la valeur de la fréquence critique foF2 de la région F2 de l'ionosphère à partir des caractéristiques du spectre décamétrique observé par des méthodes simples.

La méthode finalement retenue a consisté à considérer la fréquence du spectre comme une variable aléatoire dont la densité de probabilité est la puissance du spectre relevé. Dans un premier temps, on a observé les fréquences individuelles mais les expérimentations ont montré qu'une méthode plus simple pouvait encore être adoptée. On a finalement défini une fonction densité de probabilité par bandes de fréquences allouées. Cette densité de probabilité $P_f(fi)$ est donnée par la relation :

$$P_f(fi) = \frac{P(fi)\,B(fi)}{\sum_i P(fi)\,B(fi)}$$

dans laquelle :

- $B(fi)$ représente la largeur de bande de la Iième bande allouée centrée sur fi.
- $P(fi)$ représente la puissance totale reçue dans la bande $B(fi)$.

Cette densité n'est appliquée qu'en dehors des bandes de radiodiffusion ou uniquement sur les bandes de radiodiffusion. Ces deux options sont possibles, cependant les meilleurs résultats sont obtenus lorsqu'on applique cette relation uniquement sur les bandes hors

radiodiffusion. On définit alors le paramètres m_1F par la relation :

$$m_1F = \sum_i fi\ P_f(fi)$$

qui représente le moment du premier ordre de la variable fi. Ce paramètre m_1F se trouve être étroitement corrélé avec la fréquence foF2 de l'ionosphère.

La figure 5 représente deux enregistrements effectués par le sondeur zénithal de la station de Poitiers, et par les relevés du spectre décamétrique à la station STUDIO2. On été également représentées les prévisions à long terme de foF2 à 10, 50 et 90%. On peut noter l'excellente corrélation d'ensemble entre la fréquence foF2 mesurée à Poitiers et le facteur m_1F déduit du spectre décamétrique. On doit cependant remarquer que, lors de la transition jour-nuit, entre 15 et 20 heures, m_1F est inférieur à foF2. Ce résultat a été interprété et corrigé. En effet, si l'on reporte les points tenus pour foF2 et le facteur m_1F (figure 6) on constate que ces points sont fortement corrélés, ce qui explique la bonne concordance observée sur la figure 5. Cependant, on peut remarquer que, lors des transitions jour-nuit, les points s'écartent notablement de la droite de corrélation parfaite. En Europe occidentale (Paris) la densité des émetteurs est plus forte dans la direction Est que dans la direction Ouest. Cette constatation a été mentionnée dès 1990 [5] si bien que les résultats des mesures sont entâchés par l'inhomogéité de la transition qui fait que le nombre d'émissions non reçues de la direction Est est sensiblement liée aux fréquences foF2 des points de réflexion c'est-à-dire inférieures à la fréquence foF2 locale. Ce biais est apparu de façon systématique dans toutes les mesures qui ont été faites, si bien qu'il a pu être corrigé. Pour estimer la précision de l'évaluation ainsi faite, nous avons introduit deux paramètres :

- l'un représentant l'écart entre la fréquence critique foF2 vraie et le paramètre m_1F :

$$X = \text{foF2vraie} - m_1F$$

- l'autre représentant l'écart entre la fréquence critique foF2 et la fréquence foF2 estimée par des prévisions à 90% :

$$Y = \text{foF2vraie} - \text{foF2 à 90\%}$$

On a, pour chacune de ces quantités X et Y, calculé sur l'ensemble des mesures effectuées l'écart moyen et l'écart quadratique moyen. Les résultats ont donné :

- Ecart moyen de X : $\mu x = 0{,}313$MHz
- Ecart type moyen de X : $\sigma x = 0{,}807$MHz.
- Ecart moyen de Y : $\mu y = 3{,}654$MHz.
- Ecart type moyen de Y : $\sigma y = 1{,}346$MHz.

L'écart moyen et l'écart type de X sont accrus par les transitions jour-nuit, mais on peut noter que les résultats de l'évaluation de foF2 par le paramètre m_1F sont meilleurs que ceux des prévisions à long terme à 90%. Pour améliorer la prévision on a corrigé de façon systématique le biais qui apparaît lors de la transition jour-nuit dans l'estimation de m_1F. Après cette correction systématique, les résultats obtenus sont les suivants :

- Ecart moyen de X : $\mu' x = 0{,}071$MHz.
- Ecart type moyen de X : $\sigma' x = 0{,}418$MHz.

On peut noter que ces derniers résultats montrent un biais quasiment nul et un écart type sur X réduit à 0,4MHz ce qui correspond à une précision moyenne des sondeurs zénithaux.

La méthode de prévision apparaît donc correcte pour la fréquence critique foF2.

V. - SYSTEME OPERATIONNEL -

Afin de rendre opérationnelle la méthode, un système de prévisions passives a été élaboré. Le principe est résumé sur la figure 7. L'hypothèse prise consiste à mettre en relation un émetteur et un récepteur sans qu'une communication prélable soit établie entre ces deux points. Pour ceci, chaque extrêmité est équipée d'un récepteur permettant, à l'aide d'une antenne active large bande, de mesurer le spectre panomarique décamétrique. Ces récepteurs, couplés à un calculateur qui applique l'algorithme présenté précédemment, servent à commander le système de communication constitué par un émetteur d'une part, un récepteur d'autre part. L'objectif fixé consiste à ce que l'émetteur et le récepteur choisissent de façon indépendante la même fréquence pour établir la liaison. Ce résultat ne peut être obtenu à coup sûr, aussi un protocole de recherche simplifiée doit être mis en place dans l'hypothèse où les fréquences choisies par chacune des extrêmités de la liaison ne sont pas identiques. Pour aboutir au choix de la même fréquence par les stations d'émission et de réception, il suffit de vérifier que les spectres mesurés soient identiques. Pour cela, on a effectué des relevés de spectres décamétriques à la station STUDIO2 du LETTI située dans le Sud de la France et à la station STUDIO du LETTI située dans la région parisienne. La distance entre les deux stations est de 500 km. Pour évaluer la similarité des spectres, on a effectué des études sur la puissance relevée dans chacune des 95 bandes allouées par l'UIT. On a comparé ces résultats relevés à STUDIO à ceux de STUDIO2. La figure 8 donne un exemple de relevés de tels spectres. On peut remarquer que la droite de régression des 95 points relatifs aux 95 bandes allouées par l'UIT est une droite à 45°, décalée d'environ 9dB par rapport à la première bissectrice. Ce décalage est dû au fait que les antennes utilisées à chaque station ne sont pas identiques et ont des hauteurs effectives différentes. On doit remarquer que le décalage ne modifie pas les résultats obtenus par cette méthode, étant entendu qu'une normalisation de la puissance des spectres est effectuée par l'introduction de la notion de densité de probabilité Pf décrite précédemment. On doit

noter l'excellente corrélation de ces points, qui indique les spectres relevés à 450 km de distance demeurent très cohérents. Le coefficient de corrélation pour l'établissement de cette droite de régression étant dans tous les cas supérieur à 0,9 et la plupart du temps supérieur à 0,95.

Ces résultats ont été confirmés par une expérimentation effectuée entre deux stations distantes de 300 km. Dans cette expérimentation un lot de 30 fréquences a été alloué pour la liaison, et chaque extrêmité a recherché la fréquence allouée assurant la meilleure qualité. Une classification de 4 autres fréquences par ordre décroissant de la qualité évaluée a été faite.

Les résultats de cette campagne d'expérimentation ont montré que le contact était toujours établi. Dans 70% des cas, le premier contact était établi pour la première fréquence définie, dans 20% des cas pour la seconde fréquence et dans 8% des cas pour la troisième, quatrième ou cinquième fréquence. Ce résultat demeure valide pour un établissement des liaisons par évaluation passive de la fréquence critique foF2.

Le système a été implanté dans un récepteur incluant un microprocesseur. Le programme de traitement a été stocké dans une mémoire ROM d'une capacité de 200 koctets.

VI. - CONCLUSION -

Les interférences en ondes décamériques constituent une gêne importante pour les télécommunications. Une modélisation du spectre des interférences a été proposée et vérifiée par comparaison avec des mesures expérimentales. Il est apparu un excellent accord entre les mesures expérimentales et la modélisation théorique. L'utilisation du spectre décamétrique pour l'établissement de prévisions de propagation exploitant les modifications de ce spectre apparaît comme possible et introduit une nouvelle méthode de connexion dans les systèmes opérationnels.

La méthode élaborée est simple et son expérimentation a donné d'excellents résultats.

L'implantation d'un système de prévision passive de ce type dans les équipements de télécommunications apparaît possible et contribue à l'évaluation des canaux en temps réel. La qualité des résultats obtenus apparaît satisfaisante. Testé sur des liaisons de courtes distances, le succès de cette procédure est évident. Son extension sur des liaisons de grandes distances - quelques milliers de kilomètres - doit être envisagée, en considérant que les mesures effectuées entreraient dans un processus de correction de prévision à long terme, ou de programme universel de type IRI 90, qui permettent d'avoir une représentation de l'ionosphère à l'échelle mondiale.

BIBLIOGRAPHIE

[1] C. GOUTELARD, A. JOISEL - *Lutte contre le brouillage en ondes décamétriques dans un système d'aériens adaptatif orientable tout azimut.* 32ème AGARD COnf. : Propagation Factors affecting remote sensing by radio waves. Oberammergau (RFA), 24-28 mai 1983.

[2] C. GOUTELARD - *Frequency supportability for ECM resistant communications systems.* 5ème ARFA symp. NATO Headquarter Bruxelles, 15-18 décembre 1986.

[3] C. GOUTELARD, J.P. VAN UFFELEN - *Influence des variations ionosphériques sur les systèmes H.F. à haute fiabilité utilisant de grandes bases.* AGARD Conf. CP n° 441. Münich, 1988.

[4] C. GOUTELARD, J. CARATORI - Prévisions à très court terme par modélisation du brouillage et de l'ionosphère à l'aide de la rétrodiffusion. AGARD Conf. n° 453. San Diego, 1989.

[5] T. CANAT, J. CARATORI, C. GOUTELARD - *Modélisation spatio-temporelle des interférences électromagnétiques H.F. en Europe occidentale.* 47ème AGARD Conf. n° 486. Athènes, octobre 1990.

[6] P.J. LAYCOCK, M. MORELL, G.F. GOTT, A.R. RAY - *A model for H.F. spectral occupancy.* IEE, London, 1988.

FIGURE 1
Exemple de spectre de la bande décamétrique relevé à Paris - Europe occidentale

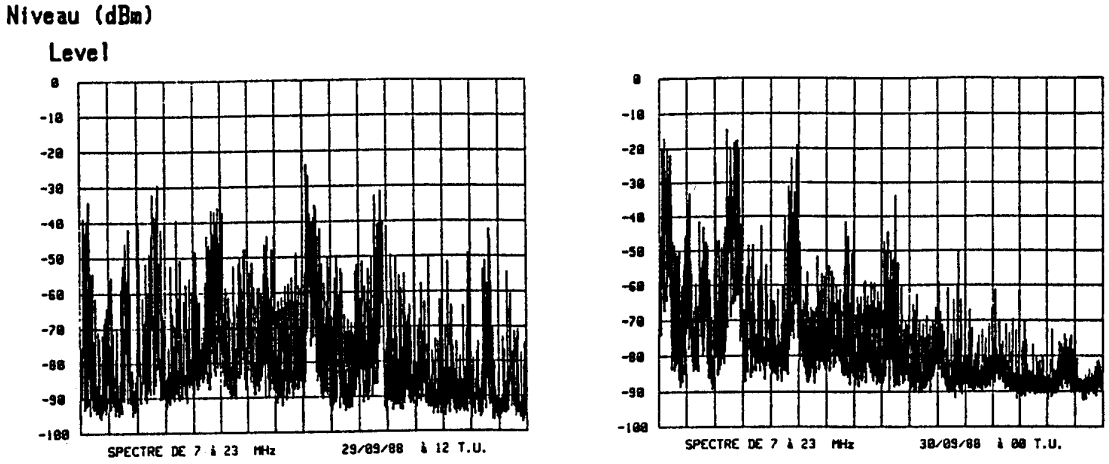


FIGURE 2
Représentation des probabilités de clarté relevées avec le système SARABANDE du LETTI

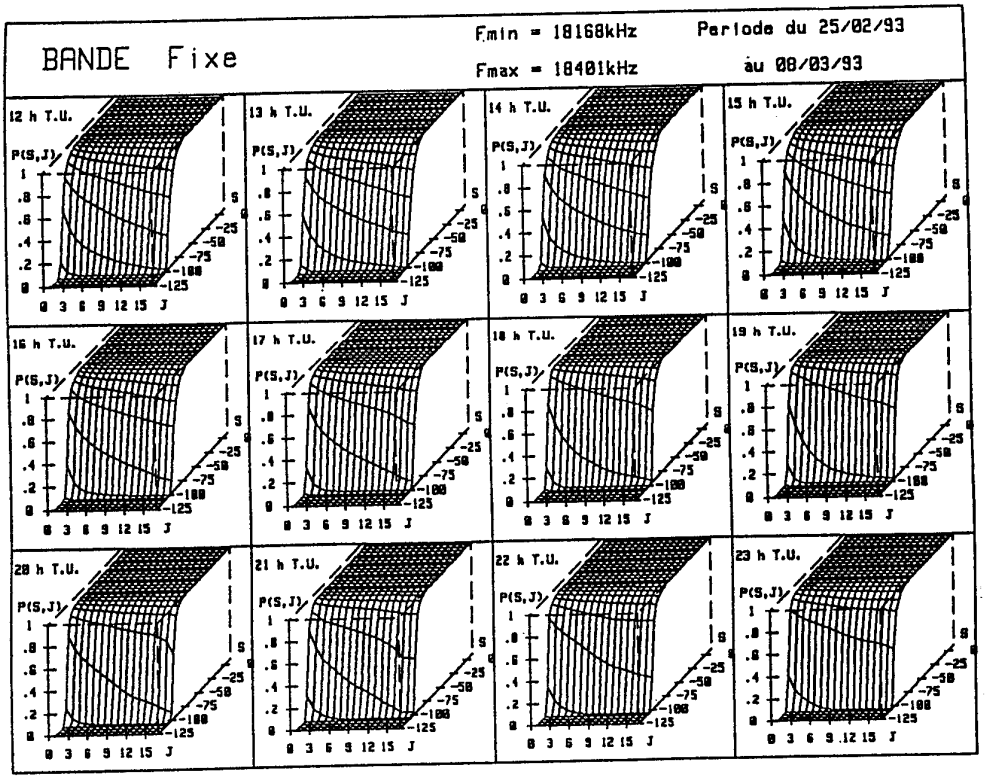


FIGURE 3
Spectres de la bande décamétrique simulés
Spectres relevés à Paris

January 1990 - Solar activity index 100
Receiving site : 48.8 N 2.3. E
Sum of the amplitudes

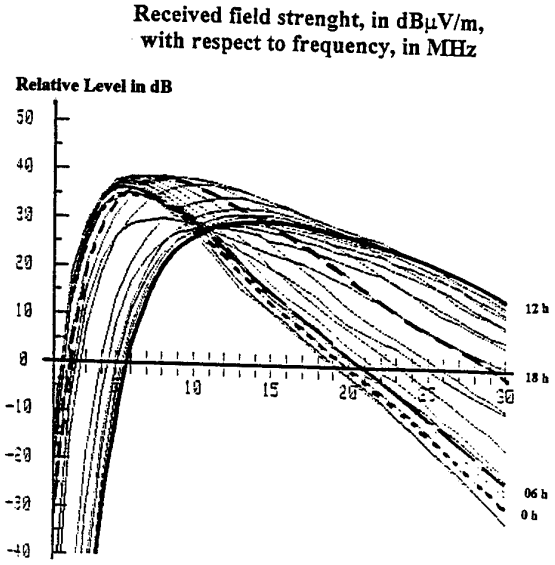
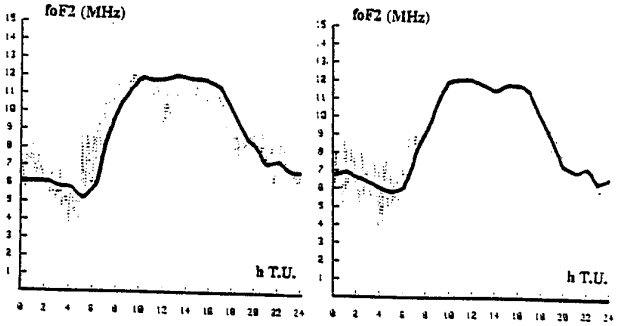


FIGURE 4
Variation de la fréquence critique foF2
relevée par sondage zénithal
et par exploitation du spectre décamétrique

RESULTS



Prediction of foF2 from the spectrum measured in "STUDIO2"
and comparison with measurements of "Poitiers"

foF2 Poitiers
foF2 STUDIO2

FIGURE 5
Comparaison de la fréquence critique relevée à Poitiers
par sondage zénithal (foF2)vraie et
celle déduite du spectre décamétrique *(m_1F)* à STUDIO2

IMPROVEMENTS BROUGHT BY THE REAL TIME CORRECTION OF IONOSPHERIC PREDICTIONS

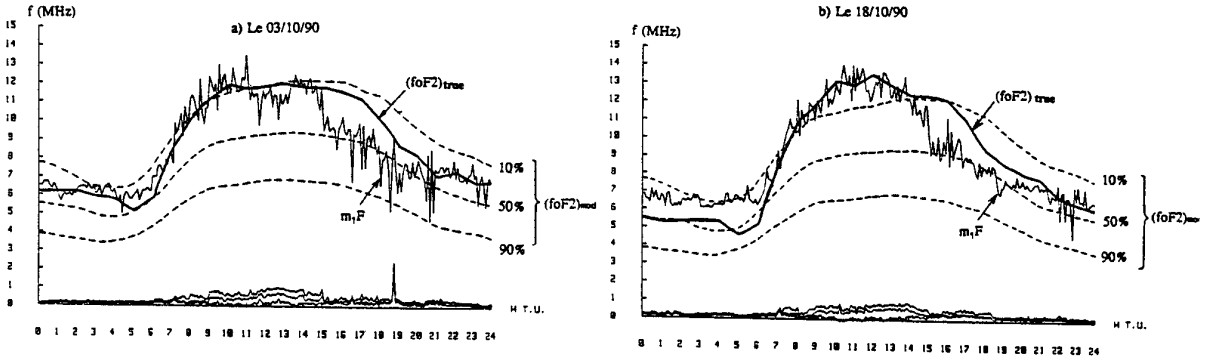
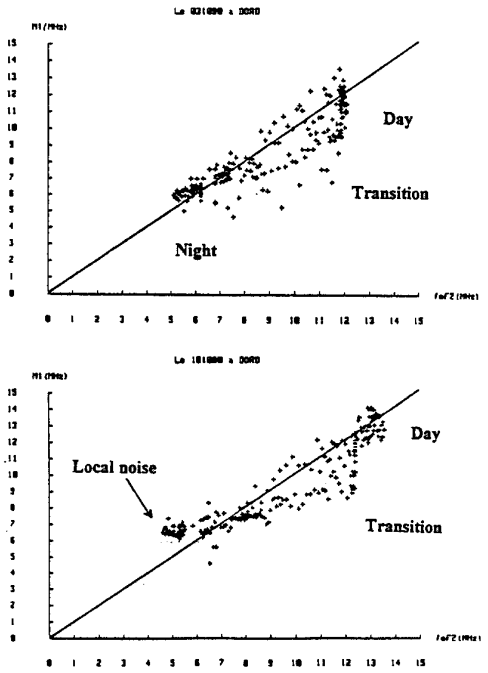


FIGURE 6
Corrélation entre la fréquence critique relevée à Poitiers (foF2) et celle déduite du spectre décamétrique (m_1) à STUDIO2

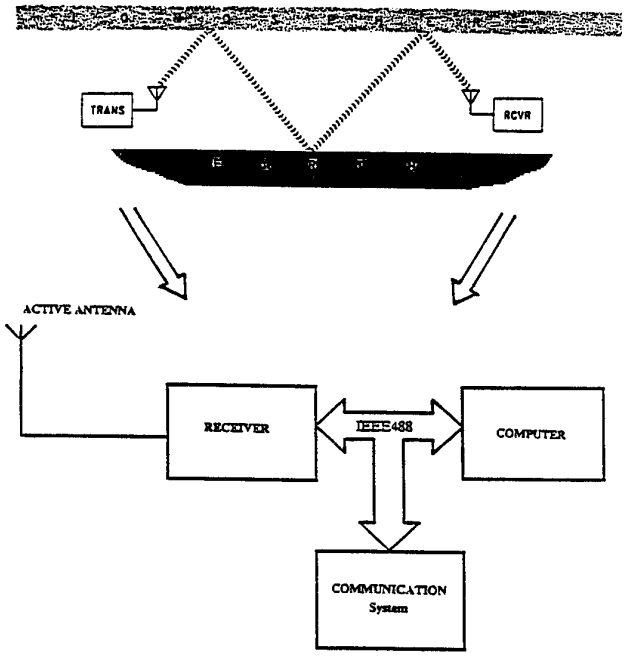
CORRELATION BETWEEN m_1F et foF2



Correlation between m_1F et foF2
"STUDIO2"

FIGURE 7
Dispositif de prévision passive dans un système de transmission

PASSIVE PREDICTION SYSTEM



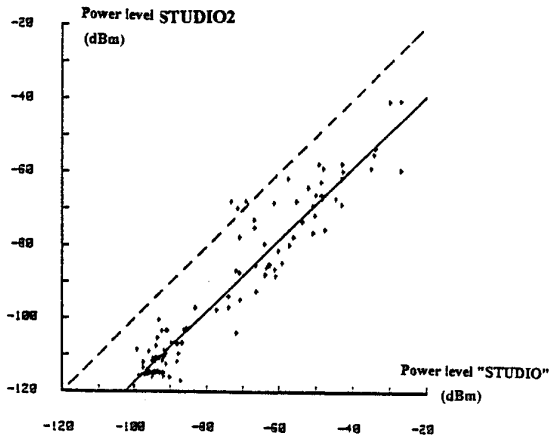


FIGURE 8
Comparaison des puissances relevées dans les 95 bandes allouées par l'UIT à STUDIO et STUDIO2 distantes de 450kms

3 October 1990 *02.00 U.T.*

DISCUSSION

Discussor's name: P. Cannon

Comment/Question:

You explained that your system operated in real-time. Could you please indicate how long the computations take?

Author/Presenter's reply:

Le système présenté a été implanté par une firme industrielle dans un prototype opérationnel qui établi la prévision en temps réel. Le temps de traitement pour déduire du spectre la fréquence de trafic la plus probable est inférieur à la seconde. Il faut ajouter le temps de relevé du spectre qui peut être très réduit puisque le système doit scruter un nombre réduit de bandes de fréquences, de l'ordre de 85, ce qui peut être effectué, compte tenu de la faible résolution nécessaire, en moins de 30 secondes.

(Translation:

The system shown was installed by an industrial company in an operational prototype which established a forecast in real-time. The processing time required in order to deduce the most likely frequency of trafic from the spectrum is less than one second. You must add the spectrum read time which can be very short since the system must examine a reduced number of frequency bands, of the order of 85, which given the low resolution necessary, can be carried out in less than 30 seconds.)

Evaluation de la bande de cohérence du canal de transmissions VHF/UHF longues distances

G. AIGNEL *, O. RAVARD *, D. SORAIS **, L. BERTEL *

* Laboratoire Radiocommunications
URA CNRS 834, Campus de Beaulieu
Université de Rennes 1
35 042 Rennes Cedex, France

** THOMSON-CSF/RGS
66, rue du Fossé Blanc
92 231 Gennevilliers Cedex, France

Abstract:

This paper deals with the problem of coherence bandwidth of troposcatter radio links in the VHF *frequency range. Partial reflections (specular and diffuse) on tropospheric sheets (feuillets) is the propagation mechanism considered. A propagation model (*ASTRAL*), which considers the appearance of such rough surfaces in the common volume, is then used. For a given* VHF *link, a numerical integration allows the determination of the received power. In this paper, we show that modelling the troposcatter propagation in such a way allows the evaluation of the coherence bandwidth of the channel by calculating the power delay profile. Some numerical examples are given and discussed.*

1 INTRODUCTION

Pour la gamme de fréquences VHF militaire (30 - 88 MHz), les liaisons longues distances ($\geq$ 100 km) peuvent être réalisées via la diffusion sur les trainées ionisées produites par les météores (80-120 km d'altitude) ou via la diffusion par les hétérogénéités de l'indice de réfraction de la troposphère (< 10 km d'altitude). Bien que la gamme VHF n'ait pas été classiquement exploitée pour effectuer des liaisons par diffusion troposphérique, nous avons expérimentalement montré que des portées inférieures à 400 km peuvent être atteintes en utilisant des antennes élémentaires (dipôles) et une puissance d'émission de 1 KW environ ([11]).

En utilisant un protocole de transmission adapté au caractère intermittent d'une telle liaison, on obtient des performances supérieures à celles du canal météorique, quelle que soit la période de l'année [13]. Ce résultat signifie simplement que la diffusion troposphérique est dominante pour les courtes distances ($d \leq 400$ km) lorsque les antennes d'émission et de réception sont placées sur des mâts suffisamment hauts (20 m par exemple).

On souhaite donc estimer théoriquement la fonction d'autocorrélation en fréquence du filtre de canal associé à la diffusion troposphérique VHF afin d'obtenir les ordres de grandeur de la bande de cohérence et de la bande de décorrélation du canal. Pour atteindre cet objectif, nous exploitons un modèle de propagation déjà publié [10]. Nous exposons ensuite un certain nombre de résultats de bandes de corrélation pour la gamme VHF. Concernant la gamme UHF, nous comparons les estimations théoriques de notre modèle aux résultats disponibles dans la littérature.

2 DEFINITION DES BANDES DE CORRELATION D'UN CANAL

2.1 Fonctions d'autocorrélation en fréquence

Classiquement, la fonction d'autocorrélation en fréquence est définie par la relation :

$$R_H(\Delta f) = \frac{1}{2}\langle H^*(f,t)H(f+\Delta f,t)\rangle \qquad (1)$$

La fonction aléatoire $H(f,t)$ représente la fonction de tranfert équivalente basse fréquence du canal. Lorsque le canal est stationnaire au sens large en temps et en fréquence (classe des canaux WSSUS [4]), la fonction $R_H(\Delta f)$ est égale à la transformée de Fourier du profil puissance - retard $u(\tau)$:

$$R_H(\Delta f) = \int_{-\infty}^{+\infty} u(\tau)e^{-j2\pi\Delta f\tau}d\tau \qquad (2)$$

Les bandes de corrélation cherchées seront définies à partir du coefficient de corrélation ρ_H lié à $R_H(\Delta f)$:

$$\rho_H = \frac{|R_H(\Delta f)|}{|R_H(0)|} \qquad (3)$$

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

Les expérimentations réalisées dans les années 70 mesuraient la fonction d'autocorrélation d'enveloppe $R_E(\Delta f)$:

$$R_E(\Delta f) = \frac{1}{2}\langle |H(f,t)|.|H(f+\Delta f,t)|\rangle \qquad (4)$$

On peut montrer ([8], [5]) que le coefficient de corrélation ρ_E associé à $R_E(\Delta f)$ s'exprime en fonction de ρ_H :

$$\rho_E = \frac{(1+\rho_H)E\left[\frac{2\sqrt{\rho_H}}{1+\rho_H}\right] - \frac{\pi}{2}}{2-\frac{\pi}{2}} \qquad (5)$$

E représente l'intégrale elliptique complète de seconde espèce.

Le développement limité de cette fonction fournit une très bonne approximation de l'expression exacte pour toutes les valeurs de ρ_H comprises entre 0 et 1 :

$$\rho_E \simeq 0.915\,\rho_H^2 + 0.0572\,\rho_H^4 \qquad (6)$$

Ce résultat justifie la relation $\rho_E \simeq \rho_H^2$ implicitement utilisée dans différents articles [6].

Pour établir la relation (5), il est nécessaire de supposer que les enveloppes complexes, liées aux différents échos discernables reçus, sont des fonctions aléatoires gaussiennes complexes et mutuellement indépendantes.

2.2 Bandes de corrélation

La bande de corrélation $B_{cor}(\alpha)$ est estimée par la relation :

$$\begin{cases} \rho_H[B_{cor}(\alpha)] = \alpha \\ 0 < \alpha < 1 \end{cases} \qquad (7)$$

D'une manière générale, la bande de cohérence représente l'écart maximum en fréquence entre deux porteuses pour lequel les évanouissements des signaux reçus demeurent suffisamment corrélés.

En présence de multitrajets, la superposition temporelle des symboles reçus (évanouissement sélectif) produit un taux d'erreur irréductible τ_{ir} qui impose une limite à la vitesse de modulation R lorsque l'on réalise une transmission numérique classique. Pour des canaux WSSUS complexes gaussiens et pour des modulations de phase à deux états, on peut estimer la valeur de R pour laquelle un taux d'erreur irréductible de 10^{-3} est atteint. ([3], [1], [7]) :

$$R_{max}(Kbit/s) = \frac{K_R}{\sigma_t(\mu s)} \qquad (8)$$
$$K_R \simeq 50 - 80$$

σ_t : étalement temporel dû aux multitrajets.

L'expérience montre que la valeur de la bande de corrélation à 0.9 est proche de la valeur de R_{max} associée à un taux d'erreur égal à 10^{-3} ou quelques 10^{-3}. Il est donc logique de définir la bande de cohérence B_c par la relation :

$$\rho_H[B_c] \simeq 0.9 \qquad (9)$$

Dans la littérature, cette bande de cohérence est fréquemment définie pour des valeurs de α égales à $1/e$ ou 0.5. En fait, la bande passante obtenue pour $\alpha = 0.5$ est suffisante pour que les techniques de diversité soient efficaces : on obtient ainsi une définition possible de la bande de décorrélation B_{de} du canal :

$$\rho_H[B_{de}] \simeq 0.5 \qquad (10)$$

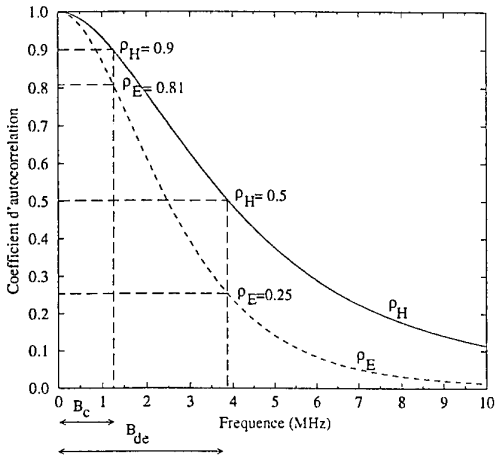


FIG. 1 - *Définition des bandes de corrélation*

Si l'on souhaite obtenir des définitions équivalentes pour des bandes estimées avec le coefficient de corrélation d'enveloppe, il est évidemment nécessaire de prendre en compte la relation $\rho_E \simeq \rho_H^2$.

3 MODELE DE PROPAGATION EXPLOITE

Les niveaux de champ reçus lors de liaisons troposphériques transhorizons peuvent s'interpréter par la présence d'irrégularités de l'indice de réfraction. Lors d'une liaison radioélectrique, seules les irrégularités dont la taille est en relation d'adaptation avec les vecteurs d'onde incident et réfléchi (analogie avec la diffusion de Bragg) contribuent à l'énergie reçue. Lorsque les fréquences de la gamme VHF sont utilisées, les irrégularités sélectionnées se situent dans la zone anisotrope du spectre de l'indice de réfraction. Cette anisotropie peut être décrite par la présence de feuillets troposphériques, caractérisés par une discontinuité du gradient de l'indice de réfraction entre deux altitudes h et $h + \Delta h$. La discontinuité s'étale, horizontalement, sur des dimensions très grandes devant l'épaisseur Δh du feuillet (figure 2).

Le principe de la modélisation, qui a été décrit dans [10], repose sur les bases suivantes : à un instant fixé t_0, on suppose que le signal reçu $S(t_0)$ est dû à des réflexions sur des

feuillets troposphériques dont la localisation est aléatoire dans le volume commun de la liaison. Le signal reçu peut donc s'écrire sous la forme suivante :

$$S(t_0) = \sum_{i=1}^{n} S_i e^{j\phi_i} \qquad (11)$$

où S_i est le module du signal associé à chaque réflexion et ϕ_i est la phase associée.

La puissance du signal reçu est donnée par la somme des puissances des signaux associés à chaque réflexion, si l'on considère l'indépendance des phases ϕ_i ainsi que leur équirépartition dans l'intervalle $[0, 2\pi]$. L'expression finale de la puissance reçue peut alors se mettre sous une forme intégrale, laquelle fait apparaître la probabilité d'apparition par unité de volume $P_a(P)$ d'un feuillet localisé par un point P du volume commun :

$$P_r = \langle SS^* \rangle = \int_V \tilde{P}_r(P) P_a(P) dV \qquad (12)$$

où $\tilde{P}_r(P)$ est la puissance réfléchie par un feuillet localisé en P, et V est le volume commun de la liaison.

Les feuillets sont considérés comme des surfaces rugueuses définies par l'écart type σ et la longueur de corrélation T. L'évaluation de $\tilde{P}_r(P)$ peut s'effectuer à l'aide de la formulation de Kirchhoff associé au problème de la réflexion d'une onde sur une surface rugueuse, tout en tenant compte de l'ensemble des caractéristiques aléatoires des feuillets troposphériques : le gradient d'indice défini par Δn, l'épaisseur Δh, et les dimensions horizontales L_x et L_y. Le résultat obtenu est le suivant :

$$\tilde{P}_r(P) = K f^2(\theta_1, \theta_2, \beta) \langle RR^* \rangle I_m \qquad (13)$$

où K est un coefficient faisant intervenir, entre autres, les gains d'antennes; $f(\theta_1, \theta_2, \beta)$ est une fonction dépendant de l'angle d'incidence θ_1 et des angles d'observation θ_2,β; R est le coefficient de réflexion de Fresnel déterminé, pour un feuillet donné, à partir de l'épaisseur Δh de celui-ci. Si l'on considère une densité de probabilité exponentielle décroissante pour Δh, de moyenne H, le coefficient de réflexion de Fresnel quadratique moyen est donné par :

$$\langle RR^* \rangle = \frac{\langle \Delta n^2 \rangle}{4Hk^2a^6} \times \left[ka \arctan(2kaH) - \frac{1}{4H} \ln(1 + 4k^2a^2H^2) \right] \qquad (14)$$

et $a = \cos(\theta_1)$.

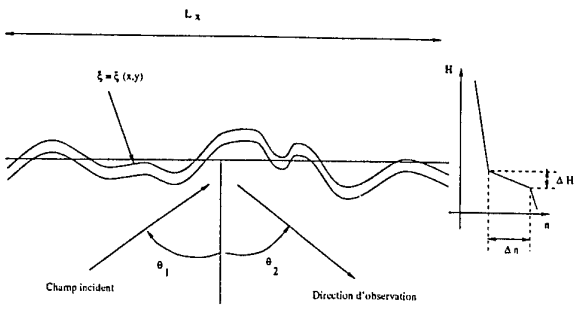


FIG. 2 - *Définition des paramètres décrivant un feuillet : épaisseur Δh, gradient d'indice défini par Δn, dimension longitudinale L_x. La surface rugueuse ξ associée est définie par l'écart type des rugosités σ et sa longueur de corrélation T. L'angle θ_1 définit l'angle d'incidence de l'onde et θ_2 l'angle d'observation dans le plan d'incidence.*

L'intégrale I_m se décompose en une contribution I_s due à la réflexion spéculaire (feuillet plan) et une seconde I_d due à la réflexion diffuse :

$$I_m = I_s + I_d \qquad (15)$$

La réflexion diffuse sur la surface rugueuse s'évalue d'une manière classique [2] tandis que le calcul de I_s doit s'effectuer au deuxième ordre pour tenir compte du fait que les dimensions des feuillets sont du même ordre de grandeur que celles des zones de Fresnel. Pour une taille de feuillet donnée, de dimensions longitudinale L_x et transverse L_y, on obtient :

$$I_{s/L_x,L_y} = \frac{\lambda^2}{16\, k_3^2 k_4^2} \Pi(k_1, k_2, L_x) \Pi(k_2, k_4, L_y) \qquad (16)$$

avec

$$\Pi(a, b, L) = \left| \Phi\left[\frac{2b}{\sqrt{\lambda}} (L + \frac{a}{2b^2}) \right] + \Phi\left[\frac{2b}{\sqrt{\lambda}} (L - \frac{a}{2b^2}) \right] \right|^2$$

où $\Phi = C + iS$, C et S sont les intégrales de Fresnel cosinus et sinus, k_1, k_2, k_3 et k_4 sont des constantes s'exprimant en fonction des angles θ_1, θ_2 et β.

La moyenne I_s s'obtient en intégrant numériquement l'expression (16) pour toutes les valeurs probables de L_x et L_y en considérant une densité de probabilité gaussienne.

Finalement, la détermination théorique de la puissance reçue lors de liaisons troposphériques longues distances s'effectue en évaluant la relation (12) par intégration numérique à travers le volume commun. Les diagrammes de directivité des antennes ainsi que certains problèmes classiques de diffraction sont pris en compte par l'intermédiaire de la constante K de la relation (13).

Ce modèle de propagation, dénommé ASTRAL, a été confronté à des résultats expérimentaux obtenus dans la

gamme de fréquence VHF basse [11]. Les trois liaisons étudiées correspondent à des distances comprises entre 300 et 500 km. La figure 3 représente le diagramme de dispersion entre les résultats numériques et les résultats expérimentaux obtenus. On observe une concordance satisfaisante entre théorie et expérience.

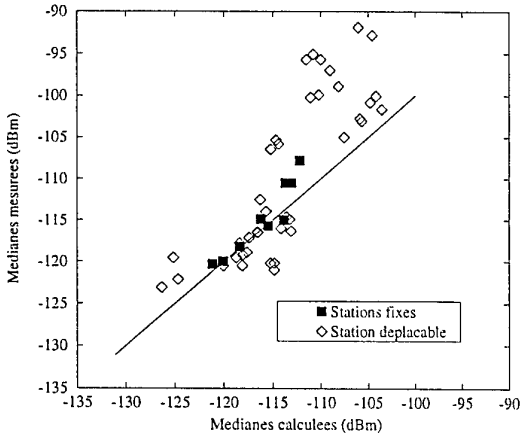


FIG. 3 - *Dispersion médianes mesurées — médianes calculées. Expérimentation du 1/10/92 au 1/10/93*

4 CALCUL DU PROFIL PUISSANCE - RETARD

4.1 Interprétation physique

L'analyse des variations temporelles de la puissance totale reçue montre que le nombre de diffuseurs simultanément présents est réduit (2 ou 3, voir figure 4) et que, sur un intervalle de temps de l'ordre d'une minute, le signal reçu n'est pas stationnaire (figure 5).

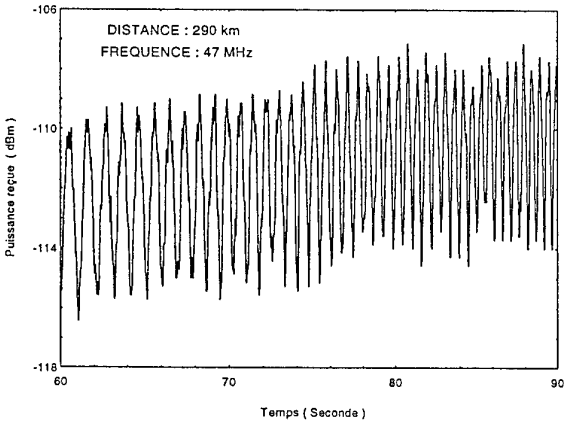


FIG. 4 - *Exemple de signal reçu*

Il est cependant possible de sélectionner des intervalles temporels, non nécessairement adjacents, pour lesquels la stationnarité au sens large est acquise. On pourrait donc réaliser une expérimentation permettant d'évaluer N_p profils puissance – retard $u_j(\tau)$ liés à ces intervalles.

Dans le cadre du modèle de canal précédemment décrit, une fonction $u_j(\tau)$ peut être vue comme une réalisation d'une fonction aléatoire décrite par un ensemble de paramètres physiques considérés comme des variables aléatoires. En supposant que les N_p profils représentent correctement les réalisations statistiques du modèle (voir paragraphe 3), la moyenne d'ensemble calculée peut être simplement estimée par la relation :

$$\langle u(\tau) \rangle = \frac{1}{N_p} \sum_{j=1}^{N_p} u_j(\tau) \tag{17}$$

Le modèle permet donc d'évaluer un profil puissance – retard moyen expérimentalement mesurable.

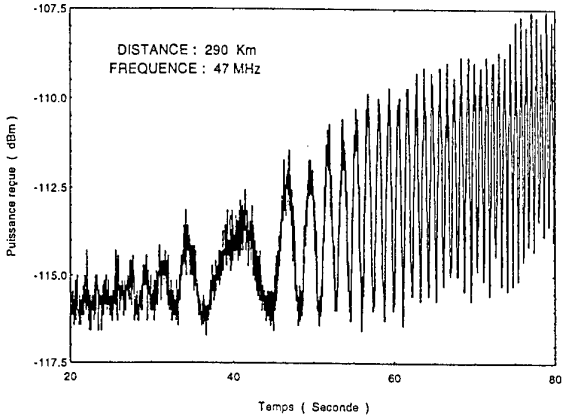


FIG. 5 - *Exemple de signal reçu*

4.2 Méthode d'évaluation

Il s'agit d'exploiter le modèle de propagation précédemment décrit en montrant que le profil puissance – retard peut être déterminé d'une manière rigoureuse. Pour cet objectif, on utilise le fait que la puissance reçue entre les instants τ et $\tau + d\tau$ est due à l'ensemble des diffuseurs situés entre les deux ellipsoïdes de foyers constitués par les points d'émission et de réception, de demi petits axes b et $b + db$ tels que :

$$b = \sqrt{\frac{\tau^2 c^2}{4} - F^2} \tag{18}$$

où F est la demi distance focale et c la vitesse de la lumière.

La figure 6 représente la géométrie du problème. Si $P_r(b)db$ est la puissance reçue due aux diffuseurs situés entre ces deux ellipsoïdes, on obtient :

$$P_r(b)db = u(\tau)d\tau \tag{19}$$

Soit V_b le volume compris entre les deux ellipsoïdes. De la même manière que dans le paragraphe 3, $P_r(b)db$ s'évalue en introduisant la probabilité $P_a(P)$ d'apparition d'un diffuseur par unité de volume :

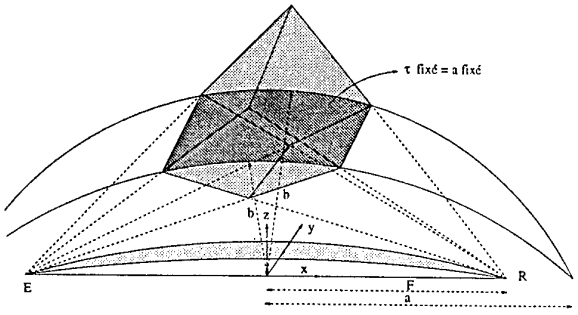


FIG. 6 - *Géométrie du problème*

$$P_r(b)db = \int_{V_b} \tilde{P}_r(P)P_a(P)dV_b \qquad (20)$$

On définit S_b la surface ellipsoïdale associée à l'ellipsoïde de demi petit axe b. La relation (20) se simplifie alors :

$$P_r(b) = \int_{S_b} \tilde{P}_r(P)P_a(P)dS_b \qquad (21)$$

L'expression du profil puissance – retard $u(\tau)$ s'exprime finalement par la relation :

$$u(\tau) = \frac{c}{2b}\sqrt{b^2+F^2}\int_{S_b} \tilde{P}_r(P)P_a(P)dS_b \qquad (22)$$

Cette expression signifie que le profil puissance – retard s'évalue en intégrant la relation (13) à travers les différentes surfaces ellipsoïdales S_b. Du point de vue numérique, pour évaluer cette expression, il est pratique d'utiliser les variables d'intégration x et y du plan S_{xy} perpendiculaire au plan du grand cercle et contenant les points d'émission et de réception (voir figure 6).

5 RESULTATS OBTENUS DANS LA GAMME VHF

Les calculs ont été réalisés pour la fréquence de 40 MHz et des antennes d'émission – réception identiques, constituées d'un dipôle placé sur un mât de 18 mètres de haut. On rappelle qu'il est nécessaire d'élever les antennes au-dessus du sol afin d'obtenir un maximum de rayonnement suffisamment bas sur l'horizon pour que les mécanismes de propagation exploités soient efficaces. Les profils puissance – retard ont été calculés pour les distances de 300, 400 et 500 km (figure 7). Ces fonctions ont été évaluées en prenant en compte les gains des antennes d'émission et de réception. La fonction $u(\tau)$ étant supposée normée pour obtenir une densité de probabilité $\breve{u}(\tau)$, on définit l'étalement temporel par l'expression :

$$\sigma_t = \left[\int_0^\infty \tau^2 \breve{u}(\tau)d\tau - \left[\int_0^\infty \tau\breve{u}(\tau)d\tau\right]^2\right]^{\frac{1}{2}} \qquad (23)$$

Les coefficients de corrélation en fréquence, directement obtenus à partir des profils puissance – retard, permettent de calculer les bandes de corrélation (figure 8). On remarque que, pour ces configurations de liaisons, les bandes de corrélation, à une distance d_2, peuvent être évaluées à partir de la connaissance des bandes de corrélation à la distance d_1 avec la relation empirique :

$$B_{cor}(d_2) = B_{cor}(d_1)\frac{d_1}{d_2} \qquad (24)$$

La figure 9 permet d'obtenir une représentation simultanée des évolutions avec la distance de l'étalement temporel σ_t et des bandes de corrélation B_c et B_{de}.

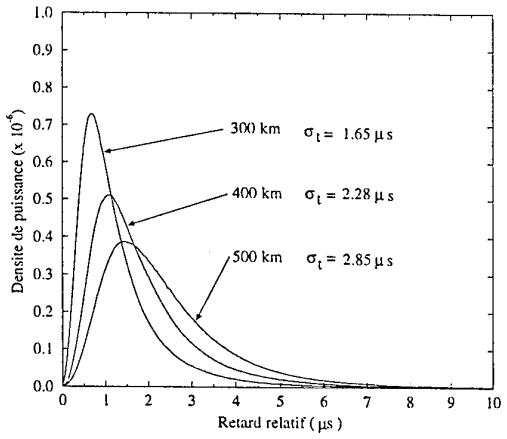


FIG. 7 - *Profils puissance – retard pour trois liaisons à 300, 400 et 500 km. La fréquence utilisée est de 40 MHz et les antennes sont contituées de dipôles placés à 18 m au-dessus du sol.*

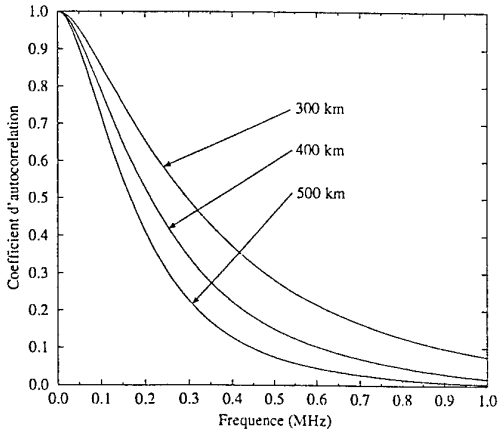


FIG. 8 - *Fonctions de corrélation en fréquence pour trois liaisons à 300, 400 et 500 km. La fréquence utilisée est de 40 MHz et les antennes sont contituées de dipôles placés à 18 m au-dessus du sol.*

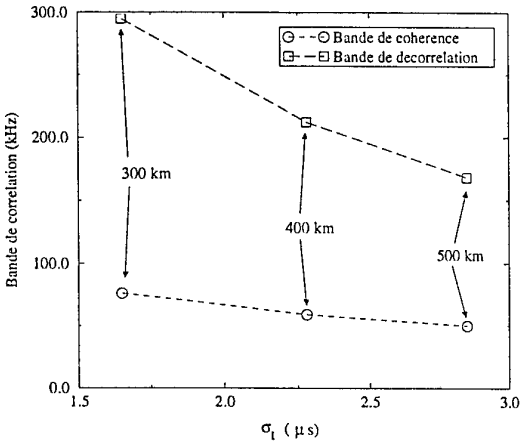


FIG. 9 - *Bandes de cohérence et bandes de décorrélation fonction de l'écart type du profil puissance – retard. La fréquence utilisée est de 40 MHz et les antennes sont contituées de dipôles placés à 18 m au-dessus du sol.*

6 DISCUSSION DES RESULTATS

Le problème de la validation des résultats obtenus est délicat à traiter puisque, actuellement, aucune mesure de bandes de corrélation n'est disponible dans la gamme VHF.

Cependant, la validation du modèle de propagation exploité dans cette gamme de fréquences donne de sérieuses indications sur la qualité des calculs de bandes de corrélation effectués. En effet, les bandes de corrélation sont déterminées à partir du calcul du profil puissance – retard, lequel est fixé par la répartition spatiale de l'efficacité des diffuseurs potentiellement présents (variation de $\tilde{P}_r(P)$ dans le volume commun). Or, cette répartition détermine (après intégration) le bilan de liaison théorique et, par conséquent, la variation paramétrique de celui-ci (variation en fonction de la distance, la fréquence, les hauteurs d'antennes, etc ...). Les études passées [10] ont montré que le modèle de propagation offre des variations paramétriques comparables à celles données par d'autres modèles (NBS 101 [12]). La confrontation du modèle à des résultats expérimentaux (figure 3) a permis de confirmer ce résultat.

On peut donc raisonnablement penser que le calcul de la répartition spatiale de l'efficacité des diffuseurs potentiels est correcte dans une gamme de fréquences au moins égale à celle considérée lors de la validation expérimentale du modèle.

Le réalisme des profils puissance – retard calculés pour la gamme de fréquences VHF apparaît donc comme une implication directe de cet état de fait.

En dehors de cette gamme de fréquences, un certain nombre de résultats expérimentaux sur les bandes de corrélation sont disponibles dans la littérature [6], [9]. Cependant, tous ne sont pas utilisables pour notre objectif de validation car il faut veiller à ce que les fréquences utilisées se situent dans le domaine de validité du modèle de propagation (réflexions partielles).

	fréquence	distance	Gain d'antenne
liaison 1	900 MHz	271 km	36.5 dBi
liaison 2	4.6 GHz	199 km	41.7 dBi

	mesures	théorie
liaison 1	1.59 MHz	2.55 MHz
liaison 2	1.8 MHz	2.75 MHz

TAB. 1 - *Comparaison théorie – mesures de la bande de corrélation $B_{cor}(0.63)$. Les résultats expérimentaux sont fournis dans [6] et [9]*

Le tableau 1 donne les bandes de corrélation, théorique et expérimentale, pour $\rho_H = 0.63$, associées à deux liaisons à 900 MHz et 4.6 GHz. Ne disposant pas de toutes les données relatives à ces expérimentations, nous ne nous intéressons qu'à l'ordre de grandeur des résultats obtenus.

Les mesures présentées dans [6] et [9] sont relatives à la corrélation d'enveloppe. Pour comparer aux valeurs théoriques de bande de corrélation, le coefficient de corrélation ρ_H a été évalué en utilisant la relation $\rho_E = \rho_H^2$. L'accord théorie – expérience est satisfaisant.

7 CONCLUSION

Pour le bas de la gamme VHF, les mécanismes de propagation présents dans les liaisons troposphériques transhorizons sont différents de ceux des gammes UHF / SHF. Afin de prendre en compte ces mécanismes, le modèle physique de canal ASTRAL a été développé et validé. Nous avons modifié ce modèle afin d'obtenir les profils puissance – retard à partir desquels les bandes de corrélation du canal peuvent être estimées. Les valeurs obtenues permettront d'optimiser les caractéristiques de transmissions numériques.

Références

[1] Bailey and Lindenlaub. Further results concerning the effects of frequency selective fading on differentially coherent matched filter receivers. *IEEE Transactions on Communications Technology*, October 1968.

[2] P. Beckman and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces*. Pergamon Press, 1963.

[3] Bello and Nelin. The effect of frequency selective fading on the binary error probabilities of incoherent and differentially coherent matched filter receivers. *IEEE Transactions on Communications Systems*, cs-11, June 1963.

[4] P. Bello. Characterization of randomly time-variant linear channels. *IEEE Transactions on Communications Systems*, 11, December 1963.

[5] H. Booker, J. Ratcliffe, and D. Shinn. Diffraction from an irregular screen with applications to iono-

pheric problems. *Phil. Transactions Royal Socity London*, 242, September 1950.

[6] C. Collin. Evaluation empirique de la bande de cohérence en diffusion troposphérique. Technical report, Revue Technique THOMSON-CSF, September 1979.

[7] Garber and Pursley. Performances of differentially coherent digital communications over frequency selective fading chanels. *IEEE Transactions on Communications*, January 1988.

[8] Jakes. *Microwave Mobile Communications*. John Wiley and Sons, New York.

[9] D. Kennedy. A comparison of measured and calculated frequency correlation functions over 4.6 and 7.6 GHz troposcatter paths. *IEEE Transactions on Communications*, com-20(2), April 1972.

[10] O. Ravard and F. Chevrier. ASTRAL : Un modèle de propagation VHF troposphérique transhorizon. In *Propagation électromagnétique dans l'atmosphère du décametrique à l'angström*. SEE, mars 1994.

[11] O. Ravard, F. Chevrier, L. Bertel, and J. L. Jannic. Etude des variations rapides et lentes de la puissance reçue lors de liaisons VHF troposphériques. In *CP-543*. AGARD, 1993.

[12] P. Rice, A. Longley, K. Norton, and A. Barsis. Transmission loss predictions for tropospheric communications circuits. Technical report, NBS, 1967.

[13] D. Sorais, O. Ravard, and L. Bertel. Analyse des résultats expérimentaux de liaisons météoriques réalisées pour différentes distances et fréquences. In *CP-486*. AGARD, 1990.

Communications and the Global Positioning System

Jules Aarons, Michael Mendillo, and Robert Yantosca
Center for Space Physics
Boston University
725 Commonwealth Ave.
Boston, MA 02215
USA

1. SUMMARY

Position data obtained by use of the Global Positioning System form basic components in a myriad of military and commercial systems. These range from navigating harbors and oceans to determining aircraft, rocket, and satellite positions. The users of GPS envisage automatic and routine communication of results of position data for control and location of resources. Propagation studies to support these needs require the evaluation of problems of finding the position of ground, naval and airborne units and then communicating the data. The area addressed by this paper outline the problems of global fading of signals for both the GPS system and the satellite communication systems at 250 MHz which are used to communicate and integrate results. For GPS frequencies (1.2 GHz and 1.6 GHz) difficulties will be encountered in the auroral region during magnetic storms during all phases of the sunspot cycle. The equatorial and polar regions will become important by the end of the millenium when the sun will exhibit maximum activity. The data base for evaluating the effects of the ionosphere on gigahertz scintillation is small but will be outlined in this paper. The effects of F layer ionospheric irregularities on communication links at 250 MHz are considerable even during the current solar minimum. The data base for ionospheric effects on satellite communication systems at 250 MHz is large and will be only briefly outlined; much of the material is in various publications. It should be noted that transmissions for the IRIDIUM series of satellites are in the same frequency band as the GPS transmissions and will be affected by F layer scattering.

2. INTRODUCTION: THE MANY USES OF THE GLOBAL POSITIONING SYSTEM

Everyone in communications and navigation is fully aware of the enormous growth of useage of GPS, ranging from surveying to plotting hiking trails in snow areas. Recently for example the Argentine government in order to control fishing in their territorial waters would require positions of fishing boats to be automatically reported to a surveillance center. These uses of GPS with relatively unsophisticated equipment might be subject to receiver problems in certain regions of the globe. In addition this paper will be relevant to INMARSAT and proposed L Band satellite systems for cellular phones; these systems might encounter similar ionospheric problems as those described.

As the Global Positioning System signals at 1.6 and 1.2 GHz traverse the ionosphere, scattering of energy takes place due to irregularities in the ionosphere, primarily at heights from 200-600 km. The scattering produces both fading and enhancements of the signal. The primary regions of the globe of concern to GPS users include auroral latitudes during severe magnetic storms, and both polar latitudes and equatorial regions during solar maximum years. The regions encumbered by the phenomenon are best shown by the updated schematic of Figure 1; significant scintillation fading depth is depicted during solar maximum years.

There is an enormous difference between the fading experienced in years of solar maximum compared to years of solar minimum. In one area, the anomaly region of the equator, there should be relatively sparse periods of significant scintillation experienced on GPS receivers during the lowest part of the sunspot cycle. Deep amplitude fading will exist in years of high solar flux and it will exist in those years at high latitudes. With the sunspot cycle having peaked in 1989-1990 and at a minimum in 1995, there is limited interest at this time. Beginning with 1999, with the advent of the third millennium, fading will once again become of importance in the equatorial and high latitude regions.

Scintillation occurrence of magnitude significant for GPS users is NOT encountered often during daylight hours in the equatorial region. In middle latitudes on occasion under specific propagation conditions, scintillation at GPS frequencies does occur; it is relatively rare.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

3. THE NATURE OF SCINTILLATION ACTIVITY

One can observe fading at GPS frequencies (1.2 and 1.6 GHz) by examining the amplitude of GPS signals and satellite transmissions at L Band and higher. In Figure 2 fading at 1.5 GHz is shown. When the fading falls into noise, bits will be lost in the data stream. How these effects will affect each GPS receiver acquisition capability and its positioning accuracy is not well known. There are two phases of reception. One is acquisition in which a stream must be matched in the receiver. For this function, loss of digits could be a problem. For evalution of the fading's impact on acquisition, the time spent below the noise level is the crucial factor; the fading spectrum must be evaluated. Once the signal is acquired, a flywheel technique may keep the signal as a data source even with amplitude fading.

In addition to the effects of amplitude fading, the rate of change of phase will at times bring the signal beyond the narrow bandwidths being used. How the phase fluctuations affect the signal of a particular receiver in tracking mode and in terms of error and position fluctuations is not known. In any event the characteristics of the receiver are quite relevant to a particular use. Small hand held receivers with relatively low signal to noise ratios may be affected differently than those with more gain in the system. Even the number of satellites used by any receiver is relevant since in some areas such as the equatorial anomaly, some ground to satellite propagation paths have little scintillation on the signal while others may produce 20 dB fades.

4. THE EQUATORIAL REGION

Figure 1 represents a late revision of a pattern that has been presented before. The areas where problems might be encountered are the polar, auroral and the equatorial regions. We shall concentrate on equatorial effects. Probably the area to be most seriously affected is the region near the magnetic equator. The equatorial region has to be divided into two areas, i.e., the area near the magnetic equator and the equatorial anomaly region. The latter encompasses a swath of over 5 degrees centered on a magnetic latitude of 15 degrees North and South of the magnetic equator. Deep fading over many hours has been experienced in the anomaly region.

The ionospheric structure principally responsible for the scattering is a plume of irregularities extending above the magnetic equator. At the magnetic equator, just before sunset, the F layer peak in electron density rises in altitude. Its descent, in the post-sunset time period, marks the time of development of the plume which contains irregularities of the order of 1 kilometer and smaller. The plume develops from a thin layer into a layer which at times extends to beyond 1500 km. Locally generated irregularities commence roughly an hour after sunset but irregularities at L band continue up to midnight for the most part. Horizontal winds bring the plumes into the field of view from regions to the west.

The turbulence generated at the magnetic equator maps along the earth's magnetic field lines to the north and to the south. Irregularities are produced at the maximum height of the F layer at latitudes distant from the magnetic equator. The effect of the plumes is felt along the lines of force of the earth's field so that at distant latitudes irregularities develop. For example what is above the magnetic equator at 700 km produces irregularities at the 300 km altitude at a distance from the magnetic equator.

Once formed the plumes move eastward with velocities initially of the order of 100-200 meters per second. At anomaly latitudes, the depletions occasionally bifurcate as will be shown in a later diagram. Even during periods of intense irregularity development the separation of irregularity patches will allow gaps in the anomaly regions.

The most striking data set on gigahertz scintillations for the anomaly region comes from Fang and Liu, 1982. In Figures 3a and 3b, fluctuations at 4 GHz up to 4 dB in fades and 7.5 dB peak to peak are shown, along with the percentage of occurrence of scintillations >2 dB for years of high solar flux.

From Hong Kong records several facets of irregularity development can be found. Figure 3a shows peak to peak excursions of received signals and include both enhancement of signal and fading. Peak to peak excursions of 6 dB at 4 GHz means that excursions at 1.6 GHz will be over 30 dB peak to peak. The top panel records the down link in Hong Kong transmissions from a satellite over the Indian Ocean; the bottom panel records are of satellite reception of a satellite that is over the Pacific Ocean. This particular night was active but the dimensions and intensity of the irregularity structure differ on the two paths.

Near 110 degrees E, Frank and Liu (1982) in Figure 3b report that the occurrence of 4 GHz scintillation activity is predominantly in the equinoctial months. From Figure 3b one can note that the percentage occurrence increases as we go to higher solar flux from 1978 to 1980. Secondly the Indian Ocean link shows lower occurrence than the Pacific Ocean Link. It should be noted that as of this time there are no absolute measurements in this field. Various

parameters, propagation angle, contamination with tropospheric activity etc. have to be evaluated before absolute levels are available.

To illustrate effects in the gigahertz portion of the spectrum, results from a campaign of 19 days to Ascension Island in the solar maximum year of 1981 will be shown. Ascension Island is located at 15 degrees West Longitude and is in the anomaly region. The aim of the campaign was to study the effects of the anomaly region during a year of high solar flux. The solar flux during the period was between 180 and 210, a high figure although later in the year there were flux levels of 250-300 reported for a period of days.

A single plume at latitudes distant from the magnetic equator appears often as a series of patches. One method of determining patch dimensions stems from all-sky recordings of the depletions of the normal sky airglow. In the case used (Figure 4a) observations were of the 6300A line of oxygen. The optical images of these observations (Mendillo and Tyler, 1983) can be related to patch parameters. In Figure 4a one can see the projection of the image to heights near the magnetic equator. It can be seen that what was a single patch on the magnetic equator has at the anomaly latitudes broken into several patches. In Figure 4b (Aarons et al., 1983) the path of GPS Satellite 9 is shown relative to the depletion patches. Figure 4c shows scintillation levels of each of the GPS satellites as recorded at Ascension Island. GPS signals were reduced in 15 minute segments. Scintillation levels were relatively high along the path (Figure 4c). Two methods of reduction are commonly used for scintillation data, i.e. the S4 method where the ratio of the mean square deviation of the signal to the average level is used and SI where peak to peak values are noted every 15 minutes. In these data an S4 measure of 1 is equivalent to 22 dB peak to peak and an S4 level of .6 is equivalent to a 10 dB peak to peak. Synchonous satellite signals were available from MARISAT transmissions at 1.54 GHz and 3.95 GHz. Levels reached over 20 dB peak to peak at 1.54 GHz.

The patches move eastward across the propagation path to the satellites with a velocity ranging from two hundred meters per second after sunset to 50 meters per second or less at the end of the night. At anomaly latitudes the periods between patches where only minor fluctuations of signal will be experienced can be used for acquisition of signals.

5. AURORAL EFFECTS

Using the data base of the International GPS Geodetic Service, the rapid fluctuations of Total Electron Content levels are ascribed to phase scintillations produced by ionospheric irregularities. At auroral latitudes Doherty et al., 1994 have shown that with this technique, phase fluctuation activity exhibits much lower levels at middle latitudes than at high latitudes. They showed that the rapid rate of change can severely limit single frequency correction capability.

Clynch (1994) noted loss of lock of signals at high auroral latitudes during the severe magnetic storm on May 10, 1992. The sites for the measurements were at 74 degrees Corrected Geomagnetic Latitudes (the South Pole) and at 80 degrees Corrected Geomagnetic Latitudes (McMurdo Sound). The South Pole receiver showed loss of lock for a 20 minute period for differing groups of satellites. The McMurdo Sound receiver with its location considered to be in the polar region did not have such a large number of outages. A less intense magnetic storm did not produce the same effects at the South Pole.

6. POLAR SCINTILLATIONS

In the polar region, intense scintillations are noted in years of sunspot maximum with little activity reported in years of very low solar flux. At the communications frequency of 250 MHz, this has resulted in deep fading for many hours (Basu et al., 1987). For GPS frequencies during years of high solar flux there is variability with maximum activity from the fall equinox to the spring equinox but with a deep minimum in summer. Bishop et al., 1994 has observed GPS scintillation for a series of days in winter and in summer in a year of high solar flux.

For the polar region, scintillation should not be a major problem in the years of solar minimum. Both the auroral and polar regions may be of importance when intense magnetic storms occur even during periods of low solar flux. However for the solar maximum years, there will be considerable fading. At polar latitudes, in particular during hours of darkness, there is intense scintillation activity during the height of the sunspot cycle.

7. RECENT OBSERVATIONS OF GPS PHASE FLUCTUATIONS

Following the work of Doherty et al., 1994, and Wanninger, 1993, the Center for Space Physics of Boston University has developed a series of studies using the data from the International GPS Service.

One approach was to obtain phase fluctuations by noting rapid changes in TEC. In Figure 5a we can note fluctuations in TEC recorded in Arequipa, Peru (three degrees from the magnetic equator). The reported parameters are first processed using orbital elements. Secondly Total Electron Content data is obtained from the 30 second values given for this parameter; in this case it can be seen that a rapid change in TEC occurs in the 02-04 UT time period (21-23 LST). Finally the rate of change of TEC per minute is calculated. This can then be contrasted with fluctuations noted in other areas for the same period or with other data such as radar returns obtained in Jicamarca, Peru.

With data from Bogota, Columbia, north of the magnetic equator, the contrast between the Northern anomaly region and the Southern anomaly region can be noted. The reduced phase fluctuation data for a relatively quiet day from the viewpoint of anomaly phase fluctuations, January 12, 1995, is shown in Figure 5b. A very disturbed day from the viewpoint of TEC fluctuations (in this case a magnetically disturbed day) is shown in Figure 5c for January 18, 1995. The phase fluctuations reach the level of over 2 TEC units/minute. In this case the fluctuations occur towards sunrise.

Fluctuations in the auroral region during magnetic storms are very high; these can be continuous over several hours. Signal levels as observed at a sub-auroral latitude such as Algonquin, Canada (57° CGL) can during quiet times experience low level fluctuation. However during magnetic storms wide fluctuations are observed. In Figure 5d, March 29, a day of low magnetic activity is contrasted with April 4 when magnetic indices were very high.

A higher latitude station, Yellowknife (Figure 5e) might show TEC fluctuation activity at midnight during quiet magnetic conditions (January 15, 1995), but during storm conditions (January 17th) show strong fluctuations. The effects of the magnetic storm after commencement move equatorwards. A lower latitude sub-auroral site such as Algonquin in Canada only shows generation and motion effects of magnetic storm activity during the storm days.

Ny Alesund is one place for almost continuous phase fluctuations of relatively high level (Figure 5f; the fluctuations are maximum in this location for the period and stations surveyed.

By putting together Total Electron Content data from two stations at different latitudes but contiguous, we can contrast magnetic equator results with those from the anomaly region. The stations chose are Arequipa, Peru and Santiago, Chile. Meshing these two stations in local time allows the development of Figure 6. This diagram illustrates the anomaly increase as a function of latitude and as a function of time for April 1, 1994, a magnetically quiet day; the values are relative rather than absolute.

The total electron content and the structure of the anomaly region may be of importance in HF propagation. It might also be of importance in forecasting effects on satellite links below the frequencies of the GPS system; there is very little time delay.

8. POSSIBILITIES FOR FORECASTING AND PREDICTION OF EQUATORIAL IRREGULARITIES

Measurements of many sets of data, such as shown for Arequipa near the magnetic equator and Algonquin, Canada at a sub-auroral latitude should allow for a global picture of scintillation activity during quiet and disturbed magnetic conditions and during various seasons. In addition it points the way for a single station to warn in real time about conditions on various paths; this could serve as forecasting for an area. One receiver can record fluctuation levels at multiple propagation paths in the overhead sky.

In the equatorial region with this technique, the plumes which produce fading develop as the sunset line goes west. Once developed the patches which maintain their integrity for several hours move eastward. If one has the entire sky covered in real time by propagation paths to GPS satellites which could be used as sensors, short term forecasting might be possible. These sensors could record amplitude and phase scintillation and phase fluuctuations.

Once the equatorial plumes develop, these patches of irregularities move eastward thru the night. On a night with active scintillation activity, if the path to the east is in the sunset region and scintillations were observed, the night might be termed potentially an active night for the development of irregularities. In the hours after overhead sunset, if scintillations were observed on a western path, then it would be reasonable to expect the wind to bring

these patches eastward. One could then forecast for the latter part of the night that there would be scintillation activity on overhead paths. This is a simplification of the output which might emerge from a warning system of this type.

Exactly what effect the fluctuations in phase have on accuracy and on time to develop and change a position is not known. It would appear that with TEC values fluctuating, at times for periods of hours, there would be an effect on jitter in position.

GPS satellites offer a unique source for measurements of amplitude and phase scintillation on a global scale. The data can be used to study ionospheric plasma structures, develop weather models of scintillation and can be scaled in frequency to support many operational systems. There is much to be gained by long term measurements. In the Pacific region there are phenomena which may be different from what have been observed in the Atlantic. There is a sparcity of data in the African and Middle East sectors. In the region between Western North America and Hawaii, there is a dramatic difference in morphology between 76° West and 160° West. It would be useful to obtain data from areas where it is lacking. Absolute intensity of scintillation is also unavailable. The absence of evaluated fading levels means that we have no way of comparing fading depths between for example Hong Kong and the Ascension Island or the area conjugate to the Ascension Island in the Northern Hemisphere.

9. ACKNOWLEDGEMENTS

The studies of Jules Aarons are supported primarily by the Office of Naval Research. The authors thank Patricia Doherty and John A. Klobuchar for their assistance in the Center for Space Physics studies of GPS.

10. REFERENCES AND BRIEF BIBLIOGRAPHY

Conference Proceedings

1. Bishop, G.J.,T.W. Builett, and E.A. Holland, "Morphology of solar maximum Total Electron Content and L-Band scintillation in the Northern Polar Cap Ionosphere: A first look", Proceedings of the Satellite Beacon Symposium, Aberystwyth, Wales, July, 1994.

2. Clynch, J.R., "Ionospheric effects on GPS and DGPS in polar regions", Proceedings of ION-GPS, Insitute of Navigation 1579-1587, 1994.

3. Doherty, P., E. Raffi, J. Klobuchar, and M.B. El-Arini, "Statistics of time rate of change of ionospheric range delay", Proc. ION GPS-94, Institute of Navigation, 1589-1598, 1984.

4. Fang, D.J., and C.H. Liu, "Fading statistics of C-Band satellite signals during solar maximum years (1978-1980)", Paper 30, in AGARD Conference Preprint 332, AGARD, Paris 1982.

5. Wanninger, L., "Ionospheric monitoring using IGS data", presented at the 1993 Berne IGS Workshop, Berne, March 25-26, 1993.

Periodical Articles

6. Aarons, J. , J.A. Klobuchar, H.E. Whitney, J. Austen, A.L. Johnson, and C.L. Rino, "Gigahertz scintillations associated with equatorial patches", Radio Sci. *18,* 421-434,1983.

7. Basu, S., E. MacKenzie, Su. Basu, E. Costa, P.F. Fougere, H.C. Carlson and H.E. Whitney, "250 MHz/GHz Scintillation parameters in the equatorial, polar and auroral environments", IEEE J. on Selected Areas in Communications, v. SAC-5, 102-115, 1987.

8. Mendillo, M. and A. Tyler, "The geometry of depleted plasma regions in the equatorial ionosphere", J. Geophys. Res., *88,* 5778, 1983.

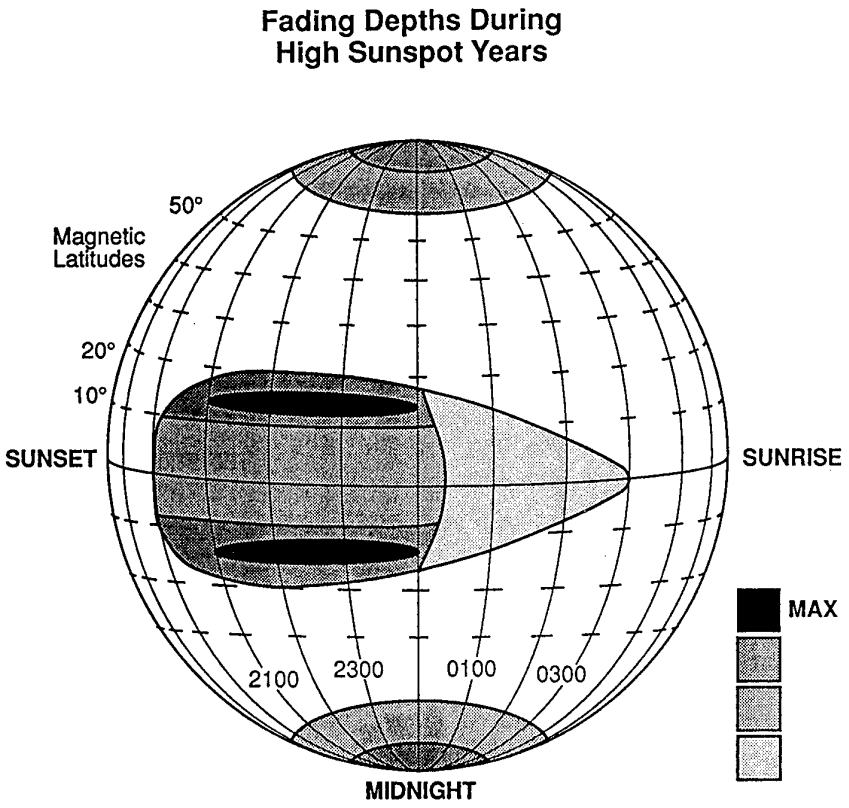


Figure 1. GPS scintillation problems will maximize during solar maximum. The fading depths to be experienced at GPS frequencies are illustrated. Bits will be lost when the fading goes below the signal to noise level.

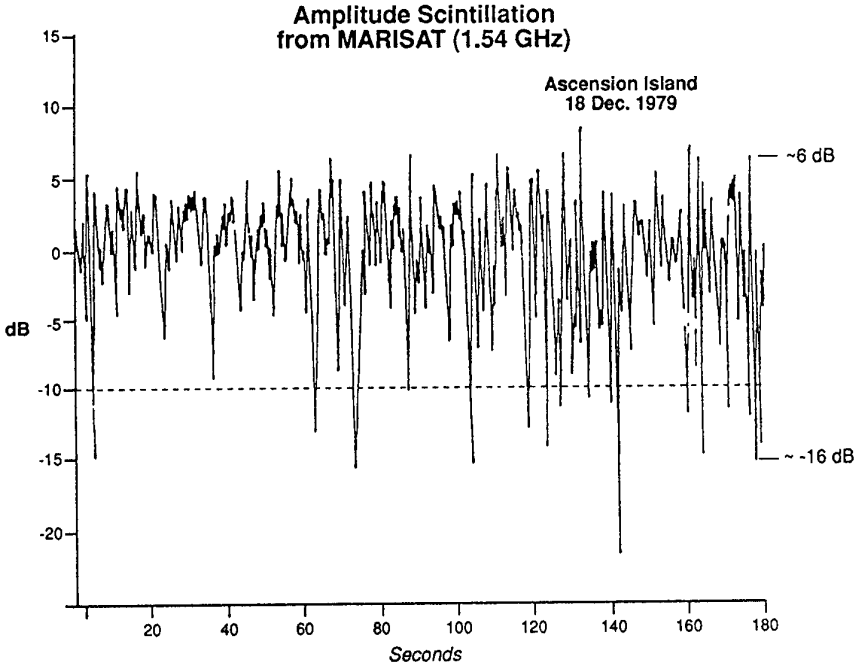


Figure 2. Scintillations as observed on the MARISAT signal at 1.54 GHz in the equatorial anomaly region. Below the signal to noise level, information will be lost in the fade.

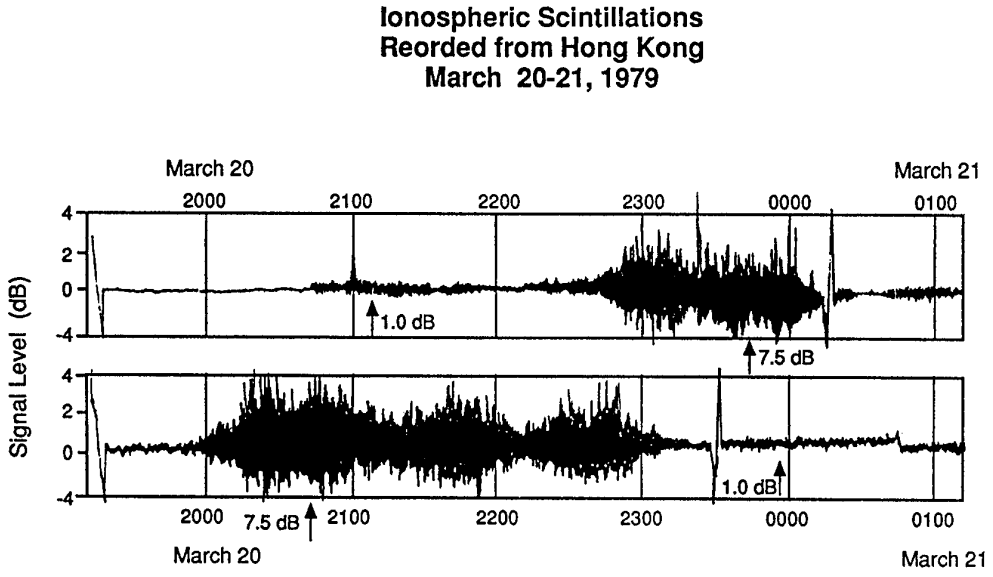


Figure 3a. Ionospheric scintillations recorded from Hong Kong, March 20-21, 1979, (Fang and Liu, 1982). Top is the 4 GHz downlink from the Indian Ocean Satellite to Hong Kong; bottom is the downlink from the Pacific Ocean Satellite.

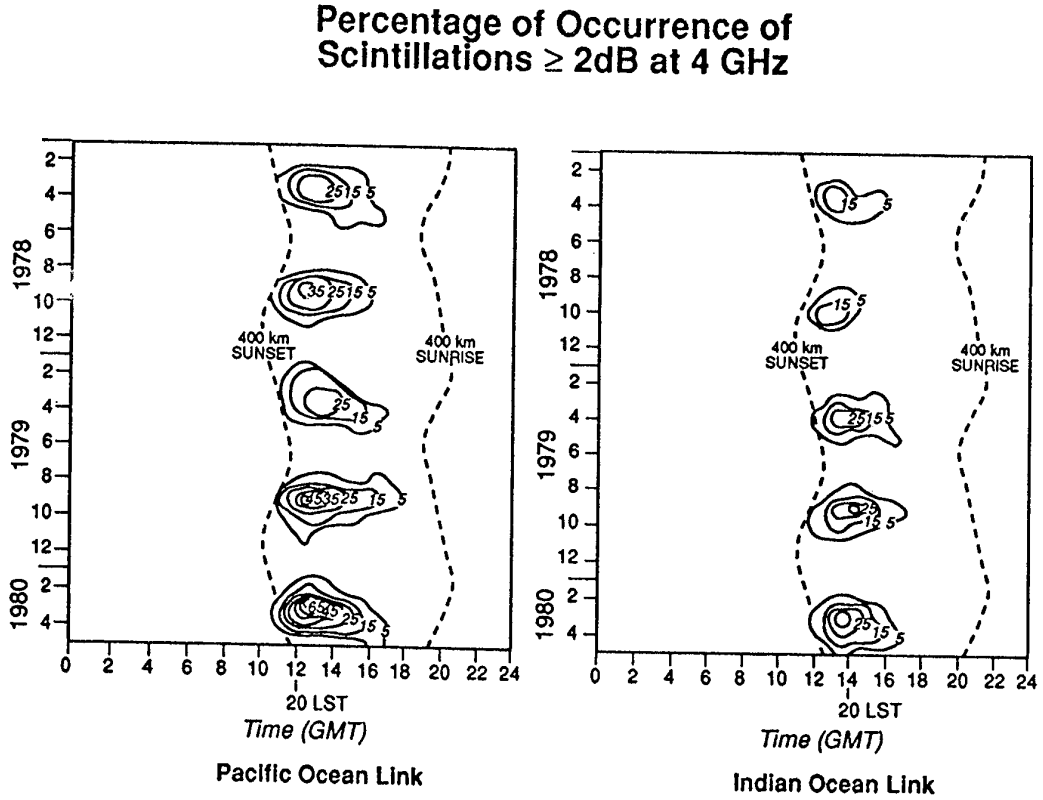


Figure 3b. Percentage of occurrence of scintillations >2 dB at 4 GHz on two links (Hong Kong data from Fang and Liu, 1982).

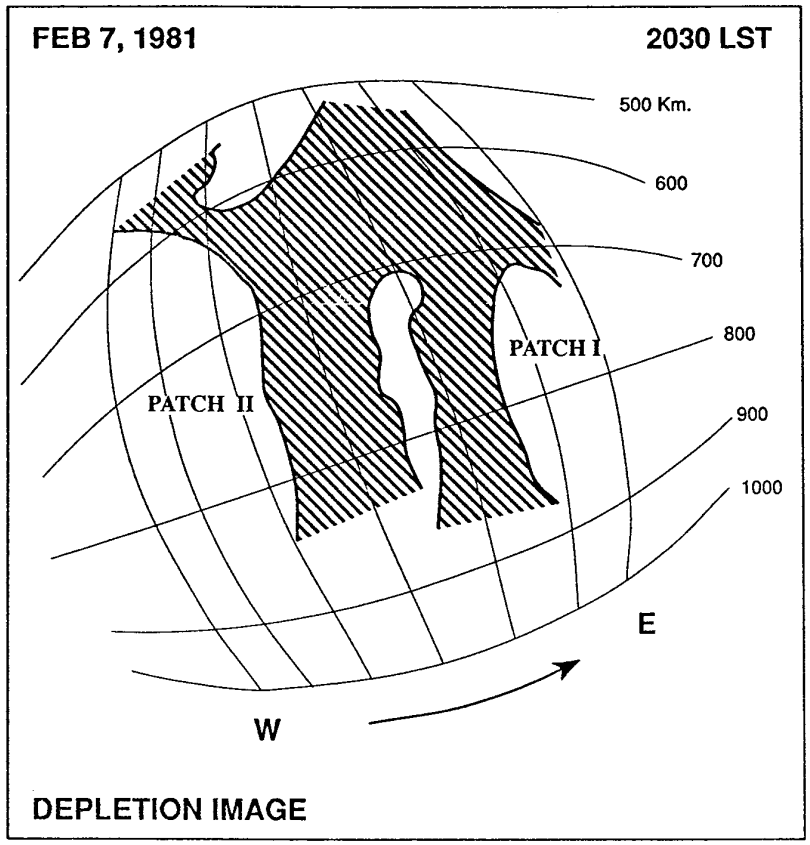


Figure 4a. Patch dimensions can be determined from all sky recordings of the depletions of airglow. In this case observations are of the 6300Å line of oxygen. Thus the optical images of all-sky observations (Mendillo and Tyler, 1983) can be related to observations of transmissions of GPS. The contour labelled 500km is the 300 projection of effects taking place at 500 km above the magnetic equator.

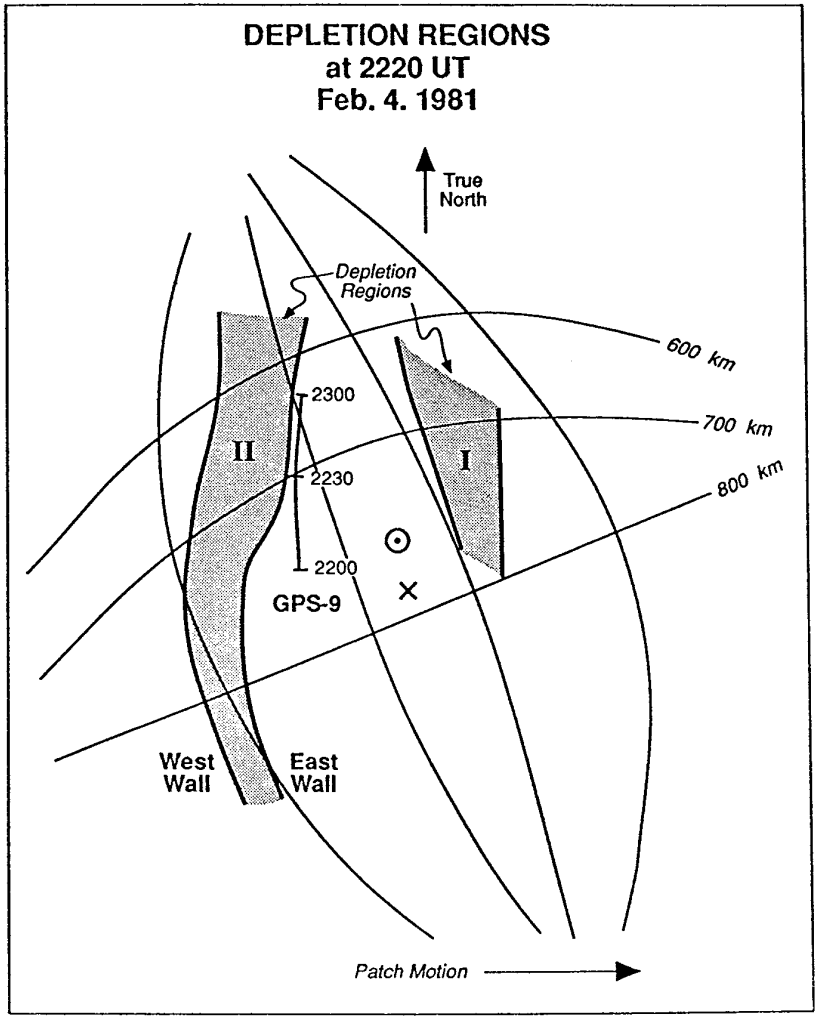


Figure 4b. Airglow depletions (Patch I and II) are shown along with the path of GPS 9 to illustrate the correlation of the two phenomena.

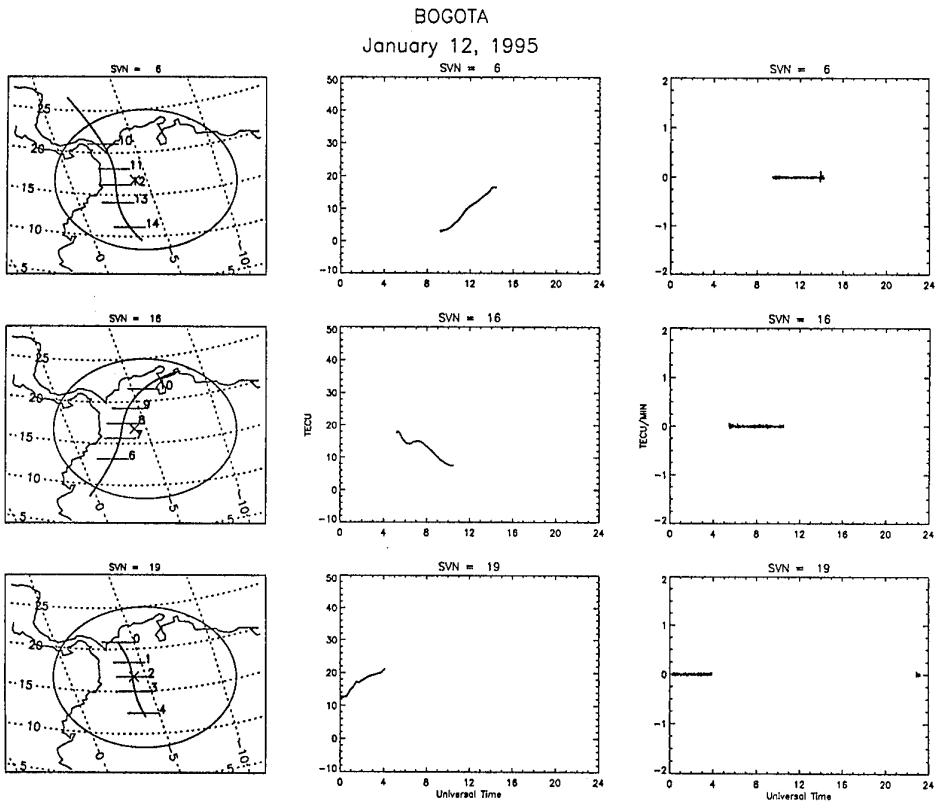


Figure 5b. On a quiet day from the viewpoint of magnetic indices and TEC fluctuations, Bogota for three satellite passes showed low fluctuation levels.

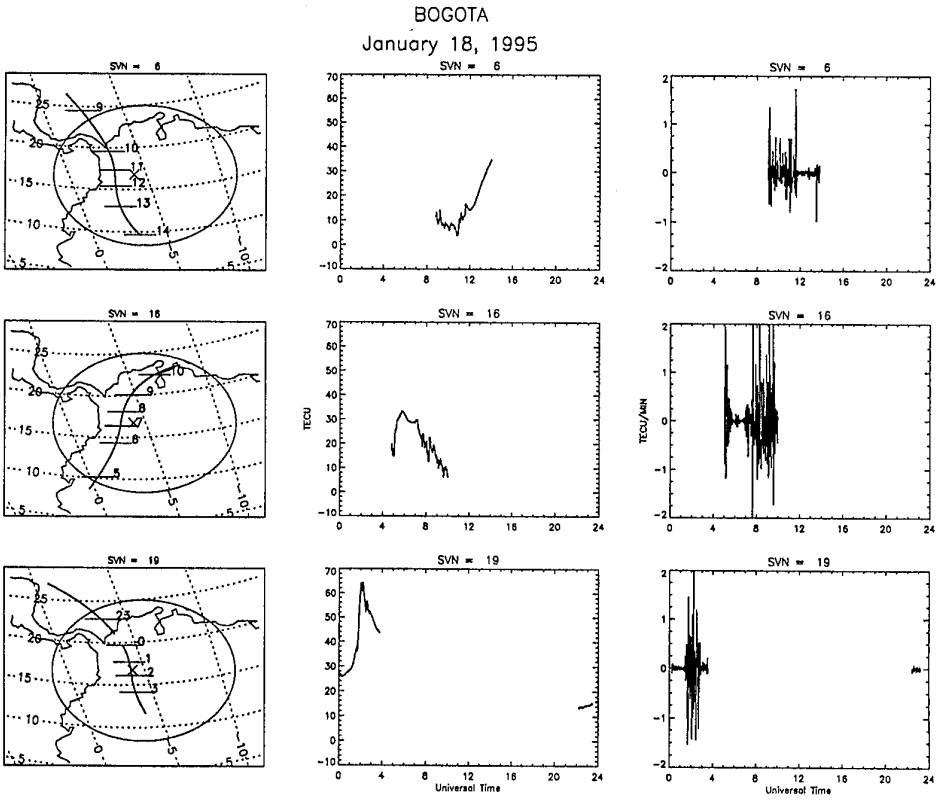


Figure 5c. On a disturbed day from the point of view of magnetic and fluctuation levels, Bogota has high levels of fluctuation.

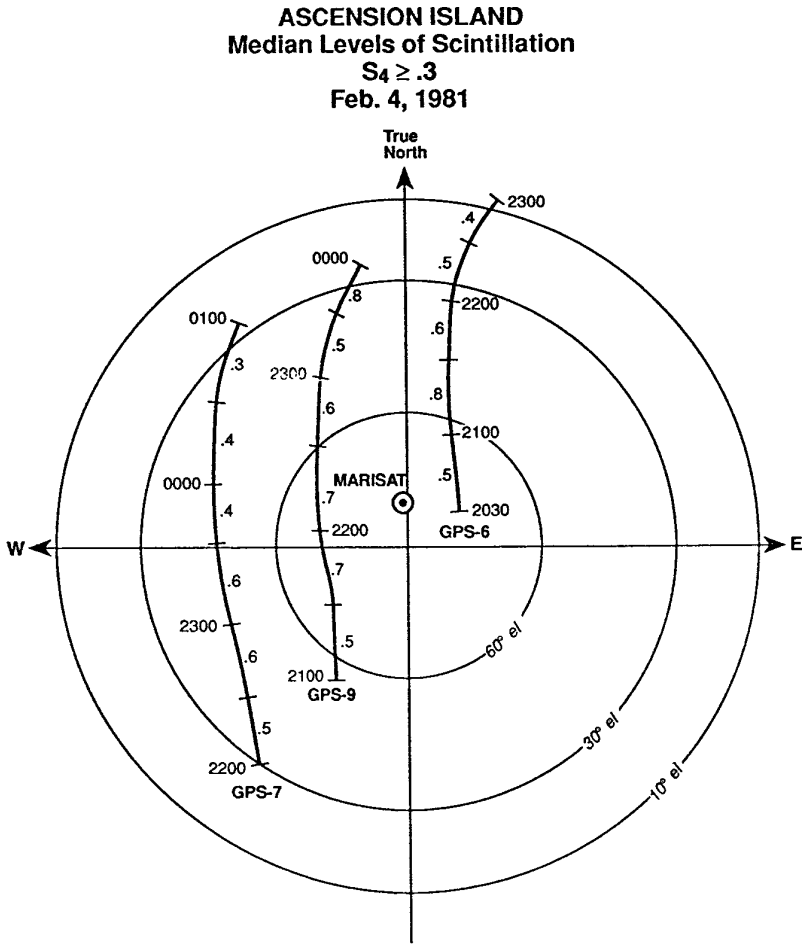


Figure 4c. GPS levels for the night of February 4, 1981 at Ascension Island.

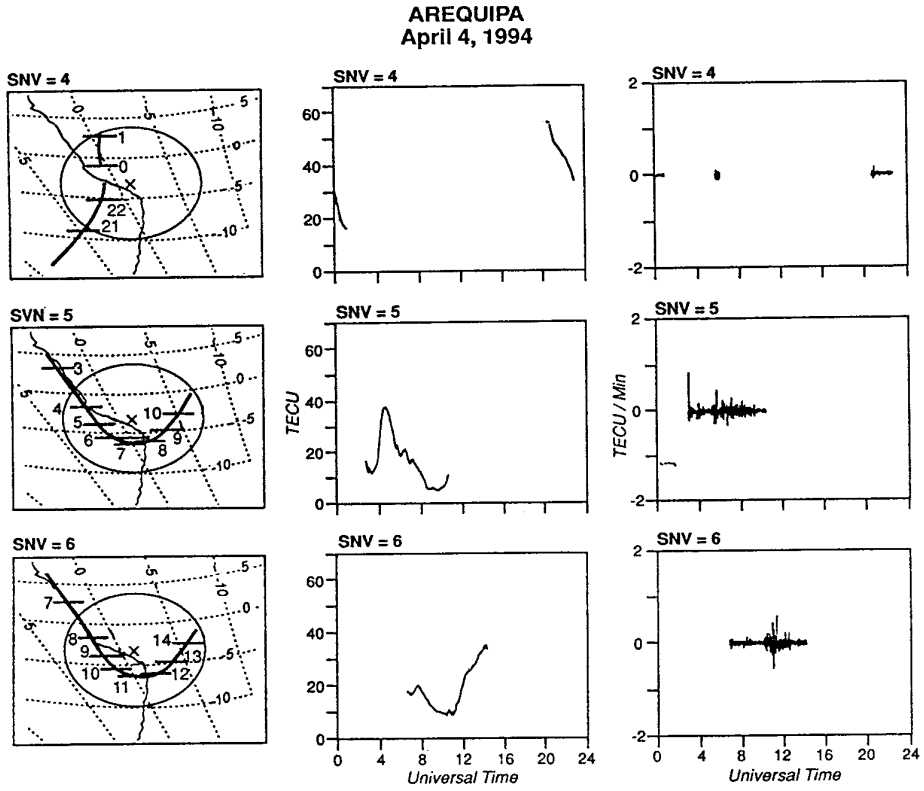


Figure 5a. Processing of the data from the International GPS Service allows for accumulating data of direct interest to GPS. The rapid rate of change in the Total Electron Content data is being used for a study of morphological and day to day variations. The diagrams shown for Arequipa, Peru, a site a few degrees from the magnetic equator, shows the fast fluctuations which can be oberved. This particular set was taken during the magnetic storm which began April 2, 1994.

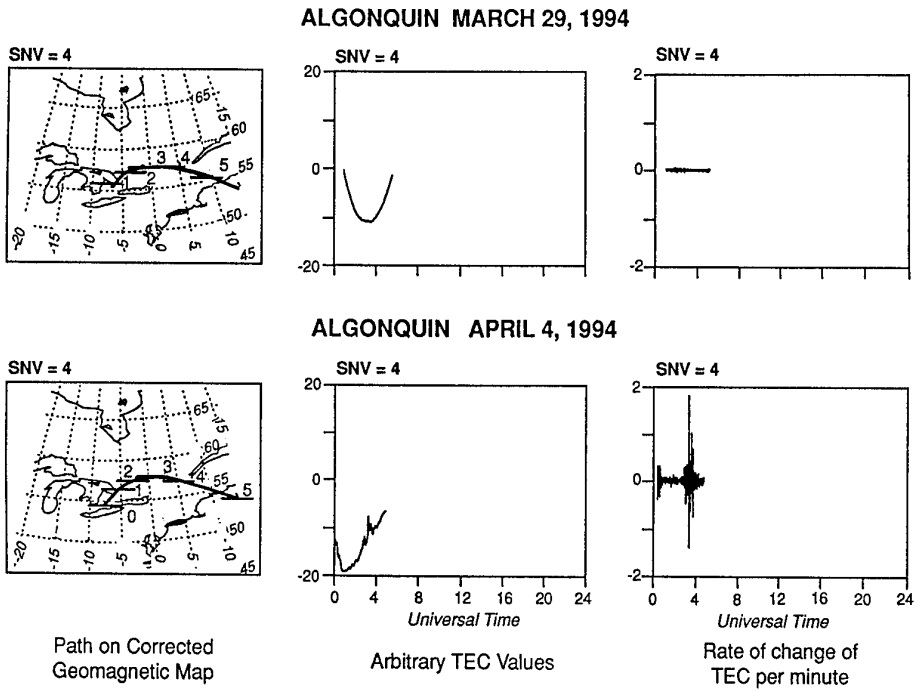


Figure 5d. A sub-auroral region is covered by the data from Alonquin, Canada. A magnetically quiet day, March 29, 1994, is contrasted with a disturbed day, April 4, 1994.

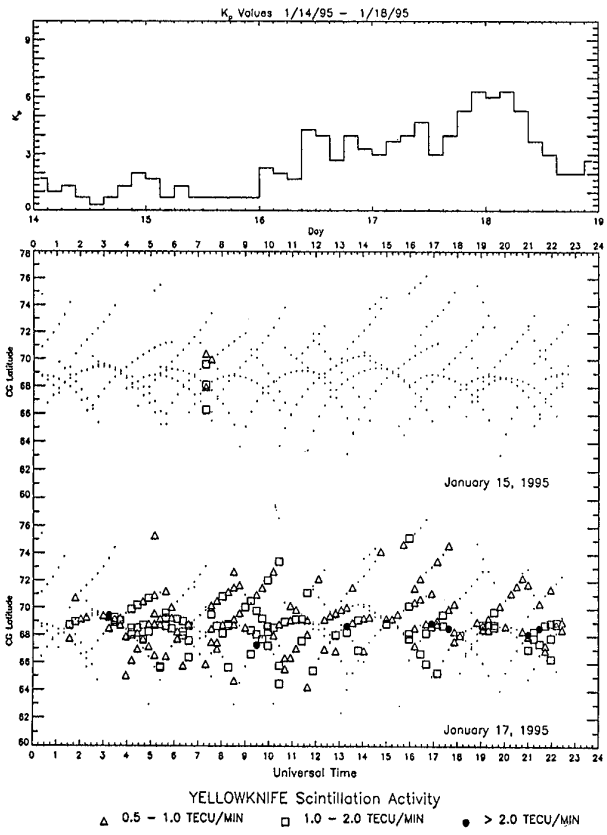


Figure 5e. TEC fluctuation levels have the same form as a function of magnetic activity as amplitude scintillations. In quiet days (January 15 in this case), the midnight time periods have amplitude scintillations but the rest of the time the activity is low. During the initial phase of magnetic storms, the irregularity region moves equatorwards as can be seen in this contrasting picture of TEC fluctuation levels on January 17, 1995.

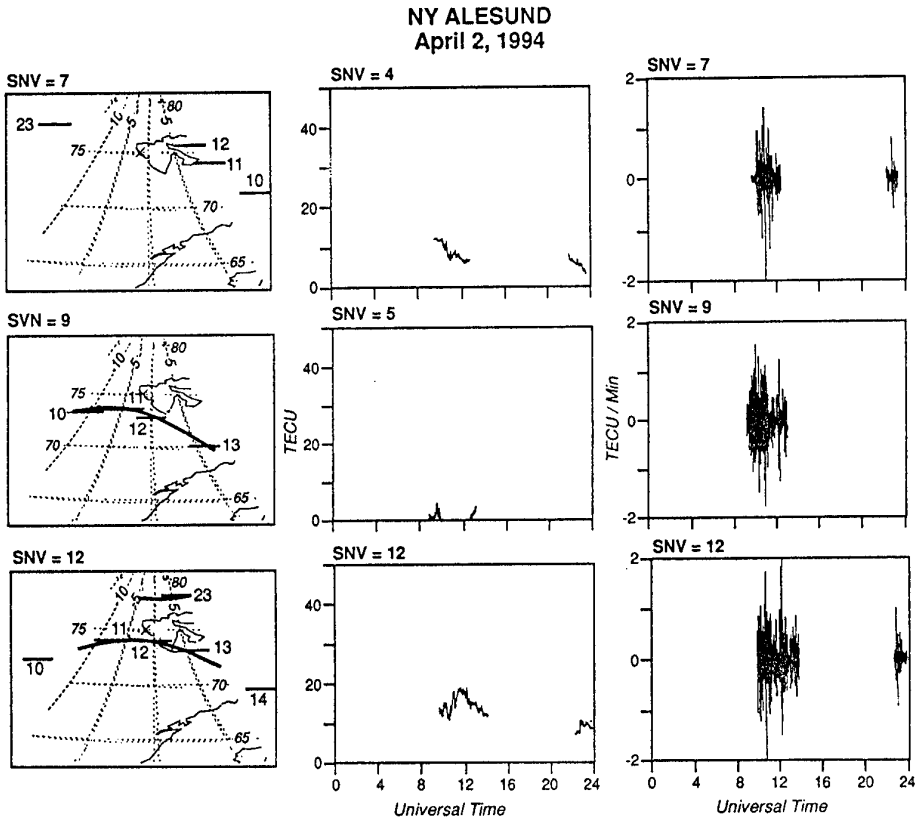


Figure 5f. In reviewing the records from NY Alesund (Spitzbergen) in the auroral region, one finds the fast fluctuations appear to be the norm. The period shown had moderate magnetic activity.

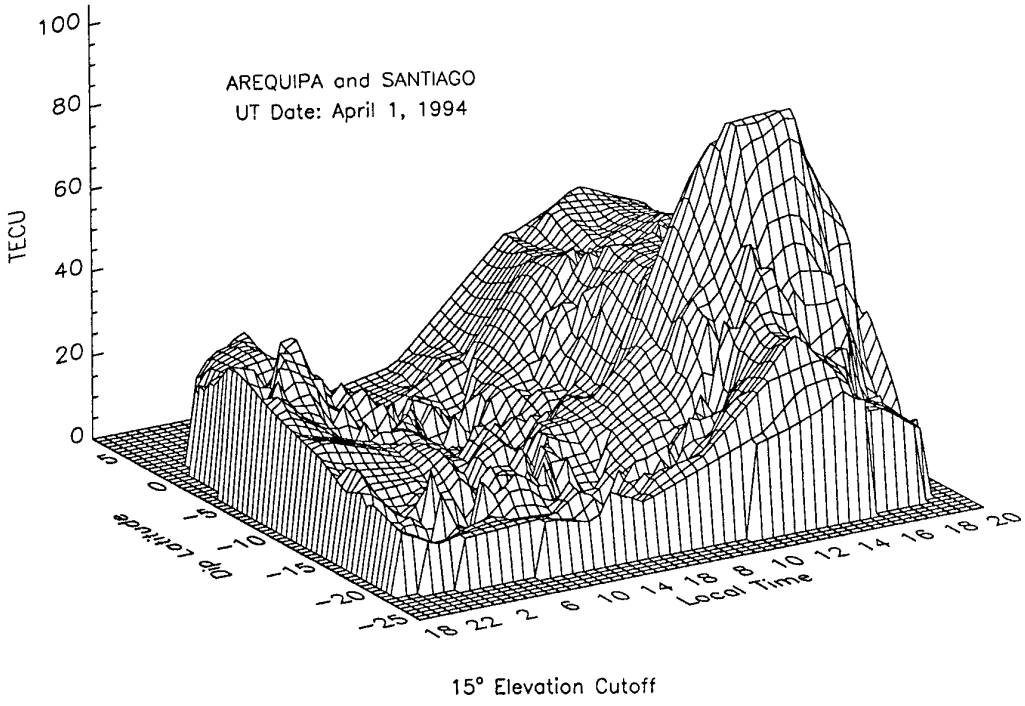


Figure 6. In putting together TEC values of two latitudinal sets near the same longitude, one can note diurnal variations and the anomaly region formation. Assumptions have been made as to the base level.

DISCUSSION

Discussor's name: D. Yavuz

Comment/Question:

Can you give an idea of the percentage of time such deep fades would be anticipated in the 1-2 GHz band in the 1990-2000 sunspot peak cycle.

Author/Presenter's reply:

During sunspot maximum ~11dB peak to peak fluctuations can occur 25-30% of the time for an hour in equinox months - at least that's what anomaly latitudes such as Hong Kong may experience.

Data taken by Basu indicates that 30-40% of the hours between 20.00 and 24.00 local time may experience 10 dB amplitude scintillations during equinox months in years of high sunspot number.

Discussor's name: H. Soicher

Comment/Question:

Is there a body of statistics for L-band scintillations (amplitude) at high latitudes? (The ITU Radio Communications Bureau is interested in such!)

Author/Presenter's reply:

There is little data at L Band at high latitudes. The Bishop paper in the Polar Cap is the only paper to measure for a very few days and for sunspot maximum only amplitude scintillation of a moderate level. Our own recordings at high latitudes indicate high levels of phase fluctuations all the time at Ny Alesund (75° CGL) and high levels in the auroral zone during magnetic storms (see Yellowknife illustration).

DISCUSSION

Discussor's name: L. B. Stotts

Comment/Question:

Would you define what you meant by a sophisticated vs simple GPS receiver? Carrier or code aiding? An Ashtech GPS receiver?

Author/Presenter's reply:

I cannot give you specifics. Simple Course Acquisition (C/A) GPS receivers for around $300.00 will be most affected. GPS receivers used by JPL for their field experiments have shown little vulnerability. These latter receivers have stronger SNRs than the simple C/A GPS receivers. Narrow bandwidth signals locking/tracking helps keep the receiver locked. However, I do not know the details. The receiver used by Clynch (reference in the paper) was of a robust variety but the magnetic storm was of maximum magnitude.

Discussor's name: C. Goutelard

Comment/Question:

You have shown the effect of equatorial plumes. Have you observed with the TEC the effects of plasma instabilities, in particular the bubbles near the terminals.

Author/Presenter's reply:

We have observed phase fluctuations from the plumes or bubbles. We are trying to determine if various horizontal gradients in total electron content are necessary conditions for the development of plumes. In addition we are looking into whether symmetry in total electron content in latitude is a necessary condition for the development of plumes (as has been suggested in several papers).

MODELISATION ET MESURE DE LA PROPAGATION RADIOELECTRIQUE SUR DES LIAISONS TERRE-SATELLITE DANS LA BANDE MILLIMETRIQUE

MODELING AND MEASUREMENT OF RADIOWAVE PROPAGATION ON EARTH-SATELLITE LINKS IN THE MILLIMETRE BAND

L. Castanet* - N. Douchin* - J. Lemorton* - R. Colas des Francs - D. Le Boulc'h** - S. Grely*****

* CERT ONERA : 2 av. Edouard Belin - BP 4025 - 31055 Toulouse CEDEX - France

** CELAR : BP n°7 - 35998 Rennes Armées - France

*** STSIE : 16 rue du Docteur Zamenhof - 92131 Issy-les Moulineaux - France

RESUME

Dans le cadre de la préparation des futurs systèmes de télécommunications par satellite, la Délégation Générale de l'Armement (DGA) a besoin de valider les modèles de prédiction de la propagation dans la bande de fréquences EHF.

Afin de déterminer le domaine de validité de ces modèles et pour améliorer la connaissance de la propagation Terre-espace à ces fréquences, la DGA envisage de mener des expérimentations avec le projet de satellite expérimental du CNES (Centre National d'Etudes Spatiales) : STENTOR.

Il est prévu que ce satellite embarque une charge utile EHF, composée d'une part d'un répéteur transparent destiné à la mise en œuvre d'essais de transmission avec différentes formes d'ondes sous différentes conditions de propagation, et d'autre part de deux balises de propagation expérimentales (à 20.7 GHz et 41.4 GHz), délivrant une zone de couverture européenne et surtout une couverture équatoriale, dans le but de vérifier les effets spécifiques rencontrés dans les climats tempérés et dans les climats soumis à de très fortes précipitations.

En attendant le lancement de STENTOR (prévu fin 1999), le CELAR (Centre d'Electronique de l'ARmement) et le CERT ONERA (Centre d'Etudes et de Recherches de Toulouse de l'Office National d'Etudes et de Recherches Aérospatiales) sont en train de mener des expérimentations sur une liaison terrestre basse élévation : la campagne CELESTE.

Quatre types d'instruments de mesures sont mis en œuvre dans cette campagne : une liaison EHF à 35 GHz, un radar (RODIN) de 6 cm de longueur d'onde, un radiomètre 35 GHz ainsi que des capteurs météorologiques.

Les premiers résultats de ces expérimentations sont présentés dans cet article ainsi que les améliorations qui pourraient être apportées à la campagne, afin de réaliser des prédictions de propagation plus précises, pour aider à la conception des futurs systèmes de télécommunications par satellites dans la bande EHF.

SUMMARY

To prepare future satellite telecommunication systems, French military administration (DGA : Délégation Générale de l'Armement) needs to validate propagation prediction models in the EHF band.

In order to assess the validity area of these models and to improve the knowledge of Earth-Space propagation at these frequencies, DGA is planning EHF experimentations in relation with the anticipated CNES (Centre National d'Etudes Spatiales) experimental satellite STENTOR.

It is planned that the satellite will carry an EHF payload which consists on one hand of a transparent repeater to implement EHF transmission experimentations with different waveforms and different propagation conditions, and on the other hand of two experimental propagation beacons (20.7 GHz and 41.4 GHz) with a European coverage and also an equatorial one in order to assess specific effects encountered in both temperate and very rainy climates.

Before the launch of STENTOR (end of 1999), CELAR (Centre d'Electronique de l'ARmement) and CERT ONERA (Centre d'Etudes et de Recherches de Toulouse de l'Office National d'Etudes et de Recherches Aérospatiales) are presently conducting measurements, on a low elevation angle terrestrial link : the "CELESTE" campaign.

In this experiment, four kinds of equipments are involved : an EHF link at 35 GHz, a 6 cm wavelength meteorological radar (RODIN), a 35 GHz radiometre and meteorological instruments.

First results of this propagation experiment are presented in this paper, and the upgradings which could be brought to the campaign, for a better efficiency on propagation prediction to plan future EHF satellite telecommunication systems are discussed.

1. INTRODUCTION

French military satellite telecommunication systems are working today with the SYRACUSE II payloads boarded on the civil TELECOM II satellites ; for the next generation of SYRACUSE satellites, DGA decided to design a self-consistent military telecommunication system, including among others an EHF payload.

Studies investigating the feasibility of such a link, are faced with the difficulty to appreciate precisely perturbing effects caused by the atmosphere on the propagation of electromagnetic waves, which prevents an accurate calculation of system margins.

These propagation impairments (attenuation, scintillation, depolarization) are highly frequency dependent and if their influence is barely noticeable at C band, it may become unbearable at EHF band.

To achieve prediction of such impairments, two kinds of models are available at CELAR : statistical ones (based on ITU-R procedures [1]) and deterministic ones such as MPM, the Millimetre-wave Propagation Model of H.J. Liebe (NTIA, USA) [2].
Since the seventies, these models have been validated up to Ka band (20/30 GHz), owing to propagation experiments

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

such as OTS, SIRIO [3] and OLYMPUS [4] in European countries, ATS, CTS and COMSTAR [5] in the U.S., and ETS [6] series in Japan.

Currently, ITALSAT propagation experiment [7] is going to study the validity of these models in EHF band (40/50 GHz), in European countries.

In order to pursue this work, DGA has decided to board an EHF experimental propagation payload on the anticipated CNES experimental satellite : STENTOR.

If the project succeeds, this payload will be constituted of two propagation beacons at 20 and 40 GHz, with two coverages : a European one (like OLYMPUS and ITALSAT) and especially an equatorial one in order to assess specific effects encountered in very rainy climates.

STENTOR satellite will not be launched before the end of 1999, so DGA decided to run a propagation experiment in the south of France, called CELESTE (Campagne d'Expérimentation sur une Liaison EHF Simulant une communication Terre-Espace), conducted by CELAR and CERT ONERA.

CELESTE is constituted of a terrestrial link on a low elevation angle between the Pic du Midi (altitude 2865 m) and the town of Lannemezan (altitude 600 m) in the Pyrenees, at a distance of 28 kilometres.

This propagation path has been chosen since it is representative of an Earth-Space worst case elevation angle (5°) and because of the melting layer presence below the Pic du Midi ten months a year.

Four kinds of equipments are involved :

- an EHF link at 35 GHz,
- a 6 cm wavelength meteorological radar (RODIN), located at Météo France in Toulouse,
- a 35 GHz radiometre, pointed toward zenith,
- meteorological instruments such as temperature, pressure, humidity and wind sensors, as well as rain-gauges, a disdrometre and a spectropluviometre.

2. STENTOR PROPAGATION EXPERIMENT

STENTOR is a technological satellite project, planned to increase French industrial competitivity in satellite telecommunications.

The program is managed by a commitee constituted by experts from CNES, France Telecom and DGA.

The three main French industrials skilled in this domain are involved in the project :

- MATRA MARCONI SPACE for structure and propulsion,
- AEROSPATIALE for thermal control,
- ALCATEL ESPACE for communication payloads.

Advanced technologies will be tested on the spacecraft, such as diphasic loops, plasma thrusters, new generation of batteries, active antennas and digital repeaters.

STENTOR is designed to a nine years duration of life with two years of active experimentations, and will be launched at the end of 1999 at 11°W on the geostationnary orbit ; it will provide a 1800 W power for a 1900 kg mass.

2.1 Objectives of the EHF program

Differently from civil payload which is planned to test new technologies for industry, the EHF payload has been decided for conducting experimentations to prepare future systems.

Like for the civil telecommunication payload, ALCATEL ESPACE is in charge of the EHF payload design and realization.

2.1.1 *Transmission part*

Because the following generation of SYRACUSE 2 satellite is planned to use on board processing and frequency hopping techniques, it seemed interesting to design a small experimental transparent repeater to test the second technique from STENTOR.

The aim of the transmission team is to test :

- new signal processing techniques such as frequency hopping,
- hardiness of waveforms in difficult propagation conditions,
- efficiency of antenna pointing systems at 44 GHz.

To achieve these goals, transmission trials need not to be managed on a continuous way : 10 % of the time would be a priori sufficient when the procedure is operational.

2.1.2 *Propagation part*

Needs for propagation experimentations are directly linked with the choice of the EHF band, which atmospheric effects such as attenuation, scintillations, depolarization and maybe dispersion, are very disturbing for.

The aim of the propagation team is to support program management in the following subjects :

- establishment of EHF possible employment conditions and determination of foreseenable telecommunication service disponibilities,
- propagation margin specification for the new system, according to covered regions and running mode (nominal, damaged),
- evaluation of the dispersion influence for 1 or 2 GHz bandwidth channels,
- building of an optimised dynamic strategy of uplink power control or other fade countermeasures,
- efficiency of antenna pointing system in bad weather conditions (strong precipitations, storms).

To obtain reliable statistics, propagation experimentations need to be managed continuously on a long period ; recent studies [8] [9] [10] have demonstrated that statistical results begin to be reliable after a period between three and seven years of measurements.

2.2 EHF Payload Description

If the project is confirmed, STENTOR will board among others a EHF payload [11] including two subsystems : the first one for a transmission task intended to signal processing trials, and the second one for propagation measurements.

The transmission subsystem is constituted by an EHF transparent repeater divided into two 40 MHz channels separated by 1 GHz ; uplink frequency is between 44 and 45 GHz and downlink frequency between 20.2 and 21.2 GHz.

The propagation experimental payload is formed by two propagation beacons at 20.7 GHz and 41.4 GHz.

The 20.7 GHz beacon (called beacon 1) is common to both missions and its main goal is to achieve more precise antenna pointing at 44 GHz ; its EIRP will be 18 dBW (end of life, edge of coverage).

The 41.4 GHz beacon (called beacon 2), is obtained by multiplication by 2 of beacon 1 ; it will supply an EIRP of 22 dBW (end-of-life, edge-of-coverage).

First link budgets indicate that a 25 dB dynamic with a 120 cm dish terminal would be available.

Two areas will be supplied, the first one will concern European countries, the second one an equatorial region, either around French Guyana, or around French West Indies or both (this last coverage will be definitely fixed before the end of 1995).

Due to the limited power supplied by the solar panels and the large consumption of the civil payload, some limitations are imposed to EHF payload users (see possible operating modes table).

N°	Beacon 1 coverage	Beacon 2 coverage	Repeater running
1	Equatorial	Equatorial	OFF
2	European	European	OFF
3	European	Equatorial	European
4	Equatorial	European	Equatorial
5	European	Equatorial	OFF
6	Equatorial	European	OFF

Table 1 : possible operating modes

The three first scenarii will likely be the most employed, on the opposite way the two last configurations will probably not be used.

After discussions between both mission teams, using times were evaluated to 10 % for scenario 3, versus 90 % for both configurations 1 and 2 ; the strategy between scenarii 1 and 2 remains to draw up.

2.3 Propagation Experiment

STENTOR will allow propagation experiments at 20.7 GHz and 41.4 GHz to manage the following activities in the two coverage areas :

- attenuation measurements, on one hand to validate models for low-fade margin systems (involving separation of rain, cloud and gaseous attenuation), on the other hand to study characteristics of events (fade slope, fade duration, interfade periods, seasonal effects, ...),
- scintillation measurements in absence and in presence of attenuation to value their impact on communication signals (reduction of the system margin, perturbations on antenna tracking systems, up-link power control),
- meteorological measurements which interest was demonstrated by the OPEX (Olympus Propagation EXperiment), managing meteorological sensors and in addition radar and radiometric measurements as in the CELESTE campaign (described here after).

To achieve a better impact on new satellite systems, the campaign will not be classified so that the participation of the scientific community could be possible.
A working group managed by DGA will be created to build a common data base in order to make exchanges easier between all members.

3. CELESTE PROPAGATION EXPERIMENT

3.1 Introduction

Since March 1995 CERT-ONERA and CRA (Centre de Recherches Atmosphériques, which belongs to the University of TOULOUSE) have been both conducting an experiment devoted to the assessment of the effects on a 34.5 GHz propagation path and supported by CELAR.

These effects are principally due to absorption and scattering by raindrops, but also by atmospheric gases (mainly H_2O and O_2) and particles in clouds, and lead essentially to attenuation and depolarization of the wave along the path. Clear air effects such as amplitude scintillations, related to atmospheric turbulences, may also be observed.

Several RF equipments and meteorological sensors are operated simultaneously and continously and will thus provide a large amount of data suitable for a statistical analysis, since CELESTE will probably last at least three years. The analysis of the data will also be performed as specific case studies in order to assess some specific effects such as depolarization, scintillations or effects due to the melting layer which are still not well known.

Another objective of CELESTE is to compare experimental data with available prediction models, either statistical ones or deterministic ones.

In the following chapter, a description of the campaign is given along with some early results obtained in the first five months.

3.2 Description of the instrumentation

3.2.1 *Description of the sites*

The experimental site of the link has been choosen according to two main criteria :

- the propagation path should be representative of an Earth-Space path with a transmitter above the 0°C level;
- meteorological equipments should be available, at least at the transmitter or at the receiver site.

Figure 1 gives a global schematic view of the experiment and figure 2 (at the end of the text) shows the geographical location of CELESTE experiment.

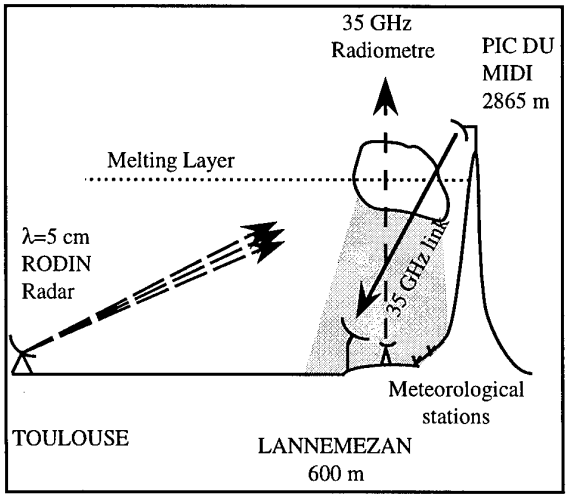


Figure 1 : CELESTE - geometry

The transmitter has been installed at the top of the Pic du Midi, in the Pyrénées, at an altitude of 2865 m which is above the 0°C level most of the time (except in July and August).

The receiver is located at CRA near the town of Lannemezan (altitude 600 m) because meteorological sensors, as described below, are available there.

Then, the choice of these two experimental sites leads to a 28 km propagation path with an elevation angle close to 5°, which is representative of a worst case geometry for an Earth-Space radiolink.

It can be noticed that Lannemezan is situated in the K-zone as defined by ITU-R, which means that the rain rate is over 42 mm/h during 0.01% of the year. Such climatic conditions are quite different from those prevailing over the north of France where most of the previous propagation campaigns (such as the OPEX) have been conducted.

It can also be noticed that Lannemezan and the Pic du Midi are situated on a same radial line viewed from a RODIN type weather radar situated in TOULOUSE 100 km far from Lannemezan. Thus the geographical locations of the transmitter and the receiver are suitable for the use of this radar as it will be further described below.

3.2.2 *The equipments*

The EHF link between the top of the Pic du Midi and the CRA-Lannemezan is operating continuously at 34.5 GHz.

The main radioelectrical characteristics of this equipment are listed in the table 2 hereafter :

Transmitter		Receiver	
Set	Ka band	Set	Ka band
Frequency	34.5 GHz	Frequency	34.5 GHz
RF Power	20.5 dBm	Sensitivity	-105 dBm
Antenna type	30 cm parabolic	Antenna type	same as transmitter
3 dB beamwidth	2.0 °	Number of channels	2
Gain	37 dB	Polarization	H & V

Table 2 : radioelectrical characteristics of the EHF radiolink

The available dynamic at receiver is over 54 dB.

Two transmitter configurations are available concerning the polarization : either a pure vertical polarization or a mixed vertical/horizontal polarization obtained by a 45° rotation of the transmitter around the radiation axis of the antenna. The first configuration allows one to assess depolarization effects while the second one is suitable for the assessment of differential attenuation. Both configurations will be used during CELESTE with a change of the transmitted polarization every six months.

At receiver, two orthogonal polarizations channels are switched on the same receiver : such a strategy is used to avoid discrepancies which may be observed between different receivers. The sampling rate of the received signal amplitude can be either 5 Hz or 50 Hz. The low value is suitable for the study of attenuation effects while the high rate is used to study scintillations effects. Most of the time the data are stored continously and automatically using the low sampling rate, but the high sampling rate is used for two periods a day during 20 minutes each time.

The meteorological equipments consist in "usual" meteorological sensors such as temperature, pressure, humidity, wind sensors and rain-gauges, as well as a disdrometre and a spectropluviometre.

At CRA, the receiver site, all the "usual" meteorological sensors are available. At the transmitter site at the top of the Pic du Midi, temperature measurement is made in order to evaluate the altitude of the 0°C level.

Three rain-gauges are installed along the propagation path : one at each end of the path and the third one near the middle point of the path. These allow one to evaluate the rain rate distribution along the propagation path and then to estimate the attenuation along it.

The disdrometre and the spectropluviometre have been set up at CRA, not far from the receiver and close to each other. They both supply the distribution of raindrops in diameter/speed coordinates, but also the rain rate at this location. Comparisons of the data obtained with these two equipments will be made, as well as comparisons of the rain rates supplied by them and by the rain-gauge. The data supplied by the disdrometre and the spectropluviometre will also be used to study the influence of the drop size distribution on the attenuation measured on the propagation path.

All the meteorological data are available with a sampling rate of 6 minutes or less and will be compiled in a single data base.

Two complementary means are also used during CELESTE :

- A weather radar (RODIN type, operating at a 6 cm wavelength) situated in TOULOUSE.

  Since the receiver and the transmitter sites are aligned, as viewed from the radar, this will supply the rain distribution along the path and allow us to estimate the attenuation along the path.

  Comparisons between rain distributions retrieved from the radar data and estimated using the three rain-gauges will be achieved, and between the attenuation evaluated from the radar data and directly measured on the propagation path.

  The meteorological radar also allows one to estimate the altitude of the 0°C level which can then be compared with the 0°C altitude estimated using temperature sensors.

- A radiometre operating at 35 GHz has also been installed at CRA.

  It looks vertically to the sky because it was impossible to point it in the link direction, due to the presence of mountains which could disturb sky brightness measurements.

  Furthermore, it is not used all the time as it is not protected against precipitations.

  Data supplied by this equipment will be used to estimate the attenuation due to clouds and atmospheric gases.

  However the estimated values should be considered carefully because they are valid only over the CRA and not all over the propagation path.

A more detailed description of the CELESTE experiment can be found in [13].

3.2.3 *First results of the campaign*

The early results presented in this paper correspond to the five first months of the campaign, i.e. the period extending from February to June 1995. They consist in cumulative distributions of attenuation in excess (i.e. the measured signal level, at receiver, as compared to the maximum of this level) and of fade durations.

a) Cumulative distribution of attenuation in excess

This has been calculated using averaged experimental data (one datum per minute).

On figure 3 the cumulative distribution of attenuation in excess is given along with the distributions obtained using the rain prediction procedure preconised by ITU-R [1] and the prediction model of R.K. CRANE [13]. In both prediction models, only the attenuation due to the rain is taken into account.

Discrepancies between experimental results and predictions can be observed :

- At low attenuation values, the experimental curve is sensibly above the curves related to prediction models, certainly because the attenuation terms due to atmospheric gases and clouds are not included into the model.

  It also can be noticed that these terms may be not negligible due to the particular geometry of the propagation path (5° elevation angle).

- Concerning strong attenuations, large discrepancies between experimental data and prediction models are observed.

  It is not surprising as one can expect some strong attenuation related to high rain rate events, for exemple storms during Summer, to change the behaviour of the measured cumulative distribution of attenuation.

  It can also be noticed that attenuation values stronger than 55 dB can not be measured at receiver ; this is another explanation for the large discrepancies observed on figure 3 at very low probability values.

  With the CELESTE very slant propagation path it could be very interesting to have a larger dynamic at receiver (at least 100 dB).

On figure 4 the experimental cumulative distribution of attenuation is compared to the distribution obtained using the CCIR prediction procedure including all the attenuation terms (gaz, clous and precipitations). The agreement between both curves is much better at low attenuation values than on figure 3. Thus it can be verified that the contribution to the total attenuation due to the atmospheric gases and to the clouds are not negligible on a very slant propagation path.

b) Cumulative distribution of fade durations

Figure 5 gives the cumulative distribution of absolute number of fade duration events for the month of May. As expected, the higher the attenuation value, the shorter the fade duration.

Cumulative distributions of fade durations are given on figure 6 for the month of May and has been established using experimental data averaged in order to retain one value per 10 seconds period. Thus one can expect scintillations to be filtered out, as was demonstrated by THE OPEX with cut-off frequencies values between 0.01 Hz and 0.1 Hz.

On figure 6, the 5 dB attenuation curve is below the 10 dB attenuation curve which itself is below the 20 dB attenuation curve. But this global behaviour is not verified by the 30 dB attenuation curve ; this can be due to the poor number of 30 dB attenuation events observed in May, which is approximately 50 as shown on figure 5.

Actually, it seems that observed fadings are not solely caused by rain events : as a matter of fact 10 dB attenuation on the path can be caused either by a large cloudy cover or by a weak localized precipitation. Consequently, the curves of figure 5 do not follow log-normal statistics as could be expected from THE OPEX and [10].

Nevertheless, it is important to remember that this study turned on one month collected data and statistics over a longer period are required before any definite trend can be established.

Furthermore, radioelectrical data should be corrected with meteorological data ; for example, rain events should be separately analysed, and RODIN and rain-gauge data will be used for that purpose.

3.3 Future orientations

First objectives of the campaign were to acquire a know-how for the different participants (CELAR, CERT and CRA) for this kind of measurement in order to prepare STENTOR experiment.

After some starting problems, the campaign is working on a nominal way and first results described here above tend to show a real interest to pursue CELESTE on a longer time.

As it is mentionned above, at least three years of measurements are necessary to get reliable statistical data, therefore it seems reasonnable to go on CELESTE experimentations until the launching of STENTOR.

In addition, it would be valuable to upgrade the instrumentation involved : EHF link, RODIN radar, 35 GHz radiometre and meteorological sensors.

Moreover one of the difficulties noted in such an experiment is to correlate different kinds of data (radioelectrical link, radar and radiometric measurements, meteorological data).

At the moment, some home-made softwares are running for CELESTE exploitation, but it would be necessary to use more powerful tool to realize more advanced analysis.

To achieve this aim, it would be necessary to dispose of power software tools such as DAPPER used for THE OPEX [14].

4. CONCLUSIONS

The objectives of this paper were to display French DGA policy about Earth-Space propagation in the EHF band.

Whereas currently available models were validated up to Ka band by previous experimentation campaigns carried out until now, such as THE OPEX managed by ESA, it appeared some needs for further experimentations, especially for new satellite telecommunication systems characterised by higher frequency bands and low service disponibilities.

French strategy in this domain has been oriented in two directions : at the end of the century, the launching of a propagation experiment in relation to the future CNES technological satellite project STENTOR, and until that, the "CELESTE" campaign since the beginning of 1995.

The characteristics of STENTOR experiment were clarified :

- launching of the satellite at the end of 1999 by Ariane 5,
- realization of a EHF payload divided into two subsystems : a transparent repeater to permit EHF

transmission experimentations with different waveforms and different propagation conditions, and two experimental propagation beacons to assess specific effects encountered in temperate and very rainy climates,

- supply of two coverage areas : European like OLYMPUS and a new kind of coverage on equatorial regions,
- experimentation propagation subsystem including two beacons at 20.7 GHz and 41.4 GHz, with an EIRP of respectively 18 dBW and 22 dBW.

In second chapter, CELESTE propagation experiment was described.

As the geometry of the propagation path used for CELESTE is particular, due to its extremely low elevation angle, one can expect non usual results from the measured data.

It has already been noticed, for example, that contributions of clouds and the atmospheric gases to the total attenuation can not be neglected on such a path.

One can also expect scintillations to be very strong on such a very slant path.

Secondly it must be pointed out that measured distributions presented here are only from early experimental data obtained during the five first months and this is a too short period for cumulative distributions to be representative and to be comparable to prediction models.

Nevertheless experimental and predicted distributions are not so far from each other, which is a promising result.

5. ACKNOWLEDGMENTS

This work was supported by the DGA ST3SII Program Management (DSTI/STSIE/TSR/TSF) and the CELAR/ITES Propagation Group.

6. REFERENCES

1. ITU-R : "Propagation data and prediction methods required for Earth-Space telecommunications systems", CCIR Recommendations, Propagation in non-ionized media, RPN serie, Rec. 618-2, Geneva 1992.

2. Liebe, H. J., Hufford, G. A., Cotton, M. G. : "Propagation Modeling of Moist Air and Suspended Water/Ice Particles at Frequencies below 1000 GHz", AGARD 52nd Specialists'Meeting of the elctromagnetic Wave Propagation Panel, Palma De Mallorca, Spain, May 1993.

3. Davies, P. J., Mackenzie, E. C., Harris, G. P. : "Measurements of SIRIO Transmissions near 11.6 GHz at Slough (UK)", Alta Frequenza, Vol. 48, n°6, June 1979.

4. OPEX : "Reference Book on Attenuation Measurement and Prediction", 2nd workshop of the OLYMPUS Propagation Experimenters, Nov. 1994, ESA ESTEC, Noordwiijk, The Netherlands.

5. Cox, D. C., Arnold, W. A. : "Results from the 19 and 28 GHz COMSTAR Satellite Propagation Experiments at Crawford Hill", Proc. of the IEEE, Vol. 70, n°5, May 1982.

6. Fugono, N., Hayashi, R., Ishisawa, Y. : "ETS II Experiments, Part.1 : Japans First Geostationnary Satellites", IEEE Trans. Aero. & Elec. Sys, Vol. AES 16, n°5, Sept 1980.

7. Paraboni, A., Mauri, M., Barbaliscia, F., Giannone, B. : "First Propagation Measurements at 20, 40 and 50 GHz Obtained in the Framework of the ITALSAT Experiment", URSI Open Symposium on Wave Propagation and Remote Sensing, Ravenscar, UK, June 1992.

8. Burgueño, A., Austin, J., Vilar, E., Puigcerver, M. : "Analysis of Moderate and Intense Rainfall Rates Continuously Recorded Over Half a Century and Influence on Microwave Communications Planning and Rain-Rate Data Acquisition", IEEE Trans. Com., COM-35, n°4, April 1987.

9. Crane, R. K. : "Estimating Risk for Earth-Satellite Attenuation Prediction", Proc. of the IEEE, Vol. 81, n°6, June 1993.

10. Vogel, W. J., Torrence, G. W., Allnutt, J. E. : "Rain Fade on Low Elevation Angle Earth-Satellite Paths : Comparative Assessment of the Austin, Texas, 11.2 GHz Experiment", Proc. of the IEEE, Vol. 81, n°6, June 1993.

11. Assemat, D., Bossuge, H. : "Specifications techniques de besoin et exigences spécifiques des chaînes fonctionnelles du satellite STENTOR", Doc. CNES, n° PTR.SP.1.500363.EPE, 26/06/1995, pp. 95-104.

12. Crane, R. K. : "A Two Composent Rain Model for the Prediction of Attenuation Statistics", Radio Science, Vol. 17, n°6, Nov.-Dec. 1982, pp. 1371-1387.

13. Douchin, N., Lemorton, J., Sauvageot, H., Bourrel, L. : "Campagne de mesure CELESTE", Doc. n° 1/1771-MO, Convention CELAR n° 0144194, June 1995.

14. OPEX : "Data Processing Reference Book", 2nd workshop of the OLYMPUS Propagation Experimenters, Nov. 1994, ESA ESTEC, Noordwiijk, The Netherlands.

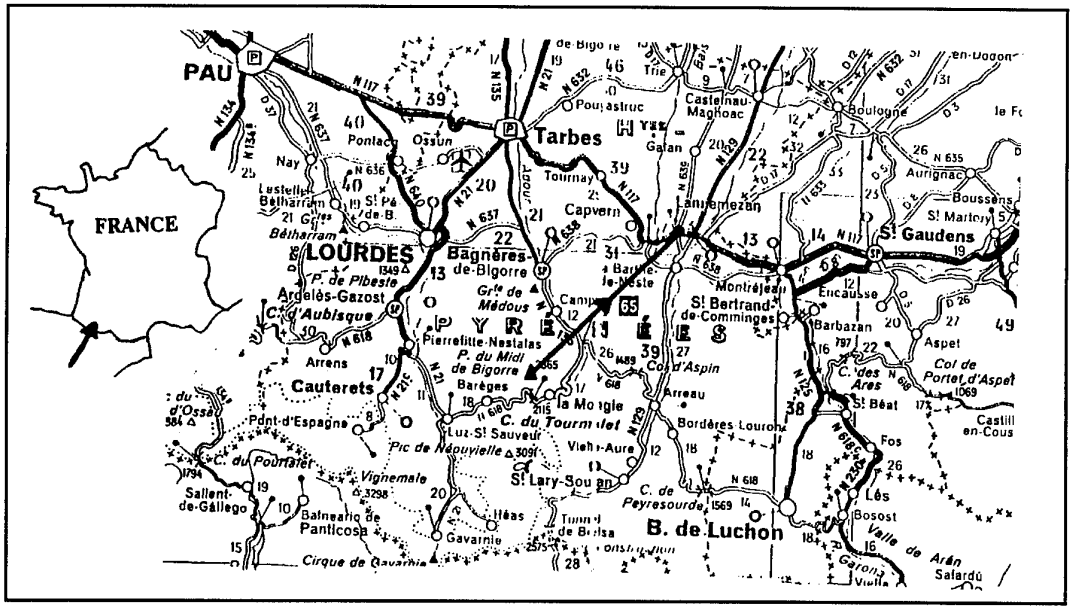


Figure 2 : Geographical location of the CELESTE experiment.

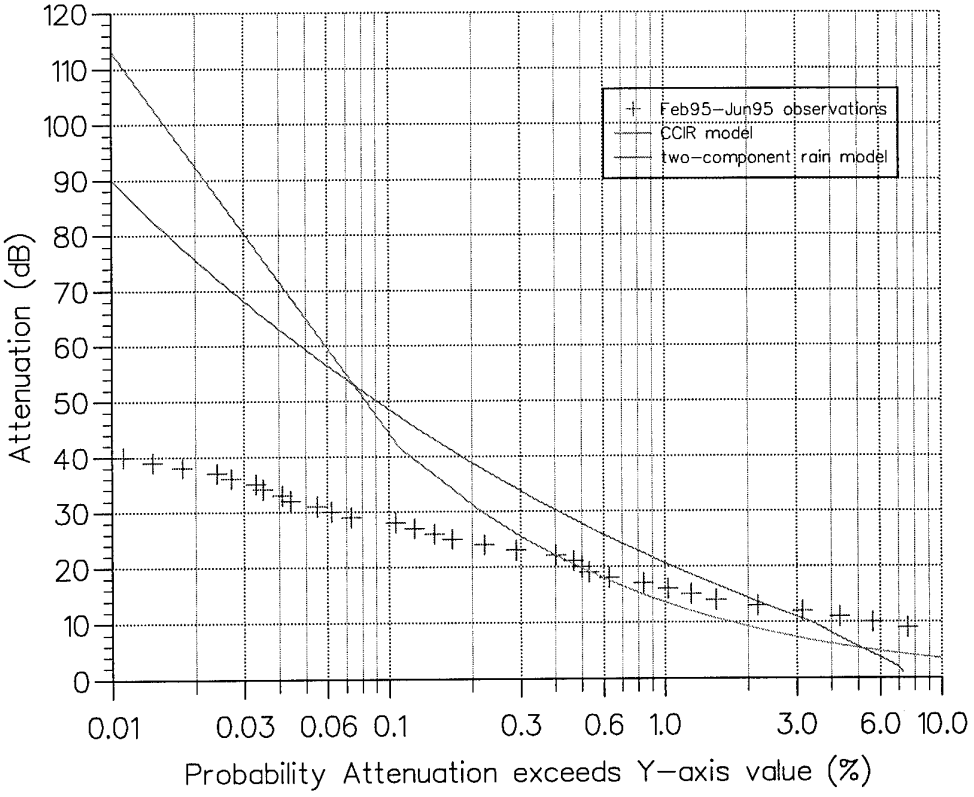


Figure 3 : Observed and predicted rain attenuation in excess distributions.

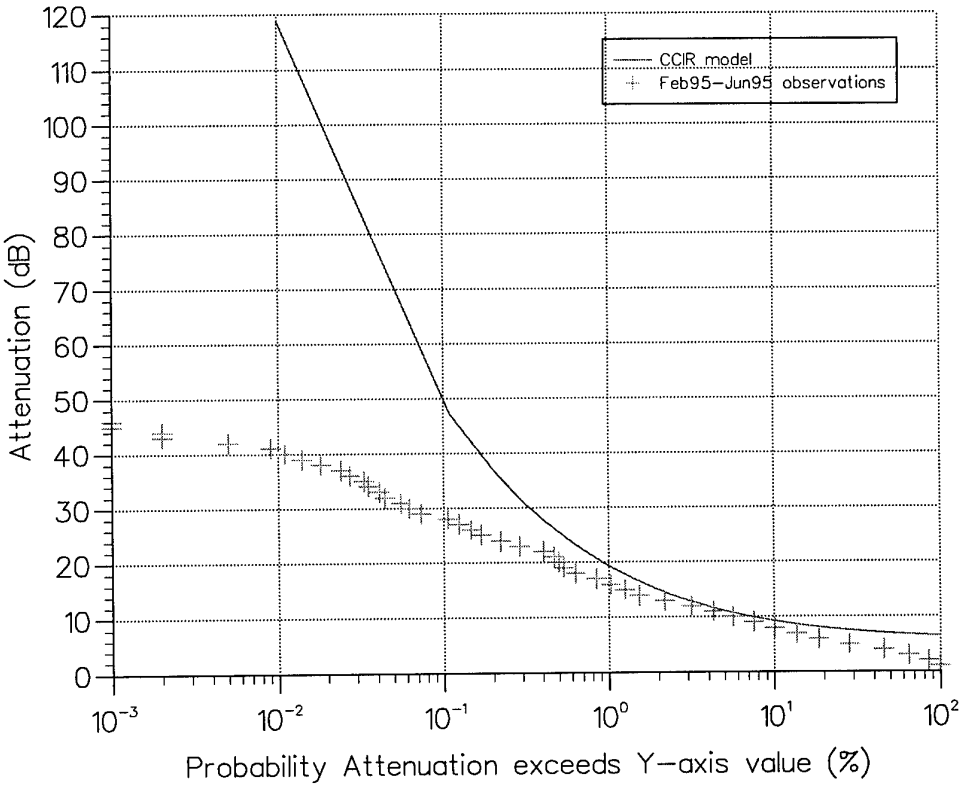


Figure 4 : Observed and predicted attenuation in excess distributions.

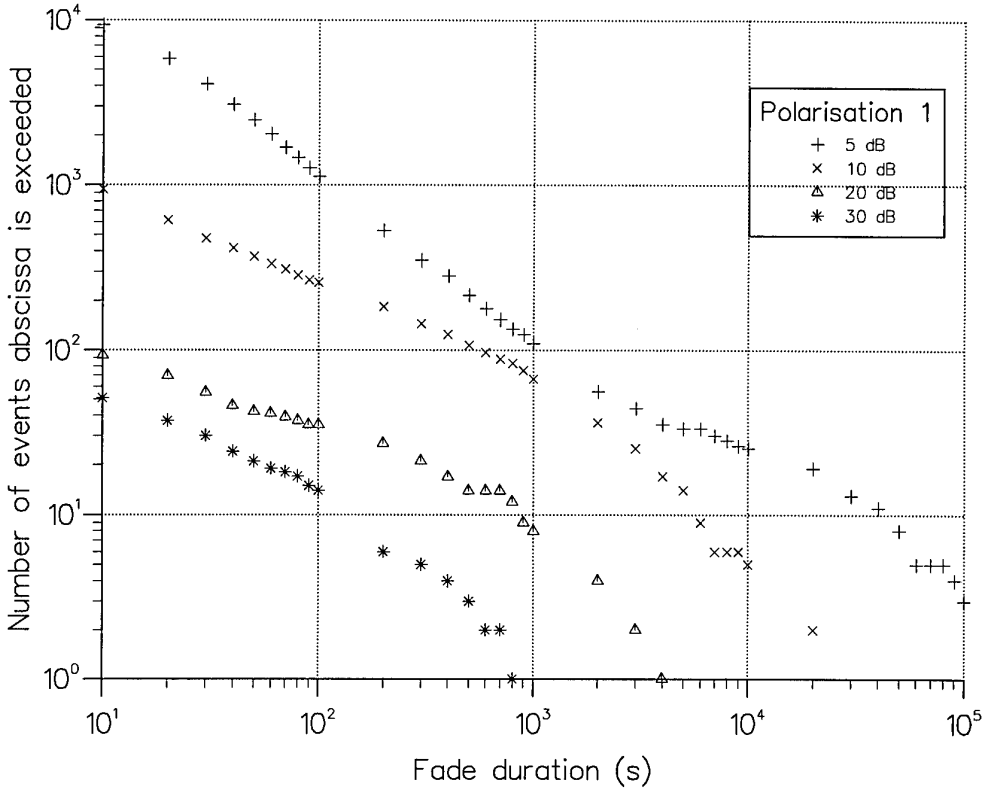


Figure 5 : Cumulative distributions of absolute number of fade duration events, May 95.

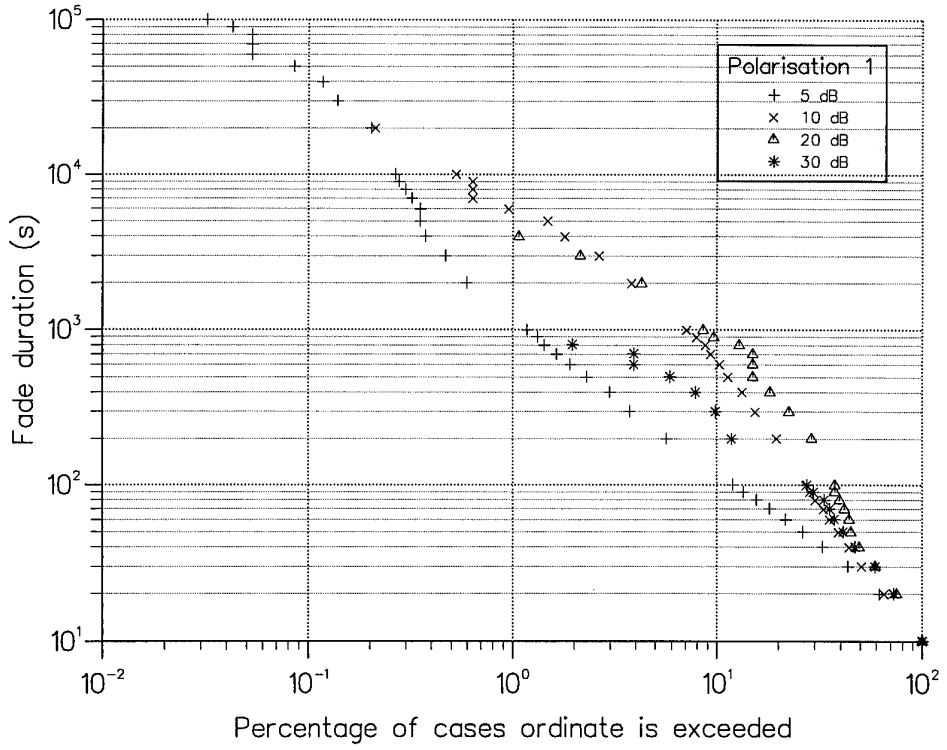


Figure 6 : Cumulative distributions of fade durations, May 95.

L'ATTENUATION DES EFFETS DE TRAJETS MULTIPLES ET DE L'INTERFERENCE DANS LES COMMUNICATIONS RADIO MILITAIRES PAR LE TRAITEMENT DU SIGNAL

G. MULTEDO
G. GOUDEZEUNE
THOMSON-CSF/RGS
66 rue du Fossé Blanc
92231 GENNEVILLIERS Cédex - FRANCE

1. INTRODUCTION

L'évolution de la technologie du numérique et la généralisation des radiocommunications numériques a ouvert au traitement du signal un champ d'application très vaste pour mettre en oeuvre la panoplie de ses outils théoriques et matériels au service de la transmission de l'information.

Les quantités d'information à transmettre sont de plus en plus importantes. Elles concernent la voix, support du commandement, mais aussi les transmissions de données, les FAX, les images ..

Les radiocommunications militaires dont la couverture va de la VLF jusqu'au millimétrique se heurtent toutes à des demandes croissantes d'automatisation et d'augmentation du débit à transmettre. Cet accroissement de performances doit se faire en conservant les spécificités propres des applications militaires, par rapport aux radiocommunications civiles, que sont la résistance au brouillage ,à l'écoute et à l'intrusion.

La résistance aux interférences est une autre de ses demandes. Elle va de pair avec l'occupation croissante du spectre de fréquences qui avec le développement conjoint des radiocommunications militaires et de la radiotéléphonie cellulaire civile, devient une denrée extrêmement rare .

Ces exigences conduisent à pousser aux "limites" l'utilisation du canal de communication pour obtenir des radiocommunications de très bonne qualité .

Parmi ces " limites" la présence de trajets multiples en particulier en zone urbaine et d'interférences volontaires ou non sur le canal de transmission sont des facteurs aujourd'hui primordiaux pour dimensionner les performances des systèmes de radiocommunication militaires et civils.

L'objet de cet article est de fournir une vision des principales solutions offertes par le traitement du signal pour adapter les radiocommunications à un canal perturbé.

2. CODAGE SOURCE : LE CODAGE DE LA PAROLE

2.1 Introduction

Les informations à transmettre proviennent de sources différentes, soit par nature numériques (fichiers,..), soit analogiques (parole, vidéo). Le souci, d'une part de sécuriser la transmission au sens cryptologie, d'autre part de résister aux aléas du canal par l'usage de techniques performantes (étalement de spectre, codes correcteurs) a été le moteur de l'essor des techniques de numérisation et de compression des sources analogiques, porté par le développement des communications numériques. Ainsi deux classes de sources numériques apparaissent : les sources de type fichier (données informatiques,fax, images,..), comprimées ou non; les sources de type "animées" (parole,vidéo) nécessitant une synthèse temps réel continue de l'information. De ce point de vue, seul ces dernières font l'objet d'un véritable codage source.

Par la suite nous ne traiterons que le codage de la parole qui est l'objet du commandement et la source d'information première ayant focalisée l'attention des transmetteurs.

2.2 Les codeurs de parole

Le codage de la parole dans la bande téléphonique (300-3400 Hz) a donné lieu au développement d'un grand nombre de codeurs sur toute une plage de débit allant de 600bps à 64Kbps. On distingue traditionnellement trois grandes techniques de codage :

- à haut débit, le codage par échantillon auxquels appartiennent le PCM64,l'ADPCM32 et les CVSD 32 & 16Kbps,
- à moyen débit, le codage par forme d'onde où l'on cherche à minimiser la distance entre la "forme d'onde" synthétisée et le signal original : l'optimisation a lieu en boucle fermée. C'est le domaine des codeurs de 4 à 16Kbps où excelle la méthode d'analyse-par-synthèse associée à la technique CELP, le codeur CELP4800 bps étant intégré dans des équipements militaires,
- à bas débit, le codage paramétrique cherchant à obtenir le meilleur rendu acoustique du signal synthétique sans forcément approcher l'original : le calcul des paramètres est mené en boucle ouverte. C'est le domaine des codeurs analyse-et-synthèse basé sur la technique LPC. Les débits offerts vont de 2400 à 800 bps (soit une compression d'un facteur 80 par rapport au PCM64 pour ce dernier!).

Ces différents codeurs sont décrits dans la littérature spécialisée. Par ailleurs les codeurs de parole font généralement l'objet de normalisations internationales pour garantir l'interopérabilité (cf figure 1).

2.3 Les caractéristiques des codeurs

Les caractéristiques d'un codeur de parole sont sa qualité de restitution de la parole, sa complexité de calcul et son délai de traitement.

La qualité d'un codeur de parole est mesurée par des tests subjectifs. On distingue les tests d'intelligibilité tels que le "Diagnostic Rime Test" (DRT) et les tests d'agrément d'écoute type "Diagnostic Acceptability Measure" (DAM) ou "Mean Opinion Score" (MOS). Une description de ces méthodes de tests peut être trouvée dans la littérature spécialisée.

La figure 1 donne une indication de la qualité des trois familles de codeurs. Si l'on fait abstraction des codeurs anciens, on constate une convergence en débit et qualité des codeurs militaires, téléphoniques et radiocoms vers le débit de 4 à 4.8 Kbps.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

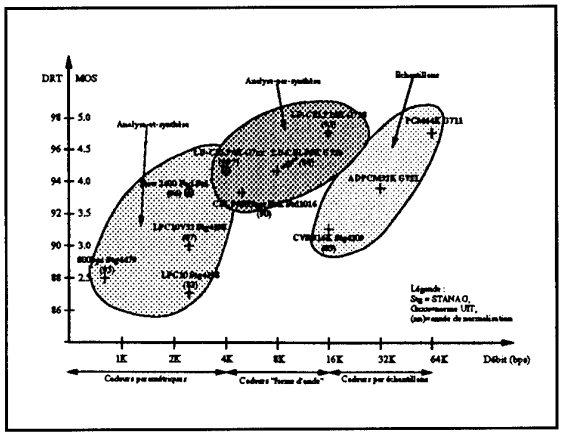


Figure 1 : Qualité des codeurs

La complexité des codeurs est aussi un paramètre important, surtout quand il s'agit de les intégrer dans un équipement portable. La figure 2 présente cette complexité en MIPS d'un DSP effectuant une multiplication-accumulation en un cycle machine.

Les codeurs par échantillons sont les moins complexes et ont fait l'objet de réalisations de composants spécialisés, ex le CVSD 16Kbps, l'ADPCM32Kbps. Pour les vocodeurs normalisés, la complexité a suivi au cours du temps l'évolution des possibilités des DSP. Ainsi la technique CELP à 4800 bps n'a été intégrable qu'au milieu de la décennie 80. Pour un débit donné, cette possibilité d'évolution de la complexité s'accompagne d'une évolution significative de la qualité. Un exemple est représenté par le vocodeur 2400bps : une évolution significative de la qualité (3 pts de DRT) a lieu tous les 7 ans avec un doublement de la complexité, cette évolution ayant d'abord eu lieu au sein de la norme STANAG4198 par amélioration des algorithmes (2400LPC10V52) et maintenant par le développement de nouvelles techniques de type harmonique ou multibande (nouveau Standard Féd. US).

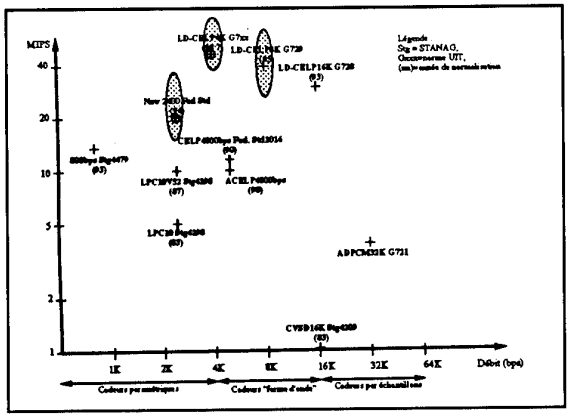


Figure 2 : Complexité des codeurs

Un autre aspect est la résistance intrinsèque du codeur aux erreurs de transmissions : celle-ci croit avec le débit, de quelques pour mille à 800bps jusqu'à 10% pour le CVSD16Kbps, en passant par quelques pour cent pour les vocodeurs 2400 ou 4800bps.

Finalement le délai de traitement d'un codeur est la dernière caractéristique à prendre en compte dans une analyse système. Les codeurs par échantillons ont un délai très faible (quelques échantillons à 8 ou 16KHz) alors que les vocodeurs 4800 ou 2400 ont des délais de 100 à 200 ms ce qui devient significatif.

On observe aujourd'hui, dans le cas de la téléphonie, le développement de codeurs "low delay" (LD) pour remplacer le PCM64K ou l'ADPCM32K. Une des conséquences de cette contrainte faible délai est un accroissement de la complexite, cf figure 2.

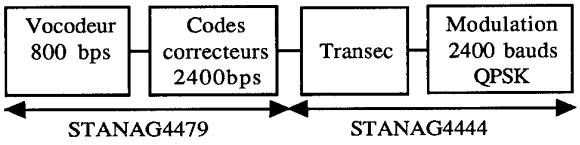
2.4 L'utilisation des codeurs

La première utilisation des vocodeurs est bien sûr la cryptophonie haute sécurité, associés à un chiffre numérique. C'est ce besoin qui a répandu l'usage du vocodeur 2400LPC10 au début des années 80 pour les CHS téléphoniques & les liaisons HF (équipement ANDVT ou STANAG4197), les modems de ces liaisons étant limités en débit utile.

L'autre intérêt de l'usage des vocodeurs repose dans le fait qu'il permet de bénéficier des techniques de protection des communications numériques face aux aléas du canal de transmission, propagation et brouillage.

Premier exemple : la normalisation OTAN HFECCM.

La transmission de la phonie sous les contraintes HF ECCM (propagation & brouillage sévère) a nécessité le développement et la normalisation d'un codeur 800 bps avec des codes correcteurs optimisés vis-à-vis de la source et de la transmission selon le Stanag4444, réf[1]. Ceci a fait l'objet du Stanag4479.



Le codeur 800bps utilise les algorithmes d'analyse et de synthèse du codeur 2400LPC10 (Stanag4198), la compression de débit étant obtenue principalement par l'exploitation des intercorrélations entre les paramètres de trames d'analyses jointives : le 800 bps code en bloc trois trames LPC10 (cf figure 3).

		2400bps	800bps
Durée trame		22.5ms	22.5ms
Durée Supertrame		-	67.5ms
Energie		5 bits	10 bits
Pitch & Voisement		7 bits	9 bits
Codage du filtre LPC	Evolution	-	3 bits
	Coefficients	41 bits	32 bits
Synchronisation		1 bit	-
TOTAL		54 bits	54 bits

Figure 3 : Tramage des codeurs 2400 & 800 bps

Ce codeur est associé à un code correcteur de type RS, un mot de code correspondant à une supertrame 800bps ce qui permet de bénéficier en réception de l'information de trame erronée et de mettre en oeuvre des techniques de lissage entre trames atténuant la perturbation du canal sur la phonie restituée. L'information codée est ensuite entrelacée sur les paliers ECCM, de manière adaptée aux paliers pour optimiser la résistance aux paliers brouillés en minimisant le délai de l'entrelaceur.

Deuxième exemple : l'introduction des vocodeurs dans le poste VHF/ECCM PR4G, réf [2] & [3].

THOMSON a mis en oeuvre dans le poste PR4G, en alternative au codeur CVSD16Kbps, des vocodeurs à 800, 2400 & 4800 bps. Ceci a été rendu possible grâce au développement technologique d'Asic à coeur de DSP.

Le codeur à 4800 bsp offre une qualité de phonie intrinsèquement meilleure que le CVSD et le débit libéré permet d'utiliser des codes correcteurs jusqu'à 16Kbps pour assurer une portée équivalente (ne pas oublier que le codeur 4800 est plus sensible aux erreurs) et surtout accroitre la résistance au brouillage. Quant à lui, le codeur 800bps, issu du Stanag4479, assure une communication phonie de dernier recours même dans les pires conditions de brouillage. Ceci a été évalué lors d'essais opérationnels, la figure 4 ci-dessous indique la qualité opérationnelle des codeurs sur une échelle de 1 à 5.

Service	Paliers brouillés			
	0%	10%	25%	50%
CVSD16K	3+	3	2+	1-
4K8 codé 16K	4	4-	3+	1+
0K8 codé 16K	3+	3	3-	2+

Figure 4 : Performances opérationnelles avec paliers brouillés

Cette performance n'a pu être obtenue que par une optimisation conjointe du codage source et du codage canal car :

- tous les bits d'une trame vocodeur n'ont pas la même sensibilité aux erreurs (ou sont plus importants pour l'intelligibilité finale de la phonie) : les bits importants bénéficient d'une redondance plus importante,
- il fallait obtenir une information de trame vocodeur effacée pour lisser efficacement les erreurs de transmission,
- dans des conditions sans interférences, la qualité phonie doit être supérieure au CVSD quand la propagation est bonne et permettre une portée équivalente.

Le compromis a été réalisé par l'usage de codes concaténés, codes en bloc binaires interne aux paliers et codes RS entrelacés en externe (cf réf. [3]).

2.5 L'évolution des codeurs militaires

Les codeurs normalisés OTAN sont le CVSD16K et le 2400LPC10, le premier utilisé pour les communications du champ de bataille ou pour les communications aéronautiques, le second surtout employé pour les CHS. L'OTAN a également normalisé un 800bps. Nous avons vu ci-dessus leurs usages.

Au delà de ces exemples, les perspectives tracées par le PG6 suggèrent à terme l'utilisation de trois codeurs :

- un codeur primaire de très bonne qualité (celle de la techno. 4800 bps d'aujourd'hui) pour l'usage général et l'interopérabilité de bout en bout des communications phonie : du téléphone au poste du champ de bataille en passant par les réseaux; ce codeur primaire serait potentiellement un nouveau codeur 2400 bps qui sera préfiguré par le résultat de la normalisation fédérale US en cours à 2400 bps (action DVPC).
- un codeur de qualité quasi-téléphonique pour les liaisons entre V.I.P.; ce codeur secondaire devrait avoir un débit objectif de 4 à 4.8Kbps, remplacer les 4800bps actuels en améliorant la qualité ce qui passe également par une réduction du délai. Ce codeur sera sans doute la passerelle vers le monde civil et la normalisation UIT à 4Kbps qui va démarrer est peut être la bonne réponse;
- un codeur pour les liaisons difficiles (mauvaise propagation HF ou sévères conditions de brouillage). Ce codeur secondaire est actuellement le codeur à 800 bps. A l'instar des efforts à 2400 bps, il conviendrait d'en améliorer la qualité.

Cette évolution pour les 5 prochaines années va entrainer également un accroissement de la complexité pour une meilleure qualité : 40 MIPS seront sans doute atteints ce qui reste compatible de la technologie des ASIC à coeur de DSP qui offrira à court terme cette puissance et à moyen terme 80 MIPS, permettant ainsi l'intégration des fonctions complémentaires codage canal et modem dans la perspective d'applications portables.

3. CODAGE CANAL : LES CODES CORRECTEURS D'ERREURS

3.1 Introduction

Par rapport aux communications analogiques, les transmissions numériques ont l'avantage de permettre en théorie des communications sans erreurs. Dans ce but on ajoute de la redondance aux informations à transmettre pour permettre de détecter et également corriger les erreurs survenues lors de la transmission; si une transmission sans erreur est requise, la détection permet de demander la répétition de tout ou partie du message. L'essor des communications numériques a vu le développement de nombreuses techniques de codes correcteurs dont les plus classiques sont aujourd'hui disponibles par le biais de composants spécialisés.

3.2 Les différents codes correcteurs

Si le débit d'information à transmettre, pour un canal donné,est inférieur au débit critique, il est théoriquement possible d'obtenir un taux d'erreur aussi faible que souhaité (à condition de trouver le codage correcteur optimal et de ne pas avoir de contrainte de délai de transmission). Le code détecteur ou correcteur modifie de façon connue la suite de symboles à transmettre de telle sorte que les messages suceptibles d'être émis soient les plus "distants" les uns des autres (cas des codes convolutionnels) ou encore soient "redondants" c'est à dire que le message peut être reconstitué à partir d'une portion correctement reçue (cas des codes en blocs).

Les codes convolutionnels sont par nature efficaces vis-à-vis de perturbations uniformes, les codes en blocs traitant plus efficacement les paquets d'erreurs. Ces différents codes sont décrits abondament dans la littérature, réf [4]&[5] par exemple.

3.3 L'utilisation des codes correcteurs

Le démodulateur fournit au décodeur soit des symboles décidés, soit des symboles associés à une note de qualité. Dans ce dernier cas, le gain de codage peut être grandement amélioré (2 dB env. pour les codes convolutionnels sur canal Gaussien) mais il convient que cette information soit statistiquement fiable; dans le cas contraire il vaut mieux travailler avec les symboles décidés.

Lorsque la liaison est affectée d'évanouissements ou de brouillage, il est souvent indispensable d'utiliser des dispositifs d'entrelacement pour que les perturbations affectant des symboles successifs à l'entrée du décodeur soient statistiquement indépendantes (canal sans mémoire) : dans ces conditions les codes convolutionnels, par ex., donnent des performances optimales.

Pour accroitre la performance d'un codage correcteur d'erreur, le concepteur peut proposer des codes concaténés.

Un premier exemple est le code utilisé pour les sondes spatiales : un code interne (proche du canal) de type convolutionnel protège l'information transmise mais à l'inconvénient de fournir des erreurs par paquets; ces paquets d'erreurs sont ensuite corrigés par un code externe de type RS. Cette concaténation apporte un gain de 2dB à TEB=10^{-5}.

Deuxième exemple : dans le cadre du poste VHF/EVF PR4G, un compromis est réalisé entre une performance en portée (erreurs uniformes) et une résistance au brouillage (paquets d'erreurs). De plus une détection des paliers brouillés est souhaitable. Le choix s'est porté sur un code en bloc binaire interne BCH par palier, corrigeant les erreurs isolées et détectant les paliers brouillés et un code externe RS, entrelacé sur les paliers pour apporter la résistance aux paliers brouillés. Ce schéma est relativement performant comme l'on montré les résultats présentés au §2.4 au sujet de la phonie vocodée.

3.4 L'exemple des modulations codées

Lors du choix d'un code correcteur, il convient d'examiner la capacité requise (en nombre de bits/s/Hz) pour la transmission. Si celle-ci est inférieure à 1, l'usage d'un codage interne de type binaire associé à une modulation binaire est satisfaisant. Par contre, pour les demandes de forte capacité on utilisera des constellations de signaux avec un alphabet de taille élevée associé à des modulations codées, apparues dans la littérature scientifique depuis 1976. Celles-ci ont été développées dans le but d'optimiser l'efficacité spectrale des transmissions, c. à d. asssurer le meilleur compromis possible entre la bande passante utilisée et les performances obtenues.

Les modulations codées reposent sur le développement conjoint du codage et de la modulation. L'optimisation est réalisée à la fois sur le codeur et sur l'affectation des signaux de modulation aux séquences de bits codés ("mapping"),ces deux étapes étant intimement liées. Le schéma de principe figure ci-après.

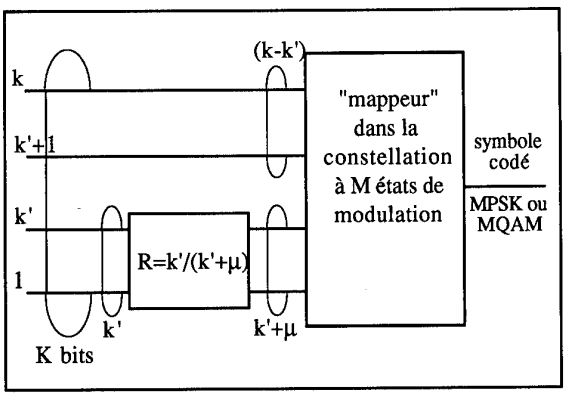


Figure 5 : Modulations codée

- k' bits passent effectivement par un codeur (généralement de type convolutionnel) et permettent de partitionner la constellation de signaux en $2^{(k'+\mu)}$ sous-ensembles,
- (k-k') bits ne sont pas traités et sélectionnent le signal de modulation parmi les $2^{(k-k')}$ signaux du sous-ensemble considéré.

Ce type de codage est représentable par un treillis où chaque branche se voit attribuer le signal de modulation résultant de la fonction de mapping.

Les modulations codées ont été d'abord optimisées pour des canaux gaussien. Elles ont été mises à l'honneur par le CCITT qui a normalisé de nombreux schémas pour les modems téléphoniques haut débit (V17, V32, V33,..).

Ce codage peut également être optimisé pour des canaux non-stationnaires tels que rencontrés en HF. En association avec son modem parallèle cohérent, THOMSON a optimisé des modulations codées pour un modem HF offrant un débit net de 4800 bps. Le gain de codage en canal Rayleigh est représenté ci-après.

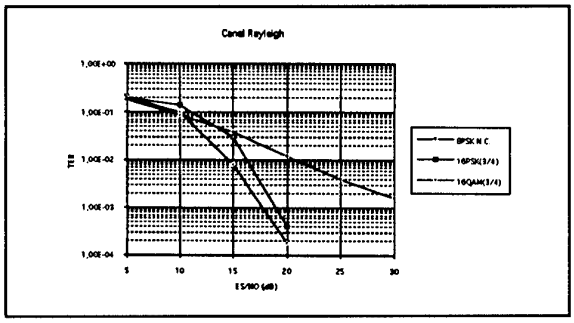


Figure 6 : Performances en HF

Ces modulations codées ont été validées sur signaux réels entre Paris et Cholet (300 km) et ont permis d'obtenir fréquemment des taux d'erreurs voisins de 10^{-4} pour une émission de l'ordre de 300W. Associées à une modulation 256QAM, des transmissions exploitables, à plus de 7Kbps, ont été obtenues dans les mêmes conditons (avec 2.7Khz de bande passante).

3.5 Les turbo-codes

Les possibilités offertes par le traitement numérique (puissance de calcul, rupture de la causalité par mémorisation) et la recherche de performances proches de l'optimum ont stimulé la recherche de nouveaux codes. Ainsi est apparue récemment (réf [6]) une nouvelle classe de codes convolutionnels appelés "Turbo-codes", dont les performances en termes de taux d'erreurs binaires (TEB) sont proches de la limite de Shannon (Eb/N0=0 dB dans le cas d'une démodulation cohérente binaire). Ces codes sont construits par une concaténation parallèle de deux codes convolutionnels récursifs et systématiques, cf figure 7

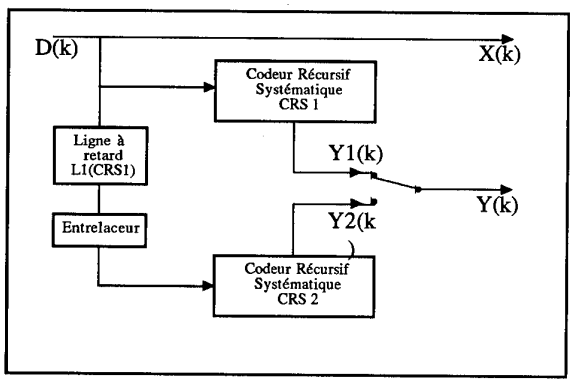


Figure 7 : Turbo-code (codeur)

Le décodage élémentaire est réalisé à l'aide de la mise en série des décodeurs relatifs à chaque codeur, tenant compte de l'entrelaceur interne et du caractère systématique. La sortie du décodeur doit obligatoirement fournir une information de qualité: des algorithmes de type SOVA ("Soft Output Viterbi Decoder") ont été développés à cet usage réf [7].

Par construction, ce décodeur fournit une information supplémentaire Wk, corrélée avec la donnée utile Dk et faiblement corrélée avec les entrées du décodeur grâce à l'entrelaceur interne. Ainsi cette variable agit comme une source de diversité pour une nouvelle tentative de décodage. Ce décodage élémentaire est donc mis en cascade, chaque étape améliorant la qualité du décodage.

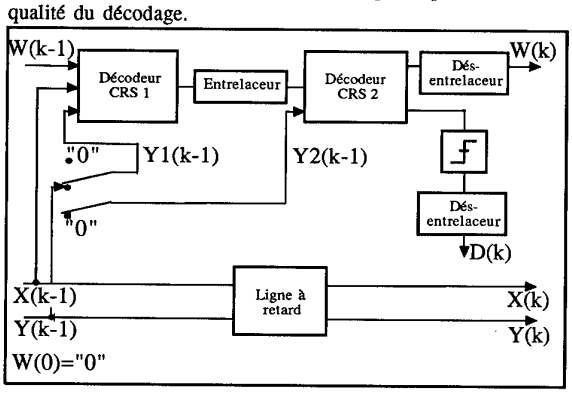


Figure 8 : Décodeur

Ces nouveaux codes permettent de s'approcher de la limite de Shannon, pour peu que la puissance de calcul soit disponible et que le délai de décodage soit compatible du type d'information transmise (la phonie numérique semble exclue). Exemple pour un code de redondance 2, réf [6] :

Eb/N0 pour un TEB de 10-5 (codage binaire)						
	Nombre d'itérations					
Shannon	0 (non codé)	1	2	3	6	18
0 dB	9.6 dB	4.6 dB	2.7 dB	1.8 dB	1.0 dB	0.7 dB

Ces nouveaux codes ont également récemment été associés à des modulations M-aires (M-PSK ou M-QAM) améliorant de l'ordre de 2 dB les performances des modulations codées (MC), en canal gaussien à TEB=10^{-5}. Ils semblent également adaptés à la transmission sur canaux perturbés type Rayleigh, HF par exemple, où ils affichent un gain de plus de 4 dB par rapport aux MC (cf réf [8]).

4. TECHNIQUES MODEM

4.1 Modulations à bande étroite

En communications numériques, le choix d'une modulation résulte toujours du compromis entre la bande disponible et le bilan énergétique atteignable, ceci pour un service donné. Savoir si un système donné doit utiliser une modulation associée à un codage pour minimiser l'énergie requise, ou dans l'autre dimension, pour minimiser la bande passante occupée, dépend du coût de chaque dimension. Les communications par satellite doivent optimiser le bilan de liaison par exemple, alors que les réseaux de radiocommunications cellulaires optimisent la bande passante.

La figure 9, extraite de réf [9], donne le positionnement dans le plan Eb/N0 (énergie) & Brf.Tb (bande passante radio normalisée par le rythme bit) des modulations les plus usitées, codées ou non. Sont présentes :

- les modulations linéaires de phase, type QPSK. Elles sont à enveloppe constante mais perdent cette caractéristique quand elles sont filtrées pour accroitre leur efficacité spectrale.
- les modulations linéaires d'amplitude et de phase, type QAM. Elles sont incontournables pour obtenir une grande efficacité spectrale (4 bits/S/Hz par ex.) et peuvent être très efficientes en énergie quand elles sont associées aux modulations codées.
- les modulations de fréquence à phase continue, dont l'exemple type est la MSK. Des développements récents ont permis de définir des CPM multi-états rélativement performantes en terme d'efficacité spectrale ce qui est intéressant pour les applications de radiotéléphonie portable ou portative où la caractéristique d'enveloppe constante permet une simplification de la radio et une minimisation de la consommation du poste (ampli en classe C, F.I. limitées,..)

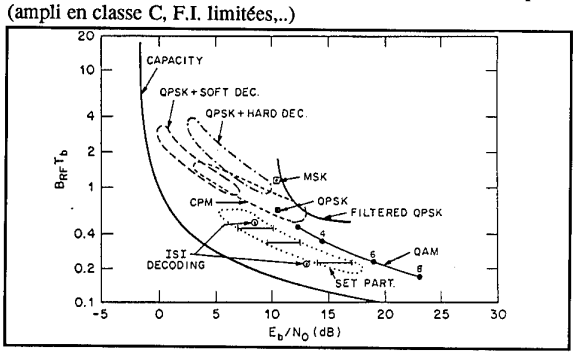


Figure 9 : Efficacité spectrale à TEB=10^{-5}

4.2 La démodulation

Dans la suite nous nous plaçons dans le cas de la démodulation cohérente de modulations linéaires.

Le démodulateur réalise les opérations suivantes (cf figure 10), dans l'ordre chronologique généralement en phase d'accrochage, en parallèle en phase de poursuite des dérives :

- filtrage adapté à la modulation
- détection du signal et synchronisation
- récupération du rythme
- récupération de la porteuse (fréquence et phase)
- démodulation proprement dite avec décision souple.

La décision souple, généralisée par l'usage des traitements numériques, permet de présenter au décodeur un canal équivalent "bruit gaussien additif" au lieu d'un canal binaire symétrique ce qui améliore de plusieurs dB le gain de codage.

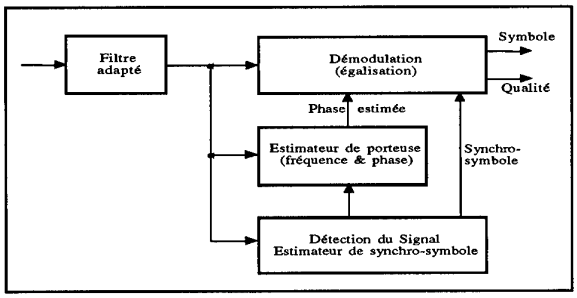


Figure 10 : Démodulateur

L'avènement des processeurs de traitement du signal ainsi que des composants spécialisés (ASIC) a permis le développement de traitements numériques par blocs au maximum de vraisemblance pour les différents estimateurs (temps, fréqeunce, phase); ainsi de nombreuses réalisations sont aujourd'hui à moins de 1 dB de la théorie.

Le cas des transmissions dans des canaux présentant des multi-trajets et des non-stationnarités (évanouissement) nécessite des traitements complémentaires : c'est le cas de la HF ou des transmissions radiomobiles en environnement urbain.

La notion de filtre adapté intègre l'adaptation au canal de transmission ce qui nécessite son estimation. Celle-ci est réalisée périodiquement par l'insertion de séquences de modulations connues dans la communication. La présence de multi-trajets de propagation implique l'utilisation d'un dispositif d'égalisation dans le démodulateur dès que l'on veut réaliser des communications haut-débit, c'est à dire que le rythme bauds est supérieur à l'inverse de l'étalement temporel du canal.

Dans le cas des transmissions HF, le canal est souvent perturbé par des interférences bande étroite vu que la propagation est difficilement maitrisable. Celles-ci posent un double problème. D'une part la mise en place un dispositif de réjection pour permettre la synchronisation initiale, préalable à tout traitement; d'autre part, l'élimination de cette interférence pendant le processus de démodulation (donc également d'estimation de canal) par blanchiment du spectre par ex. Les méthodes de réjection, rendues possibles par les traitements numériques, sont exposées aux paragraphes 4 et suivants.

4.3 L'exemple des transmissions HF

Le canal HF en propagation ionosphérique présente des multi-trajets fluctuants qui créent en réception des évanouissements sélectifs en fréquence et de l'interférence intersymbole (même phénomène interprété dans la dimension temporelle). Les progrès technologiques ont donné naissance à des sytèmes de transmission de données dans lesquels l'égalisation de canal et la correction des erreurs ainsi que des techniques de modulations efficaces autorisent des débits de 2400 à 9600 bps, cf réf [10].

Les performances reposent sur une démodulation cohérente et une égalisation performante basée sur l'estimation du canal. A cette fin des symboles de référence sont insérées dans la transmission. Deux approches duales sont possibles :

- la première historiquement, quoique utilisée en démodulation différentielle (ANDVT), est l'utilisation d'un modem parallèle où les symboles de modulation sont répartis sur un multiplex de sous-porteuses modulées à faible débit (trames ayant typiquement un rythme de 30 à 40 bauds). Ces trames comportent un temps de garde inutilisé à la démodulation pour assurer une stationarité du canal sur le symbole démodulé : ceci permet de s'affranchir de l'interférence intersymboles. Des trames comportant des symboles connus sont insérées régulièrement; elles échantillonnent en fréquence et en temps la réponse du canal HF et permettent ainsi par filtrage et interpolation la réalisation d'un dispositif d'égalisation cohérente dans le domaine fréquentiel.

- la deuxième approche est la modulation série (mono-porteuse modulée) pour laquelle l'égalisation est réalisée dans le domaine temporel à l'aide de séquences de références (Synchro-Trames & Mini-Probes). Cette approche a été aujourd'hui préférée pour les formes d'ondes normalisées présentant un débit allant jusqu'à 2400 bps.

En effet l'approche série est potentiellement plus performante que l'approche parallèle pour deux raisons principales:

- le facteur crête d'une modulation parallèle est de l'ordre de 10 à 12 dB comparés aux 4 à5 dB d'une modem série MPSK filtrée,

- en présence de multi-trajets, l'égaliseur série sait exploiter la diversité ce qui se traduit par des courbes de TEB plus pentues.

Mais l'écart de facteur crête peut être fortement réduit : dans sa réalisation d'un modem parallèle cohérent THOMSON a obtenu par écrétage un facteur crête de 4 à 5 dB sans perte de performance.

Par ailleurs, la modulation série nécessite la mise en oeuvre d'un égaliseur qui devient de plus en plus complexe avec l'augmentation du débit. Ainsi les mises en oeuvre courantes utilisent-elles des égaliseurs DFE "Decision Feedback Equalizer" ou DDE "Data Directed Equalizer" intégrant une décision sur les symboles codés dans le processus : ce faisant les performances s'éloignent significativement des performances optimales du maximum de vraisemblance car ces symboles ne bénéficient pas de la protection du code correcteur et les erreurs sont propagées dans l'égaliseur. L'approche série nécessiterait également une optimisation conjointe de l'égalisation et du décodage dont la complexité, avec les algorithmes connus, n'est pas compatible d'une mise en oeuvre avec les DPS actuels. Le modem parallèle peut quant à lui rester très proche des performances théoriques pour une complexité réduite. Cette comparaison est affichée figure 11 pour un canal à trois trajets : les modems utilisent de la 8PSK codée pour un débit utile de 2400 bps, les algorithmes sont simulés sur calculateur (égalisation & suivi des dérives temps-fréquence), les performances sont comparées aux courbes théoriques.

Il est cependant probable que dans un avenir proche, les algorithmes des modems série accomplissent des progrès significatifs. Dans l'intervalle, les formes d'ondes parallèles constituent un choix alternatif de faible complexité pour les débits moyens ($\leq$ 2400 bps), et une solution performante pour les débits supérieurs jusqu'à 9600 bps.

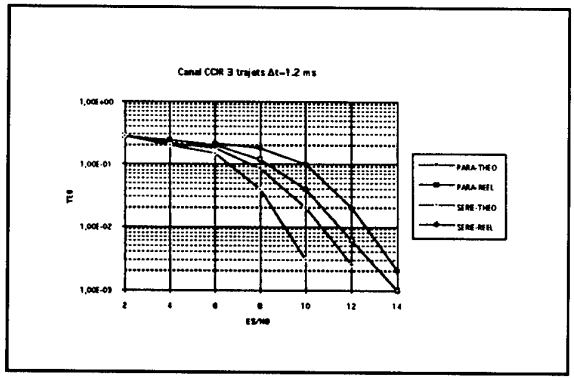


Figure 11 : Modems série & parallèle

5. TECHNIQUES DE REJECTION D'INTERFERENCES BANDE ETROITE PAR EXCISION DE SPECTRE

5.1 Méthodes classiques

Le principe des techniques de réjection d'une interférence bande étroite par excision de spectre est d'utiliser la différence de largeur de bande qui existe entre l'interférence bande étroite et le signal utile considéré comme large bande.

Il existe deux grands types de méthodes pour réaliser l'excision:

- la première méthode utilise l'estimation de la puissance du signal reçu dans le domaine fréquentiel. Un banc de filtres permet de décomposer le signal en sous bandes. La sous bande occupée par l'interférence est mise à zéro ce qui permet de reconstruire le signal traité.

Le banc de filtres peut être élaboré par transformée de Fourier ou à partir de filtres miroir en quadrature, ou **filtres QMF (quadrature mirror filter).** Le banc de filtre QMF consiste à diviser le spectre en 2 sous bandes d'égale largeur, à sous échantillonner par deux, à traiter les sous bandes (excision) et surtout reconstituer un signal de sortie transparent aux repliements spectraux que l'on a pu provoquer lors de la division. Un banc de filtres comportant 2^n filtres peut être obtenu en réitérant l'opération précédente n fois .

L'inconvénient de cette méthode est le retard induit en vue d'une estimation correcte de la puissance de chaque sous bande. La détection d'un brouilleur est assez longue et par conséquent l'adaptation de l'excision à un brouillage non stationnaire est difficile .

- la deuxième méthode utilise un filtrage adaptatif prédictif utilisant la durée de cohérence de l'interférence afin de la rejeter. On utilise le fait que l'interférence bande étroite reste prédictible plus longtemps que le signal utile de bande plus large. Un échantillonnage correct à la fréquence de Nyquist fournit des échantillons corrélés. L'échantillon courant peut être prédit à partir des échantillons passés, l'erreur de prédiction correspond au signal utile large bande.

L'excision par filtrage adaptatif consiste à :

- estimer le prédicteur du signal reçu en minimisant l'erreur quadratique moyenne entre le signal reçu et la sortie du prédicteur,c'est à dire la puissance de:

$$e_n = x_n - \sum_{k\neq 0}^{\infty} p_k x_{n-}$$

où p_n désigne la réponse impulsionnelle du prédicteur estimé .

- filtrer le signal reçu par le filtre blanchisseur $b_n=\delta_n-p_n$ où δ_r est le symbole de Kronecker qui vaut 1 pour n=0 et zéro ailleurs. C'est le filtre blanchisseur qui réalise l'excision

L'architecture du filtre est donnée sur la **figure 12.**

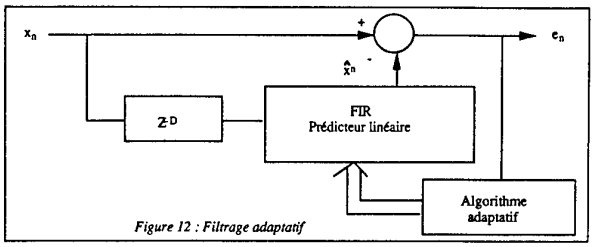


Figure 12 : Filtrage adaptatif

Pour estimer le prédicteur il existe une grande variété de critères et d'algorithmes possibles .

Le critère d'optimisation est généralement du type des moindres carrés. On peut utiliser avantageusement un algorithme des moindres carrés rapides de préférence à un algorithme du gradient. La précision de l'exciseur dépend du rapport J/(S+N) qui est le rapport brouilleur sur signal plus bruit de fond dans la bande. Pour l'algorithme adaptatif qui estime le prédicteur, plus le brouilleur est fort et plus l'exciseur est précis.

La présence de multi-trajets nuit à l'efficacité du filtrage. Une première méthode peut être de fixer un délai D assez important pour dépasser tout retard de multitrajet qui introduirait une corrélation de la contribution du signal utile dans les échantillons .

On a alors à faire un compromis entre une très bonne prédiction de l'interférence accompagnée d'une légère dégradation du signal utile et une fidèle restitution du signal utile additionnée d'un résidu de brouilleur .

Une méthode consistant à ajouter un bruit fictif avant l'adaptation des coefficients permet de masquer les trajets secondaires ou plus exactement de décorréler les multi-trajets. Une fois l'adaptation des coefficients réalisée, le bruit fictif est retiré afin d'éviter de dégrader inutilement le signal utile avant filtrage.

5.2 Le filtrage cyclique

Les deux techniques précédentes de réjection de brouilleur par excision de spectre ont été développées pour un signal stationnaire, c'est à dire un signal dont la fonction de corrélation est invariante dans le temps .

Les techniques classiques citées plus haut "stationnarisent " les signaux en n'utilisant que leur fonction de corrélation moyenne.

Or les signaux traités en transmission sont généralement constitués de signaux modulés sur porteuse :ce ne sont pas des signaux stationnaires mais des signaux dits "cyclostationnaires " dont la moyenne et la fonction d'autocorrélation sont périodiques en temps. Cette périodicité est induite outre la fréquence porteuse par les opérations de modulation et de codage de la transmission numérique.

L'autocorrélation des signaux dépendant du temps, la réponse impulsionnelle des filtres qui doivent être utilisés sur ces signaux doit également dépendre du temps .

Il a été montré dans le cas du filtrage de Wiener que des filtres à réponse impulsionnelle variable dans le temps pouvaient améliorer considérablement les performances. Dans certains cas il est possible de séparer des signaux dont les spectres se recouvrent entièrement, ce qui évidemment n'est pas possible en filtrage classique .

Pour tirer parti de la nature cyclostationnaire des signaux modulés, il faut des filtres à réponse impulsionnelle variable portant à la fois sur le signal complexe et sur son conjugué:

$$h(t,u)=\sum_{\eta}^{\infty} e^{j2\Pi\eta u}h_{\eta}(t-u)$$

$$y(t)=\int h_{Z}(t,u)z(u)du+\int h_{Z^{\cdot}}(t,u)z^{*}(u)$$

On peut concevoir des filtres cycliques ayant des réponses périodiques avec des périodes identiques à celles des signaux qu'ils traitent. La réponse impulsionnelle peut se décomposer en série de Fourier comme la fonction d'autocorrélation des signaux :

Pour $\eta=0$, on retrouve la solution stationnaire ne dépendant que de l'écart (t-u).

Le filtrage d'un signal z(t) par une telle réponse impulsionnelle donne :

$$y(t)=\int h(t,u)z(u)du=\sum_{\eta}^{\infty}\int (h_{\eta}(t-u)\,(e^{j2\Pi\eta u}z(u))du)$$

Dans cette écriture ,on voit que l'on peut écrire le filtrage cyclique comme une somme de filtres stationnaires portant sur des signaux $e^{j2\Pi\eta u}z(u)$ qui sont en fait le signal d'origine décalé en fréquence de η.

On a donc équivalence entre une approche temporelle du filtrage cyclique, dans laquelle on change la réponse impulsionnelle du filtre à chaque instant, et une approche fréquentielle, dans laquelle le filtre est réalisé par une somme de filtres stationnaires h_{η} portant sur des versions décalées en fréquence du signal d'entrée.

L'architecture d'un filtre cyclique est donnée en **figure 13**. Elle correspond à un filtre stationnaire multi-dimensionnel H(t) dans lequel on entre le signal z(t) décalé d'un certain nombres de fréquences αi, et son conjugué décalé d'un certain nombres de fréquences βi, αi et βi étant les fréquences cycliques obtenues après développement en série de Fourier des fonctions de corrélation $\Gamma zz\ (t,\tau)$ et $\Gamma zz^{*}\ (t,\tau)$.

Les applications des filtres cycliques aux problématiques d'excision nécessitent une analyse préalable des fréquences cycliques pertinentes.Une excision devra donc être précédée d'une mesure de corrélation spectrale et une détection d'énergie sur toutes les fréquences cycliques qui n'appartiennent pas à la transmission utile .

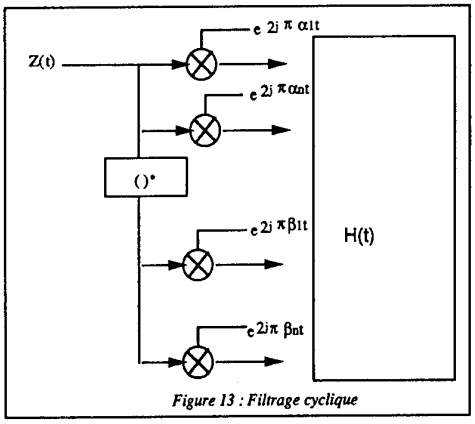


Figure 13 : Filtrage cyclique

Malgré une complexité accrue, l'application de ces méthodes de filtrage à la lutte contre les interférences sera certainement un axe fort de recherches futures. Ces techniques apparaissent également prometteuses pour resserrer les canaux dans des transmissions FDMA, ce qui pourrait déboucher sur une augmentation de la capacité des liaisons satellites par exemple.

6. TECHNIQUES DE FILTRAGE D'ANTENNES APPLIQUEES A LA LUTTE CONTRE LES INTERFERENCES ET LES EFFETS DES TRAJETS MULTIPLES

Les techniques décrites précédemment ne permettent un traitement des signaux que sur l'axe des temps ou sur l'axe des fréquences. Elles peuvent se révéler inefficaces dans des situations où les interférences sont très fortes et qu'elles occupent toute la bande de réception . Elles n'optimisent pas de manière idéale le compromis à réaliser entre le filtrage adapté au signal et la réduction des interférences intersymboles en présence de trajets multiples.

L'utilisation de la dimension spatiale par des techniques de "filtrage d'antenne" peut apporter un gain de performances important dans des applications "d'antibrouillage" lorsque ces techniques sont appliquées à la réjection d'interférence ou de "diversité spatiale" lorqu'elles sont appliquées à l'optimisation de la réception du signal en présence de multi-trajets pour lutter contre les fadings.

6.1 Techniques d'antibrouillage

L'outil privilégié est l'antenne adaptative qui a la faculté de détecter les sources interférentes et de les supprimer en construisant dans leur direction des trous dans son diagramme de rayonnement tout en améliorant la réception du signal utile. Le schéma fonctionnel d'une antenne adaptative est donné sur la **figure 14**. Elle est composée d'un réseau de N capteurs et de N filtres contrôlés par un processeur qui met en oeuvre un algorithme adaptatif. Cet algorithme vise à optimiser un critère fonction de l'information disponible sur le signal. En télécommunication le critère idéal est la minimisation de la probabilité d'erreur par bit après décodage.

Ce critère est généralement difficile à optimiser, aussi on préfère utiliser des critères plus simples à utiliser comme la maximisation du rapport signal sur bruit +interférences ou la minimisation de la puissance de bruit total sous contrainte de non annulation des pondérations .

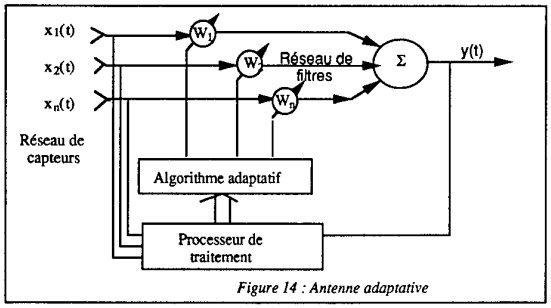


Figure 14 : Antenne adaptative

La discrimination signal utile des interférences peut s'effectuer de cinq façons différentes par : la direction d'arrivée, la forme d'onde, le temps, le spectre ou encore de manière aveugle sans information sur le signal utile.

- Filtrage sur direction d'arrivée .

Cette méthode qui nécessite la connaissance de la direction d'arrivée et donc du front d'onde au voisinage des capteurs. Cette antenne est optimale, en absence de multitrajets et de "dépointage" elle correspond au **filtre adapté spatial (FAS)** qui conduit à une maximisation du rapport **signal à bruit + interférences**. Elle est obtenue en minimisant la puissance totale de sortie en mettant un gain unité dans la direction d'arrivée du signal utile .

Cette antenne est sensible aux erreurs sur la direction d'arrivée et à la présence de trajets multiples. Pour ces raisons, il est souvent nécessaire de doter cette antenne d'une plus grande robustesse aux erreurs en introduisant des contraintes supplémentaires visant à élargir le lobe principal de l'antenne et à affaiblir la puissance du signal vue par l'algorithme d'adaptation. Une des structures de réalisation de cette antenne est donnée sur la **figure 15**.

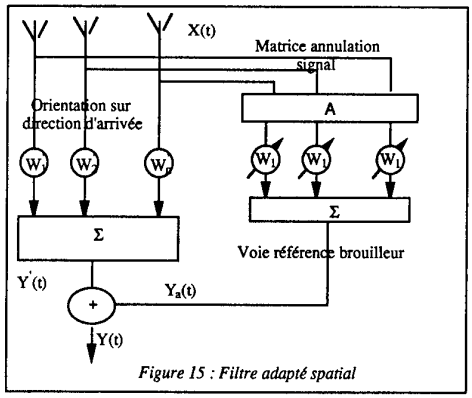


Figure 15 : Filtre adapté spatial

- Filtrage sur forme d'onde

Cette méthode est utilisée lorsque la modulation ou la forme d'onde du signal utile permet la génération d'un signal de référence, corrélé avec l'utile et non corrélé avec le bruit et les interférences. C'est en particulier le cas lorsque la forme d'onde émise comporte des séquences connues de synchronisation. Le critère utilisé est alors un critère de minimisation de l'erreur quadratique moyenne entre le signal de référence et la sortie de l'antenne , ce qui permet de rejeter toutes les sources décorrélés avec le signal de référence. Le schéma fonctionnel de cette antenne est donné sur la **figure 16**.

Une telle antenne ne nécessite pas la connaissance de la direction d'arrivée. Les multitrajets corrélés sont seuls pris en compte par cette antenne. Ce concept conduira au **modem spatio-temporel** décrit plus loin .

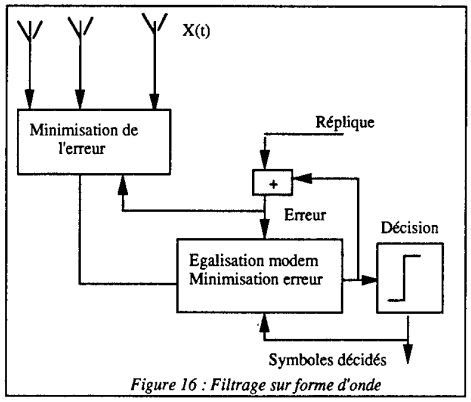


Figure 16 : Filtrage sur forme d'onde

- Filtrage par discrimination temporelle

Cette méthode peut être utilisée sur des transmissions à évasion de fréquence. L'antenne minimise alors la puissance totale de sortie en absence de signal utile, c'est à dire par anticipation d'un palier EVF, sous une contrainte non directionnelle de type inversion de puissance et élimine ainsi les brouilleurs stationnaires. La contrainte utilisée, qui consiste à laisser un des capteurs pondéré par l'unité, impose au diagramme de rayonnement de l'antenne d'être le plus omnidirectionnel possible dans les directions non brouillées de façon à éviter l'annulation du signal utile lors de sa réception. L'antenne ainsi optimisée est appelée **antenne à contrainte d'inversion de puissance et à référence bruit seul (ACIP RBS).** Cette antenne n'optimise pas le gain dans la direction de l'utile et elle est mise en difficulté sur une situation d'inteférence non stationnaire. Elle ne permet pas un traitement optimal des multi-trajets .

- Filtrage par discrimination spectrale

Cette méthode est calquée sur la précédente, elle est utilisée lorsque la bande du signal utile est plus faible que la bande des brouilleurs. L'antenne minimise alors la puissance totale de sortie dans la bande extérieure à la bande du signal utile, sous une contrainte non directionnelle de type inversion de puissance.

- Filtrage en aveugle

Dans certaines situations aucune information n'est disponible sur le signal. L'introduction d'une réplique est également souvent coûteuse en débit. Il peut donc s'avérer judicieux d'utiliser une antenne adaptative n'exploitant aucune information a priori sur le signal utile. Une telle antenne vise à séparer les signaux reçus sur les différents capteurs du réseau par un filtrage linéaire spatial multi -sorties G comme l'illustre la **figure 17** .

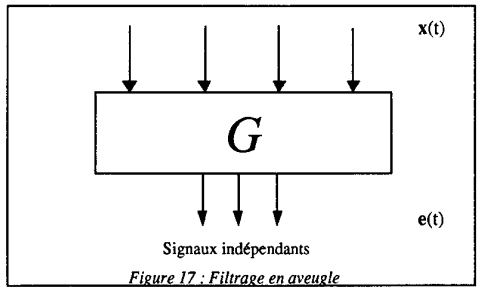


Figure 17 : Filtrage en aveugle

Il existe deux familles de méthodes , qui exploitent les statistiques des signaux à l'ordre deux et à l'ordre quatre.
Les séparateurs à l'ordre deux n'exploitent que la décorrélation supposée des signaux reçus. Ils visent à mettre en oeuvre une matrice G générant des sorties décorrélées.
Les séparateurs aux ordres supérieurs (quatre) exploitent l'indépendance statistique supposée des signaux reçus. Ils visent à mettre en oeuvre une matrice G générant des sorties statistiquements indépendantes .
Les méthodes à l'ordre deux ne permettent une séparation des sources que si celles-ci sont bien espacées angulairement et de puissances très différentes. Les méthodes à l'ordre quatre encore à l'étude sont efficaces sur des sources " **non gaussiennes** "

6.2 Techniques de diversité spatiale.

Les traitements multi-capteurs de filtrage d'antenne peuvent être appliqués à la mise en oeuvre de traitements se rapprochant de l'optimum sur un canal de propagation à multitrajets .
Différentes approches de la diversité spatiale sont possibles,l a plus simple consistant à commuter deux antennes suffisamment écartées spatialement. Une antenne adaptative telle que décrite précédemment peut également être utilisée en amont d'un modem. Mais le traitement optimal des multitrajets passe par un couplage des parties spatiales et temporelles du traitement .

Dans le cas d'une transmission série, le récepteur multicapteurs optimal de la **figure 18** est au sens du maximum de vraisemblance composé:

- du filtre adapté spatio temporel (FAST) à la forme d'onde multicapteur qui s'exprime par :

$$W(t)=R_b^{-1}(t)*G(t)$$

- un échantillonneur au rythme symbole,
- un organe de décision qui peut être réalisé par un algorithme de Viterbi .

G (t) est la forme d'onde multicapteurs reçue, constituée de la forme d'onde convoluée avec la réponse impulsionnelle du canal multicapteur. Rb(t) est la matrice de corrélation du bruit de fond supposé gaussien .

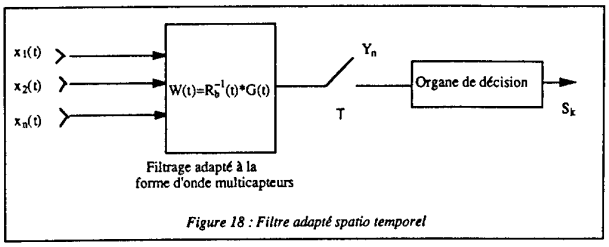


Figure 18 : Filtre adapté spatio temporel

La structure spatio-temporelle du FAST utilise le fait que les trous de fadings ne se produisent pas au même instant sur chaque antenne, elle permet en outre de bénéficier du gain d'antenne en 10 log N. Le FAST en présence de brouilleurs permet de rejeter N-1 brouilleurs, quel que soit le nombre de trajets associés à chacun des brouilleurs. En effet le FAST réalise un traitement temporel de remise en phase de tous les trajets en utilisant un seul degré de liberté spatial.

Dans le cas d'un modem parallèle, l'architecture optimale multicapteurs de la **figure 19** consiste à placer un modem parallèle mono-voie sur chacun des capteurs et à sommer les contributions sur chaque fréquence avant décodage. L'interprétation du gain en présence de multitrajets est plus facile dans ce cas.
Pour deux trajets de même amplitude le résultat du filtrage adapté multi-voies se met sous la forme :

$$|H(f)|=k(1+|S_1^{\dagger} S_2|\cos(2\pi f (t_1-t_2)+\varphi))$$

k est une constante , S1 et S2 sont les vecteurs directeurs des deux trajets sur le réseau et ϕ la phase initiale.

- si les vecteurs directeurs des deux trajets sont identiques le traitement multivoies ne change pas le fading sélectif en fréquence, le seul gain est le gain d'antenne.
- si les vecteurs directeurs sont orthogonaux, le fading sélectif en fréquence est totalement supprimé.
- dans le cas intermédiaire, en plus du gain d'antenne, la diversité spatiale apporte une amélioration des performances en présence de fading sélectif en fréquence.

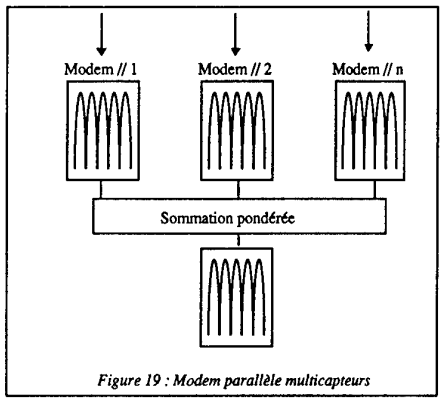


Figure 19 : Modem parallèle multicapteurs

CONCLUSION

La nécessité d'accroître les services et les performances des systèmes de radiocommunication ouvre un champ d'applications important à la mise en oeuvre des techniques de traitement du signal dont certaines ont été présentées ici.

Les solutions techniques existent pour compresser au minimum l'information à transmettre, atteindre les limites de capacité des canaux, éliminer les sources d'interférences et compenser les effets des trajets multiples, en particulier par des "antennes intelligentes" et des traitements multi-voies.

L'application de ces techniques ira de pair avec les progrès dans la technologie du numérique, dans la miniaturisation des équipements et des chaînes radio.

BIBLIOGRAPHIE

[1] "NATO STANAG 4479 : A standard for an 800 bps vocoder and channel coding in HF-ECCM system"
B. Mouy, P. De la Noue, G. Goudezeune, ICASSP 95

[2] "Combined source-channel coding for a very noisy channel"
A. J. Vigier, EUROSPEECH 91

[3] "Low rate speech coders for robust VHF radiocommunications"
A.J. Vigier, S. Le Dreau, MILCOM 93

[4] "Error-Correction Coding for Digital Communications"
G.C. Clark, J. Bibb Cain, PLENUM PRESS 1982

[5] "Error Correcting Codes"
W.W. Peterson, E.J. Weldon, MIT PRESS 1972

[6] "Near Shannon limit error-correcting coding and decoding : turbo-codes"
C. Berrou, A. Glavieux, P. Thitimajshima, ICC'93

[7] "A low complexity soft-output viterbi decoder architecture"
C. Berrou, P. Adde, E. Angui, ST. Faudeil, ICC' 93

[8] "Turbo-codes and high spectral efficiency modulation"
S. Le Goff, C. Berrou, A. Glavieux

[9] "Overview of advances in constant-envelope coded modulation"
J.B. Anderson, C.E. Sundberg

[10] "Comparaison entre les formes d'onde série et parallèle pour les transmissions HF"
D. Pirez, D. Merel, AGARD 95

[11] "Antenne adaptative d'une structure linéaire à une structure non linéaire de Volterra"
P. Chevalier, Thèse de doctorat de l'université de Paris Sud de juin 1991

[12] "MLSE Antenna diversity equilization of a jammed Frequency-Selective Fading Channel"
P. Vila, F. Pipon, D. Pirez, L. Fety, EUSIPCO 94

[13] "MMSE Antenna diversity equilization of a jammed frequency selective fading channel"
P. Vila, F. Pipon, D. Pirez, L. Fety, ICASSP 95

DISCUSSION

Discussor's name: Dr. Yavuz

Comment/Question:

1. As you know, speech coders perform differently with different languages. Have you any results for languages such as Portuguese which are known to perform quite differently with, e.g. LPC-10.

2. Your statement about the single tone (ST) and multi-tone (MT) modems is surprising! For the HF channel where time domain equalization converts multi-path to diversity combination ST modems are clearly much superior. Equalizing in frequency domain only the frequency selective fading with much less diversity gain. Your comments please.

Author/Presenter's reply:

1. The different 800 h/s and 2400 h/s have been tested in English and French. I think that for European languages the new 2400 h/s coder will provide intelligibility independent of the language.

2. A coherent parallel modem which is limited to reduce to 4/5 dB the amplitude modulations and associated with a good TCM is equivalent to a DFE modem. The progress on ST modem concerns the integration of decoding in the equalizer.

LES SEQUENCES GQ SEQUENCES Q-AIRE ORTHOGONALES A CORRELATION PARFAITE

C. GOUTELARD
LETTI — Université Paris-Sud — Bâtiment 214
91405 ORSAY — France

Résumé

La recherche de séquences présentant de bonnes fonctions de corrélation périodiques est un sujet sur lequel de nombreuses études ont été faites. Il a pu être montré qu'à l'exception d'un cas trivial, il n'existait pas de séquences binaires à corrélation parfaite. Des séquences binaires à corrélation presque parfaite, les séquences WG (Wolfmann-Goutelard), ont été proposées en 1992. Ces séquences sont telles qu'en dehors du pic central de corrélation et d'un point situé au milieu de la période de la fonction de corrélation, cette fonction est toujours nulle. Les séquences à corrélation parfaite ne peuvent donc exister que dans les séquences Q-aires. Les séquences GQ (séquences Goutelard Q-aires) sont des séquences qui présentent des fonctions de corrélation parfaites et qui peuvent être décomposées en sous-séquences orthogonales à corrélation presque parfaite.

Les séquences GQ sont construites à partir d'un ensemble de séquences orthogonales qui, par entrelacement, donnent une séquence nommée séquence complète à fonction de corrélation parfaite. L'addition terme à terme des éléments de l'ensemble des séquences de base permet de construire une séquence plus courte que la séquence complète mais possèdant les mêmes propriétés de corrélation. Cette séquence, appelée séquence analogique, est donc à corrélation parfaite.

Cette présentation démontre l'existence de ces séquences et détermine leur structure. Il est montré que ces séquences existent en grand nombre et pour une très grande variété de longueurs. Les propriétés de ces séquences sont étudiées et il est défini un algorithme de construction très simple permettant de les utiliser. Ces séquences peuvent exister pour des longueurs minimales de 4 et les longueurs ne sont pas limitées.

I. - INTRODUCTION

L'utilisation de séquences périodiques ou apériodiques à fonctions de corrélation nulles en dehors du pic central a toujours été d'un grand intérêt dans les problèmes de radiolocalisation, de télécommunication et de radar. On demande, la plupart du temps, à ces séquences d'être à amplitude constante pour des raisons technologiques dues essentiellement à des problèmes liés aux émetteurs. Dans les séquences apériodiques, des solutions ont été depuis longtemps proposées, telles que les séquences de Golay qui présentent des caractéristiques de corrélation parfaite mais qui ne sont pas des séquences à amplitude constante. Des séquences périodiques polyphases [1] [2] [3] présentent de telles caractéristiques mais elles sont en nombre limité et ne conviennent pas à toutes les applications. Dans les séquences binaires des séquences à corrélation presque parfaite, les séquences WG (Wolfmann-Goutelard) [4], la fonction de corrélation périodique est nulle en dehors du pic central exepté pour un seul point situé au milieu de la période de la fonction de corrélation. D'autres séquences telles que les séquences pseudo-aléatoires ou les séquences pseudo-orthogonales [5] ont des fonctions de corrélation qui approchent la corrélation parfaite, cependant, en dehors du pic central, ces fonctions de corrélation ne sont jamais nulles.

L'étude présentée repose sur la définition d'un ensemble de séquences orthogonales à corrélation presque parfaite. Il est démontré que de telles séquences, à l'exception d'un cas trivial, ne peuvent être binaires et que l'orthogonalité de ces séquences interdit d'avoir pour chacune des auto-corrélations parfaites. L'étude est donc orientée vers la recherche de séquences q-aires et les contraintes imposées à cette recherche consiste à définir une séquence complète formée par l'entrelacement de ces séquences. L'orthogonalité des séquences permet de définir une seconde séquence, obtenue par la somme algébrique terme à terme des symboles, constituant une séquence plus courte appelée séquence analogique qui possède elle aussi une fonction d'auto-corrélation parfaite. Il est montré qu'une fonction d'auto-corrélation parfaite pour la séquence complète entraîne une fonction d'auto-corrélation parfaite pour la séquence analogique et réciproquement. Le développement permet de déterminer la structure de telles séquences et de montrer que cette structure est unique. Les séquences sont déterminées pour toutes les longueurs possibles d'existence et la méthode de construction de ces séquences est exposée. En dernier lieu, il est montré comment l'utilisation d'une séquence complète à l'émission peut être utilisée à la réception par la construction d'une séquence analogique qui permet de réduire de façon importante la complexité du calcul.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

II. - DEFINITIONS ET FORMALISME -

II.1. Suites et séquences.

Soit $A_q = \{0, 1, \ldots q - 1\}$ corps fini si q est premier, anneau dans le cas inverse qui sera celui considéré dans cette étude.

On définit une suite q-aire ou séquence q-aire par :

$$g = (g_0, g_1, g_2, \ldots g_i \ldots g_{N-1}) \mid g_i \in A_q \qquad (1)$$

et $$A_{qN} = \{g\}$$

On définit la séquence q-aire déduit de g par :

$$b = (b_0, b_1, b_2, \ldots b_i \ldots b_{N-1}) \quad \text{où} \quad b_i = V^{gi} \qquad (2)$$

avec $V^t = 1$, t premier avec N

On peut alors représenter la séquence q-aire par le polynôme formel :

$$b(x) = \sum_{i=0}^{N-1} b_i x^i \quad \text{dans} \quad C_{(N)} = A_{(x)} / (X^N - 1) \qquad (3)$$

On établit le formalisme suivant, dont une part est démontrée dans [4].

II.2. p transformé d'un polynôme.

On appelle p transformé de $b(x)$ le polynôme

$$\tilde{b}_p(x) = \sum_{i=0}^{N-1} b_{(W_p^{-h})} x^h \qquad (4)$$

où $W_p = \exp j \frac{2\pi p}{N}$ p premier avec N

et il est simple de démontrer que :

$$b(x) = \frac{1}{N} \sum_{i=0}^{N-1} \tilde{b}_{(W_p^l)} x^l \qquad (5)$$

II.3. Polynôme de corrélation.

Soit deux séquences représentées par leurs polynômes formels :

$$b_1(x) = \sum_{i=0}^{N-1} b_{1i} x^i \ , \ b_2(x) = \sum_{i=0}^{N-1} b_{2i} x^i$$

On définit le polynôme de corrélation $c_{12}(x)$ entre les deux polynômes $b_1(x)$ et $b_2(x)$:

$$c_{12}(x) = \sum_{k=0}^{N-1} c_k x^k = b_1^*(x) b_2(1/x) \qquad (6)$$

Si $b_1(x) = b_2(x) = b(x)$ alors on définit le polynôme d'autocorrélation $c(x)$ par :

$$c(x) = b^*(x) b(1/x) \qquad (7)$$

II.4 . p transformé d'un polynôme de corrélation.

Le p transformé du polynôme de corrélation $c_{12}(x)$ est défini par :

$$\tilde{c}_{12}(x) = \sum_{h=0}^{N-1} c_{12(W_p^{-h})} x^h$$

et il est facile de montrer que :

$$\tilde{c}_{12}(x) = \sum_{h=0}^{N-1} b^*_{1(W_p^{-h})} \cdot b_{2(W_p^{-h})} x^h \qquad (8)$$

et $$\tilde{c}_{12}(x) = \tilde{b}_1^*(x) \otimes \tilde{b}_2(x)$$

où $\otimes$ figure le produit d'Hadamart et pour l'autocorrélation

$$\tilde{c}(x) = \tilde{b}^*(x) \otimes \tilde{b}(x) \qquad (9)$$

II.5. q permutée d'une séquence.

Si $b(x)$ est une séquence, on appelle q permutée de $b(x)$ la séquence :

$$b_q(x) = b(x^q) \qquad (10)$$

II.6. p transformé d'un polynôme q permuté.

Le p transformé $\tilde{b}_{qp}(x)$ d'un polynôme q permuté $\tilde{b}_q(x)$ s'exprime par :

$$\tilde{b}_{qp}(x) = \sum_{l=0}^{N-1} b_{q(W_p^{-l})} x^l$$

soit

$$\tilde{b}_{qp}(x) = \sum_{l=0}^{N-1} b_{(W^{-pql})} x^l$$

d'où l'on déduit

$$\tilde{b}_{qp}(x) = \tilde{b}_{pq}(x) \qquad (11)$$

II.7. Polynôme Z étendu, polynôme de corrélation p transformé.

On définit le polynôme Z étendu de $b(x)$ par :

$$\underline{b}^Z(x) = \underline{b}(x^Z) \qquad (12)$$

Le souligné de la lettre indiquera par la suite que le polynôme est défini dans :

$$C_{(NZ)} = A_q(x) / \left(x^{ZN} - 1\right)$$

Le polynôme de corrélation entre deux polynômes étendus $\underline{b}_1^Z(x)$ et $\underline{b}_2^Z(x)$ est défini par :

$$\begin{aligned} \underline{b}_1^{Z*}(x)\,\underline{b}_2^Z(1/x) &= \underline{b}_1^*(x^Z)\,\underline{b}_2(1/x^Z) \\ &= \underline{c}_{12}(x^Z) = \underline{c}_{12}^Z(x) \end{aligned} \qquad (13)$$

Ce résultat indique que le polynôme de corrélation d'une séquence étendue est le polynôme étendu du polynôme de corrélation de la séquence.

Le p transformé de la séquence étendue est :

$$\underline{\tilde{b}}_p^Z = \sum_{h=0}^{ZN-1} \underline{b}^Z{}_{(W_p^{-h})} x^h = \sum_{h=0}^{ZN-1} \underline{b}_h^Z x^h \qquad (14)$$

On doit noter que :

$$\underline{\tilde{b}}_h^Z = \tilde{b}_{hZ[N]} \qquad (15)$$

où $hZ[N]$ dénote la valeur de hZ modulo N.

II.8. Polynôme uniforme.

On appelle polynôme uniforme un polynôme dont la norme des coefficients est unitaire et qui s'exprime par :

$$U_{I\ H}^{P\ Q}(x) = \sum_{i=0}^{P-1} W_{PQ}^{HQi} x^{I+iQ} \qquad (16)$$

avec $$W_{PQ} = \exp\left(\frac{j2\pi p}{PQ}\right) = W_p^{\frac{N}{PQ}}$$

si $Q \neq 1$ le polynôme est un polynôme étendu
si $I \neq 0$ le polynôme est translaté.
H est un entier arbitraire.

Son p transformé s'exprime par :

$$\tilde{U}_{I\ H}^{P\ Q}(x) = \sum_{h=0}^{Q-1} P W_{PQ}^{-hIP} x^{-H+hp} \qquad (17)$$

et d'après (9) le p transformé de son polynôme de corrélation s'écrit :

$$\tilde{c}_U(x) = \sum_{h=0}^{Q-1} P^2 x^{-H+hp} = P U_{0\ H}^{P\ Q} \qquad (18)$$

ce qui entraîne pour le polynôme de corrélation l'expression :

$$c_U(x) = \sum_{i=0}^{P-1} P W_{PQ}^{HQi} x^{iQ} = P U_{0\ H}^{P\ Q} \qquad (19)$$

Toute séquence à polynôme uniforme a un p transformé, un polynôme de corrélation, et un p transformé polynôme de corrélation uniforme.

Les applications des polynômes vers les polynômes de corrélation définissent un homorphisme.

III. - LES SEQUENCES GQ -

Les séquences GQ sont des ensembles de Z séquences de longueur N définies par :

$$\begin{aligned} b_1 &= \left(b_{10}, b_{11}, \ldots b_{1i} \ldots b_{1N-1}\right) \\ b_2 &= \left(b_{20}, b_{21}, \ldots b_{2i} \ldots b_{2N-1}\right) \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ b_z &= \left(b_{z0}, b_{z1}, \ldots b_{zi} \ldots b_{zN-1}\right) \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ b_Z &= \left(b_{Z0}, b_{Z1}, \ldots b_{Zi} \ldots b_{ZN-1}\right) \end{aligned}$$

représentées par les polynômes

$$b_z = \sum_{i=0}^{N-1} b_{zi} x^i$$

et vérifiant les propriétés suivantes :

III.1. Propriété 1 : Intercorrélation entre les séquences.

Toutes les intercorrélations des séquences prises 2 à 2 sont nulles. On dira que les séquences prises 2 à 2 sont strictement orthogonales.

Cette condition se traduit par :

$$c_{z_1 z_2}(x) = \tilde{b}_{z_1}^*(x) \cdot b_{z_2}(1/x) \equiv 0$$

soit encore

$$\tilde{c}_{z_1 z_2}(x) = \tilde{b}_{z_1}^*(x) \otimes \tilde{b}_{z_2}(x) \equiv 0$$

ce qui entraîne compte tenu de (8) et (9)

$$\tilde{c}_{z_1 z_1}(x) \otimes \tilde{c}_{z_2 z_2}(x) \equiv 0 \qquad (20)$$

La stricte orthogonalité des séquences entraîne, d'après (13), la stricte orthogonalité des séquences étendues dans :

$$\underline{\tilde{c}}_{z_1 z_1}^Z(x) \otimes \underline{\tilde{c}}_{z_2 z_2}^Z(x) \equiv 0 \qquad \forall\, z_1 z_2 \in \{1, \ldots Z\} \qquad (21)$$

III.2. Propriété 2 : Séquence complète.

La séquence complète résultant de l'entrelacement des Z séquences est à fonction d'autocorrélation parfaite, c'est-à-dire nulle en dehors du pic central.
La séquence complète est définie par le polynôme :

$$\underline{B}(x) = \sum_{z=1}^{Z} x^{z-1} \underline{b}_z(x^Z) = \sum_{h=0}^{ZN-1} \underline{B}_h x^h \qquad (22)$$

Le polynôme de corrélation s'écrit :

$$\underline{C}_B(x) = \left[\sum_{z=1}^{Z} x^{z-1} b_z^*(x^Z)\right] \times \left[\sum_{z=1}^{Z} x^{-(z-1)} \underline{b}_z(1/x^Z)\right]$$

soit, compte tenu de la propriété 1 :

$$\begin{aligned} \underline{C}_B(x) &= \sum_{z=1}^{Z} \underline{b}_z^*(x^Z)\, \underline{b}_z(1/x^Z) = \sum_{z=1}^{Z} \underline{c}_{zz}(x^Z) \\ &= \sum_{z=1}^{Z} \underline{c}_{zz}^{Z}(x) \end{aligned} \qquad (23)$$

qui montre que $\underline{C}_B(x)$ est la somme des polynômes de corrélation étendus des polynômes $b_z(x)$.

Pour obtenir une corrélation parfaite de la séquence complète, il suffit que la somme des polynômes de corrélation étendus des Z séquences soit égal à ZN soit :

$$\underline{C}_B(x) = \sum_{z=1}^{Z} \underline{c}_{zz}(x^Z) = NZ \qquad (24)$$

III.3. Propriété 3 : séquence analogique.

On appelle séquence analogique la séquence associée au polynôme :

$$A(x) = \sum_{z=1}^{Z} b_z(x) \qquad (25)$$

Le polynôme de corrélation s'écrit :

$$C_A(x) = \left[\sum_{z=1}^{Z} b_z^*(x)\right]\left[\sum_{z=1}^{Z} b_z(1/x)\right]$$

soit, compte tenu de la propriété 1 :

$$C_A(x) = \sum_{z=1}^{Z} c_{zz}(x)$$

Le polynôme de corrélation de la séquence analogique est égal à la somme des polynômes de corrélation des séquences.

Le polynôme de corrélation du polynôme étendu est le polynôme étendu du polynôme de corrélation de la séquence analogique, soit de la caractéristique 2 :

$$C_A(x) = \sum_{z=1}^{Z} c_{zz}(x) = NZ \qquad (26)$$

La corrélation de la séquence analogique est donc parfaite.

III.4. Propriété 4 : autocorrélation des séquences.

La corrélation des séquences ne peut être parfaite. En effet, la corrélation parfaite impose pour les polynômes de corrélation :

$$c_{zz}(x) = \sum_{h=0}^{N-1} \left|b_z(w_p^{-h})\right|^2 x^h = N$$

ce qui entraîne:

$$\left|b_z(w_p^{-h})\right|^2 = 0 \qquad \forall h \neq 0$$

$$\left|b_z(w_p^{0})\right|^2 = N$$

Ce résultat est incompatible avec la propriété 1 qui impose, d'après (20) :

$$\left|b_{z1}(w_p^0)\right|^2 \left|b_{z1}(w_p^0)\right|^2 = 0$$

On ne peut donc obtenir des séquences à corrélation parfaite, ce qui a conduit à rechercher des séquences dont les polynômes de corrélation ont le maximum de coefficients nuls.

On a appelé presque parfaites les fonctions de corrélation correspondantes..

IV. - DETERMINATION DES SEQUENCES GQ -

IV.1. Séquence complète.

La séquence complète est à fonction de corrélation parfaite et son polynôme de corrélation défini par (24) s'écrit :

$$\underline{C}_B(x) = NZ\, U_{00}^{1\,NZ}(x) \qquad (27)$$

Son p transformé s'écrit donc :

$$\widetilde{\underline{C}}_B(x) = NZ\, \tilde{U}_{00}^{1\,NZ} \qquad (28)$$

Il s'agit d'un polynôme uniforme non étendu dont on sait d'après (23) que :

$$\underset{\sim}{\tilde{C}}_B(x) = \sum_{z=1}^{Z} \underset{\sim}{\tilde{c}}_{zz}^{Z}(x) \tag{29}$$

Un monôme de $\underset{\sim}{\tilde{C}}_B(x)$ ne peut appartenir simultanément à deux polynômes $\underset{\sim}{\tilde{c}}_{z_1 z_1}^{Z}(x)$ et $\underset{\sim}{\tilde{c}}_{z_2 z_2}^{Z}(x)$, $z_1 \neq z_2$, puisque la condition (21) ne serait pas respectée. Les coefficients non nuls de $\underset{\sim}{\tilde{c}}_{zz}^{Z}(x)$ sont donc égaux à NZ et la condition d'auto-corrélation presque parfaite impose que $\underset{\sim}{\tilde{c}}_{zz}^{Z}(x)$ soit uniforme. Il est simple de vérifier que, dans le cas contraire, les coefficients de $\underset{\sim}{\tilde{c}}_{zz}^{Z}(x)$ sont nuls, excepté pour quelques valeurs particulières de h.

$$\underset{\sim}{\tilde{c}}_{zz}^{Z}(x) = N\, \tilde{U}_{0\ z}^{Z\ N} \tag{30}$$

Il s'en déduit immédiatement :

$$\underset{\sim}{c}_{zz}^{Z}(x) = N\, \tilde{U}_{0\ z}^{Z\ N} \tag{31}$$

dont seuls Z coefficients sont non nuls.
Il s'agit d'un polynôme N étendu qui a été Z étendu par la construction de la séquence complète.

Le polynôme non étendu s'écrit donc :

$$c_{zz}(x) = N\, U_{0z}^{ZY} = N \sum_{i=0}^{Z-1} W_{ZY}^{zY_2 i} x^{Yi} \tag{32}$$

où $Y = \dfrac{N}{Z}$ doit être un entier, ce qui impose $ZY = N$

(33)

La corrélation parfaite de la séquence analogique, qui est équivalente à la corrélation parfaite de la séquence complète entraîne, d'après (26) et (32) :

$$\sum_{z=1}^{Z} W_{ZY}^{zYi} = \sum_{z=1}^{Z} W_p^{zYi} \equiv 0 \quad \forall\, i \neq 0 \tag{34}$$

Cette condition est toujours vérifiée.

Enfin, en comparant (18) et (30), il apparaît que NZ est un carré parfait et, compte tenu de (33), Y est également un carré parfait ce qui entraîne :

$$Y = Y_0^2$$
$$ZN = (ZY_0)^2 = N_0^2 \tag{35}$$

La longueur de la séquence complète est un carré parfait.

Enfin $$N = ZY_0^2 = Y_0 N_0 \tag{36}$$

IV.2. Séquences GQ.

Les séquences doivent vérifier la condition (32) qui peut être satisfaite par différents types de séquences. Les séquences recherchées sont à coefficients à norme constante. Le p transformé polynôme de $c_{zz}(x)$ qui s'exprime sous la forme :

$$\tilde{c}_{zz}(x) = N\, \tilde{U}_{0z}^{ZY} \tag{37}$$

peut s'écrire

$$\tilde{c}_{zz}(x) = \sum_{k=0}^{Y_0-1} N_0\, \tilde{U}_{0\ z+kY}^{N_0 Z} = \sum_{k=0}^{Y_0-1} \underset{\sim}{\tilde{c}}_{kz}(x) \tag{38}$$

$\underset{\sim}{\tilde{c}}_{kz}(x) = N_0\, \tilde{U}_{0\ z+kY_0}^{N_0 Y_0}$ est le p transformé du polynôme de corrélation :

$$\underset{\sim}{c}_{kz}(x) = N_0\, U_{0\ z+kY}^{N_0 Z} \tag{39}$$

qui est lui-même le polynôme Y_0 étendu de

$$c_{kz}(x) = N_0\, U_{0\ z+kY}^{N_0 1} \tag{40}$$

dont le p transformé polynôme est :

$$\tilde{c}_{kz}(x) = N_0\, \tilde{U}_{0\ z+kY}^{N_0 1} \tag{41}$$

D'après (9) les séquences, que nous appellerons sous-séquences, satisfaisant ce polynôme ont des p transformés de la forme :

$$\tilde{s}_{kz}(x) = W_{kz}\, \tilde{U}_{0\ z+kY}^{N_0 1} \tag{42}$$

où W_{kz} est un complexe arbitraire de norme constante, et la séquence associée s'écrit :

$$s_{kz}(x) = W_{kz}\, \tilde{U}_{0\ z+kY}^{N_0 1} \tag{43}$$

Il est élémentaire de vérifier que

$$\tilde{s}^*_{k_1 z(x)} \otimes \tilde{s}^*_{k_2 z(x)} \equiv 0 \quad \forall\, k_1 k_2 \in \{0, \ldots Y_0 - 1\}$$

ce qui permet d'entrelacer ces séquences et, compte tenu de (38), (39), (40) et (43)

$$b_z(x) = \sum_{k=0}^{Y_0-1} W_{kz}\, U_{k\ z+kY}^{N_0 Y_0} \tag{44}$$

qui définissent les séquences GQ.

IV.3. Structure des séquences GQ.

La structure des séquences GQ est définie par (44) et on définit les séquences de base par $W_{kz} = 1$.

Il est simple de montrer que $b_z(x)$ s'écrit sous la forme :

$$b_z(x) = \sum_{i=0}^{Y_0 N_0 - 1} W_{N_0}^{(z+Y_0 I_F) I_P} x^i$$

où $W_{N_0} = \exp \frac{j2\pi p}{N_0}$

$I_P = i[Y_0]$

$I_F = i - I_P$

qui donne une définition très simple des séquences GQ.

IV.4. Propriétés des séquences GQ.

Propriété 1 : Dimension et dénombrement des séquences GQ de base.

On définit les séquences GQ de base comme celles pour lesquelles $W_{kz} = 1$ dans (44) et pour lesquelles p = 1.
Les séquences de base GQ sont de longueur $N = Y_0 N_0$.
La séquence analogique est de longueur N
La séquence complète de longueur N_0^2 est constituée par l'entrelacement de Z séquences telles que :

. N_0 est un entier arbitraire

. $N_0 = ZY_0$

Le nombre de séquences de base GQ est donné dans le tableau 1 pour les valeurs de N_0 inférieures ou égales à 100.

Propriété 2 : Corrélation des séquences de base GQ.

Les séquences de base GQ $b_{0z}(x)$ sont définies à partir de (44) en posant $W_{kz} = 1$:

$$b_{0z} = \sum_{k=0}^{Y_0 - 1} U_{k\ z+bY}^{N_0 Y_0}$$

Le p transformé du polynôme de corrélation donné par la relation (32) montre que :

- La fonction de corrélation est nulle en-dehors de Z valeurs non nulles équidistantes de Y_0^2 symboles et dont l'amplitude prend la valeur N. Ces fonctions de corrélation sont presque parfaites.

- Les fonctions d'intercorrélation des séquences de base sont strictement grâce à la condition (30).

La séquence complète, définie par (27), est à corrélation parfaite, sa valeur, pour un décalage nul, étant égal à NZ.

La séquence analogique de longueur N, définie par (26), est à corrélation parfaite, sa valeur, pour un décalage nul étant égal à NZ. Le module de la séquence analogique n'est pas constant, et il est facile de montrer qu'il est l'échantillonné de la loi $\sin(npZ\alpha / N) / \sin(\pi p \alpha / N)$ où α = Entier (x/Y_0).

La figure 1 montre les résultats obtenus pour les séquences GQ obtenues avec $N_0 = 30$, $Z = 3$, $Y_0 = 10$, $p = 1$.
On peut noter que :

- La séquence complète de longueur 900 est à corrélation parfaite, la valeur maximum étant égale à $N_0^2 = 900$ et d'amplitude constante égale 1.

- La séquence analogique est à corrélation parfaite de longueur $N = 300$, la valeur maximum étant égale à $N_0^2 = 900$.

- Les séquences de base sont à corrélation presque parfaite et comportent $Z = 3$ valeurs non nulles équidistantes d'amplitude $N_0^2 = 900$.

- L'amplitude de la séquence analogique suit la loi de variation indiquée..

Propriété 3 : Séquences GQ primitives.

Tous les résultats obtenus pour les séquences GQ ont été établis par les p transformés et sont donc valables pour tout p [N].
La relation (11) indique que le p transformé d'une séquence q permutée est le q transformé polynôme d'une séquence p permutée.

Les résultats obtenus pour les séquences de base demeurent donc valables pour toutes les séquences p permutées, p étant premier avec N.

La figure 2 représente les résultats obtenus avec les mêmes valeurs de $N_0 = 30$, $Z = 3$, $Y_0 = 10$, que pour la figure 1, mais avec des séquences de base p = 7 permutées. On peut noter que les résultats ne sont pas modifiés, à l'exception de l'amplitude de la séquence analogique dans laquelle une opération d'addition classique modifie les résultats.

Cette propriété montre que le nombre de séquences GQ existant pour chaque triplet (N_0, Z, Y_0) est toujours élevé. Le nombre p étant premier avec N, le nombre d'ensembles de séquences GQ que l'on peut construire est donné par :

$$\phi_{(N)} = N\left(1 - \frac{1}{a}\right)\left(1 - \frac{1}{b}\right)\ldots\left(1 - \frac{1}{e}\right)$$

où a, b, ... e premiers tels que

$$N = a^{\alpha}\, b^{\beta} \ldots e^{\varepsilon}$$

Le tableau 1 donne le nombre de séquences GQ primitives que l'on peut construire à partir du nombre minimum de séquences que l'on peut utiliser pour chaque valeur de N_0. Le nombre de séquences primitives que l'on peut construire devient très vite important dès que N_0 croît, ce qui montre la richesse de ces séquences.

Il est également possible de permuter les éléments de la séquence complète sans modifier ses propriétés.
Le p transformé polynôme $\underline{\widetilde{C}}_B(x)$ étant un polynôme uniforme (28) où $P = 1$:

$$\underline{\widetilde{C}}_B(x) = NZ \quad U_{00}^{1NZ}$$

toute p permutation le laisse invariant. Le polynôme de corrélation est donc également invariant.

Propriétés 4 : Séquences GQ translatées.

Si $b(x)$ est un polynôme, son polynôme translaté est $b^t(x) = x^t b(x)$. Il est simple de voir que, si on translate un polynôme, son polynôme de corrélation est inchangé. En effet, (7) s'écrit alors :

$$c(x) = b^{t*}(x)\, b^t(1/x) = b^*(x)\, b(1/x)$$

Si on translate une séquence $s_{kz}(x)$ (43) de tk son polynôme de corrélation (41) est inchangé et par suite de (39) le polynôme de corrélation de la séquence $c_{zz}(x)$ demeure invariant :

- Les séquences GQ demeurent strictement orthogonales.
- Les autocorrélations des séquences GQ demeurent invariantes.
  Il s'ensuit que les séquences complète et analogique demeurent invariantes.

Les propriétés des séquences GQ sont conservées pour toute translation arbitraire de chaque sous-séquence $s_{kz}(x)$.

La figure 3 montre un exemple de séquences GQ translatées pour les paramètres pris dans le cas des figure 1 et 3 : $N_0 = 30$, $Z = 3$, $Y_0 = 10$, les séquences étant non permutées, p = 1, et les translations sur les 3 séquences étant respectivement $t_1 = 11$, $t_2 = 251$, $t_3 = 37$.

On note que les résultats ne sont pas modifiés, à l'exception de l'amplitude de la séquence analogique dans laquelle une opération d'addition classique modifie les résultats.

Propriétés 5 : Séquences GQ décalées.

Le polynôme décalé de $b_z(x)$ est $b_z^r(x) = W_{N_0}^r b_z(x)$.
Les relations (48) et (41) montrent que W_{kz} ne modifie pas $\widetilde{c}_{kz}(x)$ donc $c_{zz}(x)$.

On peut donc donner à W_{kz} une valeur arbitraire sans que l'on modifie les propriétés des séquences GQ.
Il est pratique de choisir :

$W_{kz} = W_{N_0}^{r_k}$ où r_k est un entier arbitraire.

Tout décalage arbitraire des sous-séquences laisse invariantes les propriétés des séquences GQ.

La figure 4 montre les résultats obtenus pour des décalages arbitraires de chaque sous-séquence pour des séquences GQ définies par les mêmes paramètres que pour les figures 1, 2 et 3 : $N_0 = 30$, $Z = 3$, $Y_0 = 10$. On note, comme pour les autres cas, une invariance des résultats, à l'exception de la séquence analogique, pour les raisons citées précédemment.

Propriété 6 : Séquences GQ primitives, translatées, décalées.

Il est simple de vérifier que les propriétés des séquences GQ sont invariantes si on applique simultanément les transformations décrites précédemment :

- Translations primitives.
- Translations.
- Décalages.

La figure 5 montre les résultats obtenus pour les séquences GQ définies par $N_0 = 30$, $Z = 3$, $Y_0 = 10$ et qui sont simultanément :

- Primitives : $p = 7$.
- Translatées : $t_1 = 11$, $t_2 = 251$, $t_3 = 37$.
- Décalées avec les mêmes valeurs que dans le cas de la figure 4.

Les résultats, à l'exception de l'amplitude de la séquence analogique, demeurent invariants.
On peut noter que, dans toutes ces transformations, les amplitudes de la séquence complète et des séquences GQ demeurent constantes, égales à l'unité.
Le tableau 2 montre, pour un cas simple, $N_0 = 6$, $Z = 2$, $Y_0 = 2$, différentes suites obtenues par les transformations précédemment décrites.

V. EXTENSION DES SEQUENCES GQ -

Il est possible, sous réserve d'admettre une modification des propriétés des séquences, de modifier la longueur des séquences GQ pour une valeur donnée de N_0.
La relation (22) donne la constitution des séquences GQ complètes à partir des séquences GQ.

Si on forme le polynôme :

$$\underline{\underline{B}}(x) = \sum_{z=1}^{Z} x^{E(z-1)} \sum_{e=0}^{E-1} x^e b_z(x^z) W_E^e$$

où

- $\underline{\underline{B}}(x)$ indique que le polynôme est défini sur

$$A(x) / (x^{ZNE} - 1)$$

- $$W_E = \exp\left(\frac{j2\pi pe}{E}\right)$$

Il est simple de vérifier que :

- La fonction de corrélation de la séquence complète est à pointe centrale unique et que la fonction de corrélation est nulle en dehors de cette pointe centrale.
- La séquence analogique est nulle.

Il est particulièrement intéressant de choisir $\frac{N_0}{E}$ entier car, dans ces conditions :

$$W_E = W_{N_0}^{\frac{N_0}{E}}$$

ce qui n'introduit pas de nouvelles valeurs de symboles dans la séquence.

La figure 6 montre un exemple d'extension de séquence GQ pour $N_0 = 30$, $Z = 3$, $Y_0 = 10$, $p = 7$, $t_1 = 11$, $t_2 = 251$, $t_3 = 37$ et pour des décalages identiques à chaque sous-séquences de chacune des 3 séquences respectivement égales à 0, 23, 16 et pour une extension $E = 15$. On obtient donc ainsi une séquence complète d'une longueur de 13500. On peut observer sur ces résultats les propriétés énoncées ci-dessous :

- Séquence complète d'amplitude unitaire.
- Corrélation de la séquence complète à pointe unique.
- Séquence analogique nulle.

VI. - APPLICATIONS DES SEQUENCES GQ -

Les séquences GQ sont utilisables dans les systèmes nécessitant l'utilisation de séquences longues traitées par corrélation.

L'émission de la séquence complète, de longueur NZ correspond au critère, imposé dans la majorité des cas, d'utiliser une séquence à amplitude constante.

A la réception, la périodicité de la séquence complète permet de former sur le signal reçu la séquence analogique.

La complexité du calcul de la fonction de corrélation par la méthode directe, nécessite N^2Z^2 multiplications pour la séquence complète alors qu'elle ne nécessite que N^2 multiplications pour la séquence analogique. Le gain sur la complexité, égal à Z^2, peut être très important par un choix convenable de Z lié aux objectifs de détection. Dans les exemples présentés dans les figures 1 à 5, le gain est de 100.

VII. - CONCLUSION -

Les séquences GQ, constituées par des ensembles de Z séquences orthogonales de longueur N et de fonction de corrélation presque parfaite, permettent la construction de deux séquences ayant des fonctions de corrélation parfaites.

La séquence GQ complète de longueur $NZ = N_0^2$ est d'amplitude constante.

La séquence GQ analogique de longueur N est d'amplitude variable.

Ces séquences existent pour toutes les valeurs de N_0 et pour tout couple NZ tel que $NZ = N_0^2$.

Il existe, pour chaque couple NZ, un très grand nombre de séquences obtenues par des transformations primitives, de translation ou de décalage. Outre les propriétés de corrélation parfaite et d'orthogonalité des séquences, l'un des intérêts de ces séquences vient de la réduction de la complexité de calcul de la fonction de corrélation par la reconstitution de la séquence analogique à partir de la séquence complète.

Ces séquences trouvent des applications dans les systèmes de télédétection et de télécommunication.

BIBLIOGRAPHIE

[1] R. SIVASWAMY - *Multiphase complementary Codes*. IEEE Trans. Infor. Theory, Vol. IT-24, n° 5, Sept. 1978.

[2] C. GOUTELARD - *Séquences d'étalement à synchronisation optimale par commutations d'opérateurs incommutables*. Conf. AGARD-EPP, n° 442, Paris, Oct. 1988.

[3] R. HOHOLDT, J. JUSTEGEN - *Ternary sequences with perfect periodic autocorrelation*. IEEE Trans. Infor. Theory, Vol. IT-29, pp 597-599, 1983.

[4] C. GOUTELARD - *The W.G. Sequences. Quasi perfect binary sequences*. Conf. AGARD-EPP, n° 528, Londres, Juin 1992.

[5] C. GOUTELARD, F. CHAVAND - *Codes pseudo-orthogonaux adaptés aux transmissions dans les canaux à caractéristiques aléatoires*. Annales des Télécomm., n° 11-12, Tome 30, Sept. 1975.

TABLEAU 1

N_0	$N^2_0=ZN$	Nbre de séquences de base	Nbre de séquences primitives pour $Z\min\neq 1$		
			$Z\min$	N	$\Phi_{(N)}$
2	4	2	2	2	1
3	9	2	3	3	1
4	16	3	2	8	4
5	25	2	5	5	1
6	36	4	2	18	6
7	49	2	7	7	1
8	64	4	2	32	16
9	81	3	3	27	18
10	100	4	2	50	20
12	144	6	2	72	24
14	196	4	2	98	42
16	256	5	2	128	64
18	324	5	2	162	72
20	400	5	2	200	80
25	625	3	5	125	100
30	900	8	2	450	120
35	1225	5	5	245	168
40	1600	8	2	800	320
45	2025	6	3	675	360
50	2500	6	2	1250	500
55	3025	4	5	605	440
60	3600	13	2	1800	480
65	4225	4	5	845	624
70	4900	8	2	2450	980
75	5625	6	3	1875	1000
80	6400	10	2	3200	1280
85	7225	4	5	1445	1088
90	8100	12	2	4050	1080
95	9025	4	5	1805	1368
100	10000	9	2	5000	2000

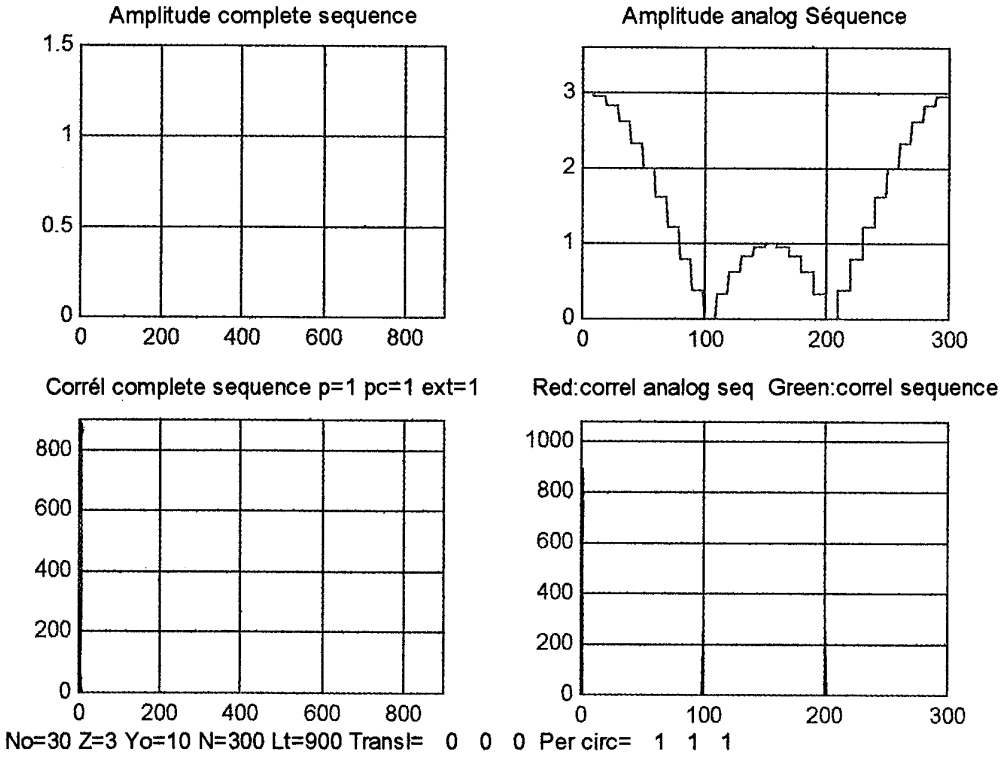


FIGURE 1: SEQUENCES DE BASE GQ
No=30 Z=3 Yo=10 p=1

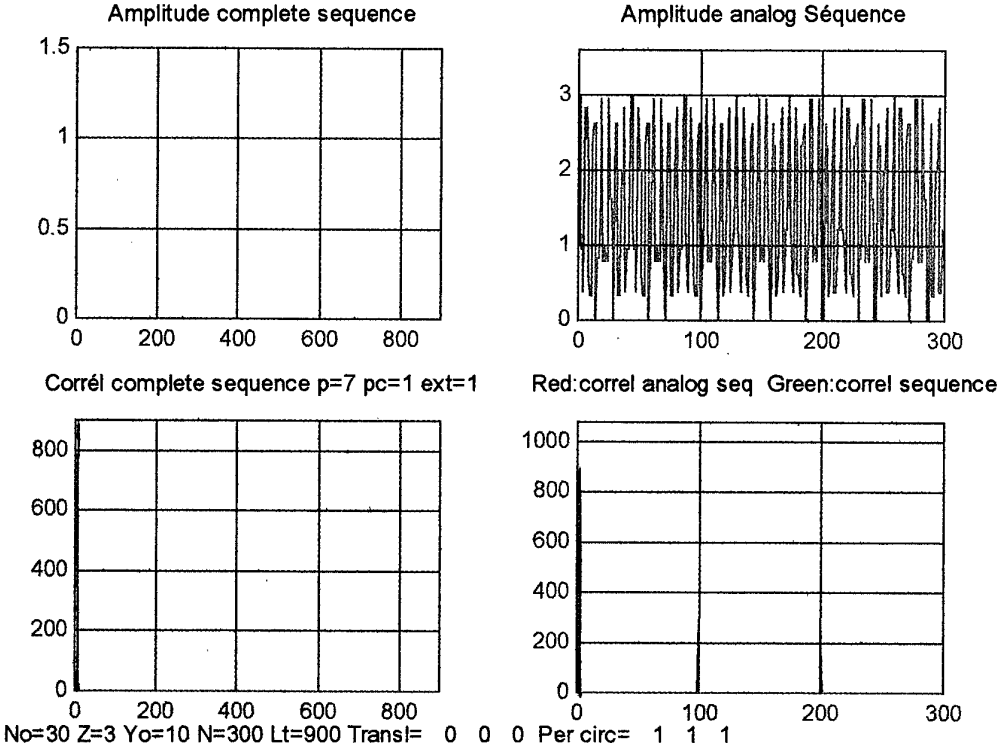


FIGURE 2 : SEQUENCES GQ PRIMITIVES
No=30 Z=3 Yo=10 p=7

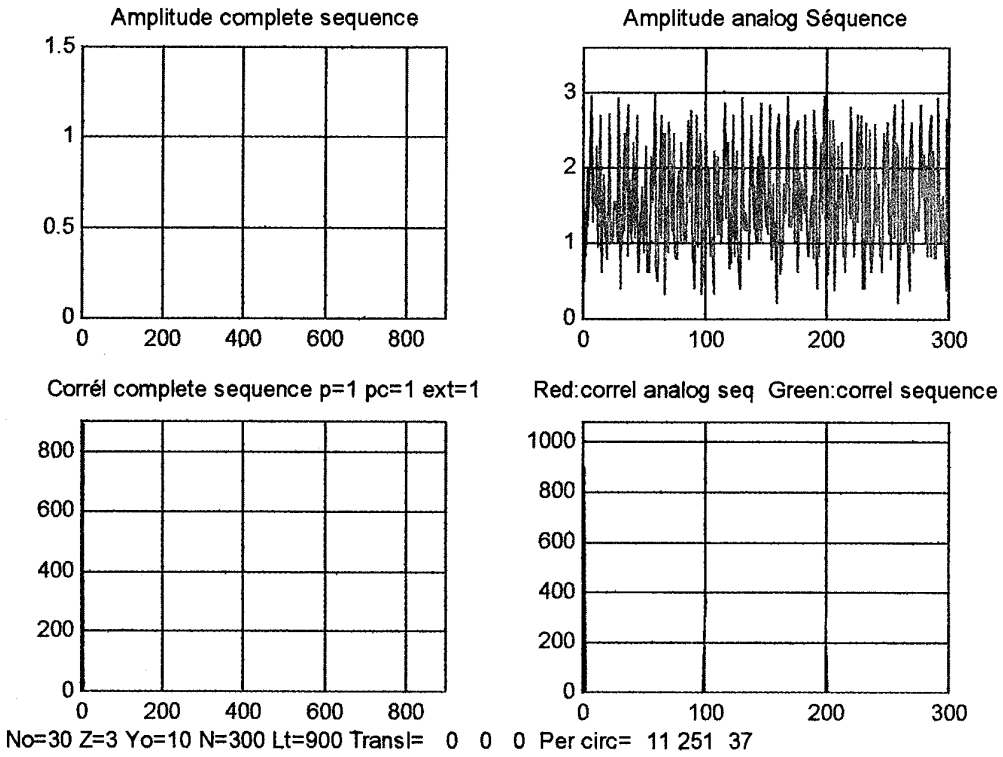


FIGURE 3 : SEQUENCES GC TRANSLATEES

No=30 Z=3 Yo=10 p=1 Translations des séquences : t1=11,t2=251,t3=37

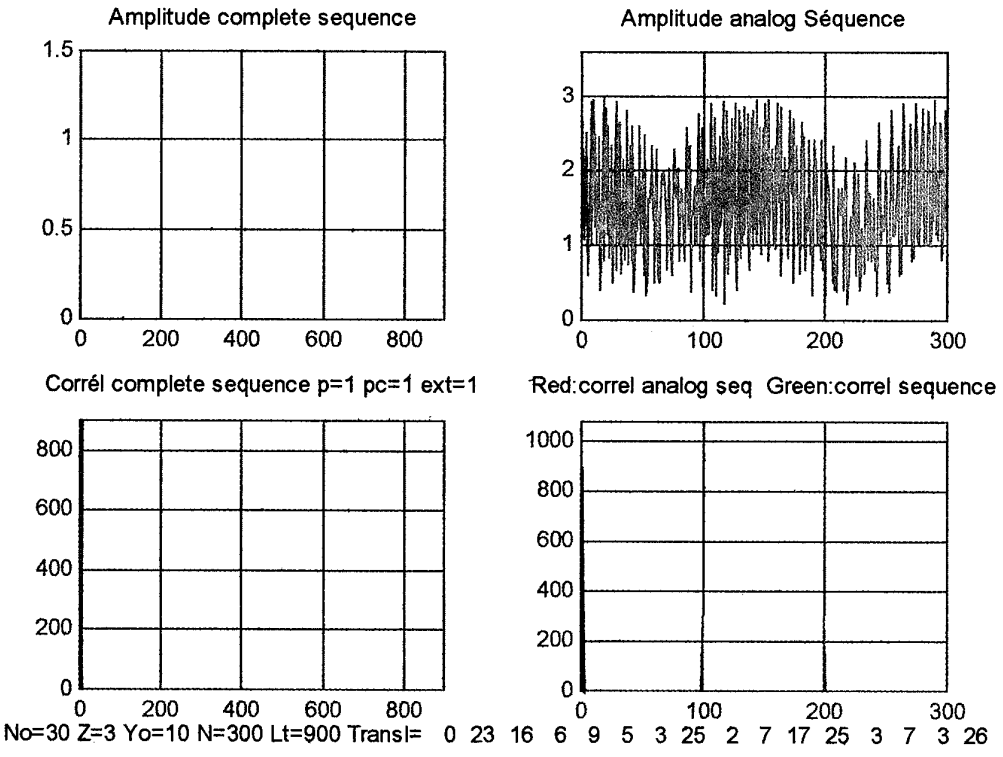


FIGURE 4 :SEQUENCES GQ DECALEES

No=30 Z=3 Yo=10 p=1 Decalages des sous séquences : 0,23,16,6,5,3,25,3,7,3,26,21,2,17,...

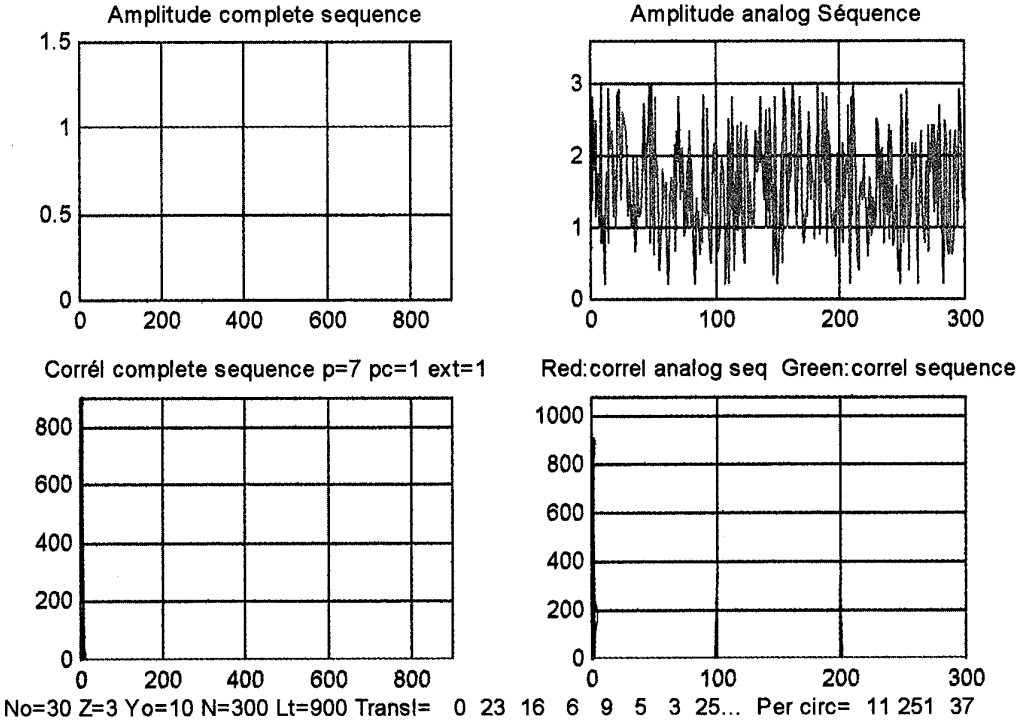


FIGURE 5 : SEQUENCES GQ
PRIMITIVES,TRANSLATEES,DECALEES

No=30 Z=3 Yo=10 p=7 Décalages des sous séquences :
0,23,16,6,9,3,25,3,7,3,26,21,2,17..
Translation des séquences:t1=11,t2=251,t3=37

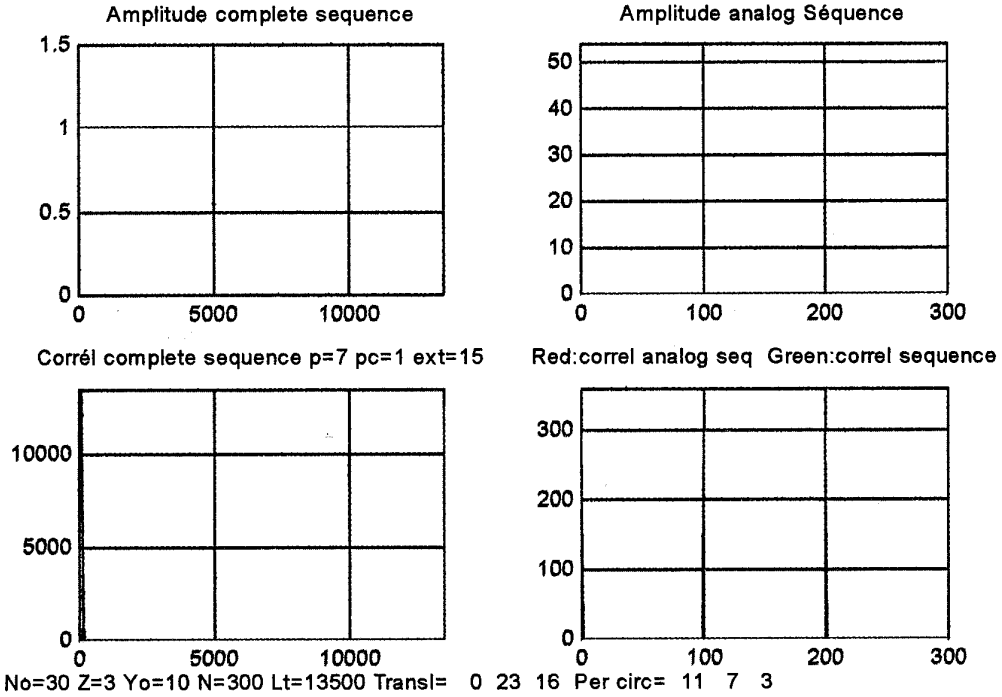


FIGURE 6 :EXTENSION DE SEQUENCE GQ
Extension par un facteur E=15 d'une séquence primitive,étendue,décalée

SEQUENCE COMPLETE DE BASE

0	0	0	0	0	0	0	1	2	3	4	5
0	0	4	0	2	4	0	3	0	3	0	3
0	4	2	0	4	2	0	5	4	3	2	1

SEQUENCE COMPLETE DECALEE
Decalages 1 5 3 7 2 4

1	3	2	5	1	4	1	4	4	2	5	3
1	5	0	5	3	2	1	0	2	2	1	1
1	1	4	5	5	0	1	2	0	2	3	5

SEQUENCE COMPLETE PRIMITIVE DECALEE
p=7 Decalages 1 5 3 7 2 4

1	3	2	2	1	1	1	4	4	5	5	0
1	5	0	2	3	5	1	0	2	5	1	4
1	1	4	2	5	3	1	2	0	5	3	2

SEQUENCE COMPLETE PRIMITIVE DECALEE TRANSLATEE
p=7 Decalages 1 5 3 7 2 4 t1=11 t2=7 t3=3

5	1	0	1	1	0	2	5	5	1	2	2
5	3	4	1	3	4	2	1	3	1	4	0
5	5	2	1	5	2	2	3	1	1	0	4

SEQUENCE COMPLETE PRIMITIVE DECALEE TRANSLATEE EXPANSEE
p=7 Decalages 1 5 3 7 2 4 t1=11 t2=7 t3=3 E=3

5	1	3	1	3	5	0	2	4	1	3	5
1	3	5	0	2	4	2	4	0	5	1	3
5	1	3	1	3	5	2	4	0	2	4	0
5	1	3	3	5	1	4	0	2	1	3	5
3	5	1	4	0	2	2	4	0	1	3	5
3	5	1	1	3	5	4	0	2	0	2	4
5	1	3	5	1	3	2	4	0	1	3	5
5	1	3	2	4	0	2	4	0	3	5	1
1	3	5	1	3	5	0	2	4	4	0	2

TABLEAU 2 : EXEMPLES DE SEQUENCES GQ PRIMITIVES , TRANSLATEES , DECALEES,EXPANSEES
No = 6 , Z = 3 , Yo = 2 , N = 12 , No^2 = 36

DISCRETE HILBERT TRANSFORM

by

A. Marguinaud and S. Gienger
ALCATEL ESPACE
5, rue Noël Pons
92737 Nanterre
France

1. SUMMARY

Analytic signal processing has gained wide acceptance in the electronics community and it is important to define the transform family which allows regular analytic samples of a signal to be calculated from its real sampling. This paper is composed of three main parts.

The first part shows that band limited continuous signals are well approximated by trigonometric polynomials whose frequencies lie within a frequency band of the same width, and that we may thus represent these signals by analytic samples.

The second part is devoted to the derivation of an explicit formula for the weighting coefficients used to calculate the two cartesian coordinates of the complex samples.
The third part reviews some properties of this transform from the signal processing standpoint.

2. INTRODUCTION

The analytical signal representation has been devised by VILLE [1] in 1950 in order to give a sound foundation to the notions of phase and frequency for a signal to be processed.

For practical purposes radio frequency signals are translated to an intermediate frequency band by mixing them with the output of a local oscillator. For instance $\cos(\omega t+\varphi)$ mixed with $\cos(\omega_o t)$ gives two signals which can be separated by adequate bandpass filtering.

$$2\cos\omega_o t\cos(\omega t+\varphi)=\cos\left[(\omega_o-\omega)t-\varphi\right]+\cos\left[(\omega_o+\omega)t+\varphi\right]$$

We must however impose $\omega_o>\omega$ to keep the proper phase rotation, for otherwise $\omega_o-\omega$ is negative and the signal may be interpreted as $\cos[(\omega-\omega_o)t-\varphi]$. To avoid the above restriction and also translate to base band, it is usual to mixe the RF signal with $\sin\omega_o t$ and combine this result with the $\cos\omega_o t$ translation.

$$2\sin\omega_o t\cos(\omega t+\varphi)=\sin\left[(\omega_o-\omega)t-\varphi\right]+\sin\left[(\omega_o+\omega)t+\varphi\right]$$

The lowest frequency term of this last equation appears as the **quadrature** of the corresponding term of the first $\cos\omega_o t$ translation, so that $\cos(\omega t+\varphi)$ can be represented in the complex plane by $\exp j(\omega t+\varphi)$. In this last form, there is no ambiguity on the sign of ω since negative frequencies are now well defined.

The above method is commonly used to derive the two components (real and imaginary) of the analytic representation of real signals. The corresponding implementation shown in fig 1 is justified by the fact that locally, any real signal can be uniformly approximated by a trigonometric polynomial as will be shown in the following chapter.

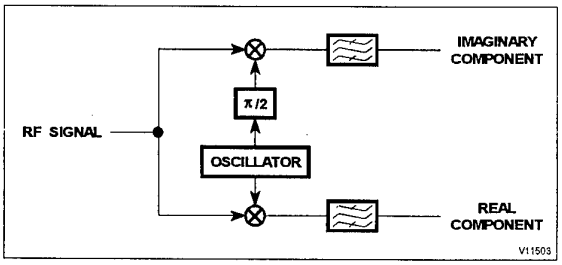


Figure 1 - Analog signals in quadrature

For subsequent digital processing, the components of the analytic representation of a signal must be sampled and digitized. The usual way to go from analog to digital consists of keeping the same functional architecture and digitizing successively each elementary identified module.

This procedure seems technically secure.

However it is far from producing an optimum design as one can see by comparing it with a digital analytic sampler designed for the purpose.

In this case, the signal to be digitized at an intermediate frequency is digitized by real sampling, and the complex components are then calculated by processing with an appropriate filter function.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

3. HEURISTIC DESIGN OF AN ANALYTIC DIGITIZER

3.1. Definition of the signal in quadrature

Every textbook on signal processing claims that the quadrature signal associated to r(t) is obtained by convolving it with $\frac{1}{t\sqrt{\frac{2}{\Pi}}}$ provided that $\int_{-\infty}^{+\infty} r(t)dt = 0$.

(1) $$r(t) \rightarrow q(t) \overset{\Delta}{=} \frac{1}{\Pi}\int_{-\infty}^{+\infty} \frac{r(\tau)}{t-\tau}d\tau$$

When r(t) is a sequence of real samples with frequency F_e, one should expect to get a good result by using the sampled Hilbert kernel $\frac{1}{t\sqrt{\frac{2}{\Pi}}}$. Evaluation of this method with sinewaves as shown on figure 2 gives a result which is far from satisfactory for a reasonable complexity (i.e. number of coefficients) even with the interpolation method.

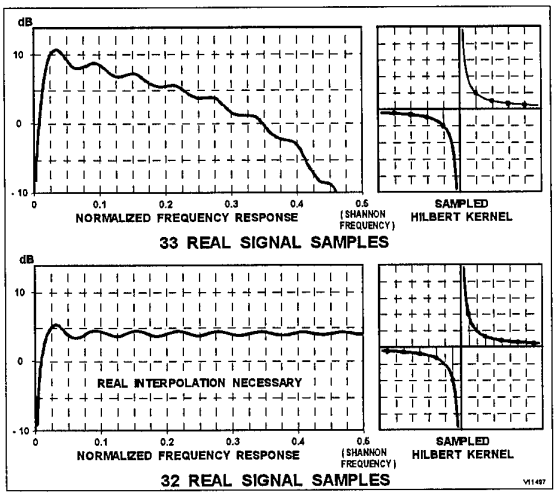


Figure 2 - Quadrature from sampled Hilbert kernel

We are therefore obliged to look for a better idea. We consider the fact that the quadrature of $\cos(\omega t+\varphi)$ is $\sin(\omega t+\varphi)$ and that any periodic continuous function r(t) can be approximated **in the mean** by its Fourier Series.

(2)

$$r(t)\text{ continuous for } O \leq t \leq T \Rightarrow r(t) = \frac{a_o}{2} + \sum_{n\geq 1}\left[a_n \cos 2\Pi n\frac{t}{T} + b_a \sin 2\Pi n\frac{t}{T}\right]$$

with $r(0) = r(T)$

$$\begin{cases} a_n = \frac{2}{T}\int_o^T r(\tau)\cos 2\Pi n\frac{\tau}{T}d\tau \\ b_n = \frac{2}{T}\int_o^T r(\tau)\sin 2\Pi n\frac{\tau}{T}d\tau \end{cases}$$

By straightforward trigonometric manipulations, it is possible to transform $a_n \cos 2\Pi n\frac{t}{T} + b_n \sin 2\Pi n\frac{t}{T}$ into $c_n \cos\left[2\Pi n\frac{t}{T} + \varphi_n\right]$, so that r(t) takes a more convenient form for defining the signal q(t) in quadrature with r(t).

(3)
$$\begin{cases} r(t) = \sum_{n\geq 1} c_n \cos\left(2\Pi n\frac{t}{T} + \varphi_n\right) \\ q(t) \overset{\Delta}{=} \sum_{n\geq 1} c_n \sin\left(2\Pi n\frac{t}{T} + \varphi_n\right) \\ z(t) \overset{\Delta}{=} r(t) + jq(t) \end{cases}$$

In this last expression, we have taken $a_o=0$ which is justified in the case of radio frequency signals which have no zero frequency component.

Although the definition of the analytic signal z(t) associated with the real signal r(t) as given by (3) is formally correct, its derivation is questionable. The main objection concerns the type of approximation of the real signal r(t). In fact a theorem due to WEIRSTRASS (1885) establishes the existence of **uniform** (over a limited time interval) approximation of continuous signals by trigonometric polynomials.

To make things more precise, we sketch the proof as given by Courant and Hilbert [2].

For the interval [0,1] the polynomial $[1-(u-\xi)^2]^n$ in u takes its maximum value 1 when $u=\xi$, and decreases rapidly toward zero outside.

The function $f(\xi)$ continuous in the interval $0 < a \leq \xi \leq b < 1$ is continuously extended to the interval $\alpha \leq \xi \leq \beta$ with $0 < \alpha \leq a \leq b \leq \beta < 0$.

For $a \leq \xi \leq b$ it can be prouved that the polynominals of degree 2n in ξ

$$P_n(\xi) \overset{\Delta}{=} \frac{\int_\alpha^\beta f(u)\left[1-(u-\xi)^2\right]^n}{\int_{-1}^{1}(1-u^2)^n du}$$

approaches **uniformally** the (uniformally) continuous function f(ξ). Uniform convergence with n signifies that given any $\varepsilon > 0$, one can find an integer N independent of ξ such that $n > N \Rightarrow \left|P_n(\xi) - f(\xi)\right| < \varepsilon$.

This uniform approximation is then extended to two dimensions.

$P_n(\xi, n)$ given by

$$\frac{\int_{\alpha_1}^{\beta_1}\int_{\alpha_2}^{\beta_2} f(\xi,n)\left[1-(u-\xi)^2\right]^n\left[1-(v-\eta)^2\right]^n du dv}{\left[\int_{-1}^{1}(1-u^2)^n du\right]^2}$$

$f(\xi,\eta)$ is continuous in the square defined by $0 < a_1 \le \xi \le b_1 < 1$ and $0 < a_2 \le \eta \le b_2 < 1$.

By linear change of coordinates, we see that any function continuous over a finite interval of the plane can be approximated uniformally by polynomials in two variables.

We are now in a position to derive the uniform approximation by trigonometric polynomials of the continuous function r(t). The periodic condition r(0)=r(T) can be considered as being always satisfied by modifiying r(t) outside the useful interval and changing the coordinate scaling.

The function $\varphi(\xi,\eta)=\rho r(t)$ where $\xi = \rho\cos 2\Pi\frac{t}{T}$ and $\eta = \rho\sin 2\Pi\frac{t}{T}$ is continuous and coincides with r(t) on the unit circle $\xi^2+\eta^2=1$. According to previous results, $\varphi(\xi,\eta)$ can be uniformally approximated by polynomials in ξ and η. By setting ρ=1 in these polynomials, we get a uniform approximation by polynomials in $\cos 2\Pi\frac{t}{T}$ and $\sin 2\Pi\frac{t}{T}$ which can be written in the form (2) and then (3) as before.

3.2. Frequency interpretation of the Hilbert transform for band limited signals

Under reasonable conditions, the complex functions z(t) and Z(f) of the real variables t and f are a Fourier pair according to (4).

$$(4)\quad \begin{cases} Z(f) = \frac{1}{\sqrt{2\Pi}}\int_{-\infty}^{+\infty} z(t)\left[\exp -2\Pi jft\right]dt \\ z(t) = \frac{1}{\sqrt{2\Pi}}\int_{-\infty}^{+\infty} Z(f)\left[\exp 2\Pi jft\right]df \\ z(t) \supset Z(f) \text{ or } Z(f) \subset z(t) \end{cases}$$

With above notations, the convolution of the two functions f(x) and g(x) is defined (cf (3)) by the integral $(f*g)(t) \overset{\Delta}{=} \frac{1}{\sqrt{2\Pi}}\int_{-\infty}^{+\infty} f(t-\tau)g(\tau)d\tau$ which has the property that its Fourier Transform equals the product F(f)G(f) of their Fourier transforms in accordance with (4).

Expression (3) gives Fourier transform pairs, and (1) defines the Hilbert transform which can be interpreted as a convolution product.

(5)

$$\begin{cases} \frac{1}{t}\sqrt{\frac{2}{\Pi}} \supset -j\,\mathrm{sgn} f \text{ with} & \mathrm{sgn} f = 1 \text{ if } f > 0 \\ & \mathrm{sgn} 0 = 0 \\ r(t) \supset R(f) & \mathrm{sgn} f = -1 \text{ if } f < 0 \\ q(t) = \left[\frac{1}{\bullet}\sqrt{\frac{2}{\Pi}} * r(\bullet)\right](t) \supset -jR(f)\,\mathrm{sgn}(f) \end{cases}$$

From (3) and (5) one can evaluate Z(f) as a function of R(f).

$$(6)\quad \begin{cases} Z(f) = 2R(f) & \text{for } f > 0 \\ Z(f) = 0 & \text{for } f \le 0 \end{cases}$$

Remarks :

- It should be noted here that the frequency variable f in (6) is not related in a simple way to the discrete frequencies $\frac{n}{T}$ appearing in (2) and (3).
- The approximation r(t) as given by (3) is analytic by construction and thus justifies the expression **analytic representation** of real signals.

3.3. Shannon interpolating formula for band limited signals

The complex signal z(t) considered here has its Fourier transform Z(f) which equals zero outside the positive open interval $B-\frac{\Delta B}{2}, B+\frac{\Delta B}{2}$.

If both Cartesian coordinates of Z(f) are **of bounded variation** in the open interval defined above, the function $\bar{Z}(f)$ defined by the Fourier Serie of Z(f) converges pointwise to the average of the limits of the function Z(f) from the right and left (Dirichlet condition given in (4)) in the interval $B-\frac{\Delta B}{2},B+\frac{\Delta B}{2}$.

$$(7)\quad \begin{cases} \bar{Z}(f+B) = \sum_{n>0} z_n \exp -2\Pi jn\dfrac{f}{\Delta B} \\ \bar{Z}(g) \quad = \sum_{n>0} z_n \exp -2\Pi jn\dfrac{g-B}{\Delta B} \end{cases}$$

By the Fourier inversion formula (4) it is possible to evaluate z(t) from $Z(f) \subset z(t)$ up to a multiplicative constant C.

$$z(t) = C\int_{B-\frac{\Delta B}{2}}^{B+\frac{\Delta B}{2}} \bar{Z}(g)[\exp 2\Pi jgt]dg$$

$$= C\sum_{n>0} z_n \exp 2\Pi jn\frac{B}{\Delta B}\int_{B-\frac{\Delta B}{2}}^{B+\frac{\Delta B}{2}} \exp 2\Pi jg\left(t-\frac{n}{\Delta B}\right)dg$$

$$= C'\sum_{n>0} z_n[\exp 2\Pi jBt]\sin\frac{\Pi(t\Delta B-n)}{\Pi(t\Delta B-n)}$$

To remove constants C' and z_n, $z\left(\frac{n}{\Delta B}\right)$ is evaluated which gives the Shannon interpolating formula below.

$$(8)\quad z(t) = \sum_{n>0} z\left(\frac{n}{\Delta B}\right)\frac{\sin\Pi(t\Delta B-n)}{\Pi(t\Delta B-n)}\exp 2\Pi jB\left(t-\frac{n}{\Delta B}\right)$$

This relation shows that, provided the Dirichlet condition is met, a real signal r(t) of positive band-width ΔB centered at B is fully represented by its analytic samples of periodicity $\frac{1}{\Delta B}$ in time.

3.4. Band limited almost-periodic interpretation of signal (5)

The samples of $\cos(2\Pi ft+\varphi)$ for $t=\frac{n}{\Delta B}$ are invariant when f is changed into f+ΔB. Thus sampling at frequency ΔB creates an equivalence mod ΔB on the frequency axis, so that the component frequency of a WEIRSTRASS expension of a band limited signal can be considered as belonging to the same band $\left(B+k\Delta B-\frac{\Delta B}{2},B+k\Delta B+\frac{\Delta B}{2}\right)$ where k is any integer.

This result is valid even when the Dirichlet condition is not met because there is a finite number of discrete frequencies lines.

The presence of these critical lines is important because the Shannon interpolation formula diverges for a single sinewave, whatever its frequency. However for normal processing this interpolation formula is not necessary, and only the existence of a uniform approximation by a trigonometric polynomial in the physical band is required. One verifies afterwards that an analytical sampling at rate ΔB is sufficient to represent the original real signal r(t).

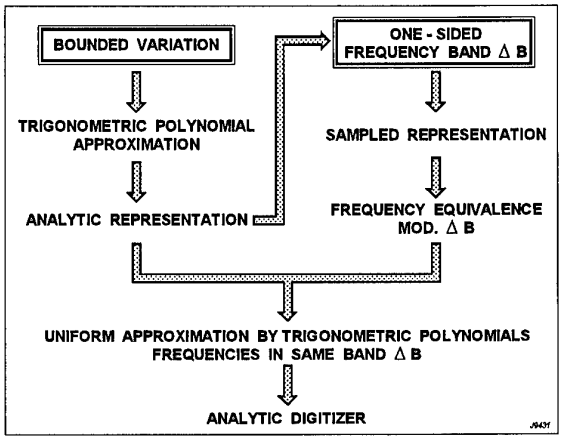


Figure 3 - Concept derivation

4. OPTIMUM ANALYTIC SAMPLER OF GIVEN COMPLEXITY

4.1. Definition of the Discrete Hilbert Transform

Let F_s be the real sampling frequency.

A discret Hilbert transform is defined by the n coefficients applied to the same number of samples taken from the real signal r(t) to be digitized.

The significantly better performance obtained with the second discretized Hilbert kernel suggests the use of an antisymetric weighting centered in the middle of two consecutive time samples. This weighting requires to either resampling r(t) at the corresponding time or the use of interpolation. In fact interpolation is preferable as it is a good method also to correct some defects or limitations of real digitizers as it will be evaluated in paragraph 5.2.

The Discrete Hilbert transform is defined by (9) which gives the 2 Cartesian coordinates of z(t), the analytic representation of r(t).

(9)

$$\begin{cases} x(t) \overset{\Delta}{=} a_o\left[r\left(t-\frac{1}{2Fs}\right)+r\left(t+\frac{1}{2Fs}\right)\right]+\ldots+a_m\left[r\left(t-\frac{2m+1}{2Fs}\right)+r\left(t+\frac{2m+1}{2Fs}\right)\right] \\ y(t) \overset{\Delta}{=} b_o\left[r\left(t-\frac{1}{2Fs}\right)-r\left(t+\frac{1}{2Fs}\right)\right]+\ldots+b_m\left[r\left(t-\frac{2m+1}{2Fs}\right)-r\left(t+\frac{2m+1}{2Fs}\right)\right] \\ z(t) \overset{\Delta}{=} x(t)+jy(t) \end{cases}$$

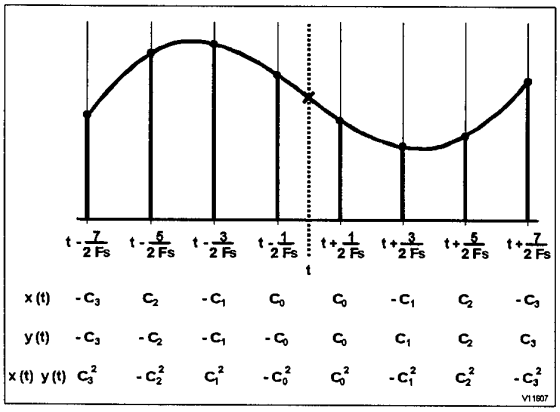


Figure 4 - Analytic interpolating coefficients

The transform r(t) → z(t) defined in (9) is linear and r(t) can be approximated by a trigonometric polynomial whose frequencies are in the spectral band. As a consequence we determine the coefficients a_i, b_i in such a way that z(t) obtained by (9) is nearly exact about the central frequency F.

To obtain this result, the standard method is to equate the first derivatives of r(t)=cos(2Πft+φ) to zero for f = F.

For this test signal, we evaluate (9) thus highlighting its basic structure.

(10)

$$\begin{cases} \theta \overset{\Delta}{=} \Pi\frac{f}{F_s} \\ x(t) = 2A(\theta)r(t) \text{ with } A(\theta) \overset{\Delta}{=} \sum_{k=0}^{m} a_k \cos(2k+1)\theta \\ y(t) = 2B(\theta)q(t) \text{ with } B(\theta) \overset{\Delta}{=} \sum_{k=0}^{m} b_k \sin(2k+1)\theta \end{cases}$$

The technical problem is reduced to the solution of a linear system of the form A(θ)=B(θ) or A(θ)=-B(θ) and $\partial A=\partial B=0$, $\partial^2 A=\partial^2 B=0$ etc, where the derivatives are taken with respect to θ.

4.2. Structure of the optimum coefficients set

As this system is homogeneous in a_k, b_k we can divide by two the number of equations, by introducing the 2 variables a=cosθ and b=sinb.

By recurrence, it can be proved that cos(2k+1)θ and sin(2k+1)θ can be expressed by the same set of polynomials $Q_k(u)$ of degre k in u.

(11)

$$\begin{cases} \cos(2k+1)\theta \overset{\Delta}{=} (-1)^k Q_k(a^2).a & Q_0(u)=1 \\ \sin(2k+1)\theta = Q_k(b^2).b & Q_1(u)=3-4u \\ \text{with } 1 = a^2+b^2 & Q_2(u)=5-20u+16u^2 \\ a = \cos\theta & Q_3(u)=7-56u+112u^2-64u^2 \\ b = \sin\theta & \text{etc...} \end{cases}$$

The solution in a_k, b_k can be reduced to the solution of C_k in the system defined by (12).

(12)

$$P(v) = v\sum_{k=0}^{m} C_k Q_k(v^2) \text{ and } \frac{d^k P(v)}{dv^k} = 0 \text{ for } k = 1,2,\ldots,m$$

Effective determination of A(θ) and B(θ) is deduced from solution of (12) by the substitution below.

(13)

$$\begin{cases} \bar{A}(a) = \bar{A}(\cos\theta) \overset{\Delta}{=} A(\theta) \text{ and } \bar{B}(b) = \bar{B}(\sin\theta) \overset{\Delta}{=} B(\theta) \\ \bar{A}(a) = a\sum_{k=0}^{m} a_k(-1)^k Q_k(a^2) \Leftarrow a_k = (-1)^k C_k \Leftarrow P(a) \Leftarrow a = \cos\theta \\ \bar{B}(a) = b\sum_{k=0}^{m} b_k Q_k(b^2) \Leftarrow b_k = c_k \Leftarrow P(b) \Leftarrow b = \sin\theta \end{cases}$$

In general the method indicated by (13) requires optimization of P(v) for v = a and v = b.

Optimisation of P(v) gives the same output $\{c_k\}$ provided that $v^2 = a^2 = b^2$. This condition is equivalent to $\theta = \pm\frac{\Pi}{4} \bmod \Pi$ for $f = F$.

From now on, we suppose that this condition is fulfilled which yields a new relation between the sampling frequency F_s and the mean frequency F of the signal.

(14)

$$2\theta_0 = 2\Pi\frac{F}{F_s} = \pm\frac{\Pi}{2} + 2k\Pi \Rightarrow Fs = \frac{4F}{1+2k} \text{ with } k \geq 0$$

Thanks to the presence of the integer k in (14), the variety of solutions in Fs is satisfactory in practice, as will be explained later.

The set of C_k solution of (12) for $\theta=\theta_0$ verifies (15).

(15)

$$\begin{cases} A(\theta) = (\cos\theta)\sum_{k=0}^{m}(-1)^k C_k Q_k(\cos^2\theta) \\ \quad = \sum_{k=1}^{m}(-1)^k C_k \cos(2k+1)\theta \\ B(\theta) = (\sin\theta)\sum_{k=0}^{m} C_k Q_k(\sin^2\theta) \\ \quad = \sum_{k=1}^{m} C_k \sin(2k+1)\theta \end{cases}$$

One verifies directly on (15) that the two gains $A(\theta)$ and $B(\theta)$ are related by symmetry about the half Shannon sampling frequency.

(16) $$|A(\theta| = \left|B\left(\frac{\Pi}{2} - \theta\right)\right| \text{ and } |B(\theta)| = \left|A\left(\frac{\Pi}{2} - \theta\right)\right|$$

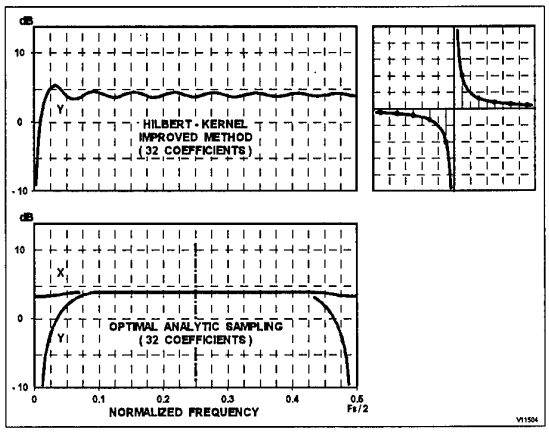


Figure 5 - Optimal analytic sampling (32 coefficients)

4.3. Derivation of the optimum coefficients values

From the first definition (10), it is clear that in the optimisation process, the Van der Mond structure will show up. For this purpose, we formulate the optimization with the (new) complex function $P(\theta)$.

(17)

$$P(\theta) \overset{\Delta}{=} A(\theta) + jB(\theta)$$

$$(P(\theta))^{(k)} \overset{\Delta}{=} \frac{d^k P(\theta)}{d\theta^k} = 0 \quad \text{for } k = 1, 2, \ldots, m$$

$$P(\theta) = \sum_{k=0}^{m}(-1)^k C_k\left[\cos(2k+1)\theta + j(-1)^k \sin(2k+1)\theta\right]$$

$$= \sum_{k=0}^{m}(-1)^k C_k e^{js_k\theta} \text{ with } s_k \overset{\Delta}{=} (-1)^k(2k+1)$$

The h^{th} derivative of $P(\theta)$ with respect to θ takes the form below within a constant, which is irrelevant for the optimisation.

(18)

$$\begin{cases} P(\theta) = C \\ [P(\theta)]^h = \sum_{k=0}^{m}(-1)^k C_k (s_k)^h e^{js_k\theta} \text{ with } h = 1,2,\ldots,m \\ \quad = \sum_{k=0}^{m} d_k (s_k)^h = 0 \text{ with } d_k \overset{\Delta}{=} (-1)^k C_k e^{js_k\theta} \end{cases}$$

Up to an irrelevant constant, the solution of this homogeneous system in d_k is given by the following determinants obtained by removing the first row and k^{th} columm of the determinant of the original system (18).

$$d_k = (-1)^k \begin{vmatrix} s_0 & s_1 \ldots\ldots s_{k-1} s_{k+1} \ldots\ldots s_m \\ s_0^2 & s_1^2 \ldots\ldots (s_{k-1})^2 \ldots\ldots (s_m)^2 \\ s_0^m & s_1^m \ldots (s_{k-1})^m \ldots\ldots (s_m)^m \end{vmatrix}$$

$$\Rightarrow C_k = (-1)^k e^{-j s_k \theta} d_k$$

Since we have an explicit formula to evaluate a Van der Mond determinant, we divide each column by a convenient factor to get one at each entry of the first line.

$d_k = (-1)^{k-1}\ s_{k-1}\ G_{k-1}$ and $d_{k-1} = (-1)^k\ s_k\ G_k$

with

$$G_\lambda = \begin{vmatrix} 1 \ldots\ldots\ldots 1 \ldots\ldots\ldots 1 \ldots\ldots\ldots 1 \ldots\ldots\ldots 1 \\ s_0 \ldots\ldots\ldots s_{k-2} \ldots\ldots\ldots s_\lambda \ldots\ldots\ldots s_{k+1} \ldots\ldots\ldots s_m \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ s_0^{m-1} \ldots\ldots s_{k-2}^{m-1} \ldots\ldots s_\lambda^{m-1} \ldots\ldots s_{k+1}^{m-1} \ldots\ldots s_m^{m-1} \end{vmatrix}$$

$$= f(s_0, \ldots, s_{k-2}, s_{k+1}, \ldots, s_m) \cdot w$$

with

$$w \overset{\Delta}{=} \left[\prod_{i=0}^{k-2} (s_\lambda - s_i)\right] \left[\prod_{j=k+1}^{m} (s_j - s_\lambda)\right]$$

From this two columns expansion, we are able to give a compact expression for the ratio of two consecutive d_k.

$$\frac{d_k}{d_{k-1}} = -\frac{2k-1}{2k+1} \frac{G_{k-1}}{G_k}$$

$$= -\frac{2k-1}{2k+1} \left[\prod_{i=0}^{k-2} \frac{s_{k-1} - s_i}{s_k - s_i}\right] \left[\prod_{j=k+1}^{m} \frac{s_j - s_{k-1}}{s_j - s_k}\right]$$

For $\theta = \pm\frac{\Pi}{4}$ we draw from (19) an expression for $\frac{C_k}{C_{k-1}}$.

(19)

$$\frac{C_k}{C_{k-1}} = -(-1)^k \frac{d_k}{d_{k-1}}$$

$$= -(-1)^k \frac{2k-1}{2k+1} \left[\prod_{i=0}^{k-2} \frac{s_{k-1} - s_i}{s_k - s_i}\right] \left[\prod_{i=k+1}^{m} \frac{s_{k-1} - s_i}{s_k - s_i}\right]$$

As such the expression can be significantly simplified.

$$S(k,i) \overset{\Delta}{=} \frac{s_{k-1} - s_i}{s_k - s_i} = -\frac{2k - 1 - (-1)^{k+i}(2i+1)}{2k + 1 - (-1)^{k+i}(2i+1)}$$

Direct evaluation gives the following results.

$$\begin{cases} k+i \text{ odd} & S(k,i)\ S(k,i+1) = 1 \\ k \text{ even} & S(k,0) = -1 \end{cases}$$

$$\frac{C_k}{C_{k-1}} = (-1)^{k-1} \frac{2k-1}{2k+1} \left[\prod_{i=0}^{k-2} S(k,i)\right] \left[\prod_{j=k+1}^{m} S(k,i)\right]$$

When k is **even**, k-1 is odd, and the first bracket equals $-1=(-1)^{k-1}$, the second bracket equals 1, if k+n is even the second bracket equals 1, and S(k,m) otherwise.

When k is **odd**, the first bracket equals $1=(-1)^k$, and the second bracket is different from one only when k+m is odd and then equals S(k, m).

(20)

$$\frac{C_k}{C_{k-1}} = \begin{cases} \dfrac{2k-1}{2k+1} & k+m \text{ even} \\ \dfrac{2k-1}{2k+1}\ \dfrac{1+m-k}{1+m+k} & k+m \text{ odd} \end{cases}$$

$k = 0, 1, \ldots, m$

The number of signal samples is n=2(1+m), $0 < \frac{C_k}{C_{k-1}} < 1$, the last ratio is always $\frac{C_m}{C_{m-1}} = \frac{2m-1}{2m+1}$.

n=4 or m=1

$$\frac{C_1}{C_0} = \frac{1}{3}$$

$$\begin{cases} -1 & 3 & 3 & -1 \\ -1 & -3 & 3 & 1 \end{cases}$$

n=6 or m=2

$$\frac{C_1}{C_0}=\frac{1}{6},\frac{C_2}{C_1}=\frac{3}{5}$$

$$\begin{cases} 3 & -5 & 30 & 30 & -5 & 3 \\ -3 & -5 & -30 & 30 & 5 & 3 \end{cases}$$

n=8 or m=3

$$\frac{C_1}{C_0}=\frac{1}{3},\frac{C_2}{C_1}=\left(\frac{3}{5}\right)^2,\frac{C_3}{C_2}=\frac{5}{7}\Rightarrow\frac{C_3}{C_0}=\frac{3}{35}$$

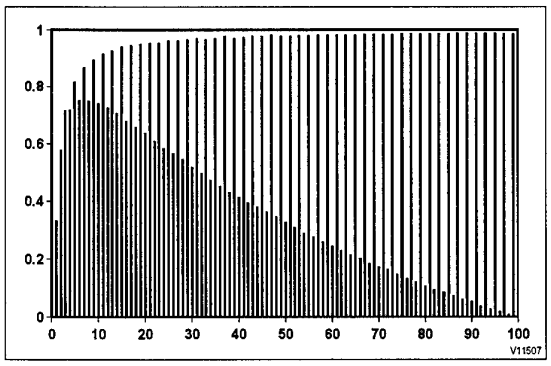


Figure 6 - Coefficient ratios for 200 samples

5. PRACTICAL FEATURES OF DHT

5.1. Quadrature noise ratio -ρ_Q

For any finite number of samples used to evaluate an analytic sample from a real signal, there exists a difference between the exact analytic sample and the complex sample obtained by (20).

Since trigonometric polynomials constitute a complete set and the modulus of the above difference is proportional to the sinewave amplitude, it is sensible to characterise the discrete Hilbert Transform by a signal to noise ratio, or quadrature noise ratio ρ_Q, given by (21).

test signal $\quad r(t)=\cos(2\Pi ft+\varphi)$

exact quadrature calculated $\quad q(t)=\sin(2\Pi ft+\varphi)$

analytic sample $\bar{z}(t) = A(\theta)r(t) + jB(\theta)q(t)$

Q is a measure of the DHT approximation :

$$Q=\min_{Greal}\left[(A(\theta)-G)^2+(B(\theta)-G)^2\right]=\frac{1}{2}\left[A(\theta)-B(\theta)\right]^2$$

$$\rho_0(\theta)\overset{\Delta}{=}\left[\frac{A(\theta)+B(\theta)}{A(\theta)-B(\theta)}\right]^2=\rho_Q\left(\frac{\Pi}{2}-\theta\right) \quad (21)$$

$$\text{with}\,\theta=\Pi\frac{f}{F_s}$$

ρ_Q is a function of the number n of real samples.

Digital exploration of (20) and (21) shows that DHT with n a multiple of 4 always performs better. For this favorable situation, the relation $\frac{C_1}{C_0}=\frac{1}{3}$ holds. Figure 6 gives ρ_Q as a function of the number n of real samples, and as a function of bandwidth normalised by $\frac{Fs}{2}$ half the real sampling rate (Shannon frequency).

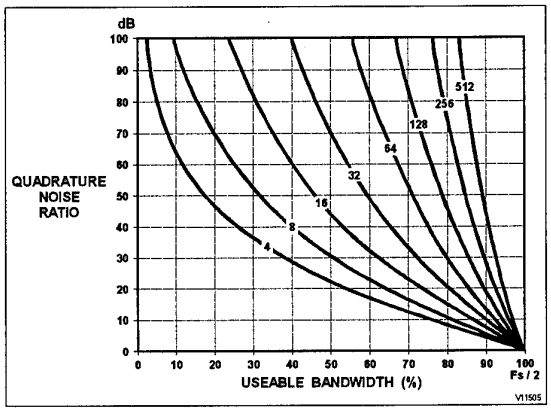


Figure 7 - Performance complexity chart

5.2. Dynamics of a finite analytical digitizer

Let the real sample r(t) be perturbed by an additive noise n(t) such that $\langle n(t)\rangle=0$ and $\langle n(t)n(t')\rangle=\sigma^2\delta tt'$.

This noise may be generated by linearity defects of the real digitizer, as each cartesian component of an analytic sampler has better dynamics than the actual real digitizer.

$$x(t)=\sum_{k=0}^{m}a_k\left[r\left(t+\frac{k}{Fs}\right)+r\left(t-\frac{k}{Fs}\right)\right]$$

$$\text{with}\ r(t)=\bar{r}(t)+n(t)$$

Let $\bar{r}(t)$ be the exact value of r(t).

For a sinewave whose frequency corresponds to $\theta = \frac{\Pi}{4}$, with a mean power of R^2 the output noise is $\left(2\sum_{k=0}^{m} a_k^2\right)\sigma^2$ and the mean useful power is $\left(2\sum_{k=0}^{m}|a_k|^2 R^2\right)$.

We can thus express the output signal to noise ratio ρ as a function of the input signal to noise ratio $\frac{R^2}{\sigma^2}$.

$$\rho = \frac{2\left(\sum_{k=0}^{m}|a_k|\right)^2}{\sum_{k=0}^{m} a_k^2} \quad \frac{R^2}{\sigma^2} \tag{22}$$

For m=3, corresponding to eight real samples, the improvement factor $I_n \overset{\Delta}{=} 2\left(\sum_{k=0}^{m}|a_k|\right)^2 \Big/ \left(\sum_{m=0}^{m} a_k^2\right)$ is 4 or 6 dB, which means that x(t) is equivalent to a real digitizer with one more bit than the first real digitizer.

$$I_8 = 2\frac{\left(1+\frac{C_1}{C_0}\left(1+\frac{C_2}{C_1}\left(1+\frac{C_3}{C_2}\right)\right)\right)^2}{1+\left(\frac{C_1}{C_0}\right)^2\left(1+\left(\frac{C_2}{C_1}\right)^2\left(1+\frac{C_3}{C_2}\right)^2\right)}$$

$$= 2\frac{\left(1+\frac{1}{3}\left(1+\left(\frac{3}{5}\right)^2\left(1+\frac{5}{7}\right)\right)\right)^2}{1+\left(\frac{1}{3}\right)^2\left(1+\left(\frac{3}{5}\right)^4\left(1+\left(\frac{5}{7}\right)^2\right)\right)}$$

5.3. Statistical properties of the noise in quadrature

The Analytic Samples developed here (cf (13) and (20)) are linear, meaning that we can study its behaviour separately in the case of deterministic signals and of additive noise $n_r(t)$.

The structure of the solution given in (13) is sufficient to prove that the noise in quadrature $n_q(t)$ is decorrelated from $n_r(t)$ and has the same variance, provided the n real samples are decorrelated, as we can see from Figure 4.

$$< n_r(t) >=< n_q(t) > = 0, < (n_r(t))^2 >=< (n_q(t))^2 > = \sigma^2.$$

5.4. Maximum likelihood principle made analytic

According to the maximum likelihood principle, the most probable hypothesis is the one which has the greatest probability to produce what has been observed. In signal processing where perturbation can be modelized as an additive noise of zero mean and specified variance, this principle can be interpreted in the following non-parametric manner : choose the hypothesis which gives the lowest value for the calculated energy of the noise.

As an example of this procedure, we try to recognize a symbol (in the complex plane) belonging to a finite alphabet A subject to rotation and scaling. The analytic samples of the symbol $H \in A$ are noted $z_H(t)$.

The actual analytic samples z(t) are of the form below, where G is an unknown complex factor and n(t) the complex additive noise.

$$z(t) = Gz_H(t) + n(t) \text{ with } \sum_t z_H(t) z_H^*(t) = 1$$

For the hypothesis H, the most probable value of G is the one which minimizes $\sum_t n(t)n^*(t)$, which we define as G_H.

(23)

$$Q_H = \min_G \sum_t \left[z(t) - Gz_H(t)\right]\left[z^*(t) - G^* z_H^*(t)\right]$$

$$G_H = \frac{\sum_t z(t) z_H^*(t)}{\sum_t z_H(t) z_H^*(t)}$$

$$\Rightarrow Q_H = \sum_t z(t) z_H^*(t) - \frac{\left(\sum_t z(t) z_H^*\right)\left(\sum_t z^*(t) z_H(t)\right)}{\sum_t z_H(t) z_H^*(t)}$$

The expression of Q_H shows that the most probable gain G corresponds to the greatest G_H for H fixed and the most probable hypothesis H corresponds to the greatest G_H or **estimated carrier**.

The application of (23) to the identification of sinewaves of frequencies f_1, f_2, etc leads us to calculate the module of $G_i = \frac{1}{N}\sum_t z(t) \exp -2\Pi j f_i t$ which is an

element of a Discrete Fourier Transform but **without any windowing.**

5.5. Band pass filtering

For sampled analytic signals, the time samples of a band-pass centered at F can be obtained by multiplying the samples of a real low-pass exp-2ΠkΔt as we now prove.

Let r(t)=cos[2Πft+φ] and a_k be the real coefficients of a low pass filter. The filtered r(t) is defined as $\bar{r}(t)$.

$$\bar{r}(t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} a_k \cos\left[2\Pi f(t+k\Delta t)+\varphi\right]$$

Its signal in quadrature $\bar{q}(t)$ is given by the following expression.

$$\bar{q}(t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} a_k \sin\left[2\Pi f(t+k\Delta t)+\varphi\right]$$

$$\bar{z}(t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} a_k \exp j\left[2\Pi f(t+k\Delta t)+\varphi\right]$$

$$z(t) = \exp j(2\Pi ft+\varphi)$$

Z(t) is defined as the translation of z(t) to the band centered at F.

$$Z(t)=z(t)\exp 2\Pi jFt$$

Filtering this last signal by a filter whose coefficients are $A_k=a_k \exp -2\Pi jFk\Delta t$ verifies the identity below.

$$(24)\quad \bar{Z}(t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} A_k Z(t+k\Delta t) = \bar{z}(t)\exp 2\Pi jFt$$

q.c.d.

Equation (24) is valid for any analytic signal Z(t) because frequency translation and bandpass filtering are linear transformations.

6. CONCLUSION

Our motivation to investigate and use sampled analytical arithmetics began in 1986 in order to develop an all digital receiver for SARSAT system. The difficulties at that time were the very low nominal $\frac{E_b}{N_0}$ and the Doppler range which required a great variety of band-pass filters **[6].**

Applications to other domains, such as spread spectrum, localization and antenna arrays necessitated to master the analytical sampling theory which has been accomplished at least for practical purposes. The next important theoretical step will be the direct synthesis of discrete real low-pass filters.

BIBLIOGRAPHY

[1] VILLE (J) Signaux Analytiques à Spectre Borné Câbles et Transmission 4:9 (1950)

[2] COURANT (R) and HILBERT (D) Methods of Mathematical Physics - Interscience Publishers N.Y. - Vol 1 (1965) P65 to 69

[3] GRADSHTEYN (IS) and RYZHIK (IM) Table of Integrals, Series, and Products - Academic Press (1980)

[4] BOROWSKI (EJ) and BORWEIN (JM) Dictionary of Mathematics - Collins (1989)

[5] WIENER (N) The Fourier Integral and certain of its Applications - Dover (1958)

[6] MARGUINAUD (A) - QUIGNON (T) - ROMANN (B) Demodulation of Messages Received with Low Signal to Noise Ratio - ICDCS 24-28 April 1989.

DISCUSSION

Discussor's name: C. Goutelard

Comment/Question:

Votre procédé est bien large bande?

(Translation:

Is this really a wide band procedure)

Author/Presenter's reply:

Le procédé est bien à large bande puisqu'assymptotiquement la bande utilisable tend vers la moitié de la fréquence d'échantillonnage Fe. Pour un nombre d'échantillons égal à 4 on a 23dB pour $\frac{Fe}{8}$ et plus to 30 dB pour $\frac{Fe}{4}$, la bande de Shannon étant de $\frac{Fe}{2}$.
Le procédé est utilisable à des Fi élevées car le bruit de quadrature n'est fonction que du rapport $\frac{Fe}{\text{bande filtrée}}$.

(Translation:

The procedure is in fact wide band, since asymptotically the usable band tends towards half of the sampling frequency Fe. For a number of samples equal to 4 one has 23dB for $\frac{Fe}{8}$ and more than 30 dB for $\frac{Fe}{4}$, if the Shannon band is $\frac{Fe}{2}$.
The procedure can be used at high Fi as quadrature noise is only a function of the relationship $\frac{Fe}{\text{filtered band}}$.

Effects of bandwidth reduction of transmitted motion picture sequences on human recognition performance

Klaus-Peter Gärtner
Frank E. Schneider
Research Institute for Human Engineering of the FGAN e.V.
Neuenahrer Str. 20, 53343 Wachtberg, Germany
Tel: +49-228-94 35 481
e-mail:gae@fatvax.fgan.fat.de

1. INTRODUCTION

The Research Institute for Human Engineering (FAT) has been working in the field of land telerobot systems for the last six years. The main goal of current research work is the optimization of remote control and supervision of semiautomatic vehicles from a single control station for various missions [1]. The research work done on a particular aspect of robot deployment in a driving task will be described in this paper. Robots are likely to be positioned at particular locations or to be moving in a terrain overlooking areas of military interest, such as likely enemy positions, contested areas, or supply lines (Figure 1). During the moving operation the robot is autonomous and ready to detect obstacles in a selected part of the landscape in front of it. After the robot is programmed with its mission there is no need for further radio communication until an obstacle is sensed which it cannot interpret because the automatic recognition system cannot recognize all threatening object situations for the driving task. "Upon request", the vehicle requires help and support from humans. In this case the human acts as a remote controller and supervisor. In a remote control station he receives still pictures or picture sequences obtained from the vehicle over a communication channel.

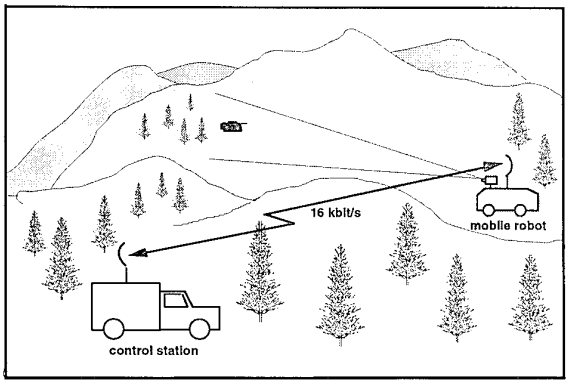


Figure 1: Example of a surveillance mission.

Land telerobot vehicles mostly operate out of sight, that is, at distances beyond normal visual range, usually greater than a kilometer from the control station and frequently with various obstacles inbetween. For that reason and also to prevent enemy interception of friendly communications for military purposes, communications between the robot and the control station are on a short wave band. Because the available communications bandwidth is always limited on short wave radio, wide band transmission of standard TV signals is impossible. Instead, it is necessary to strongly compress the picture contents. Such a strongly reduced band width causes a loss in picture quality on the receiving end.

The goal is to minimize data exchanges and communication loads between the control station and the telerobot subsystems.

The aim of the ergonomic experiment is to answer the following question: How narrow can the communication band be, with a picture refresh rate of 16-18 images/s, in order for the remote controller to extract the detail information necessary to drive the remotely controlled vehicle?

Similar studies have been done by Ludgat [2].

2. TEST FACILITY/TRIAL ROUTES

The test facility consists of a robot vehicle in a test environment and an operator in a control station. To verificate the interactions and effectiveness of the equipment, for the first tests an indoor experiment was chosen. It is based on a vehicle that operates in a factory hall. The factory hall measures approximately 18 x 14 m. It is equipped with cylindric obstacles of about 0.50 m diam-

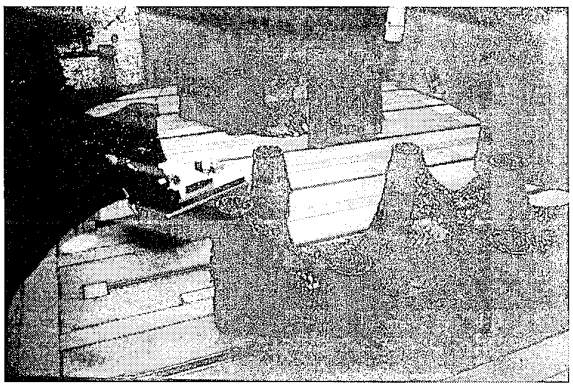


Figure 2: Test vehicle in experimental set-up.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

eter and 1.50 m height. The walls and the cylindric obstacles are covered with camouflage nets to simulate bushes and trees (Figure 2). Also included are facilities for data transmission to and from the experimental vehicle and a TV link. The hall is equipped with a number of safety systems to insure safe operation of the vehicle. The vehicle itself is a four wheel platform of less than 2 m in length. It is 1.30 m wide and about 1.60 m high.

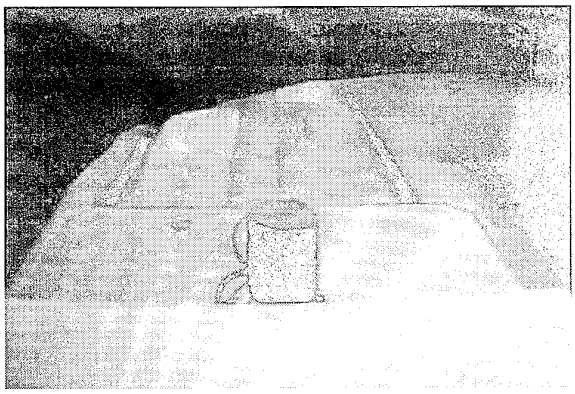


Figure 3: Camera front view.

The vehicle has two electric motors that power the two back wheels. It has a pair of front wheels which are steerable. The maximum speed is about 2.3 m/s. The vehicle is equipped with a camera which surveys the terrain directly in front of the vehicle (Figure 3). The possible driving route in the hall is essentially circular and describes an endless eight. The vehicle is driven with a joystick from a remote teleoperated driving console on which the camera image from the vehicle is displayed [3].

3. COMPRESSION FACILITY

The compression hardware used in the experimental set-up is based on the JPEG-Standard [4]. JPEG is a picture compression standard (ISO/IEC CD 10918-1), developt by the ISO/CCITT-Group (Joint Photographic Experts Group - JPEG). It defines a standard for the compression of grey and colour pictures.

The JPEG-Method is not only used for still-picture compression but also for Video or motion-picture-compression. The method is a lossy compression, which means information loss is accepted to achieve higher compression rates [5].

If the difference between the original and the compressed picture should be as less as possible for a human beholder, the compression method should use a technique that reduces the picture only in a manner that is suitable for human recognition [6].

For example brightness changes, areas in the picture with high frequency are less critical for human recognition as areas with low frequencies. The picture may look a little unsharp with reduced or without high frequency, but it still remains perceptible.

The JPEG-Method uses a transformation that allows the directed removal or reduction of "unimportant" information. The DCT (Discrete Cosinus Transformation), a special form of the Fourier-Transformation (DFT/FFT) supports this requirements [7].

The JPEG standard allows different methods for compression that can be divided into three groups:

- base process (Baseline Sequential Process, DCT)
- extended DCT-Based Processes
- lossless Processes

All JPEG-Variants must support the Base-Process, the other processes can be used application dependent. The paper will only focus one the DCT-Processes.

The whole compression can be divided into three steps:

- removal of the DCT redundancies
- quantisation of the DCT-Coefficients with a special weighting-function, taking into account human recognition aspects
- encoding with a Huffmann based algorithm

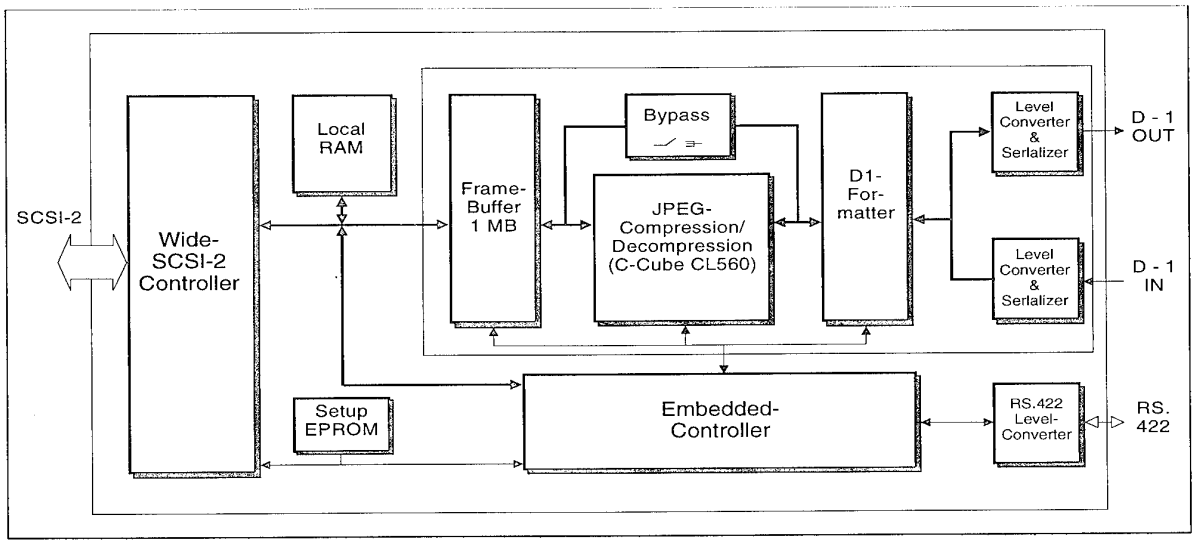


Figure 4: Scheme of the C-Cube CL 560 (M-JPEG) compression chip.

Logically the whole picture in one should be transformed before the information reduction, but, to keep the computing costs low, usually the picture is divided into blocks of 8x8 pixels. These blocks will be transformed and compressed seperatly and independent from each other.

The result from the DCT, is a 8x8 floating point matrix that represents a measure for the details present in this picture section. The coefficient in the upper left corner of the matrix is called the DC-Coefficient, because it represents the average brightness of the entire block. It is the most important information and is stored without reduction. As more as the other coefficients fade away from the DC-Coefficient, the picture structure gets more precise.

Until this point the method works with out major loss and compression. The compression and the loss of information is introduced through the next step, the quantisation.

The 8x8 matrix is converted into to a vector. In this vector only the first few coefficients are unequal zero, most of the rest are near or equal zero. To decide which of the coefficients are important and which are not, they are divided by a constant. Each of the coefficients in the 8x8 matrix has a individual constant. To control the quality loss the first constants are small and getting larger to the end. As a fact larger areas with the same brightness (and/or colour) can be compressed higher than areas which high frequency (a lot of details in different colour and brightness).

After this quantisation the whole picture information is encoded with a Huffmann coder.

To restore the compressed picture (decompress) first a Huffmann decoding is done. Second the coefficients are multiplicated with there constants and a inverse DCT is done.

On colour pictures, JPEG treats colour as a separate component, so different colour picture formats can be used (i.e. RGB, YCbCr and CMYK). The best results can be achieved with independent colour components like YCbCr where the brightness- and color-information are separate components. The Chip used in our hardware (a Hardware implementation of the Base-Process) requires the colour signals in the 4:2:2 (also called D1) format of YCbCr. RGB-components (like used in the experiment) can be transformed into YCbCr with linear transformation.

Because the JPEG-Algorithm divides the picture in 8x8 pixel blocks and each of these blocks is compressed independent of the others, errors occur at the borders to the next blocks [8]. These errors result in edges between the blocks after the decompression and can be seen very well if there is a high contrast and edges in the picture.

With the JPEG-Method compression rates between 2:1 and 100:1 can be achieved. A special extension to JPEG, called M–JPEG (Motion-JPEG), allows compressing video signals on-the-fly with 1 to 25 pictures/s. This extension is based on the powerful C-Cube CL560 compression chip (Figure 4). A user-selectable quality index offers the user highest flexibility to adapt the picture quality to the application as well as to his working environment.

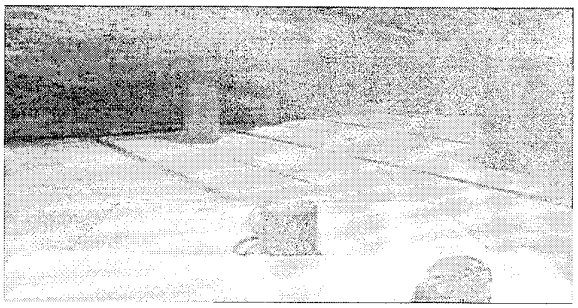


Figure 5a: Picture quality 15%

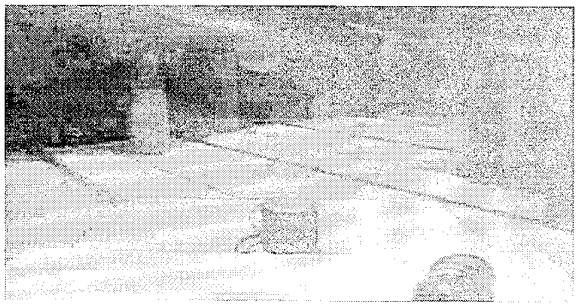


Figure 5b: Picture quality 11%

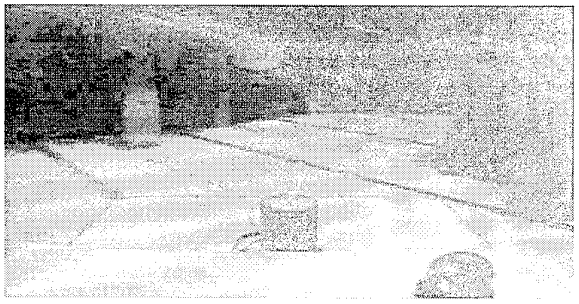


Figure 5c: Picture quality 9%

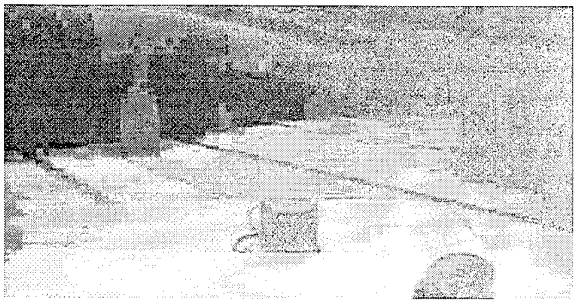


Figure 5d: Picture quality 5%

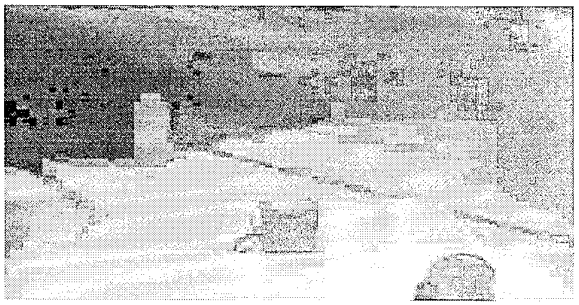


Figure 5e: Picture quality 3%

Due to the hardware restrictions we had a delay of two frames measured from the time the camera picture is received until the quality reduced picture is presented on the monitor screen watched by the controller. The delay is caused primary by the SCSI-Bus.

4. TRIALS SPECIFICATIONS

Drivers were instructed to drive the route with best accuracy while keeping nominally to the center of the track. The speed for each run was at a fixed value and could not be changed by the driver during the driving task.

The test person could activate the brakes by pressing a button if necessary. The driving speeds were 75 cm/s and 100 cm/s. The test persons drove first at a picture quality level of 100% to get used to the vehicle dynamics and the trial course. For the following tours, the picture quality was subsequently reduced to 15%, 11%, 9%, 5%, and 3% (Figure 5a-e).

5. RESULT AND CONCLUSION

The test results show that the task was made too easy for the test persons. The course was learned in a very short time. The results show, the test persons' mistakes do not differ significantly at reduced picture quality levels (Figure 6) [9]. The relatively good driving performance can be attributed mainly to the clearly visible contrast between ground and camouflage netting. Hanging the nets higher, which leads to a reduced contrast level, wasn't considered an added difficulty by the test persons (Figure 7)

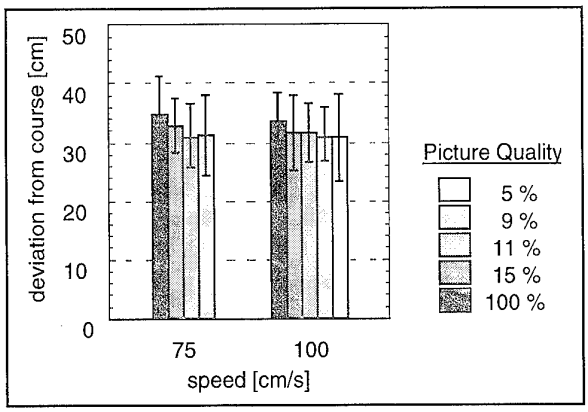


Figure 6: Deviation from command course depending on driving speed and picture quality

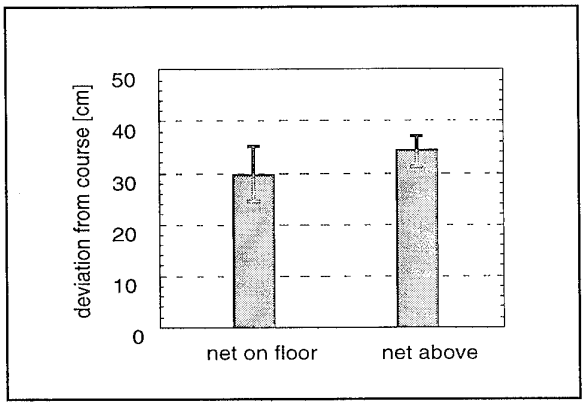


Figure 7: Deviation from command course depending on (camouflage-) net height

In summary it can be stated:

The M-JPEG-System can be of advantageous use for additional driving or observation tests. Further tests will have to take place in open terrain.

6. REFERENCES

1. Gärtner, K.-P., Holzhausen, K.-P., Krüger, W., Pitrella F. D., Wolf, H.-L., "Identification of field objects in reduced quality TV pictures transmitted from telerobots to a remote control station", Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 93), Yokohama, Japan.
2. Ludgate, D., "An Investigation to Assess the Ability of a Driver to use a Low Bandwidth Image to Control an Unmanned Vehicle through a Teleoperated Link", Robotics Department, Fighting Vehicles and Systems Sector, Military and Aerospace Group, DRA Chertsey, Note 08/92.
3. Holzhausen, K.-P., Wolf, H.-L., "Ein experimentelles Robotersystem für die anthropotechnische Forschung, Mensch-Maschine-Systeme", Research Institute for Human Engineering, August 1994, ISBN3-86073-3354-4.
4. ISO DIS 10918-1, CCITT recommendation T.81, "Digital Compression and coding of continuous-tone still images" (The JPEG standard).
5. Pennebaker, W. B. and Mitchell, J. L., "JPEG Still image data compression standard" (Contains [1] as an appendix),, Van Nostrand Reinhold ISBN 0-442-01272-1.
6. Nesenbergs M., "Image Data Compression Overview: Issues and Partial Solutions", U.S. Dept. of Commerce, October 1989.
7. Jain, A.K., "Digital Image Processing", Prentice Hall, 1989.
8. Musmann, H.-G., Werner, O., Fuchs, H., "Kompressionsalgorithmen für interaktive Multimedia-Systeme", it+ti 2/93, Informationstechnik und Technische Informatik 35(1993)2, p. 4-18, R. Oldenbourg Verlag.
9. Martinek, H., Zarin, A., "The Effects of Bandwidth Compression on Image Interpreter Performance", U.S. Army Research Institute for the Behaviroal and Social Siences, August 1979.

DISCUSSION

Discussor's name: F. Davarian

Comment/Question:

1. How do you define picture quality? What does 10% quality mean?

2. Comment: You can use a modem with higher through-put for signal transmission. Therefore rather than cutting picture quality one can use a more sophisticated modem. For example, use 8-PSK rather than BPSK.

Author/Presenter's reply:

1. The picture quality index 10% i.e. means the compression mode of the JPEG system. The compression factor is possible from 2:1 until 100:1, so 100:10 means 10% picture quality. The quality index represents the pixel with its actual gray step and a certain cut of higher frequencies of picture details in the frequency domain. In the standard version we have used, the basic subparts of the picture were determined by the standard compression algorithm. It is possible to change the weight of each matrix element. In that case you will get a different picture index.

2. Thank you for your comment.

Analysis of propagation effects and non-linear distortions on digital modulation schemes

Kurt Metzger and Rolf Valentin

Telekom AG
Forschungs- und Technologiezentrum
Postfach 100003
D 64276 Darmstadt, Germany

1 SUMMARY

The paper deals with the analysis of digital transmission systems which are corrupted by thermal noise, linear channel dispersion and nonlinear distortions which are caused by the transmit amplifier. The central part of this analysis is a recursion model for the calculation of the probability density function (pdf) of Intersymbol Interference (ISI). The nonlinearity introduces statistical interdependencies between the interfering symbols and these dependencies are implicitly taken into account in a trellis-structured recursion rule. The procedure was applied to 4-level modulation schemes such as 4-QAM, offset 4-QAM and linearized GMSK. The results were verified by time consuming Monte-Carlo simulation and show, e.g., that the nonlinear characteristic of a high power amplifier (HPA) reduces in some cases the error probability caused by the linear dispersion. Surprisingly, the ISI due to the nonlinear amplifier is increased in the case of offset modulation.

2 INTRODUCTION

The performance of digital radio transmission systems is corrupted by thermal noise, linear channel dispersion, which detoriates the spectral shaping of the signal, and by the nonlinear characteristics of HPA's. For systems with threshold decision the symbol error probability can be evaluated if the probability density function (pdf) of the signal at the decision element is known. This procedure offers the possibility of evaluating fairly small error probabilities with a significant reduction of computation time. In addition our method yields some insight into the statistical properties of the received signal and it is conceptually easy to understand.

The nonlinearity has two effects on the transmission system; the first is the spectral widening of the signal which leads to adjacent channel interference (ACI). The second effect is the distortion of the signal in the nominal frequency band which causes ISI. This paper deals with the second phenomenon.

In the following we are concerned with QAM, offset-QAM and eventually GMSK. In section 2 the baseband equivalent system model is presented. For a realistic description the influence of the transmission channel and of the nonlinearity on timing- and carrier synchronization has to be taken into account. There are different solutions for the sampling instant and the recovered carrier phase if offset or non-offset modulations are used. In section 4 we present a trellis-structured recursion rule for determining the pdf of ISI which is generated by a frequency selective channel and a memoryless nonlinearity. The recursive calculation is an extension of a previous work on the pdf of ISI for linear channels [1] and, in order to make possible a compact presentation in section 4, a short review of the linear case is included in section 3. The results are verified by time consuming Monte-Carlo simulations in section 5. The results show, e.g., that system performance can be improved by the nonlinearity of a travelling wave tube amplifier (TWTA) during a two-path fading event.

For reduction of nonlinear distortions offset modulation schemes are sometimes favored. But our results in section 5 show that there is no improvement -even an impairment- for the offset modulation case due to enhanced ISI.

GMSK is a nonlinear continuous phase modulation (CPM) scheme, which in principle does not suffer from nonlinear memoryless amplitude distortions. But in practice the continuous phase property of the GMSK signal is disturbed by an anti-aliasing filter or by the linearized implementation of GMSK. Therefore, we include results for GMSK revealing that a nonlinearity can diminish the linear channel distortions also for this modulation method.

3 SYSTEM MODEL

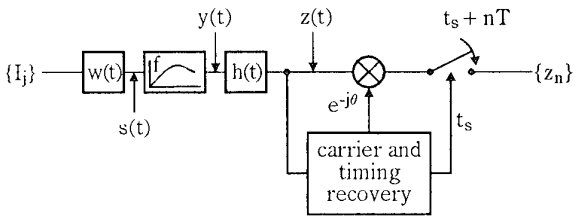


Figure 1: Baseband equivalent system model

Fig. 1 shows the baseband equivalent model for the system which is to be analyzed. A series of complex symbols $\{I_j\}$ enters the system. The examples in the following hold for independent symbols of a QAM scheme, but the extension to a redundant symbol set is no problem, because statistical dependencies between the symbols can be dealt with in the trellis-structure of the recursion rule in section 5. The transmit filter $w(t)$ equals the receive filter $w(t)$, both

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

are assumed to have a square-root cosine roll-off characteristic, which yields ISI-free matched filter reception in the absence of a nonlinearity and of channel dispersion. The nonlinear distortion of a HPA is described at baseband by a memoryless nonlinearity f because of the bandlimited nature of the transmission [2]. This means that the function f defines the envelope of the distorted rf-fundamental. The transmission channel is described in the following by a two-path propagation and the channel together with the receive filter $w(t)$ constitutes the filter $h(t)$ in Fig. 1. The phase of the received complex signal is normalized to the recovered carrier phase θ. The signal is sampled at the timing instants $t_s + nT$ resulting in a time discrete sequence $\{z_n\}$. We will be interested in the pdf of the sample z_0, but first the values of θ and t_s have to be determined. The recursion rule of section 5 does not depend on the particular synchronization algorithm, although we have to model the synchronization in order to get realistic results when the method is applied to the calculation of bit error probabilities caused by nonlinear distortion, linear channel dispersion and thermal noise.

3.1 Carrier and Timing Synchronization

Frequency selective transmission channels cause a shift of the optimal sampling instant t_s as well as an offset of the recovered carrier phase θ. These two parameters can be determined by the methods described in [3].

3.1.1 *Non-Offset Systems*

In [3] analytic expressions for θ and t_s were obtained dependent on the overall system impulse response. These analytic expressions were evaluated for non-offset systems from

$$t_s = -\frac{T}{2\pi} \arg \mathcal{F}\{< |z(t)|^2 > -0.1 < |z(t)|^4 >\}_{f=1/T}. \tag{1}$$

and

$$\theta = \arg < z(t_s) I_0^* > . \tag{2}$$

$< x >$ denotes the expectation of a random variable x with respect to the input symbols $\{I_j\}$, $\mathcal{F}\{x(t)\}$ is the Fourier transform of $x(t)$ and $\arg x$ is the phase of a complex number x. The received signal envelope $|z(t)|$ is transformed in eq. 1 by a nonlinear characteristic with Taylor coefficients k_2 and $k_4 = -0.1k_2$ and this transformation results in a spectral line at $f = 1/T$ [3]. Eq. (2) refers to the remodulation scheme for carrier recovery. With the nonlinearity f being present in Fig. 1 there is no overall impulse response and eq. (1) and (2) cannot be evaluated analytically. Therefore, we start first a short Monte-Carlo simulation run with a length of about 1000 symbols in order to determine t_s and θ from eq. (1) and (2). Subsequently, we can replace $w(t)$ by $w(t + t_s)e^{-j\theta}$ and arrive at a system model without synchronisation loop. The latter replacement holds because for HPA's (e.g. a TWTA) the phase shift of the output signal y does not depend on the phase of the input signal s in Fig. 1, but rather on its absolute value :

$$f(se^{j\phi}) = e^{j\phi} f(s). \tag{3}$$

3.1.2 *Offset Modulation*

For offset modulated systems the quadrature channel is delayed by half a symbol period $T/2$ relative to the inphase channel. In this case the time function $< |z(t)|^2 > -0.1 < |z(t)|^4 >$ in eq. (1) has no spectral line at $f = 1/T$. The timing recovery has therefore to be performed at baseband (see [3]). With

$$z_{BB}(t) = \Re\{z(t)e^{-j\theta}\}. \tag{4}$$

we obtain

$$t_s = -\frac{T}{2\pi} \arg \mathcal{F}\{< z_{BB}^2(t) > -0.1 < z_{BB}^4(t) >\}_{f=1/T}. \tag{5}$$

and

$$\theta = \arg < z(t_s)\Re\{I_0\} - jz(t_s - T/2)\Im\{I_0\} > . \tag{6}$$

Note that t_s depends on θ due to eq. (4) and θ in eq. (6) depends on t_s. For simultaneous solution of eq. (5) and (6), we calculate first with one relatively short Monte-Carlo simulation of 1000-2000 symbols a set of solutions $\theta(t_s)$, $-T/2 < t_s < T/2$ from eq. (6). Furthermore, we split up $z(t) = z_r + jz_i$ into real and imaginary part, introduce $z_{BB} = z_r \cos\theta - z_i \sin\theta$ from eq. (4) into eq. (5) and obtain an expression for t_s dependent on θ. The terms $< \mathcal{F}\{z_r^m(t) z_i^n(t)\}_{f=1/T} >$ for $m + n = 2$ and $m+n = 4$ in this expression are estimated in the same simulation run and we obtain $t_s(\theta)$. Having thus two sets of solutions $\theta(t_s)$ and $t_s(\theta)$ the actual values of θ and t_s are determined iteratively via $\theta^{(0)} = \theta(t_s = 0)$; $t_s^{(1)} = t_s(\theta^{(0)})$; $\theta^{(1)} = \theta(t_s^{(1)})$ and so on. This iteration converges rapidly.

The GMSK signal can be approximated by a linear offset modulated signal (see section 5). The carrier- and timing synchronization can thus be obtained according to eq.s (4)-(6) with the appropriate impulse response for the linear approximation.

4 PDF OF ISI DUE TO LINEAR DISPERSION

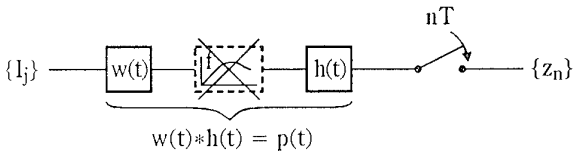


Figure 2: Simplified baseband equivalent model for linear channels

If the actual values of t_s and θ are incorporated into the transmit filter impulse response $w(t)$ we end up with the simplified baseband equivalent system model of Fig. 2. In order to enable an easier understanding of the procedure in the next section 5, a short review of the calculation of the pdf of ISI is presented for linear transmission channels according to [1]. This means the nonlinearity is now deleted in Fig. 2 and we can describe the system by the overall impulse response $p(t) = w(t) * h(t)$. The received signal at time 0 is then obtained by

$$z_0 = \sum_{j=-K}^{K} I_j p_{-j}, \tag{7}$$

if p_j denotes $p(jT)$ and if the overall impulse response $p(t)$ has finite memory $2KT$. Eq. (7) consists of a sum of independent random variables and the pdf of such a sum is calculated by the convolution of the pdf's of the single terms :

$$g_{z_0}(x) = g_{I_{-K}p_K}(x) * \ldots * g_{I_K p_{-K}}(x). \qquad (8)$$

In eq. (8) $g_z(x)$ denotes the pdf of a random variable z and the convolution of the $2K+1$ densities can be performed recursively. The procedure is displayed in Fig. 3 for binary transmission and for real values of the impulse response $\{p_j\}$. We start with a δ-distribution, which is numerically represented by the thin uniform density in Fig. 3a. In the first step we take account of the term $I_{-K}p_K$, which has the value $+p_K$ or $-p_K$, if $I_{-K} = \pm 1$. Its density thus equals $1/2[\delta(x-p_K)+\delta(x+p_K)]$, which means that

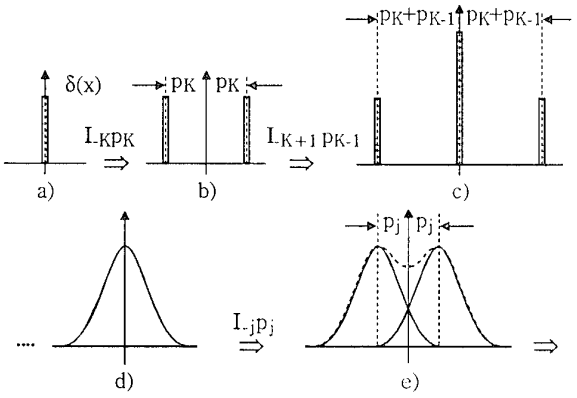


Figure 3: Recursive calculation of the pdf for linear channels

the original pdf is shifted by an amount p_K either to the right or to the left and the shifted densities are superimposed (Fig. 3b). In the next step we take account of the term $I_{-K+1}p_{K-1}$ and shift the second pdf by $\pm p_{K-1}$. In the example Fig. 3c p_K equals p_{K-1} and the shifted densities overlap at the centre. After a number of steps we usually obtain some density, which is continuous with respect to the numerical resolution. For infinite sums in eq.(7) the story becomes more complicated. The different types of distribution functions for this case can be found in [4]. There are rare cases, where the pdf of ISI can be calculated analytically, and these analytic solutions are also continuous functions (see [5], [1]). The continuous density of Fig. 3d is shifted in Fig. 3e by $\pm p_j$ yielding the dashed curve as result of the superposition. In this way we have to perform a simple transformation

$$g^{(j)}(x) = \frac{1}{2}[g^{(j-1)}(x+p_j) + g^{(j-1)}(x-p_j)] \qquad (9)$$

for each interfering symbol, which implies that the computation effort increases only linearly with the number of interfering symbols. Complex values of the impulse response $\{p_j\}$ and multilevel QAM-schemes can be easily included by taking the real part of eq. (7) for the in-phase channel and shifting the densities in eq. 9 by all amounts of the possible levels $\Re\{I_j p_{-j}\}$.

5 INCLUSION OF A MEMORYLESS NONLINEARITY

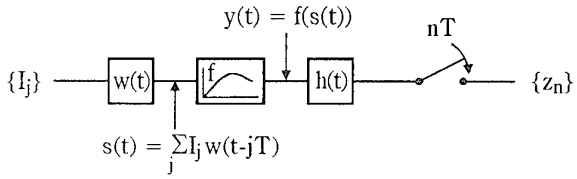


Figure 4: Simplified baseband equivalent model including the nonlinearity

Now the nonlinearity is included into the system model Fig. 4. The signal s at the output of the transmit filter is a linear combination of transmitted symbols. The signal y at the output of the nonlinearity is some nonlinear function of s and the sample z_0 of the received signal is obtained by convolution of the filter h with the signal y at its input:

$$z_0 = \int_{-VT}^{VT} y(t)h(-t)dt. \qquad (10)$$

We assume finite memory of the transmitter and receiver filters. Thus, the integral in eq. (10) extends only over the length $2VT$ of the channel/receiver filter impulse response $h(t)$. We split up the integral into different time sections with the length of a symbol duration and obtain in this way again a sum of random variables (r.v.) δz_m :

$$\begin{aligned} z_0 &= \int_{-VT}^{(-V+1)T} y(t)h(-t)dt + \ldots \\ &\quad \ldots + \int_{(V-1)T}^{VT} y(t)h(-t)dt \\ &= \delta z_{-V}(I_{-V-W+1}, \ldots I_{-V+W}) + \ldots \\ &\quad \ldots + \delta z_{V-1}(I_{V-W}, \ldots I_{V+W-1}) \qquad (11) \end{aligned}$$

Now the single terms δz_m are statistically dependent. Each signal section $y(t)$, $mT \le t \le (m+1)T$, depends on a collection of source symbols $(I_{m-W+1}, \ldots I_{m+W})$, whose number is determined by the memory $2WT$ of the transmit filter $w(t)$, and the next term δz_{m+1} depends on a shifted collection $(I_{m+2-W}, \ldots I_{m+1+W})$ of $2W$ symbols. (For time-symmetric transmit filters with a memory equal to an odd number of symbol durations $(2W+1)T$, we should define a signal section for the interval $(m-1/2)T \le t \le (m+1/2)T$ dependent on $2W+1$ symbols $(I_{m-W}, \ldots I_{m+W})$).
The sequence of tuples of symbols can be represented in the usual trellis diagram. Instead of dealing with the joint probability densities of the δz_m, which is impossible due to the rapid increase of dimensions, we now work with densities conditioned on the trellis states. Let each state in a trellis be defined by a tuple of $2W-1$ symbols. The transition to the following state then defines the following symbol. State and transition, therefore, determine $2W$ source symbols and, thus, a section of the transmit signal $y(t)$ in eq. (11), $mT \le t \le (m+1)T$. In this way each state and transition at stage m in the trellis structure determines the value of a term δz_m in eq. (11). The statistical dependencies between the δz_m are now inherent

in the trellis structure with the possible transitions. If the results are conditioned on a state in stage m then the symbol collection $(I_{m-W+1}, \ldots, I_{m+W-1})$ is fixed and the r.v.'s δz_j right and left of the conditioning trellis state become independent. We therefore define a set of density functions, one for each trellis state. In mathematical notation let $g_j^{(m)}(x)$ define the pdf of $\sum_{l=-V}^{m-1} \delta z_l$ conditioned on the tuple of source symbols $(I_{m-W+1}, \ldots I_{m+W-1})$ which is defined by the trellis state $s_j^{(m)}$, $0 \leq j \leq M^{2W-1} - 1$ for a M-level signal alphabet. If state $s_i^{(m+1)}$ is a successor state at stage $m+1$ of the states $s_{j_0}^{(m)}, \ldots s_{j_{M-1}}^{(m)}$ at stage m, then the updated density $g_i^{(m+1)}(x)$ at state $s_i^{(m+1)}$ of stage $m+1$ is obtained by

$$g_i^{(m+1)}(x) = \sum_{l=0}^{M-1} g_{j_l}^{(m)}\{x - \delta z_m[i, j_l]\}. \qquad (12)$$

Note that the value of δz_m is determined via eq.(11) by the predecessor state $s_{j_l}^{(m)}$ and the successor state $s_i^{(m+1)}$. Normalization of the densities can be performed at the end of the recursion.
In Fig. 5 we want to illustrate the procedure for a very simple trellis where each state defines a pair of binary symbols. The triangular and the uniform

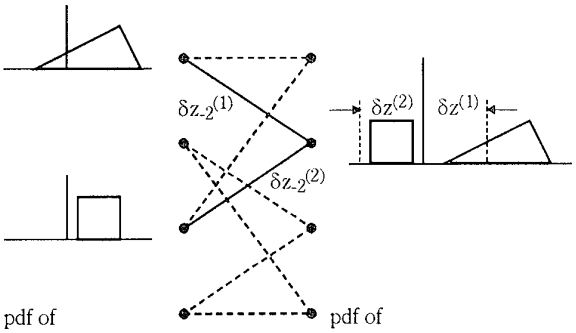


Figure 5: Recursion rule for updating densities in a trellis structure

density at the left side of Fig.5 are the conditional densities of the partial sum of δz's at the bottom of the figure. The pdf's refer to the corresponding trellis states. Each one of the two transitions indicated by a continuous line defines a definite value of the following term δz_{-2}. For the subsequent state we have, therefore, to shift either the triangular density by $\delta z^{(1)}$ or we have to shift the uniform density by $\delta z^{(2)}$. Superposition then yields the updated density at the subsequent state. This recursion deals with a collection of densities which are updated in a trellis structure until all terms δz_j, $j = -V, \ldots V-1$, have been taken into account. At the end of the procedure the pdf's of all states are summed up and yield the unconditioned density of z_0. The number of trellis states, however, depends exponentially on the number of symbols which uniquely determine a section of the transmitted signal. For large memory of the transmit filter we get therefore an enormous number of states and the method becomes impractical. This problem is attacked in the next subsection.

5.1 Decomposition of the Nonlinearity

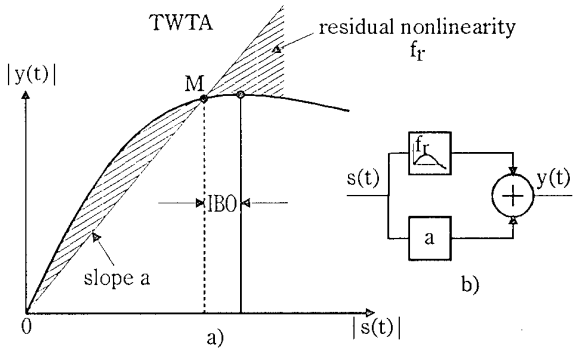


Figure 6: a) Amplitude distortion of TWTA and decomposition of the nonlinear characteristic b) equivalent block diagram

We make a partition of the nonlinear characteristic, which is similar to the perturbation method in electromagnetic field theory. In Fig. 6 a) the fat line shows the amplitude distortion of a Travelling Wave Tube Amplifier (TWTA) according to [2]. The value of the input back-off (IBO) defines the mean point of operation M. We now separate the function f into a linear part given by the slope a of the line $\overline{OM}$ and into the dashed residual nonlinearity f_r. Thus we obtain the equivalent block diagram of Fig. 6 b), which is not an approximation.

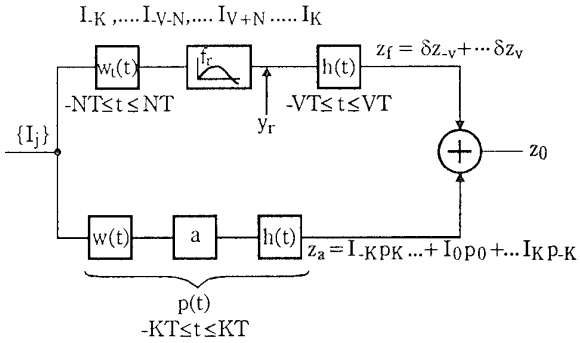


Figure 7: Modified block diagram of the transmission system

For the transmission system the modified block diagram of Fig. 7 is obtained. The lower branch takes account of the linear part of the characteristic and can be described by an overall impulse response $p(t)$. It yields the contribution z_a to the r.v. z_0. The upper branch contains the residual nonlinearity f_r, which is small compared to the total signal. For this small contribution we now truncate the transmit filter $w_t(t)$ in the upper branch:

$$w_t(t) = \begin{cases} w(t) & |t| \leq NT \\ 0 & \text{otherwise} \end{cases} \qquad (13)$$

The number of trellis states is now determined by the memory of the truncated filter w_t and becomes

manageable. The nonlinear upper branch produces the part z_f of z_0. The terms δz_j in Fig. 7 are statistically dependent and they also depend on the inner terms $I_j p_{-j}$, $|j| < N + V$ of z_a. On the top of Fig. 7 we have the total sequence which determines z_0. The outer symbols I_j, $|j| \geq V + N$ pass only the lower linear branch, because the filter w in the lower branch is not truncated. The inner symbols I_j, $|j| < V + N$ pass both branches, and they define a path through the trellis, where each transition determines one of the terms δz_j and as well a corresponding term $I_j p_{-j}$ in the inner part of z_a. The statistical dependencies between the δz_j and the terms $I_j p_{-j}$ have to be taken into account in a hybrid way during the trellis-structured recursion procedure. The density at the starting states at stage $-V$ in the trellis is given by the pdf of $\sum_{|k| \geq V+N} I_k p_{-k}$ which is the ISI caused by the outer symbols at the top of Fig. 7. For these symbols the ISI consists of statistically independent terms and its starting pdf $g_s(x)$ can be calculated as in section 3, [1]. Each of the starting states $s_i^{(-V)}$ at stage $-V$ defines a tuple of symbols $(I_{-V-N+1}, \ldots I_{-V+N-1})$ and therefore the starting ISI

$$z_{s,i} = \sum_{l=-V-N+1}^{-V+N-1} I_l p_{-l}. \qquad (14)$$

So, we have to position the starting pdf's for each state $s_i^{(-V)}$ at the value of $z_{s,i}$:

$$g_i^{(-V)}(x) = g_s(x - z_{s,i}) \qquad (15)$$

The densities $g_i^{(-V)}(x)$ are transformed in the trellis structure according to eq.(12), where δz_{-V} is defined by replacing $y(t)$ by $y_r(t)$ in eq.(11). This means that

$$\delta z_\mu = \int_{\mu T}^{(\mu+1)T} y_r(t) h(-t) dt \qquad (16)$$

now takes account only of the residual nonlinear signal contribution y_r in Fig.7. At the successor states $s_j^{(-V+1)}$ the symbol I_{-V+N} and its linear contribution

$$z_{a,-V+N} = I_{-V+N} p_{V-N} \qquad (17)$$

to the ISI is fixed. For this reason the total transformed densities at the successor states $s_j^{(-V+1)}$ have to be shifted additionally by an amount of $z_{a,-V+N}$

$$g_j^{(-V+1)}(x) := g_j^{(-V+1)}(x - z_{a,-V+N}). \qquad (18)$$

The procedure is explained in the following by a simple example. Fig. 8 holds for a binary system with a memory $20T$ of the overall impulse response $p(t)$, a memory of $10T$ of the channel/receive filter $h(t)$ and -for ease of explanation- a very short truncated transmit filter w_t with memory $2T$. We begin with the linear pdf of the outer symbols I_j, $|j| > 5$ at each of the starting states s_1 and s_2. The states s_1 and s_2 define the symbol I_{-5}, this fixes the interference from this symbol via the sample p_5 of the impulse response. Therefore, the starting pdf is situated at $+p_5$ in the upper state for $I_{-5} = 1$ and at $-p_5$ in the

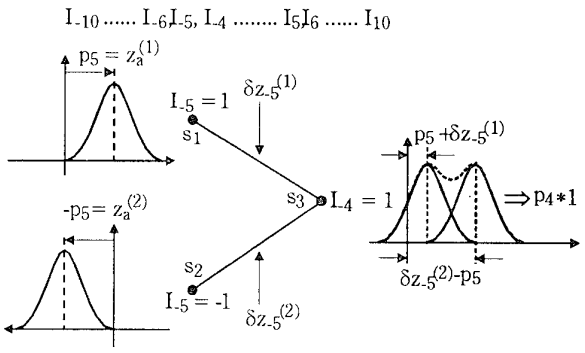


Figure 8: Hybrid transformation in the trellis structure

lower state for $I_{-5} = -1$. The transition to the following state s_3 yields the nonlinear contribution δz_{-5} dependent on the starting and subsequent state :

$$\delta z_{-5}^{(\nu)} = \int_{-5T}^{-4T} y_r(s_\nu, s_3, t) h(-t) dt, \quad \nu = 1, 2, \quad (19)$$

where y_r denotes the transmit signal in the upper nonlinear branch of Fig. 7. Therefore, at state s_3, we have the density of state s_1 shifted by $\delta z_{-5}^{(1)}$ and superimposed the density of state s_2 shifted by $\delta z_{-5}^{(2)}$. Additionally, there is now at state s_3 a definite value of $I_{-4} = 1$, which yields a linear contribution $p_4 * 1$. This means that the total density of state s_3 is to be shifted by p_4. We get again with each step through the trellis an updated collection of densities and having performed all recursion steps up to I_5 we can sum up the densities of all states and obtain the unconditioned pdf of z_0.

6 RESULTS

The following results for 4-QAM and offset 4-QAM (4-OQAM) hold for signals, which are spectrally shaped by a cosine roll-off characteristic (roll-off factor $\alpha = 0.5$) distributed equally on the transmit and receive filter. The transmission channel is modelled by a two-path propagation which introduces ISI. The transfer function of such a channel is given by

$$H_c(f) = 1 - b \exp\{-j2\pi(f - f_N)\tau\}, \qquad (20)$$

where b denotes the relative echo amplitude, the notch frequency f_N defines the phase of the delayed ray and τ is the echo delay relative to the direct path. For the nonlinearity we assume the Saleh-model of a TWTA [2] :

$$f(s) = \frac{2s}{1 + |s|^2} \exp\{j\phi_0 \frac{2|s|^2}{1 + |s|^2}\} \qquad (21)$$

with $\phi_0 = \pi/6$.

If we assume a memory length of 6 symbol durations for the truncated transmit filter w_t , we obtain $4^5 = 1024$ different trellis states and the same number of pdf's in the recursion procedure. The memory length of the channel/receive filter $h(t)$ as well as the memory of the untruncated transmit filter $w(t)$ is not critical and can be taken as 10-20 symbol durations. The densities are numerically represented at $L = 128$ equally spaced discrete points.

6.1 Examples

For illustration we show in Fig. 9 a scatter plot of the received signal for 4-QAM with a TWTA (IBO = 0 dB) transmitted over a two-path propagation channel with the indicated parameters. This figure was calculated by Monte-Carlo simulation with rather few symbols in order to give an impression of the situation. The four clusters of signal samples in each quadrant are due to the echoes of the two-path propagation. All samples right of the ordinate correspond to a transmitted symbol 1 in the inphase channel, the samples left correspond to -1. The pdf of the received signal in the inphase channel is the marginal density of these points dependent on the abscissa. This pdf

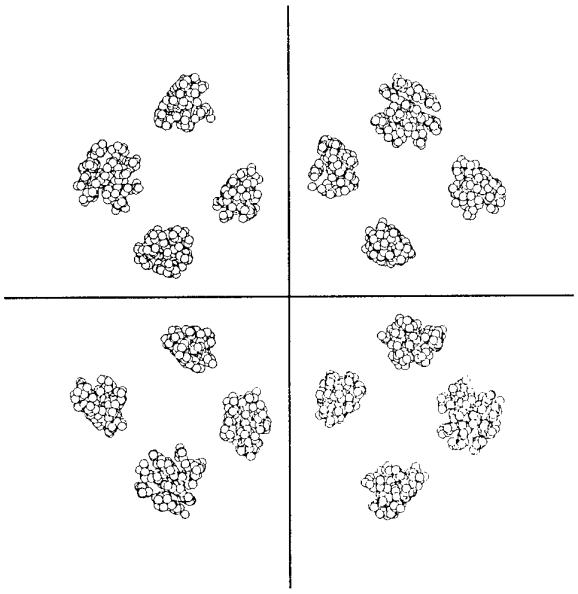


Figure 9: Scatter plot of the received signal for 4-QAM, with a TWTA, IBO = 0 dB, two-path propagation with
$b = 0.9$ relative echo amplitude
$f_n T = 0.1$ notch frequency
$\tau = 0.1\ T$ echo delay

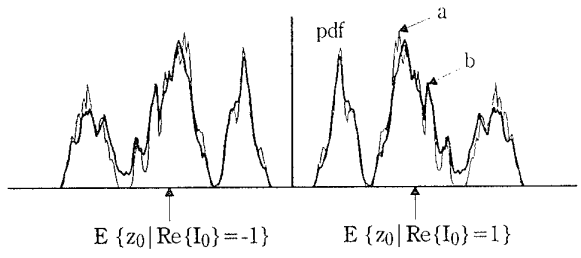


Figure 10: pdf of ISI for the system of Fig. 9
thin line a: brute force Monte-Carlo simulation
fat line b: result of recursive algorithm

is plotted in Fig. 10 with the decision threshold in the centre. The fat line b refers to our recursive algorithm and the thin line a was calculated as a test by brute force Monte-Carlo simulation with 60 000 symbols. The coincidence of both curves is rather good for such a complicated recursion procedure and the recursive algorithm is more accurate than the Monte-Carlo simulation. The saving of computation time by the recursion is about a factor of 10 for comparable accuracy. Having calculated the pdf of ISI we can determine the bit error probability caused by noise and the distortions, if threshold decision is applied. The abscissa in Fig. 11 counts the ratio of energy

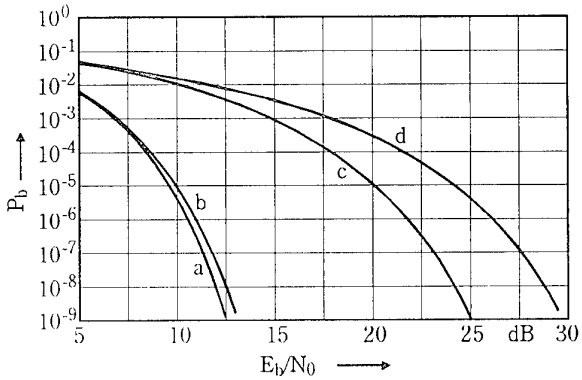


Figure 11: Bit error probability for the 4-QAM system of Fig. 9
a: AWGN only
b: AWGN and TWTA (IBO = 0 dB)
c: AWGN, TWTA and two-path propagation according to Fig. 9
d: AWGN, two-path propagation, no TWTA

per bit E_b to the one-sided spectral density of noise N_0. At the leftmost curve a in Fig. 11 the system is corrupted only by additive white Gaussian noise. This means no nonlinear distortion and no ISI due to the cosine roll-off filtering. The second curve b shows

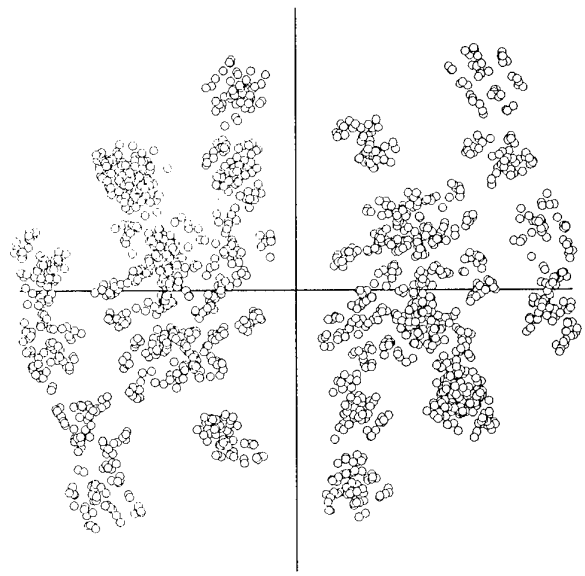


Figure 12: Scatter plot of the received signal for 4-OQAM
roll-off factor $\alpha = 0.5$
two path propagation as in Fig. 9
TWTA with IBO = 0 dB

a slight impairment if a TWTA with IBO = 0 dB is included. The rightmost curve d holds for the dispersive two-path channel without a nonlinearity, and the curve c yields a very interesting result: The nonlinearity reduces the error ratio which is caused by the channel dispersion. This effect is already known from soliton theory [6]: a nonlinearity can diminish the dispersion which is caused by a linear channel.
For digital transmission systems sometimes the offset modulation scheme is favored because of its reduced variation of the signal envelope. This should lead to smaller distortions by nonlinearities. In order to test this statement, we show corresponding results for offset 4-QAM (4-OQAM) revealing that the statement is not true in all respects.
Fig. 12 is the scatter plot for the 4-OQAM system at the sampling instants of the inphase channel. For Fig. 12 the same channel dispersion and the same nonlinearity as in Fig. 9 was assumed. The marginal density of the real part of the signal is shown in Fig. 13, there is again a very good coincidence between the proposed algorithm (fat line b) and the Monte-Carlo simulation (thin line a).

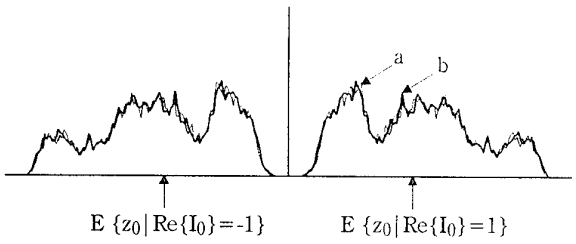


Figure 13: pdf of received signal for 4-OQAM, TWTA with IBO = 0 dB, two path propagation as in Fig. 9
thin line a: brute force Monte-Carlo simulation
fat line b: recursive algorithm

The resulting error probabilities are plotted in Fig. 14. The left curve (a) without channel dispersion and without a nonlinearity is the same as in Fig. 11 for 4-QAM. The next curve (b) shows an increased impairment due to the nonlinear distortion as compared to the non-offset system. At the worst line (d) for the frequency selective channel there is a severe degradation relative to the corresponding curve in Fig. 11. It is well known that the offset-modulation leads to an increased interference from the quadrature channel if dispersion is present. At curve (d) the peak distortion has exceeded the decision threshold because of the latter effect and this causes a residual error floor for large E_b/N_0. The reduction of the error ratio by the nonlinearity at curve (c) seems now to be dramatic. But nevertheless this result for offset modulation with TWTA and two-path propagation remains 2-3 dB worse than the corresponding curve in Fig. 11 for non-offset modulation. The above mentioned argument for advantages of nonlinearly distorted OQAM, which was based on the reduced envelope variation of the signal, does thus not hold as far as the error probability in the nominal frequency band is considered. But besides this ISI the nonlinearity also causes a spectral widening of the signal

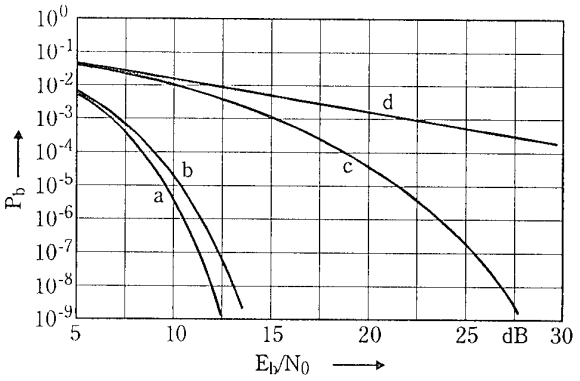


Figure 14: bit error probability for 4-OQAM according to Fig. 14
a: AWGN only
b: AWGN and TWTA (IBO = 0 dB)
c: AWGN, TWTA and two-path propagation according to Fig. 9
d: AWGN, two-path propagation, no TWTA

and this spectral widening leads to Adjacent Channel Interference (ACI). For offset modulation the ACI is reduced at the expense of increased ISI.

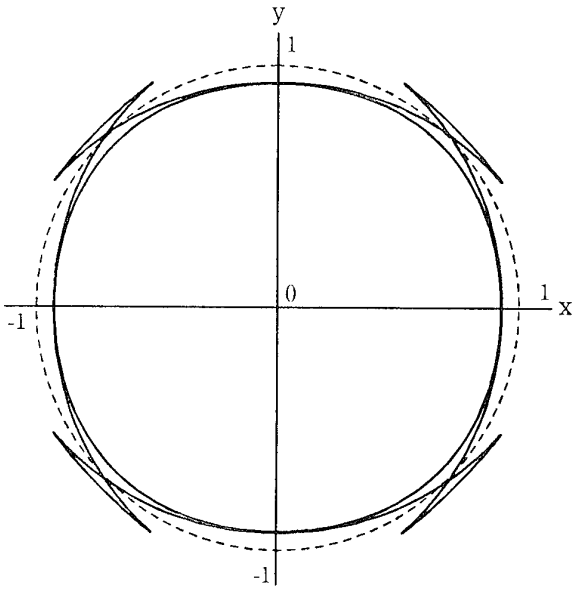


Figure 15: Envelope of the linearized GMSK signal

The algorithm can also be applied for the GMSK-modulation scheme used in the European digital mobile GSM network. The exact GMSK-modulation possesses a constant envelope so that non-linear amplitude distortions do not degrade the performance. Some implementations, however, are based on the linear approximation where slight variations of the envelope are observed (Fig. 15). The linearized version of the GMSK signal $s(t)$ is given by [7]

$$s_{GMSK}(t) = \sum_{i=-\infty}^{\infty} \exp\{j\frac{\pi}{2}\sum_{l=-\infty}^{i} d(l)\}c_0(t-iT) \qquad (22)$$

with $d(l) \in \{+1, -1\}$ and
$c_0(t)$: impulse response as in [7].

This modulation scheme doesn't fullfill the first Nyquist criterion so that inherently a certain amount of ISI is present also in the case of an AWGN channel as shown in the eye diagram in Fig. 16.

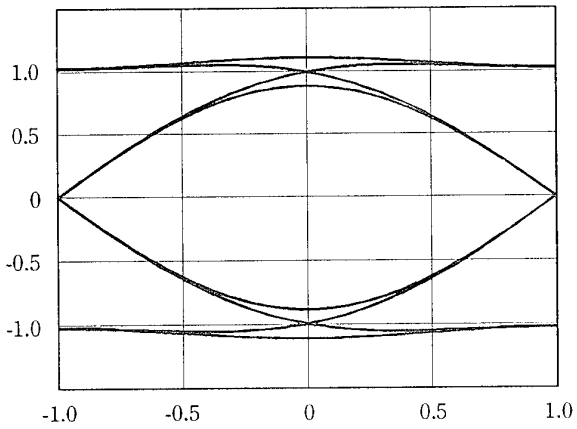


Figure 16: Eye diagram for the linearized GMSK scheme

The linearized GMSK modulation can be interpreted as offset QAM resulting in a block diagram as shown in Fig. 17.

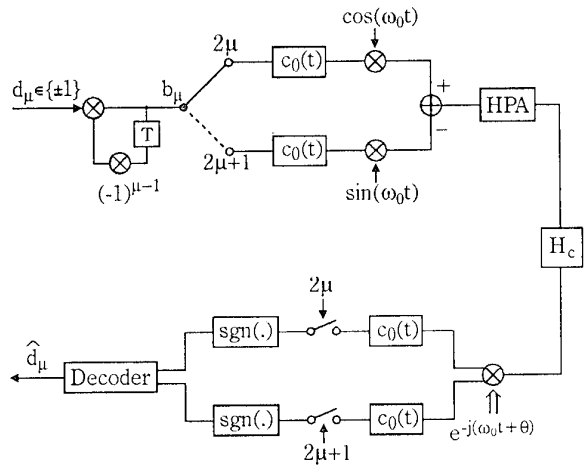


Figure 17: Block diagram of the linearized GMSK scheme

Thus, the pdf of ISI and the BER can be calculated using the results for timing- and carrier-phase offset described in Sec. 3.1. The performance of the GMSK system is depicted in Fig. 18 for the same channel conditions and non-linearities as for the QAM modulation. Again it can be observed that the nonlinearity improves the perfomance under frequency-selective fading conditions.

CONCLUSION

A recursive algorithm for the determination of the pdf of ISI after nonlinear distortions and linear channel dispersion was described. In this paper the feasibility of the method for system analysis was shown.

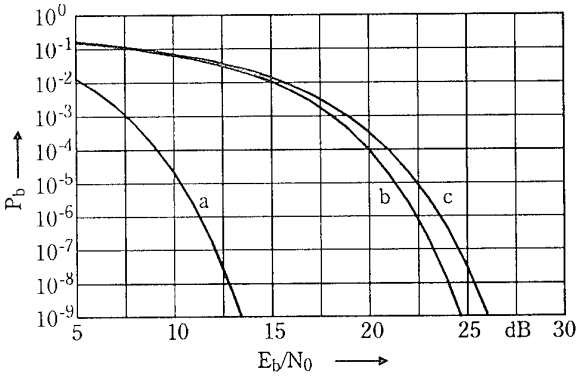


Figure 18: BER versus E_b/N_0 for the linearized GMSK scheme
a: AWGN channel only
b: AWGN, TWTA and two-path propagation, according to Fig. 9
c: AWGN, two-path propagation, no TWTA

The main results are the occasional improvement of channel dispersion by the characteristic of a HPA and the increase of nonlinear inband distortion with offset modulation. The linearized approximation of GMSK can also be analyzed with nonlinear distortion and linear channel dispersion. Similar as for the QAM the nonlinearity counteracts the effect of linear channel dispersion.

References

[1] Metzger, K., "On the Probability Density of Intersymbol Interference", IEEE Trans. Commun., vol. COM-35, pp.396-402, April 1987.

[2] Saleh, A.A.M., "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers", IEEE Trans. Commun., vol. COM-29, pp. 1715-1720, Nov. 1981.

[3] Metzger, K., Valentin, R.,"An Analysis of the Sensitivity of Digital Modulation Techniques to Frequency-Selective Fading", IEEE Trans. Commun., vol. COM-33, pp. 986-992, Sept. 1985.

[4] Wittke, P.H., Smith, W.S., Campbell, L.L.,"Infinite Series of Interference Variables with Cantor-Type Distributions", IEEE Trans. Inform. Theory, vol. IT-34,pp. 1428-1436, Nov. 1988.

[5] Hill, F.S., Blanco, M.A.,"Random geometric series and intersymbol interference", IEEE Trans. Inform. Theory, vol. IT-19, pp. 326-335, May 1973.

[6] Govind, P., Agrawal, E.,"Nonlinear fiber optics", Academic Press, San Diego, 1989

[7] Laurent, P.A., "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses (AMP)",IEEE Trans. Commun., Vol. COM-34, pp. 150-160, 1986.

Integrated Data and Channel Estimation for Selective Fading Channels

Luca D'AMBROSIO, Rossano MARCHESANI and Marina RUGGIERI

Rossano MARCHESANI
DED/STAS, ALCATEL-TELSPACE, 5 rue Noel Pons, 92734 Nanterre, France

Luca D'AMBROSIO, Marina RUGGIERI
Università di L'Aquila, Dipartimento di Ingegneria Elettrica, 67040 Poggio di Roio - L'Aquila, Italy

ABSTRACT Per-Survivor Processing is a general approach including in the survivors of the Viterbi Algorithm trellis the relative estimation of unknown parameters; this expensive method better approximates the optimum decoder in some conditions. The method is applied to the case of a typical HF channel and performance is evaluated. Some reduced complexity implementations are also evaluated and relative performances reported.

1. INTRODUCTION

The Viterbi algorithm (VA) is the well known optimum decoder in the presence of both Gaussian noise and Intersymbol Interference (ISI) [1]. To operate the VA requires the knowledge of the overall channel impulse response (CIR), which is usually computed using an adaptive estimator driven by either least squares (LS) or least mean squares (LMS) algorithms [2][3][4].

When multipath and fast fading rates are present, the optimum solution, to approach the theoretical performance, is the introduction in the VA trellis of the additional estimation of unknown parameters, of course resulting in a computational heaviness [5][6][7].

This method, per-survivor processing (PSP), has been independently introduced, recently, by several authors in various application fields [5][6][8]. In this paper, the PSP method is applied to a data transmission over a typical HF channel, affected by both multipath and Doppler spread.

A per-survivor channel estimator and linear equalizer are introduced in the VA trellis, where the selection of the least cost data sequence also assures the selection of the associated equalizer. That increases the complexity related with the single survivor, but considerably reduces possibility of divergence of the algorithm, in the presence of fast fading. The complexity of the proposed method is evident, as well as the difficulty of a practical implementation.

First the length of the required trellis is evaluated, then some simplifications of the method are examinated, that allow to keep its intrinsical robustness although reducing the computational effort.

A possible symplification consists in a reduced state sequence estimation (RSSE), a suboptimum decoder, keeping only the most probable survivors on the basis of the least distance; this solution is effective when the CIR is minimum phase or properly equalized [2]. This approach leads to the use of a trellis with a reduced dimension state and is already used for standard VA decoders.

The channel used in all simulations consists of two independent Rayleigh fading paths, with same average magnitude, 1Hz Doppler spread each and 1ms relative delay. The signal consists of i.i.d. data, with periodic retraining, transmitted over a 3kHz voiceband channel, using 8PSK modulation scheme and 2400 symbols per second. At the receiver the signal is supposed correctly sampled at symbol rate. Both the transmitter and the receiver are equipped with a raised cosine filter.

The structure of the PSP decoder is explained in section 2; section 3 describes the complexity of the algorithm as a function of the decoding delay and of possible symplification; simulation results are in section 4 and, finally, in section 5 conclusions and practical implications are considered.

2. THE PSP DECODER

The transmitted signal consists of i.i.d. data, with periodic retraining, passed through a typical HF channel, both the transmitter and the receiver are equipped with shaping filters. During the training

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

sequence, the received signal is sampled at a frequency higher than symbol rate, then the optimum sampling instant is chosen and is supposed constant all over the data frame.

The starting point, to understand the PSP decoder, is the well known decision feedback equalizer (DFE), whose classical structure is reported in figure 1. It consists in a linear equalizer G=$\{g_q\}$, a decision device and a feedback equalizer F=$\{f_p\}$, with q=0,...L1 and p=0,...L2. The equalizer G compensates the ISI of the noncausal (before the synchronisation point) part of the CIR, while the equalizer F is an estimator of the causal (after the synchronization point) part of the CIR.

The complex envelope of the received signal r(n) at the input of the decoder is given by:

$$r(n) = \sum_{p=0}^{L} h_p(n)\, x_{n-p} + w_n , \qquad (1)$$

where $\{w_n\}$ represents white Gaussian noise with variance σ_w^2, independent of data, $\{x_n\}$ is the 8PSK symbol sequence and H(n)=$\{h_p(n)\}$ is the time varying channel impulse response.

At time n, the signal y(n) at the input of the decision device is:

$$y(n) = \sum_{p=0}^{L1} g_p(n-1) r(n+p) - \sum_{q=1}^{L2} x_{n-q} f_q(n-1). \qquad (2)$$

Eight possible values of x_n are inserted in the equation (2) and eight corresponding values $\varepsilon_i(n)$ of error are obtained:

$$\varepsilon_i(n) = y(n) - x_{n,i} f_0(n-1). \qquad (3)$$

The value of $x_{n,i}$ originating the least value of $\varepsilon_i(n)$ is retained and the equalizers G(n) and F(n) are updated, according to the classical LMS equations:

$$G(n) = G(n-1) + \mu\varepsilon(n) R^*(n), \qquad (4)$$

$$F(n) = F(n-1) + \mu\varepsilon(n) X^*(n), \qquad (5)$$

where R(n)=$\{r(n),...r(n+L1)\}$ and X(n)=$\{x_n,...x_{n-L2}\}$ are the vectors of received and decoded signal stored in the taps delay lines of the linear equalizer and feedback equalizer respectively.

The first step towards PSP consists in delaying the final decision for x(n) and keeping eight different DFE structures, one for each possible value of $x_i(n)$. The new structure is illustrated in figure 2.

At time n+1 eight possible values of $x_i(n+1)$ are inserted in each independent branch of the structure in figure 1, originating in 8^2 possible values of total error $\varepsilon^2_{Tij}(n+1)$ at time n+1:

$$\varepsilon^2_{Tij}(n+1) = \varepsilon^2_j(n+1) + \varepsilon^2_i(n). \qquad (6)$$

At this point the value $x_i(n)$ and the relative DFE structure are retained, corresponding to the least value of $\varepsilon^2_T(n+1)$. The value of $x_i(n)$ is fed back and 8 new branches of the structure in figure 2 are built, depending on the 8 possible values of $x_i(n+1)$. Then the decoding procedure evolves with a new sample.

Instead of decoding x(n) at time n+1, it is possible to further delay the decision by keeping 64 different branches, each corresponding to all possible values of $\varepsilon_i(n)$ and $\varepsilon_i(n+1)$. The resulting structure is illustrated in figure 3.

At next step, 8^3 possible values of the total error $\varepsilon^2_{Tijk}(n+2)$ exist, in correspondence of each possible decoded triplet $x_i(n)$, $x_j(n+1)$, $x_k(n+2)$:

$$\varepsilon^2_{Tijk} = \varepsilon^2_k(n+2) + \varepsilon^2_j(n+1) + \varepsilon^2_i(n). \qquad (7)$$

Again the possibility exists of either decoding x(n) or further delaying the decision, still increasing of a factor 8 the number of branches in the decoder.

When all the taps in the feedback equalizer F are occupied by a possible unresolved value of symbol $x_i(n,...\ n+L2)$, it is possible to further delaying the decision and build a trellis, like in the case of a standard VA decoder. The difference lays in the fact that at each branch, survivor of the trellis, is associated a different couple of equalizers G(n) and F(n), which is generated by a different evolution of possible transmitted data.

The structure of the trellis, for the case L2=3, is illustrated in figure 4 where, for clearity, a 2PSK modulation scheme has been assumed.

3. COMPLEXITY

Performances of the method have been evaluated for different values of the decoding delay: 1, 2 and 3 samples. Then a Viterbi-like trellis has been built of length 3*L2, also corresponding to the decoding delay.

It is assumed that the main cost is associated to the multiplications, therefore the computational complexity will be expressed in terms of complex multiplications. This is not quite true, expecially if digital signal processing microprocessors are used, because, in most cases, an addition requires the same

computing time than a multiplication. Nonetheless the number of complex multiplications gives a good idea of complexity and a linear correction factor could be added to include additions cost.

In the case of a DFE, the computational complexity to calculate y(n) can be estimated, from equation (2), in L1+1+L2 complex multiplications. From equation (3), we have to add M (where M is the number of symbols in the modulation scheme, 8 in the case of 8PSK) multiplications to calculate $\varepsilon(n)$. From equations (4) and (5), we get L1+1 complex multiplications to update G(n) and L2+1 complex multiplications to update F(n). The total DFE cost, in terms of complex multiplications for each decoded symbol, is:

$$C_{DFE} = 2 * L1 + 2 * L2 + M + 3. \qquad \textbf{(8)}$$

When a decoding delay of one symbol is introduced, like in the case of figure 2, M independent DFEs are required and the resulting cost is $M*C_{DFE}$. In the same way we can say that the cost of the structure in figure 3 is M^2*C_{DFE}, and so on.

In the case of the trellis in figure 4, the number of branches is M^{L2}, requiring M^{L2} different DFEs to be run in parallel. The resulting cost is approximately $M^{L2}*C_{DFE}$.

A solution adopted to reduce the complexity of the Viterbi algorithm consists in keeping only a reduced number of survivors, the most probable ones [2]. This reduced state sequence estimation is effective in reducing the complexity of the algorithm, without degrading performance, when the channel impulse response presents minimum or almost minimum phase.

A reduced state sequence estimation has also been attempted in the case of the per-survivor processing decoder. At each step of the trellis only the 64 survivors are kept which correspond to the 64 least values of the total cost. The resulting computational complexity is $64*C_{DFE}$.

4. SIMULATION RESULTS

Simulations have been carried out considering 8PSK transmission at 2400 symbols/s; both transmitter and receiver models are already equipped with a fixed shaping filter, roll-off = 0.2. Data are transmitted in frames of 256 symbols, of which the first 80 are known data and the following 176 are information data. Each transmission consists of a continuous flux of 40 frames.

The channel utilized in simulations consists of two independent Rayleigh fading paths, with equal average amplitude, 1 ms relative time delay and 1 Hz Doppler spread each one. The time dispersion of the considered channel is over 4 symbols, but, due to the mismatching of the Nyquist filters, the overall sampled channel impulse response has a significant magnitude over 9-10 symbols.

The average channnel impulse response for each data frame, presents an equal distribution between minimum phase and non-minimum phase. Even if this condition is indifferent for the complete PSP decoder, all the other decoders either have degraded performance, or completely fail to operate with non-minimum phase CIR. Therefore, to obtain comparable results, an analysys of the CIR is made in the known data packets at the beginning of each frame and at the end (the beginning of the next frame), then the frames are decoded backward in the case of average non-minimum phase.

Figure 5 reports the symbol error rate at different SNR values of four different decoders: a classical DFE, curve n.1, and the resulting structures when delaying the decision of 1, 2 and 3 symbols, curves n.2, n.3 and n.4, respectively.

Figure 6 reports the symbol error rate at different SNR values of three different decoders. The curve n.1 has been obtained using a classical VA decoder with 8^4 states, where a first step decision is used for channel estimate. The curve n.2 has been obtained by delaying the decision of the DFE of 3 symbols, the same decoder of curve n.4 in figure 5. The curve n.3 has been obtained with a full PSP decoder with a decoding delay of 12 symbols (3*L2) and a trellis with 512 states.

The results obtained with the simplified decoder with trellis length 12 and only 64 survivors are not reported, but they were practically coincidents with the results obtained with the complete PSP decoder and 512 survivors.

Simulations marked the most important improvement of the PSP method with respect to the standard VA decoder: sometimes the standard VA decoder falls in an unrecoverable error burst. Due to the contemporaneous presence of errors and a fast time varying channel, sometimes the adaptive equalizers, G and F, fail. As a consequence, the CIR is completely loss and un error burst is originated in the decoder. That never happened in the PSP decoder. As an example, figure 7 and figure 8 report the error locations of the same data packet, respectively for the systems of curves n.1 (classical VA decoder) and n.3 (complete PSP decoder) in figure 6. The classical system in figure 7 shows an evident failure of the channel estimate algorithm, while the PSP decoder is intrinsically able to recover the estimation capability, by this reason the PSP approach has also been suggested for blind equalization [5].

It is worth noting that the two simulations of figure 7 and figure 8 are relative to the same data frame, with the same value of SNR. The presence of error burst in the classical decoder is at the origin of the slight difference between the curves n.1 and n. 3, in figure 6. This important difference in behavior does not originates a large difference in the graphs of the symbol error rate because of the data frame structure, that allows a periodic retraining of the estimators. The analisis of the channel impulse response showed that the error bursts happened when the channel was more rapidly varying.

5. CONCLUSIONS

In the performed work, the application of PSP to the case of a typical HF channel, in the presence of both fading and multipath, has been evaluated. The computational cost, as well as some symplification, have been analyzed.

Although, at present, the complexity and heavyness of the proposed method make its implementation quite difficult, in the case of high order modulation schemes, some solutions could be examinated in the case of 2PSK modulation schemes.

Nonetheless, the capability to operate in a selective fading environment and the intrinsical robustness make the use of PSP attractive when a reliable communication is required, especially when more powerful processors will be available. The possibility to operate with long unknown data sequences could probably allow, after further investigations, to reduce the presence of known data probes in the present standard waveforms and increase the useful bit rates.

REFERENCES

[1] G.D.Forney Jr., "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference", *IEEE Trans.Inform.Theory*, vol.IT-18, pp.363-378, May 1972.

[2] A.P.Clark, *Adaptive Detectors for Digital Modems*, London, England: Pentech Press, 1989.

[3] S.J.Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1991.

[4] N.Benvenuto, G.Lubello and R.Marchesani, "Channel Estimators in Time Varying Multipath Environments", *proc.6th Int.Conf.on HF Radio Systems and Techniques*, York, England, July 1994.

[5] R.A.Iltis, "A Bayesian Maximum-Likelihood Sequence Estimation Algorithm for a priori Unknown Channels and Symbol Timing", *IEEE Journal Sel.Areas in Commun.*, vol.SAC-10, pp.579-588, Apr. 1992.

[6] R.Raheli, A.Polydoros and C.K.Tzou, "Per-Survivor Processing: a General Approach to MLSE in Uncertain Environments", *submitted to the IEEE Trans.on Commun.*, June 1992.

[7] D.Williamson, R.A.Kennedy and G.W.Pulford, "Block Decision Feedback Equalization", *IEEE Trans.on Commun.*, vol.COM-40, pp.255-264, Feb. 1992.

[8] A.N.D'Andrea, U.Mengali and G.M.Vitetta, "Approximate ML Decoding of Coded PSK with no Explicit Carrier Phase Reference, *submitted to the IEEE Trans.on Commun.*, Nov. 1991.

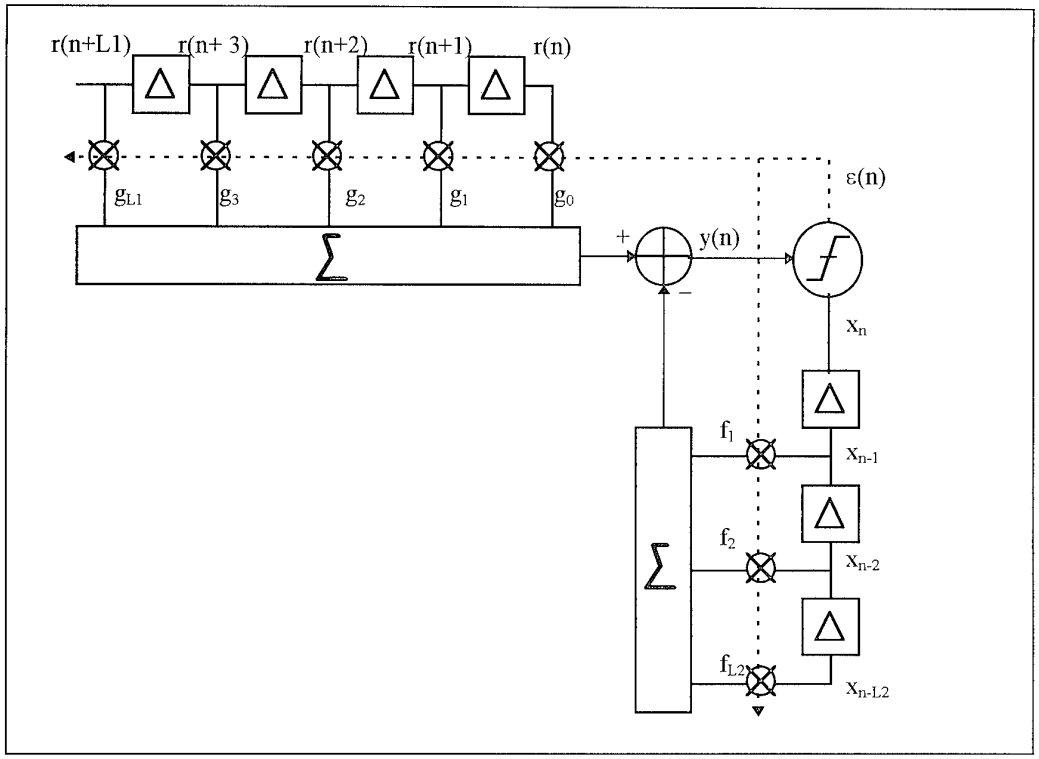


Figure 1 - Classical DFE structure.

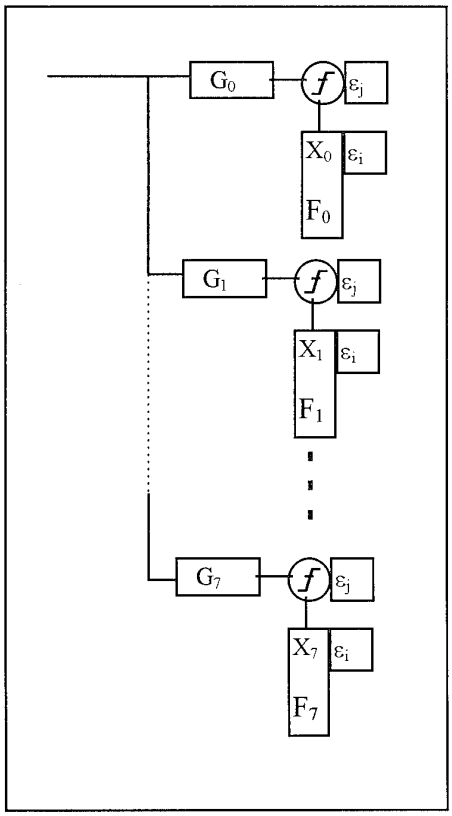


Figure 2 - DFE with 1 symbol decision delay.

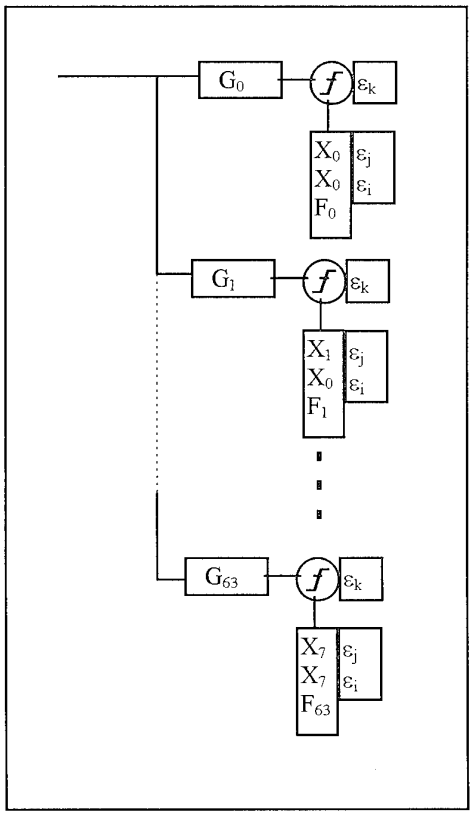


Figure 3 - DFE with 2 symbols decision delay.

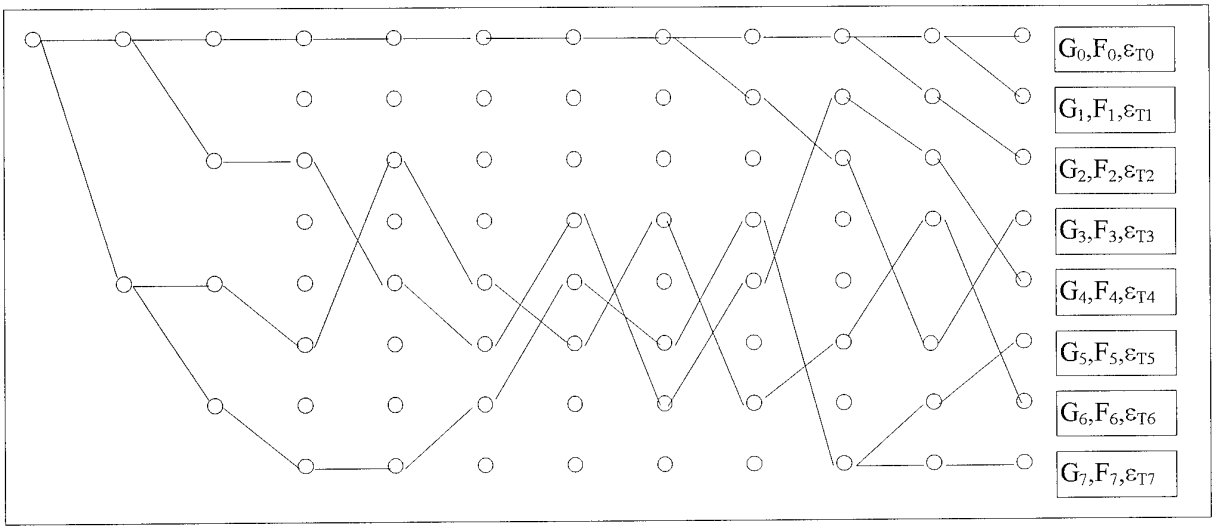


Figure 4 - PSP decoder with trellis length=12

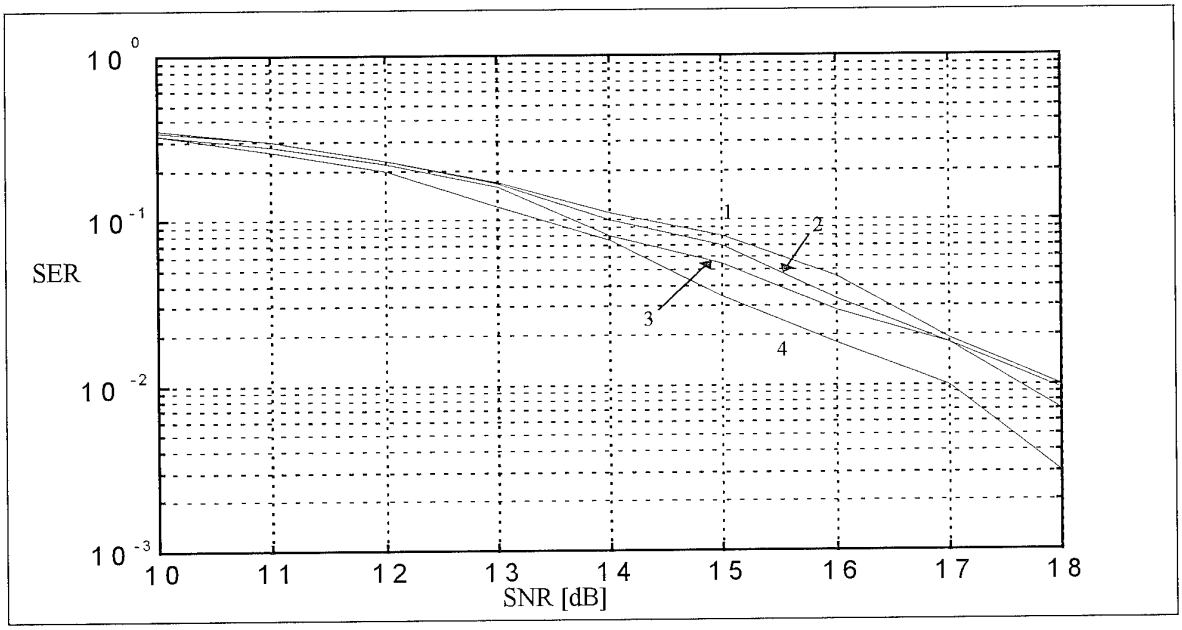


Figure 5

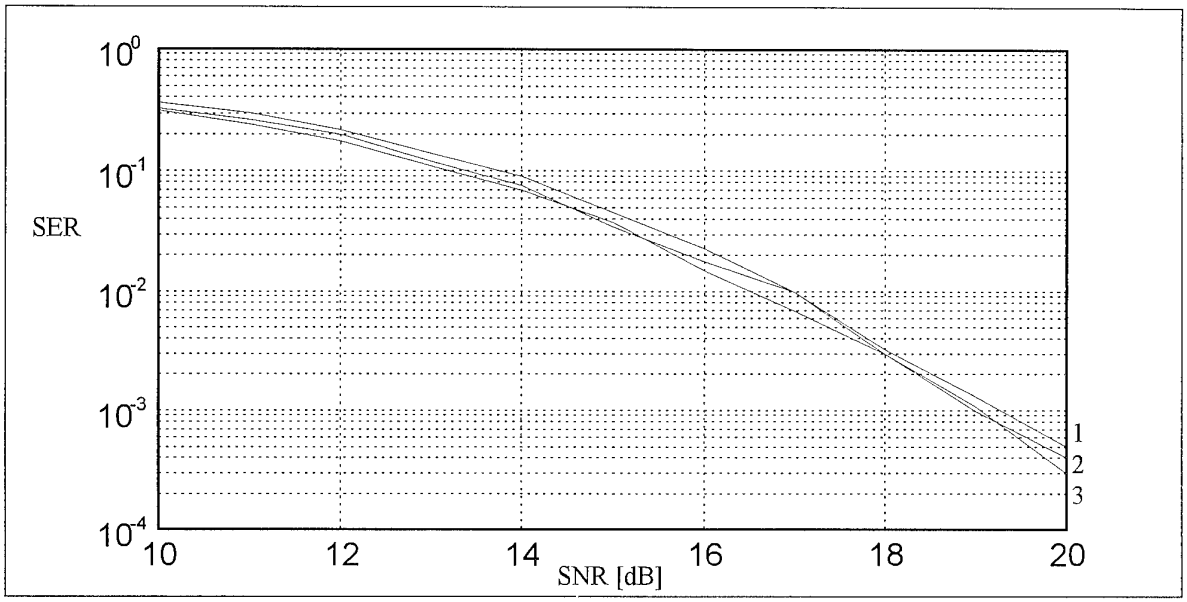


Figure 6

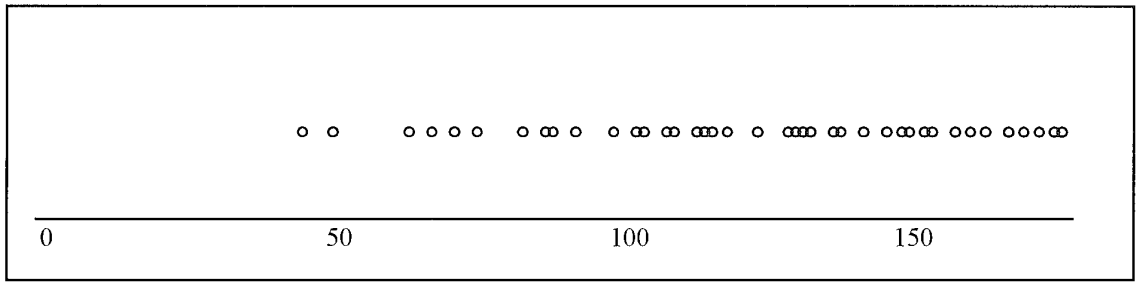


Figure 7 - Error locations for classical VA decoder

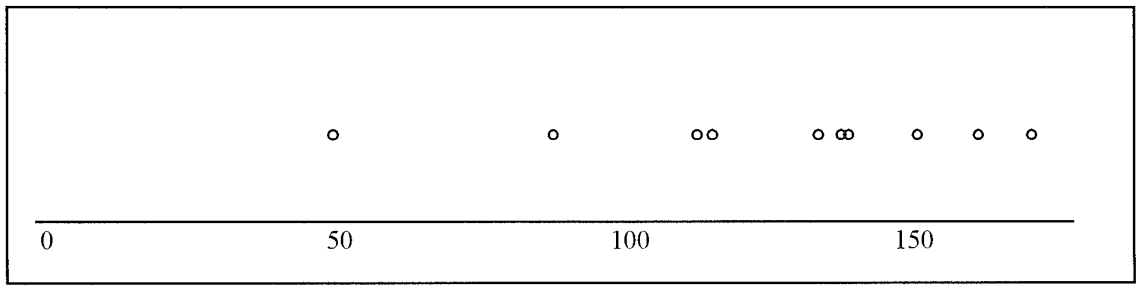


Figure 8 - Error locations for PSP decoder

Link Level Protocols for Radio Channels

Dr. Davras Yavuz STC/COMD

POB 171, Den Haag 2501 CD, The Netherlands
Tel: 31 70 314 2422 Internet: yavuz@stc.nato.int

Abstract: *Describing and/or modelling the characteristics of transmission channels is the most important aspect of performance estimation for most communication systems. Many analytic techniques for determining performance under well defined conditions such as additive Gaussian noise, Poisson impulsive noise, exist. However, some channels are too complex in terms of noise effects, interference and time variability to model with fidelity. The most outstanding example of this is the HF sky wave channel for which "some good channel simulator performance" is required prior to fielding of a system. However, in general this is only a necessary but not sufficient condition for acceptable on-air performance.*

Link level or layer two protocols provide an excellent means to reduce the effects of such uncertainties. The basic approach is to use packetization (packet radio) to allow an automatic request for repeat (ARQ) process to continuously trade-off throughput with channel conditions. Many other facilities such as link establishment/maintenance and adaptive modem speed control can also be incorporated. With such an approach the source information (data, message, fax, image and packet digital voice) will always be communicated to the destination with no errors but the transmission (or delivery) time will vary with channel conditions; the worse the channel the longer it will take to transmit the information. Packet radio techniques were originally developed by DARPA for networking and survivable multi-homing applications over terrestrial/fixed links. However, the link level applications over complex channels are now increasing significantly.

This paper/presentation will present examples of link level protocol work that has been performed at SHAPE Technical Centre. Specifically, link level ARQ protocols developed for HF sky wave and, meteor burst communications will be covered. More recent work on a packet radio protocol for data communications over analog frequency hopping radio systems will also be described.

Introduction: Channel Characterisation

Characterisation of radio propagation channels is a complex, many faceted topic with a history as old the first radio transmissions more than a century ago. In the initial stages of development the estimation of the performance and ultimately the quality of actual on-the-air service provided by a new and/or improved radio system is very important. This is highly dependent on our understanding of the details of the radio channel over which the system is to operate. There is however a trade-off between the level of detail required for modelling the channel and the resulting improvements of predicting on-air quality of service. It is possible to argue that chaotic effects due to the non-linearities in the propagation processes, particularly for non-LOS links, involve "butterfly effects" that preclude accurate estimation **[1]** after a certain level of detail has been reached in the modelling process.

There are of course various different types of radio systems; line-of-sight (LOS) (including satellite communications), scatter, ionospheric reflection/scatter, meteor trail reflection, surface wave over land and sea, passive man-made and/or natural reflectors, etc. For example, LOS channels are the most straight forward to model especially for non-mobile users. Scatter and ionospheric reflection channels provide inherent complexities/non-linearities due to the basic mechanisms by which signals are propagated over long distances even for non-mobile users. In particular, the HF sky-wave channel is the most complex as it depends on electromagnetic wave - ionospheric plasma interaction. As a result it is strongly influenced by diurnal, seasonal, solar cycle effects on the upper atmosphere (ionosphere). The meteor burst (also called meteor scatter) channel results from reflections from plasma trails formed by disintegrating microscopic cosmic dust particles as they enter the upper atmosphere. This channel behaves like a reasonably good passive reflector LOS channel during the random short intervals that it is available. The duty cycle of these intervals is of the order of few percent with average durations few hundred msec. A great deal of statistical information is available which permits prediction of link performance **[2]**.

It is possible to approximately model such radio channels with additive Gaussian noise, multiplicative noise (Eg. Rayleigh fading), Poisson impulsive noise and various other random/stochastic techniques. This paper is not intended to cover these approaches and many references **[3]** may be found on these analytic approaches. The intention is to give a system level overview of this large area from the viewpoint of link level protocols which employ packetisation with automatic repeat request (ARQ) techniques that permit an effective exchange of throughput with channel conditions. In other words, The resultant data communications is always error free but the duration of transmission will depend on channel conditions, **Fig. 1**.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

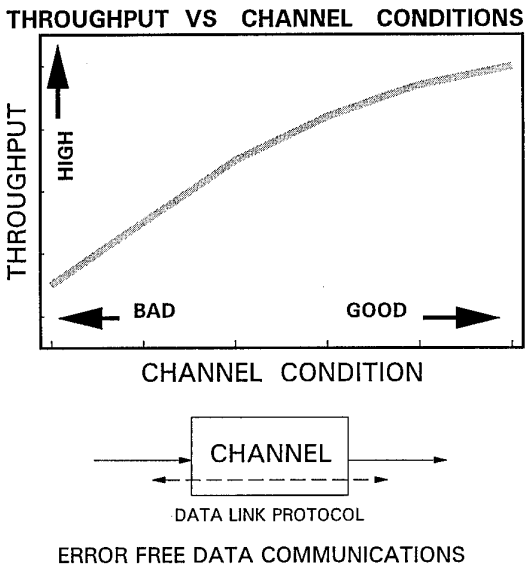


Figure 1. Generic throughput vs channel condition trade-off relationship with link level protocols. A smooth exchange of throughput with channel conditions takes place with error free data communications.

Data Link Control Layer

The Open Systems Interconnection Reference Model (OSI-RM) specifies 7 layers from physical (1), through link (2), network (3) to application (7) **[4]**. The basic function of the Layer 2 (=Link Layer or the Data-Link Control Layer) is the arrangement of the data bits into frames and their management **[5]**. Historically the most common protocol is the ISO High-Level Data Link Control Procedures (HDLC, ISO 3309).

In general, the link layer performs the following functions:

a) Establishes and releases one or more link connections

b) Assembles, keeps track of and exchanges data-link service data units (DLSDUs) or frames

c) Identifies end points of management and data frames

d) Properly sequences DLSDUs/frames as required

e) Controls data flow including requesting repeats (ARQ) and/or repeating relevant frames when errors are encountered

f) Notifies the network layer when errors are detected

g) Selects optional quality of service (QoS) parameters. These may be passed onto the physical layer and network layer

Since the advent of packetized data communications in the late sixties through pioneering studies by DARPA, much work has been done on data link protocols. The most well known data link/network protocols is the X.25 which has been and is still used for data communications over fixed terrestrial links/networks. The US Amateur Radio community must be credited with the definition/implementation of the first open packet radio protocol, the AX.25 **[6]**. This protocol was primarily intended for LOS links, including SATCOM, but could also be used for non-LOS links. It is in use since early eighties and is still the major **open** (=nonproprietory) protocol for data communications over radio links/networks.

HF Packet Radio (PR) Protocols

HF data communications has unfortunately been one engulfed by proprietary standards. To contribute to the opening of the HF data communications, in 1991 STC defined an published an open HF packet radio (PR) protocol which has since been implemented in various forms by government and industrial organisations. More recently the US has also promulgated a new protocol, FED STD 1052p **[7]** with exactly the same aims.

AX.25 suffered from a number deficiencies on non-benign channels such as HF :

a) The carrier sense multiple access (CSMA) scheme used suffers from collisions and also introduces unnecessary delays

b) Go-back-N ARQ scheme used performs badly on poor channels such as HF

c) Flow control and address overheads are high

d) Has no provision for automatic adaptivity (Eg. modem speed control) to varying channel conditions

The STC protocol **[8]** removes these deficiencies:

a) Inherent collision-avoidance and collision recovery mechanisms

b) Selective ARQ instead of go-back-N

c) Single byte connection reference number

d) Adaptive modem speed control supported

These provide on-the-air throughput increases of typically four to five times (same single-tone modem for both protocols). In particular, the STC HF PR protocol is designed to work with the new generation of adaptively equalised single-tone modems **[9]**. The comparison of the two protocols with the same single-tone modem (Rockwell MDM 2001) is given in **Fig. 2**. This set of results is with a channel simulator (≈CCIR "Poor HF channel") for repeatability [8]. Slow channel corresponds to two path, equal power, 2 msec path delay and 0.2 Hz fading rate on both channels. The results for the "fast" channel with fade rate of 2 Hz are the same for the STC protocol but lower for

AX.25 and hence are not given [8]. The improvements provided by the STC protocol are on the average, significantly more with on-air comparisons. The combination of single-tone modems with such an efficient protocol has been implemented in a number of test and operational systems including the six nation Communications Systems Network Interoperability (CSNI) project. The CSNI project has implemented a full duplex version of the STC protocol for a cross Atlantic link STC to CRC Ottawa **[10]**. The 1052p protocol may be anticipated to provide similar throughput improvements.

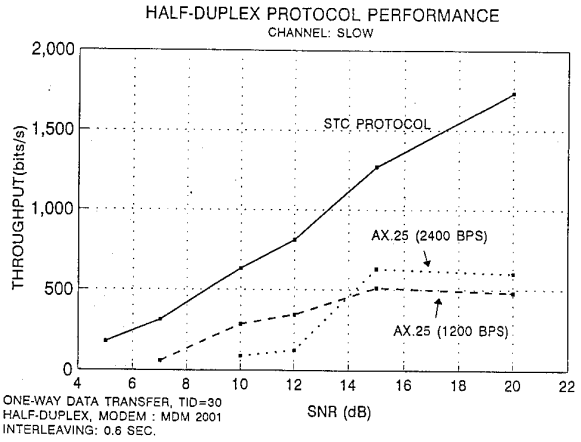


Figure 2. Throughput versus signal to noise ratio for AX.25 and STC data link protocols. Slow channel: 2 path, equal power, 2 msec delay, 0.2 Hz fade rate (mean values, both channels, uncorrelated). For the fast fading channel (2 Hz) the STC protocol results are the same, AX.25 results lower. Since AX.25 has no speed adaptation feature the performance at two rates (2400 & 1200 bps) is given [8].

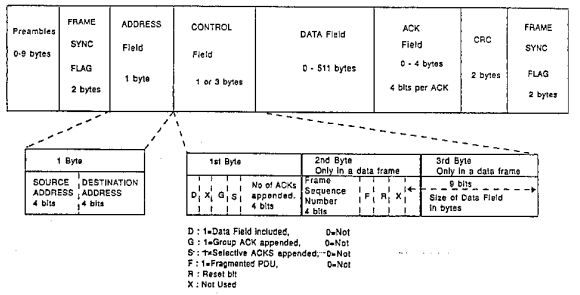


Figure 3. STC Meteor Burst Data Link Protocol packet/frame structure. The packet overhead is about 2 % excluding the preamble when data field is used to full capacity [11].

Bursty Channel (Meteor Burst Channel) Protocols

As briefly described above, the MB channel results in links that are available at random times and with random durations. These are dependant on the geographic location, orientation of the link and also on time, season, etc. In spite of these apparent complexities it has many practical applications and a unique property in terms of covertness [2]. MB technology is essentially equivalent to VHF radio + PC but is still considered esoteric due to novel nature of the medium.

Because of the random nature of the channel, the link level protocol must implement some form of channel probing to determine the instant at which the ionised trail (tangent to the oblate spheroidal surface formed by the transmitter and receiver as foci) is formed. A basic class of MB protocols probe with very short signals; when the handshake takes place after the initiation of a suitable meteor trail data packets begin to be transmitted. The second basic class uses the first data packet in the transmit queue as the probe. When the existence of a trail is acknowledged by a response the following data packets/frames are transmitted. This latter technique eases implementation and is the basis of an open STC protocol **[2,8].**

The generic properties of this data link protocol are similar to the HF protocol summarised above except that it incorporates efficient means for channel probing to determine availability . The frame structure for the STC MB PR protocol is shown in **Fig. 3** as an example; the packet overhead excluding the preamble is only about 2%, when the full capacity of the data field is used, as shown. Further details may be found in [11].

Frequency Hopping Channel

For the purposes of this paper we consider a frequency hopping radio link which is being fed data packets as a "frequency hopping (FH) channel." The STC work in this area was initiated by a requirement to provide data communications over the analog UHF frequency hopping radio system known as HaveQuick (HQ) (NATO Stanag 4246). Since about two years HQ is in widespread use for protected voice communications in most NATO nations. STC has defined/tested and prototyped a data link protocol that permits throughputs up to 12 Kbps over HQ radios with a modem speed of 19.2 Kbps in the standard 25 Khz channel in the UHF band (225 - 400 Mhz). The protocol architecture is again similar to the HF data link protocol with selective ARQ; only frames (subsegments of packets) not received are repeated while transmission of new packets/ frames continues. Some modifications are made to minimise and/or accommodate the effects of frequency hopping. Further details can be obtained through appropriate national channels from STC. The protocol does not require any modifications to the radio.

UHF SATCOM Channel

UHF SATCOM is in extensive use for mobile (land, maritime and airborne) military communications requirements. Although it is highly vulnerable to jamming and prone to interference, its low cost, ease of use due to wide antenna beams still make it attractive for mobile use. Because of the proliferation of UHF sources and the low

propagation losses in the UHF band, many UHF SATCOM channels suffer from some form of interference. In particular, some NATO UHF channels also have this problem.

STC has defined a data link protocol for communications over UHF SATCOM channels which is again derived from the above protocol architectures. This protocol allows the packet radio ARQ process to overcome the effects of interference, again trading off throughput with channel interference levels. On air tests with the protocol have shown that with 2400 bps basic signalling rate the protocol provides reliable data service throughput of typically around 1500 bps (up to 2000 bps) on 5 Khz channels that would not normally support 2400 bps rates due to interference.

Mixed-Media Protocol

It can be shown that for highly survivable communications the most cost effective approach is the combination of different media or media diversity rather than attempts of optimization towards technology boundaries of a single media [1]. A mixed-media protocol (or mixed-media controller) efficiently combines diverse links into a "virtual link" which offers all the survivability advantages of multi or mixed media operation.

Fig. 4 gives an example of mixed-media operation for which a protocol was developed and tested at STC **[12]**. Although these tests focused on the complementary properties of HF and MB communications for a specific application at the time, the protocol can readily be interfaced to other media including SATCOM. In fact it can be shown that COTS standard products for each of the three media HF, MB, SATCOM can be combined through such a protocol to provide links/networks that are survivable virtually under almost all possible stressed conditions.

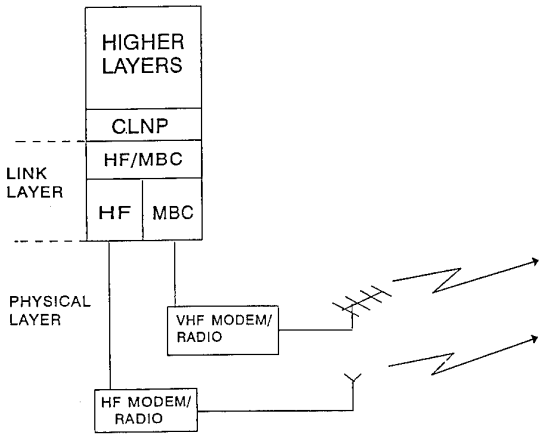


Figure 4. STC Mixed-Media Data Link Protocol. The present implementation was tested on a two media (HF and MB) link but the protocol can be readily interfaced to additional media such as SATCOM.

Conclusions

Describing and/or modelling the characteristics of transmission channels is the most important aspect of performance estimation for most communication systems. Many analytic techniques for determining performance under well defined conditions such as additive Gaussian noise, Poisson impulsive noise, exist. However, some channels are too complex in terms of noise effects, interference and time variability to model with fidelity. The most outstanding example of this is the HF sky wave channel for which "some good channel simulator performance" is required prior to fielding of a system. However, in general this is only a necessary but not sufficient condition for acceptable on-air performance.

Link level or layer two protocols provide an excellent means to reduce the consequences of communications channel effects, including additive & multiplicative noise, fading, other propagation anomalies, natural/man-made interference, hostile jamming and nuclear effects. Some of these effects can be successfully modeled under certain simplifying assumptions for line-of-sight (LOS) radio links. For non-LOS links and particularly for HF links such models give only a partial picture in the form of a "necessary but not sufficient condition" for successful communications. The basic approach is to use packetization (packet radio) to allow an automatic request for repeat (ARQ) process to continuously trade-off throughput with channel conditions. Many other facilities such as support to link establishment/maintenance, adaptive modem speed control can also be incorporated. With such an approach the source information (data, fax, image and under certain conditions voice) will always be communicated to the destination with no errors but the transmission (or delivery) time will vary with channel conditions; the worse the channel the longer it will take to transmit the information.

This paper has presented examples of link level protocol work that has been performed at SHAPE Technical Centre. Specifically, link level ARQ protocols developed for HF sky wave, meteor burst, frequency hopping channel, UHF SATCOM and a mixed media controller have been briefly described. There is no doubt that efficient link-level protocols are the highest pay-off in data communications over radio links/networks.

Acknowledgements

Many past and present members of the STC Communications Division have contributed to the work described in this paper. Their efforts are gratefully acknowledged.

References

[1] D. Yavuz, F. Eken, N. Karavassilis, "Highly Survivable Communications: Complementary media packet-switched networks", AGARD EPP-53, Rotterdam, October 1993.

[2] D. Yavuz, "Meteor Burst Communications", IEEE Communications Magazine, Sept. 1990.

[3] B. Goldberg (Editor), Communication Channels: Characterisation and Behaviour, IEEE Press, 1976

[4] OSI Basic Reference Model - ISO 7498

[5] ISO Data Link Service Definition - ISO 8886.3

[6] T. Fox, AX.25 Amateur Packet-Radio Link-Layer Protocol, Version 2, October 1994, ARRL.

[7] Proposed Federal Standard 1052 "Telecommunications: HF Radio Modems", General Services Administration, Washington DC. *(Appendix B describes the ARQ data link protocol; the main body describes the serial/single tone modem. Appendix A describes the 39 tone modem)*

[8] F. Eken, D. Clark, STC OSI HF PR Data Link Protocol, STC TN-492, 1992 (NU). *See also STC TM-930, TN-507 (implementation Details), TN-506 (Laboratory and On-air Test Results), all NU.*

[9] NATO Stanag 4285, Characterization of 1200/2400/3600 b/s single-tone modulators/demodulators for HF radio links & US MIL-STD-188-110A, Interoperability and performance standards for HF data modems 1991.

[10] Proceedings of the CSNI Symposium, SHAPE Technical Centre, Sept 1995 (NU).

[11] N. Karavassilis, STC OSI MBC Data-Link Protocol, STC TN-547, 1994 (NU).

[12] N. Karavassilis, STC OSI Mixed Media Protocol, STC TN-548, 1994 (NU).

DISCUSSION

Discussor's name: P.A. Kossey

Comment/Question:

You mentioned VHF Meteor-Burst-Communications (MBC) specifically in the context of its propagation survivability for certain nuclear scenarios. Was (is?) that the primary (or only) feature of MBC that was (is?) of interest to STC/NATO?

Or (to ask the question in another way), are the features of MBC, such as LPI, AT, low power, light--weight, etc., attractive enough in their own right to warrant the use of MBC for evolving NATO military needs?

Author/Presenter's reply:

Our work in MBC was in the context of multi-radio medium systems in support of highly survivable communications in hostile environments. For example, we implemented an MBC/HF broadcast system which was tested between STC and Latina in Italy for many years in the late eighties. This work was documented and put in the "freezer" as a result of political developments.

All the features of MBC, particularly difficulty of jamming, are of interest.

DISCUSSION

Discussor's name: C. Goutelard

Question/Comment:

L'expérience que vous avez faite entre le STC et le CRC donne dites-vous des débits satisfaisants. Cependant, sur le trajet nocturne, le débit devient très faible, ce qui est une caractéristique des communications HF. Quelle solution adoptez-vous dans cette situation?

(Translation:

You say that the experiment carried out by STC and CRC gives satisfactory rates. However, on the nocturnal path the rate becomes very poor, which is a characteristic of HF communications. What solution do you adopt in this situation?

Author/Presenter's reply:

The ALE implemented for this link was one based on fixed intervals since the only existing standard does not support ALM (Automatic Link Maintenance). Furthermore, the ALE is also very slow as it was designed more than a decade ago.

With an improved ALE linked to the protocol ARQ rate some improvements could be expected. However, with HF there will always be some times when it will be unavailable. It is free beyond LOS communications (it can be shown that per message, costs are many hundreds of times smaller than SATCOM in mobile applications). We have combined HF and INMARSAT C for a data terminal that can povide connectivity 24 hr; ~90% of the time on HF 10% on SATCOM, etc.

DEFINITION OF A COMMUNICATION SYSTEM USING THE TROPOSPHERIC CHANNEL

G. Vivier, P. Brélivet, J. Richard, P. Sehier

ALCATEL TELSPACE
5, rue Noël Pons
92734 Nanterre, France

1. INTRODUCTION

Many transmission channels, like the radio channel or the tropospheric channel, are frequency selective and time variable. Receivers designed for this type of transmission have to fight against intersymbol interference generated by the multiple paths and at the same time adapt to the time variations of the channel. Only with progress in digital technology, leading to the application of new concepts of advanced equalization and error correcting codes, will good quality, high bit rate transmission become achievable on this type of channel.

This article covers high bit rate communications (of the order of 8 Mbit/s). The first part discusses the special features of the tropospheric channel and system aspects that must be considered when attempting a complete definition of a communication system on this channel.

The second part describes two waveforms, associated with their respective receivers, designed to meet the requirements identified in the first part. The first, designed for a single-carrier modulation system, is based on an architecture combining equalization, decoding and deinterleaving. This method, introduced by R. Mehlan and H. Meyr in [2] and [3], is called Combined Equalization and Decoding (CED) in the remainder of this article. The second is a single-carrier alternative known as Coded Orthogonal Frequency Division Multiplexing (COFDM), adopted for digital audio broadcasting (DAB), and envisaged for digital terrestrial TV broadcasting [4-6].

In the second part, we attempt to compare these two methods in terms of Bit Error Rate (BER), obtained by simulations in conditions approximating a real link, and in terms of difficulty of implementation.

2. TROPOSPHERIC CHANNEL TRANSMISSION SYSTEM

2.1 The tropospheric channel

Over-the-horizon communication on the tropospheric channel highlights the phenomena of scatter brought about by atmospheric heterogeneity. The receiving beam captures a fraction of the energy of the scattered transmitting beam, in the volume common to both beams, known as the scatter volume (or "common volume"). Propagation loss is on average much higher than on a line-of-sight link and subject to rapid, major variations due to the instability of the scatter volume. As this behaves like a transmitter comprising a large number of individual sources of virtually identical amplitudes but random phase, it can be shown that the tropospheric channel conforms to a Rayleigh statistic.

Assuming Wide Sense Stationary Uncorrelated Scatter (WSSUS), the channel can be fully characterized by the Scattering Function [1], which gives the power spectral density of each of the delayed paths. This function can be summed up by two parameters: coherence band B_c and coherence time T_c.

The coherence band is the minimum frequency separation between two signals affected by uncorrelated fading. It therefore characterizes the frequency selectivity of the channel. For a single-carrier signal occupying a band W, the spread of intersymbol interference and, because of this, the complexity of the receiver, is inversely proportional to B_c/W.

The coherence time is defined as the time interval in which channel variations are strongly correlated. When high, deep interleaving must be used, which introduces significant delays, or latencies. When low, the receiver must be able to follow rapid variations in the channel's impulse response.

Elsewhere, the B_c and T_c parameters which characterize the short term statistical behaviour of the channel, are subject to slow variations at different times of the year. Measurements taken on real channels have shown that the variation ranges of B_c and T_c could be as high as two orders of magnitude.

This long term variation phenomenon brings with it a major difficulty when it comes to dimensioning a transmission system that is required to offer a high level of availability (typically 99.9% of the hours in a year). The waveform and associated signal processing devices must, in practice, provide a specified level of performance across the entire range of parameter variations on the channel.

2.2 System considerations

To overcome fading, troposcatter transmission entails the use of diversity techniques. It has been shown that BER performance improves with the order of diversity, tending asymptotically towards that of the gaussian channel when this channel tends to infinity.

The diversity techniques routinely used in troposcatter transmission are explicit:

- frequency diversity, which entails sending the same signal simultaneously on a number of frequencies.
- space diversity, in which the signal is received on a number of sufficiently distant antennas. This costly solution does save on frequency spectrum. It is often linked to frequency diversity.
- angle diversity, which enables the signal to be captured from two different scatter volumes using a special antenna. This method is beneficial with high gain antennas.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

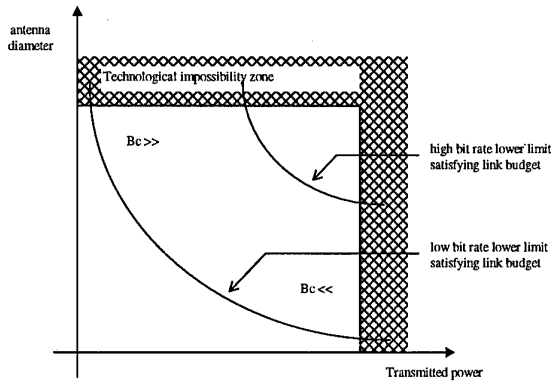


Figure 1 : Antenna/power selection zone

Diversity techniques associated with the waveform, obtained by spread spectrum, encoding and interleaving, implicit in that they do not require large, costly equipment, seem promising. This is why we have chosen to explore these techniques in this comparative study.

System engineering consists in choosing the tropospheric station architecture in terms of component dimensioning, to create conditions favourable to reception. Thus, reducing the antenna diameter reduces the coherence band and enables the implicit diversity associated with the multiple paths to be increased. However, in this case, the transmitted power must be increased to maintain an EIRP which satisfies the link budget. Amplifier technology then becomes a limiting factor.

So, conditions favourable to transmission on the tropospheric channel can be created by adjusting the antenna/power pairing, but only within the limitations imposed by link budget and technological constraints. Figure 1 gives a qualitative illustration of this trade-off. For high bit rates, it can be seen that the choice is very limited.

2.3 Channel model

Many theoretical models of the tropospheric channel were developed in the 1970s ([7] and [8]). These models are used to predict channel parameters on the basis of physical data (antenna diameter, distance, etc), but do not give long term variations of coherence band and time. This omission was corrected by C. Collin and A. Marguinaud in [9] and [10], by proposing an empirical method for predicting fluctuations in these parameters. The table below shows how this method can be applied in the case of a typical 200 km link with a bit rate of 8 Mbits/s. In this table, $B_{0.1}$, B_{50} and B_{99} are defined by the probabilities:

$p(B<B_{0.1})=0.1\%$, $p(B<B_{50})=50\%$ and $p(B<B_{99.9})=99.9\%$.

Ø ant.(m)	P (kW)	$B_{0.1}$ (MHz)	B_{50} (MHz)	$B_{99.9}$(MHz)
3	8	0.64	2.25	8
4.5	2	1.1	3.8	13.4
6	1	1.5	5.2	18.4
8	.5	2	6.9	24

The power distribution of the delayed paths is given by the Multipath Intensity Profile (MIP), defined in greater detail in [1].

In simulations, a "Gamma" function MIP was used [11]. This widely used function is a good approximation of the expression given in [7], particularly at low carrier frequencies (up to 2 GHz). The variance of delayed paths is given by the following equation:

$$\sigma^2(\tau) = 8B_c^2\ \tau \exp\left(-\sqrt{8}B_c\ \tau\right) \qquad (1)$$

3. DESCRIPTION OF THE SOLUTIONS CHOSEN

3.1 Single-carrier system

3.1.1 History of CED

The first systems developed for digital transmission were based on single-carrier waveforms which were the subject of widespread research into their definition (modulation, encoding, interleaving) and the receiver structure (equalization and synchronization).

These (de)modulation, (de)coding and (dis)interleaving operations were initially researched and optimized separately. A new line of research was how to combine these functions to improve overall performance.

Trellis coded modulation (TCM) techniques are one result of this research. By combining modulation and coding, error ratio performance has been improved without compromising spectral efficiency. [12] and [15] describe particularly powerful TCM techniques for fading channels.

Having combined coding and modulation, why not combine TCM equalization and decoding, since the code and the channel have similar convolutional effects on the signal, which can be processed by the Viterbi algorithm? This is the principle of "supertrellis" coding [16] and of its derivatives with fewer states [18] and [19]. Unfortunately, this process is less effective for fading channels, because its structure does not allow for the insertion of an interleaving device between the encoder and the modulator, so the decoder operates on highly correlated symbols. More recently, R. Mehlan and H. Meyr proposed a method combining equalization, deinterleaving and decoding (CED) [2, 3]. Because its performance characteristics are good and it is relatively simple to implement, this was the method adopted for our evaluation.

3.1.2 Description of the single-carrier solution

Figure 2 shows a functional block diagram of the transmission system. On transmission, the symbols are interleaved in blocks after convolutional encoding, and then modulated. On reception, the signal is converted to baseband and filtered, then input to a whitening matched filter and forwarded to the CED function where it is deinterleaved and further equalized and decoded.

On output from the filter, the intersymbol interference is causal and the gaussian noise samples are independent. This architecture is akin to that proposed by Forney in [17]. Figure 3 contains an equivalent representation of the transmission subsystem, in which F is the equivalent filter obtained by concatenating the channel and the whitening matched filter.

On transmission, the symbols are written in rows in the interleaving matrix which measures PxQ and read in columns. Each column comprises N_r reference symbols and P-N_r coded information symbols. The first N_r rows of the matrix therefore contain known symbols.

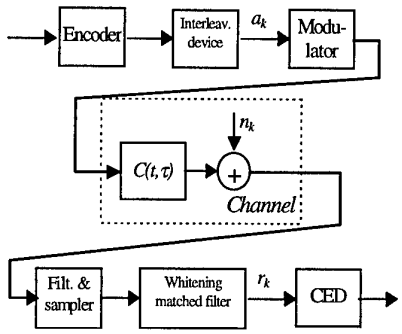


Figure 2 : Functional block diagram of the single-carrier system

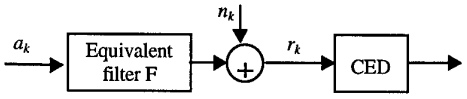


Figure 3 : Equivalent diagram

Figure 4 shows the decoding matrix, before the start of the algorithm, the basic principle of which is to equalize in columns and decode in rows. Equalization entails knowing, for each column, the values of the equivalent filter's coefficients f_i. These are estimated using the known symbols in the interleaving matrix.

The channel estimation frequency must be based on coherence time. Since the channel's rate of change is slow with regard to the transmitted bit rate, it has been assumed that this estimation is performed once every interleaving frame, which leaves a high degree of freedom in choosing the size of the matrix (PxQ < 16.10^4 in the worst case).

If a_k is the post-interleaving symbol obtained at time $k.T_s$ (where T_s is the symbol time) and $F=\{f_0, ..., f_{L-1}\}$, the impulse response of the equivalent channel, the signal sample received at time $k.T_s$, can be written:

$$r_k = f_0 a_k + ISI + n_k \quad \text{with} \quad ISI = \sum_{i=1}^{L-1} f_i a_{k-i} \qquad (2)$$

where n_k is the gaussian white noise sample obtained at time $k.T_s$. Since the a_k are introduced in the CED matrix column by column, the ISI can be calculated from the L-1 symbols in the same column, but on the previous rows and on which decisions have already been taken. This ISI term is subtracted from r_k and the resulting signal is output to the Viterbi decoder which, after a decoding cycle, outputs decision $\hat{d}_{k-T-1}$, where T is the decoder's truncating length. The decoded symbol is then recoded and returned to the deinterleaving matrix at the point where it was stored r_{k-T-1}. The process is then repeated for all the samples in the matrix. Inasmuch as Q is greater than T, it is possible to prove that the decisions needed to calculate the term ISI are always available.

The effectiveness of Mehlan's algorithm can be understood intuitively. The algorithm is a decision feedback equalizer (DFE), in which the term reinjected before the decision unit is computed from symbols tested by the decoding operation (and not, as in a conventional DFE, from decisions). Error propagation, which is a major source of impairment at low signal-to-noise ratios, particularly when the channel is highly selective, is greatly reduced.

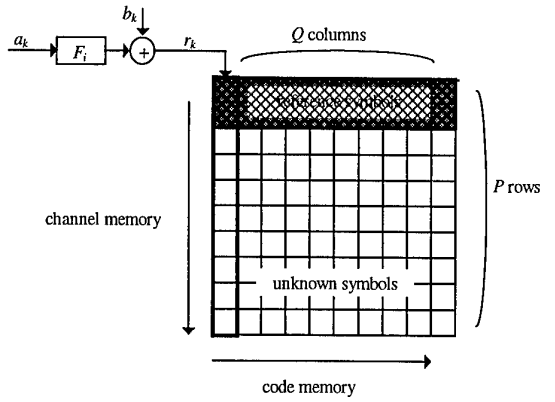


Figure 4 : CED matrix before the start of decoding

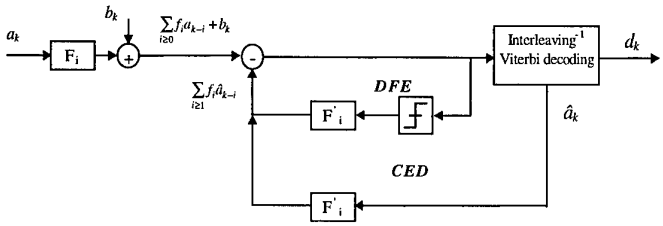


Figure 5 : Diagram comparing DFE and CED

Figure 5 shows the parallels between the two methods.

Equation (2), adapted from [2], gives, in terms of transition probability for a channel without memory, the asymptotic performance figures of the CED based on the length of the shortest error event d, the signal-to-noise ratio γ_b and the number of paths of the equivalent channel L :

$$P(\mathbf{a},\hat{\mathbf{a}}) \approx C.\left(\eta\ \gamma_b\right)^{-L.d} \qquad (3)$$

in which C and η are corrective terms which depend on the code and the path power distribution. This equation shows that the equivalent order of diversity, given by the slope of the curve $(L.d)$, increases with the selectivity of the channel and that very high equivalent orders of diversity can easily be obtained (typically 192, taking d=6 and L=32). Simulations have shown that the approximation given by (3) was very optimistic (in particular because of the constant C which can take very high values).

3.1.3 Determining the main parameters

The parameters of the system (mainly the size of the filters, the interleaving matrix and the reference pattern) must be determined to satisfy the "worst case" configuration obtained when coherence band and time are minimum.

Simulations have shown that the minimum number of coefficients for the whitening matched filter (excluding oversampling) ought to be at least equal to $L=\lceil Bc.Ts \rceil$. In these conditions, the equivalent channel also has a spread of L symbols. Nr, which is the number of reference symbols in each column of the interleaving matrix, must therefore be at least equal to L. Also, the number of columns Q is determined by the code constraint length ($Q > 5.Lc$). The reference insertion loss is $P/(P-Nr)$. This loss can be made arbitrarily low by increasing the value of P, but at the cost of the risk of error propagation. In addition, setting P and Q high increases the depth of interleaving and enhances performance. Nevertheless, the rate of change of the channel and the transmission delay caused by interleaving limit the

size of the CED matrix. The trade-off chosen for the simulations, after a number of evaluations is: N_r = 31, P = 100 and Q = 50.

3.2 COFDM

3.2.1 Definition

Coded Orthogonal Frequency Division Multiplexing (C)OFDM was adopted by the CCETT for its digital audio broadcasting (DAB) requirements [4-6]. The transmission channel is a radio channel, affected by fading due to the multiple paths and similar in quality to the tropospheric channel.

The principle of OFDM is to divide information over a large number of carriers modulated at low speed, for which the channel is non-selective. The carriers are spaced so that they are at right angles. The absence of a guard band between carriers ensures good spectral efficiency. A guard time, at least equal to the length of the impulse response of the channel, is used to absorb intersymbol interference and, in addition, facilitates synchronization. The structure of the OFDM signal allows for the use of fast Fourier transform (FFT) algorithms in transmit and receive modes.

To obtain good performance on a fading channel, it is vital to use coding and interleaving. The purpose of interleaving is to optimize the spread of the symbols in time and frequency, so that the decoder can operate in favourable conditions, in other words, conditions close to those of a channel without memory.

Figure 6 is a functional block diagram of the transmission subsystem for the COFDM system used.

3.2.2 Principle of demodulation and decoding

If the guard interval satisfies the conditions set out in the section above, there is no intersymbol interference. Assuming this, the signal sample recovered after the receive FFT for block k and on carrier i can be expressed:

$$r_i(k) = h_i(k).c_i(k) + n_i(k) \qquad (4)$$

where $h_i(k)$ is the complex coefficient of attenuation introduced by the channel (equal to the Fourier transform of the channel impulse response, calculated at the frequency of the i^{th} carrier), $c_i(k)$ is the k^{th} symbol transmitted on carrier i and $n_i(k)$ is the gaussian noise sample on carrier i for block k.

In coherent demodulation, the optimal decoding metric applicable to phase modulations can be expressed:

$$m(z,a,h) = \mathrm{Re}\left(z.h^*.a^*\right) \qquad (5)$$

where a is a coded symbol, z is the output of the channel, and h is the attenuation coefficient of the channel for the output concerned. The weighting operation can be used to favour reliable symbols (those that are weakly attenuated) over unreliable symbols when decoding. This Channel State Information is estimated by the receiver using reference symbols periodically inserted into each carrier.

Weighting can be performed outside of and uplink from the Viterbi decoder, so the metric (5) allows for the use of a conventional decoding component. This means that the sequence of processes can be performed as indicated in Figure 6.

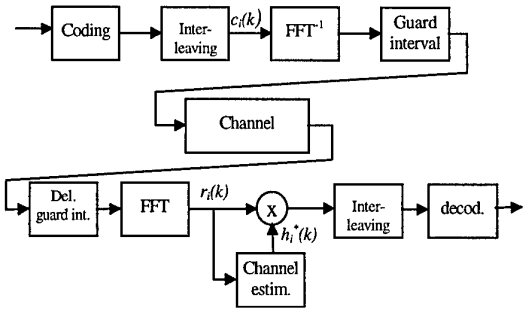


Figure 6 : Block diagram of a multi-carrier system

3.2.3 Determining the parameters

The guard time τ is determined from the channel's maximum impulse response spread. If N_p is the number of carriers active, and D_S the total symbol bit rate, the loss due to the insertion of guard times is equal to:

$$\eta = \left(1 - \tau.D_s/N_p\right)^{-1} \quad \text{with} \quad N_p \geq \tau.D_s \qquad (6)$$

This loss can be made arbitrarily low by increasing the number of carriers. However, the maximum value is limited by the coherence time of the channel according to the relation: $N_p << T_c.D_s$. The compromise must also take into account that carrier sync losses (frequency difference and phase noise) increase sharply when the bit rate per carrier is reduced (see [6]). All of these considerations have led us to take N_p=128, which causes an insertion loss of 0.9 dB, when τ=3 µs (=1.5/B_{cmin}) and D_s=8.10^6 symbols per second.

The dimensioning of the interleaving function is crucial. It must be set so that two consecutive symbols are affected by uncorrelated fadings. To remain close to the single-carrier solution, we will consider an interleaving matrix of size PxQ. Also, we will assume that the channel remains constant over an interleaving block. This is a pessimistic assumption, which is tantamount to assuming that temporal diversity cannot be used. However, this assumption is realistic, bearing in mind the limitations on the transmission times, bit rates and coherence times involved. A number of disastrous interleaving situations have been identified:

- P = kN : two consecutive symbols are always on the same carriers.
- N = kP : the same carrier supports symbols spaced by k. To be avoided when k is less than the code constraint length L_c.
- $Q < L_c$: symbols spaced by Q are sent on two adjacent carriers.

Figure 7 shows the influence of the interleaving parameters on the bit error ratio for a fixed channel. It appears that there are many substantially equivalent combinations.

4. COMPARISON OF THE TWO METHODS

The two methods are compared in terms of sensitivity to non-linearities and complexity of implementation.

4.1 Influence of non-linearities

Troposcatter transmission systems employ high power amplifiers that should be able to be used at maximum output.

These amplifiers, when used at their saturation level, introduce non-linear distortions, the magnitudes of which depend on fluctuations in the envelope of the transmitted signal. For a signal with constant (or quasi-constant)

envelope obtained by single-carrier phase modulation, the amplifier can be used at its saturation power level without significantly impairing performance. In OFDM, because of the wide variation range of the signal envelope, it is important to allow a significant back-off. The OFDM solution is therefore worse affected than the single-carrier solution in real-life operating conditions. These losses are explained in the section on simulations.

For our evaluation, we have assumed the use of a TWT. Using Saleh's model [20], the amplitude-amplitude and amplitude-phase conversion formulae are:

$$g(A) = \frac{\alpha_A A}{1+\beta_A A^2} \quad \text{and} \quad \varphi(A) = \frac{\alpha_\varphi A^2}{1+\beta_\varphi A^2} \qquad (7)$$

where:

$\alpha_A = 1$ $\beta_A = 0.25$ $\alpha_\varphi = 0.26$ and $\beta_\varphi = 0.25$

The Output Back-Off (OBO) is defined by:

$$OBO = 10\log\left(\frac{P_s}{P_0}\right) \qquad (8)$$

where P_S is the saturation power of the amplifier and P_0 the average output power. By increasing the OBO, in other words by converging towards the linear zone, signal distortions can be reduced, but with a loss of transmitted power. There is therefore an optimum back-off which minimizes the combination of degradation due to distortion and degradation due to power loss.

Figure 8 represents the bit error ratios of the two system on a gaussian channel, with or without non-linearities. As predicted, it can be seen that the COFDM system is far more sensitive than the single-carrier system. However, the respective degradations of COFDM systems with 64 and 1024 carriers are significantly the same, while the dynamic range of the 1024-carrier signal is greater.

The table below gives the minimum degradations due to non-linearities and the optimum back-off on a gaussian channel, for a BER of 10^{-4}.

	Single	COFDM 64	COFDM 1024
OBO_{opt} (dB)	1.2	2.1	2.5
losses (dB)	1	3.5	3.5

These results were obtained in the following configuration: QPSK modulation, cosine filtered, boosted with 0.5 roll-off (in single-carrier mode), (7,1/2) code.
We will assume in the discussion below that the losses incurred by non-linearities for the single-carrier waveform are less than 1 dB. It is important to remember that single-carrier mode losses can be made arbitrarily low, even zero, by changing the signal filtering arrangements. For COFDM, we will try to identify the best operating point.

4.2 Complexity

The complexity of each method is estimated by the number of complex multiplications for each demodulated symbol. This method of comparison, recommended by its very simplicity, is clearly imperfect since it does not take into account differences in hardware architecture, which are at least as important in an estimation of complexity/cost as pure computation power.

In this comparison, we will exclude those functions assumed to be of equivalent complexity, such as synchronization, channel estimation and decoding. The term operation is used below to mean a complex multiplication.

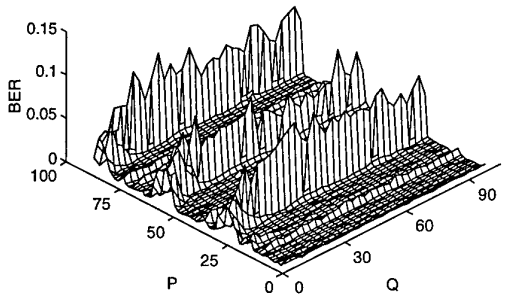


Figure 7 : BER as a function of interleaving

In the single-carrier solution, the whitening matched filter (of L coefficients) requires L operations per symbol, and computing its coefficients requires $2L^2+1$ operations per interleaving block. Computation of the ISI term also entails L operations per symbol. The total number of operations per symbol is therefore $C_{SINGLE}(L) = 2L+(2L^2+1)/(P.Q) \approx 2L$.

In the single-carrier solution, the processing of a block of N_p symbols entails calculating two FFTs (transmit + receive) of size N_p. In reality, the FFT size is normally between $1.5.N_p$ and $2.N_p$, taking into account the guard carriers provided to lighten the specifications of the anti-aliasing filter and because it is necessary to use a power of 2. Also, weighting before decoding entails one operation per symbol. The total number of operations per symbol is therefore: $1+2.(\frac{1}{2}.N_p.\log_2(N_p))/N_p \approx \log_2(N_p)$. The number of carriers is determined according to permitted insertion losses. If losses of 1 dB are allowed, as in (6) this should give: $N_p \approx 5.L$ (with, as before, $L \approx 1.5\ D_s/B_c$). Because of this, the complexity, expressed as a function of L, is: $C_{COFDM}(L) = \log_2(5.L)$.

The ratio $C_{SINGLE}(L)/C_{COFDM}(L)$ is therefore always greater than 1 (provided that $L>1$), and increases sharply with channel selectivity. This is the conventionally recognized advantage of COFDM when compared with a time domain equalizer. However, as suggested in [6] for a linear equalizer, filtering in the single-carrier method can also be implemented in the frequency domain. In this case, the complexity of the single-carrier solution is in $\log(L)$, and the complexity of the two methods becomes comparable.

4.3 Performance simulation

The two waveforms were compared for three coherence band values: B_C = 0.5, 3 and 10 MHz. The bit rate is taken to be 8 Mbits/s. A convolutional code (7,½) with a polynomial generator (133, 171) in octal was used. The modulation method was QPSK. The bit rate after coding and modulation is therefore 8.10^6 symbols per second. In this article, we have not placed any emphasis on optimizing the code in order to base the comparison of the two waveforms on a conventional code.

A long coherence time has been assumed compared to the duration of an interleaved block, for computation time reasons. The simulation principle, for both methods, was based on taking a random channel for each new block of data (the channel of block N is independent of the channel of block N-1). The performance figures obtained are therefore

very close to those for a channel with arbitrarily long coherence time and can therefore be qualified as "worst case".

The table below gives the values of the key waveform parameters, determined in accordance with the considerations set out in the third part.

Single	N_r	31
	PxQ	100x50
	L	35
	η (dB)	1.17
COFDM	N_p	128
	Guard	3 µs
	PxQ	14x64
	η (dB)	0.9

N_r is the number of reference symbols per column, PxQ is the interleaving size, L is the number of whitening filter coefficients and η is the insertion losses.

Figures 8, 9 and 10 show the performance levels for both waveforms for all three coherence band values considered. To evaluate the losses due to error propagation in the CED, figure 12 also contains an ideal CED in which the symbols used to calculate the ISI term are error-free. Asymptotically, the CED performance levels tend towards those of a perfect equalizer. However, when the channel is selective (B_c = 0.5 MHz), for normal BER levels there remains a difference of approximately 0.7 dB: since the channel memory is significant, an error on one estimated symbol can result in a very long packet of errors. This figure shows that CED performance levels improve with channel selectivity, in line with equation (3).

On a linear channel, the performance levels of the two methods are similar. In reality, the COFDM method behaves like an ideal CED in which ISI is perfectly corrected, while the CED is affected by error propagation, particularly at low signal-to-noise ratios. On a non-linear channel, it can be seen that the COFDM method (simulated with the optimum OBO value) suffers very significant degradation while the degradation of the single-carrier method is around 1 dB. The performance differs by between 5 and 7 dB according to the value of B_c, in favour of the single-carrier method.

5. CONCLUSION

Having reviewed the specific features of the tropospheric channel, we have demonstrated the importance of the choice of antenna/power pairing in defining a tropospheric station. We then went on to describe and analyze two waveforms particularly well suited to the constraints identified in the preceding part. The first, based on single-carrier modulation, is associated with a receiver which handles both equalization and decoding. The second is a single-carrier alternative, COFDM, adopted for European digital audio broadcasting.

The comparative simulations performed in similar conditions, showed that the two methods were virtually equivalent in the case of a linear channel. In more realistic conditions, taking into account non-linearities of amplification, it appears that the OFDM method is broadly affected by its high sensitivity, while losses in single-carrier mode are negligible. To obtain the same BER performance levels it is essential to use an amplifier with a power of approximately 6 dB greater in COFDM mode.

A quick comparison of complexity reveals that the COFDM method is more advantageous, particularly for very selective channels. However, if the very realistic possibility of filtering in the frequency domain is taken into account for the single-carrier solution, the two methods become equivalent.

Finally, when the channel is not very frequency selective, the performance levels of the two methods are limited by the implicit absence of diversity. In this case, it becomes necessary to use an additional diversity technique such as Frequency Hopping, which is compatible with both methods.

6. ACKNOWLEDGEMENTS

Part of the work was financed by the DRET (Direction des Recherches Etudes et Techniques).

7. REFERENCES

[1] J. Proakis, *Digital Communications*, McGrawHill, 1983.

[2] R. Mehlan, H. Meyr, '*Combined Equalization/Decoding of Trellis Coded Modulation on Frequency Selective Fading Channels*', Tirrenia International Workshops on Digital Communication, September 1991.

[3] S. A. Fechtel, H. Meyr, '*Equalization for Coded Modulation on Fading Channels*', IEEE Workshop on Synchronization and Equalization in Digital Communications, Gent, May 1995.

[4] B. Le Floch, R. Halbert-Lassalle, D. Castelain, '*Digital sound broadcasting to mobile receivers*', IEEE Trans. on Consumer Electronics, vol 35, n°3, August 1989.

[5] J.F. Hélard, B. Le Floch, '*Trellis coded orthogonal frequency division multiplexing for digital video transmission*', Globecom 91, Conf. Rec. vol 2, Phoenix, AZ, December 1991, pp 785-791.

[6] H. Sari, G. Karam, I. Jeanclaude, '*Transmission techniques for digital terrestrial TV broadcasting*', IEEE Communication Magazine, February 1995.

[7] P. A. Bello, '*A troposcatter channel model*', IEEE Trans. on Communication technology, April 1979.

[8] L. D. Daniel, R. A. Rienman, '*Performance prediction for short range troposcatter links*', IEEE Trans. on Comm., June 1976.

[9] C.Collin and A. Marguinaud, '*Evaluation empirique de la bande de cohérence en diffusion troposphérique*'. Rev. Tech. Th-CSF Vol 11.3, 1979.

[10] C.Collin and A. Marguinaud, '*Evaluation expérimentale de la sélectivité d'un liaison par diffusion troposphérique*'. Rev. Tech. Th-CSF Vol 11.1, 1979.

[11] S.J. Howard, M. S. Wallace, '*Performance of wideband signaling for slowly fading frequency selective channels*', MILCOM 1992.

[12] G. Ungerboeck, '*Channel coding with multilevel/phase signals*', IEEE Trans. on Information Theory, January 1982.

[13] D. Divsalar, M. K. Simon, '*The design of trellis coded MPSK for fading channels: performance criteria*', IEEE Trans. on Communications, Vol. COM 36, no.9, June 1988.

[14] D. Divsalar, M. K. Simon, '*The design of trellis coded MPSK for fading channels: set partitioning*

for optimum code design', IEEE Trans. on Communications, Vol. COM 36, no.9, June 1988.

[15] E.Biglieri, D.Divsalar, P.J.MacLane, M.K.Simon, *Introduction to trellis-coded modulation with applications*, Maxwell MacMillan International Editions, 1991.

[16] M. V. Eyuboglu, S. H. U. Qureshi, '*Reduced-state sequence estimation for coded modulation on intersymbol interference channels*', IEEE Journal on Selected Areas in Communications, Vol 7, August 1989.

[17] G. D. Forney, '*Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference*', IEEE trans. on Information Theory, vol IT 18, n°3, pp363-378, May 1972.

[18] A. Duel-Hallen, C. Heggard, '*Delayed decision feedback sequence* estimation', IEEE Trans. on Communications, Vol COM 37, n°5, May 1989.

[19] M. V. Eyuboglu, S. H. U. Qureshi, '*Reduced-state sequence estimation with set partitioning and decision feedback*', IEEE Trans. on Communications, Vol COM 36, n°1, January 1988.

[20] A. A. M. Saleh, '*Frequency-independent and frequency dependent nonlinear models of TWT amplifiers*', IEEE Trans. on Communication, Vol COM 29, pp. 1715-1720, November 1981.

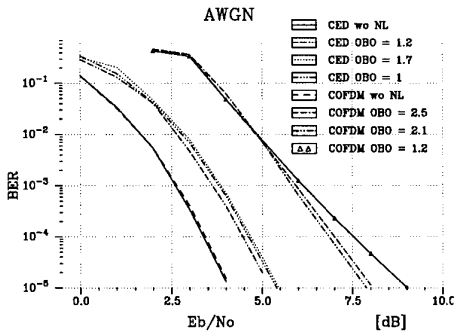


Figure 8 : Influence of non-linearities

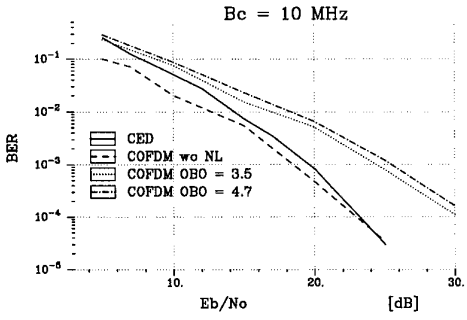


Figure 9 : COFDM / Single-carrier comparison

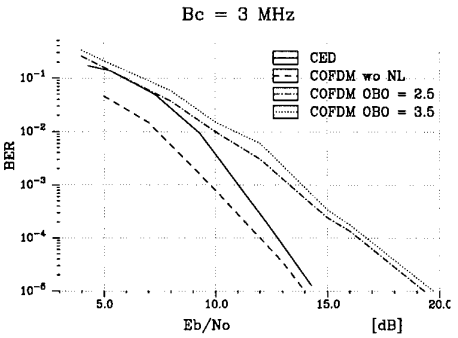


Figure 10 : COFDM / Single-carrier comparison

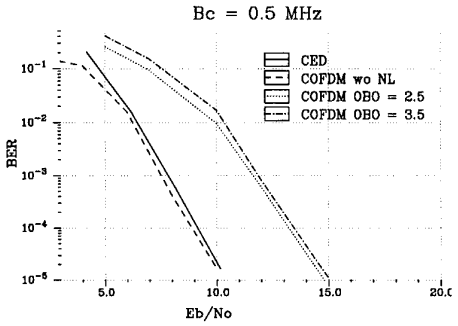


Figure 11 : COFDM / Single-carrier comparison

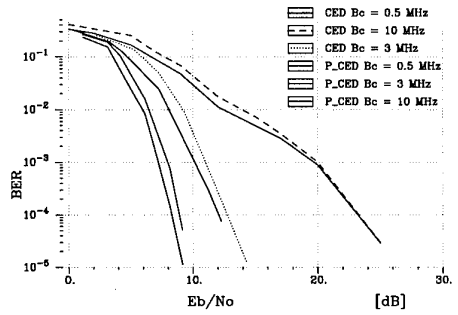


Figure 12 : CED behaviour

DISCUSSION

Discussor's name: L. Hoff

Comment/Question:

1. Have you investigated how long your adaptive equalization circuits require to adapt to the channel?

2. Are you primarily concerned with point-to-point full-period termination communication links?

Author/Presenter's reply:

1. No. Computer simulations evaluated the performance after circuits had adapted.
In fact, DFE (in the LED) is immediately adapted to the channel due to known training symbols and perfect channel estimation.

2. Yes.

COMPARISON OF RLS ADAPTIVE ARRAY ALGORITHMS IN A STRONG MULTIPATH PROPAGATION ENVIRONMENT

D. Peeters
E. Van Lil*
K.U.Leuven
Div. ESAT-TELEMIC
Kard. Mercierlaan 94
B-3001 Heverlee
Belgium

Summary

Most of the algorithms for spatial multiplexing are tested under simple multipath conditions.
In this paper a more complete and realistic model for the communication channel is described, taking into account local dispersion, Doppler-shift, multipath propagation and thermal noise. Based on this model, a program has been written to actually compute the effect of the complex environment on the signals. The numerical data for the multipath effect were taken from the GSM Recommendation 05.05 'Radio Transmission and Reception'. With the results of this propagation simulation program, algorithms of the RLS type are compared. It is found that the difference in performance between the S-RLS algorithm and the QR-RLS algorithm is negligible. Because of the limited length of the training-sequence, the number of coefficients to be determined must be kept low. Additional time-equalization does not improve the reception substantially. Also, due to the Rayleigh fading, the amelioration of the reception at decreasing noise levels stagnates.

1 Introduction

Antenna diversity has been introduced to improve the reception of signals in wireless and mobile communication systems [1]. The signals on the different antennas were supposed to be statistically uncorrelated, so that no deep fades occured on all antennas at the same time.

Antenna adaptivity differs from early antenna diversity in that the correlation of the signal on the different antennas is exploited to our benefit. Unwanted signals are namely made to interfere destructively by choosing the appropriate (complex) weights in the weighted sum of the antenna outputs. The advantage over time-equalization is that signals uncorrelated in time with the wanted signal can be suppressed too. Hereby we think of multipath reflections with excess equalizer length delay, and co-channel interferences.

As an extension, adaptive beamforming makes a supplementary kind of multiplexing possible : Space Division Multiple Access (SDMA). Different mobile or wireless stations using the same frequency and time slot can still be discerned by virtue of their spatial[1] separation.

Several algorithms have already been tested, neglecting the correlation [2], or using simple propagation circumstances [3][4][5]. The least-squares constant modulus algorithm described in [6] is not quite suited for spatial multiplexing. Purpose of this paper is to investigate the performance of RLS algorithms in a more realistic propagation environment. In the following section, the channel model of [7] is adapted with slightly less statistical behaviour, so that the algorithms described in section 3 can be evaluated and compared in section 4. Section 5 concludes with some remarks.

*National Fund for Scientific Research (NFWO), Belgium.

[1] "Angular" separation would be more appropriate.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

2 Signal and channel model

The vector channel model of [7] is briefly reviewed and slightly simplified, focussing on short term fading. The model resembles the Parsons and Turkmani channel model [8], with the different signal paths somewhat grouped.
We consider the up-link of several mobile stations to one multiple antenna base station. The base station antennas have omnidirectional antenna patterns in azimuth. Only azimuthal angles are further considered in the propagation geometry.
The complex baseband signal $S(t)$ to be transmitted is gmsk modulated. The only difference with OQPSK lies in the pulse shaping, which is Gaussian instead of cosinusoidal [9]. Before the transmitted signal reaches the base station, it is first reflected from local features (index q, from 1 to L). Afterwards, it is reflected for a second[2] time from larger objects called dominant reflectors (index p, from 1 to M). This gives rise to the following double summation [10]:

$$\vec{X}(t) = \sum_{p=1}^{M} S(t-T_p)e^{j\omega_c t}A_p e^{j\phi_p} \cdot \sum_{q=1}^{L} A_q^{(p)} e^{j\phi_q^{(p)}} e^{j\omega_{p,q}t}\vec{a}(\theta_{p,q}) \quad (1)$$

The mobile station is moving with velocity V. The local reflectors (1..L) are uniformly distributed in space on a circle with radius r around the mobile [11], Figure 1. Magnitudes $A_p^{(q)}$ and phases $\phi_q^{(p)}$ are uniformly distributed. The Doppler-shifts $\omega_{p,q}$ equal :

$$\omega_{p,q} = 2\pi f_c \frac{V}{c}\cos\epsilon_{p,q} \quad (2)$$

with $\epsilon_{p,q}$ the angle between the velocity vector of the mobile and the position vector of the local reflector with index q. The latter position vector has its origin in the centre of the circle with radius r (Figure 1).
The array vector $\vec{a}$ is evaluated for angles $\theta_{p,q}$:

$$\theta_{p,q} \in [\theta_p - \triangle\theta_p, \theta_p + \triangle\theta_p]$$

The summation over q leads to the well known Rayleigh fading. Rice fading is not treated here since the attention is focussed on urban propagation. The magnitudes A_p and multipath delays T_p are taken from GSM Recommandation 05.05. The phases ϕ_p are uniformly distributed.
Every mobile station gives rise to an expression as in (1). The contributions of all mobiles are superposed and noise is added. This signal is subsequently low-pass filtered and sampled at the optimum[3] time instant. These samples are further processed by the algorithms described in the next section.

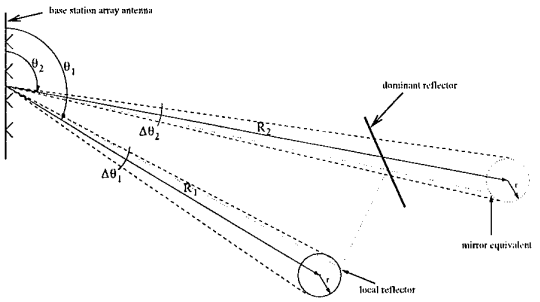


Figure 1: Definition of the angles and parameters used in the propagation model. A direct path and a reflected path are depicted.

3 Algorithm

For every mobile user and on a burst by burst basis, given the desired outputs[4] $d(t_k)$ and the baseband signal samples $x_i(t_k)$, an overdetermined set of equations can be formulated with the complex antenna weights w_i as unknowns[5] :

$$X^T(t_k) = [x_1(t_k)\ldots x_m(t_k)] \quad (3)$$

$$\begin{bmatrix} d(t_1) \\ \vdots \\ \vdots \\ d(t_k) \end{bmatrix} = \begin{bmatrix} X^T(t_1) \\ \vdots \\ \vdots \\ X^T(t_k) \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} \quad (4)$$

$$D(k) = X(k)W \quad (5)$$

With m equal to the number of antennas.

[2] Eventually, not always.

[3] The transmitter symbol clocks of the mobiles are not necessarily in phase.

[4] The 26 mid ambule bits of a GSM 'normal burst'.

[5] $X(t_i)$ is a vector, whereas $X(k)$ is a matrix.

The total squared error is defined as[6] :

$$\mathcal{E} = \sum_{i=1}^{k} e^2(t_i) = \sum_{i=1}^{k} (d(t_i) - X^T(t_i)W)^2 \quad (6)$$
$$= (D(k) - X(k)W)^\dagger (D(k) - X(k)W)$$

The weights w_i are chosen to minimize this error :

$$W = W(k) = [X^\dagger(k)X(k)]^{-1} X^\dagger(k)D(k) \quad (7)$$
(deterministic)

If the mid ambule or training sequence length would be unlimited, averaging of (7) over time would lead us to the following well-known statistical expression in terms of auto- and cross-correlations :

$$W = W(k) = R_{XX}^{-1} r_{Xd} \quad \text{(statistic)} \quad (8)$$

Expression (7) can readily be implemented. Other (recursive) algorithms exist however. They are known to be numerically more robust as they avoid calculating and inverting the matrix $P^{-1}(k) = X^\dagger(k)X(k)$ explicitly. In these algorithms, the matrix $P(k)$ is initialized and further updated recursively, using the samples $X(t_k)$ directly. This leads to better conditioned calculations [12]. The stabilized RLS algorithm (*S-RLS*,[13]), is summarized in Table 1. The QR-RLS algorithm(*QR-RLS*,[12]), which calculates and updates the QR decomposition of the matrix P, is summarized in Table 2. These algorithms can be easily extended to perform equalization in time as well as in space, and are further evaluated in section 4.

Instead of a training sequence, the constant modulus property can be used to equalize the received signal. A kind of decision feedback is then used. This works well for low signal to noise levels, so that most of the decisions are made correct. With spatial multiplexing however, the power of the co-channel interference is as high as the power of the wanted signal. Mere decision feedback without *any* initial training to discern the different users will not perform very well. The following can be done : assume all possible sequences of 142 bits[7], use each of these as the desired signal $D(k)$ in formula (7) and calculate the error (formula (6)). If the number of mobile users is f.i. three, the three bit-sequences generating the smallest error are most probably the three transmitted bit-sequences. This reduces the number of training bits to a minimum. Less than 142 bits can be used combined with a Viterbi-like elimination procedure to limit the search-complexity.
Other blind algorithms exist which perform projections on the finite alphabet sequences [14]. Whatever procedure used, the most optimal weights are obtained by using the whole transmitted sequence as a reference. This corresponds to no projection or decision errors at all, and will be referred to as the Optimal Least Squares (OLS) result.

$P(0) = \delta$ [I]
W(0) = [0]
for k = 1 to 26 do
$$\chi(k) = P(k-1)X(t_k)$$
$$\alpha(k) = \frac{1}{\lambda + X^\dagger(t_k)\chi(k)}$$
$$P(k) = \frac{1}{\lambda}[P(k-1) - \alpha(k)[\chi(k)\chi^\dagger(k)]]$$
$$W^*(k) = W^*(k-1) + \alpha(k)\chi(k)[d(t_k) - X^T(t_k)W(k-1)]^*$$
$$e(t_k) = d(t_k) - X^T(t_k)W(k-1))$$
end (for k)

Table 1: Stabilized RLS algorithm

4 Simulation results

A scenario with three mobile stations was envisaged, all moving at the same speed of 100 km/h and at a distance of 2 km from the base station. The number of multipaths (dominant reflections, M) equals 12, each of them consisting of 10 (L) local reflections.
The corresponding delay, power and angle spread for each multipath and user is given in the appendix, Table 4. Delay and power spread values are taken from GSM Recommendation 05.05 [15] (typical case for urban area).
After calculation and superposition of the signals of all three users on all (maximum 10) antennas, thermal noise is added and the resulting composite signal is low pass filtered and sampled at the appropriate time instant. A linear array is assumed with the antennas spaced have a wave-

[6] The forgetting factor λ is set to 1, and we work on a burst by burst basis.

[7] user bits+training bits in one GSM normal burst

P(0) = [0]
W(0) = [0]
for k = 1 to 26 do
  for i = 1 to N do
    if $(x_i(t_k) = 0)$ then
      $c = 1$
      $s = 0$
    else
      $r_{ii}(k) = \sqrt{r_{ii}(k-1)r_{ii}^*(k-1) + x_i(t_k)x_i^*(t_k)}$
      $c = r_{ii}(k-1)/r_{ii}(k)$
      $s = x_i(t_k)/r_{ii}(k)$
    end (if then else)
    for j =(i+1) to (N+1) do
      $(x_{N+1}(t_k) = d(t_k))$
      $r_{ij}(k) = s^*.x_j(t_k) + c.r_{ij}(k-1)$
      $x_j(t_k) = c.x_j(t_k) - s.r_{ij}(k-1)$
    end (for j)
  end (for i)
end (for k)
$W = -1.R^{-1}(1..N, 1..N).R(1..N, N+1)$

Table 2: QR-RLS algorithm

length from each other.
All users are transmitting at equal power. The mid ambule, which corresponds to the Base Station Colour Code (BSCC), is used to determine the weights. Every mobile station uses a different mid ambule (the term BSCC is then not appropriate anymore). An equalized burst typically looks like the one in Figure 2.
To have an idea of the bit error rate, the average signal amplitude s and the square deviation σ_s is calculated for a whole burst (i.e. over 142 bits) and substituted in the following formula :

$$BER = \frac{1}{2} erfc(s/\sqrt{2\sigma_s}) \quad (9)$$
$$erfc(x) = \frac{2}{\pi}\int_x^\infty e^{-t^2} dt$$

Care has been taken to use this formula only for 'stationary' burst, i.e. for burst which don't show too much fading, as the example plotted in Figure 3 makes clear. For such high faded bursts the square deviation would be estimated too high and Formula (9) would be an inappropriate estimate for the BER.
The convergence speed of the Stabilized RLS algorithm is plotted in Figure 4 and 5 for 1 and 2 taps respectively (tap spacing 3.692 μs), with

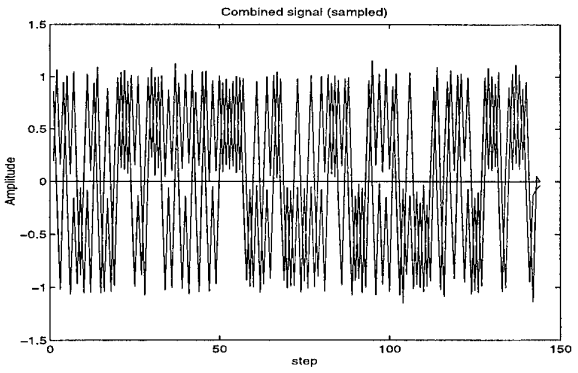


Figure 2: Typical equalized burst : mobile 1, speed 100 km/h, 2 co-channel interferers of equal power, 10 antennas, 1 tap, S/N=-50db. I- and Q-channel. (The peaks represent the sampled values.)

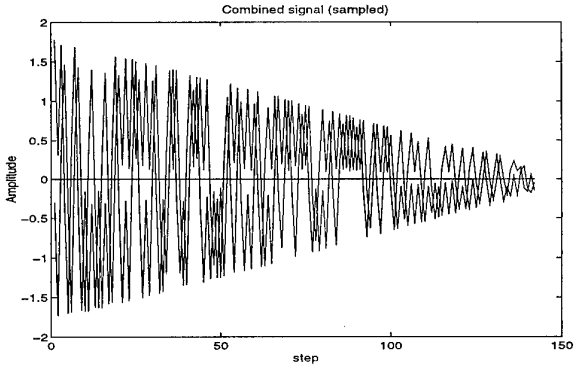


Figure 3: Equalized burst under severe fading : mobile 2, speed 200 km/h, no co-channel interferers, 10 antennas, 1 tap, S/N=-50db. I- and Q-channel. (The peaks represent the sampled values.)

the number of antennas 'm' as a parameter. The 'zero-crossings' are exploited as extra information. Mind that in practice (for GSM) the maximum number of iterations equals 26. The Optimal Least Squares result is the result after 142 iterations. The curves for m=2 and 3 at the bottom of the graphs are not discernable. The higher the number of antennas, the better the output S/N ratio after 100 iterations. It can be seen thatt the three mobile users can already be resolved with 3 antennas.
Additional time-equalization decreases the convergence speed and does not improve the output S/N ratio very much, especially when the number of antennas is high (m > 6). For a high

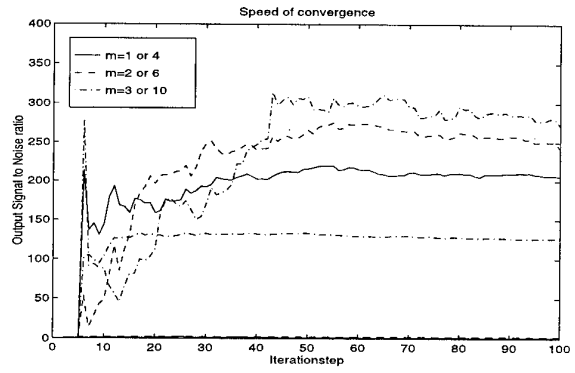


Figure 4: Speed of convergence : mobile 1, speed 100 km/h, 2 co-channel interferers of equal power, m antennas, 1 tap, S/N=-20dB. The ratio of the output Signal power to thermal plus co-channel Noise power is plotted on a linear scale as function of the iteration step. The Stabilized RLS algorithm was used. The curves for m=2 and 3 are not discernable.

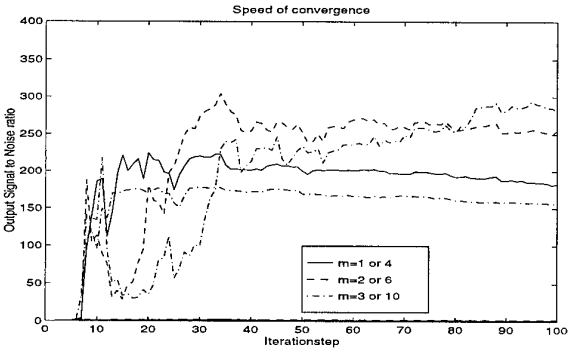


Figure 5: Speed of convergence : same scenario as in Figure 4 but with the number of taps n=2. Tap spacing = 3.692 μs.

number of coefficients (m×n > 8) convergence is not reached after 26 iterations.

The difference between the output S/N ratio computed with the QR-RLS algorithm and the S-RLS algorithm is only 1% after 15 iterations and 0.01 % after 26 iterations (computed with *MATLAB* on a HP Apollo 715/80). Henceforth, the Stabilized RLS algorithm is used. If the precision of the calculations is decreased, the difference in performance might be more pronounced. The bit error rate (BER) as function of the ratio of the signal power to the thermal noise power at the input (S/N) is plotted in Figure 6 for the Stabilized RLS algorithm after 26 iterations and in Figure 7 for the Optimal Least Squares result. Parameter is the number of antennas 'm'. The number of taps 'n' equals 1. The signal amplitude and the thermal plus co-channel noise power at the equalized output are substituted in Formula (9). A signal to noise ratio of zero means that the *average* noise power is equal to the *average* signal power. As the signal is consequently sampled at its *maximum* value, the power of the sampled useful signal is still approximately 3dB higher than the power of the sampled noise.

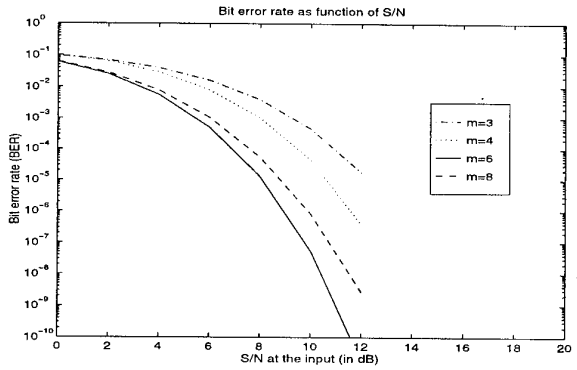


Figure 6: Bit Error Rate as function of the Signal to thermal Noise ratio at the input : mobile 1, speed 100 km/h, 2 co-channel interferers of equal power, m antennas, 1 tap. The number of iterations equals 26. The Stabilized RLS algorithm was used.

INPUT S/N	0	10	20	30	40	50
OUTPUT S/N	2.4	8.7	17	21	23	23

Table 3: Tabulation of the input Signal power to thermal Noise power ratio (first row) and the ratio of the output Signal power to thermal plus co-channel Noise power (second row)(S/N in dB, m=10, n=1).

For better comparison of the Optimal Least Squares result with the S-RLS algorithm, the ratio of the output S/N ratio of the OLS result and the output S/N ratio obtained with the S-RLS algorithm is plotted in Figure 8 ($[S/N]^{output}_{OLS}/[S/N]^{output}_{S-RLS}$) as function of the input S/N ratio. The improvement fluctuates around a factor of 1.5, and is more pronounced for a higher number of coefficients as convergence is then slower (see also Figure 4). The output S/N ratio after 26 iterations seems quite satisfactorily compared to the optimum.

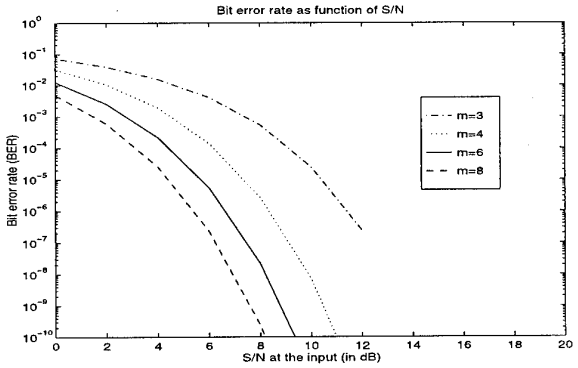


Figure 7: Bit Error Rate as function of the Signal to thermal Noise ratio at the input. Same scenario as in Figure 6. The Stabilized RLS algorithm was used but with the whole bit-sequence as a reference, thus corresponding to the Optimal Least Squares result.

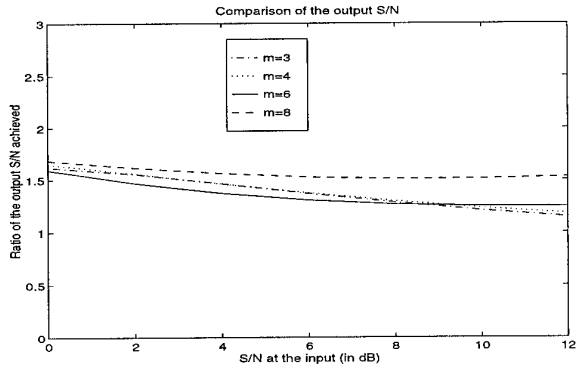


Figure 8: Comparison of the Optimal Least Squares result with the S-RLS algorithm results. $[S/N]^{output}_{OLS}/[S/N]^{output}_{S-RLS}$ as function of the input S/N ratio (averaged out over 5 bursts).

A less complex environment was simulated, still with 3 co-channel users, but each with only 3 dominant multipaths (M), each in turn consisting of 10 local reflections (L). The evolution of the output signal to noise ratio as function of the input signal to noise ratio is given in Table 3 (m=10 and n=1). It can be concluded that even for a sufficient number of antennas (m=10) the output S/N does not decrease very much at low input S/N due to fading.

5 Conclusions and Remarks

Blind identification algorithms based on the finite alphabet property of the transmitted symbols could be good alternatives [14], as they are not hindred by the limited length of training sequences. The nonstationarity of the channel however will still limit the performance. A short symbol sequence known in advance will still be necessary to discern the different spatially multiplexed mobile users.

Additional time-equalization does not improve the output S/N ratio substantionally. A more fractional tap spacing smaller than 3.692 μs would be more suited to the particular delay spread of the channel (cfr. Table 4).

Due to the high precision of the calculations, the S-RLS algorithm performs as well as the QR-RLS algorithm. A more limited precision might pronounce the difference more. A training sequence of length 26 bits suffices to equalize the desired signal.

Carrier tracking and timing reconstruction in a multi user environment with Space Division Multiple Access causes new problems. Residual carrier tracking should be performed on the digital baseband level, as the carriers cannot yet be discerned on the analog RF level. A certain oversampling will be necessary for the symbol timing reconstruction, or a synchronisation of the multiple users.

In the foregoing simulation, symbol timing was taken to be known. The knowledge of this exact timing already discriminates the co-channel interferers from the desired useful signal. The results obtained with such a priori timing knowledge are thus better than can be expected in an even more realistic simulation.

6 Acknowledgements

I wish to thank IMEC for the financial support, and my collegues for the interesting discussions and suggestions.

7 Appendix

The angular spread and power spread for the three mobile users is visualized in Figures 9,10 and 11 and tabulated together with the delay spread in Table 4.

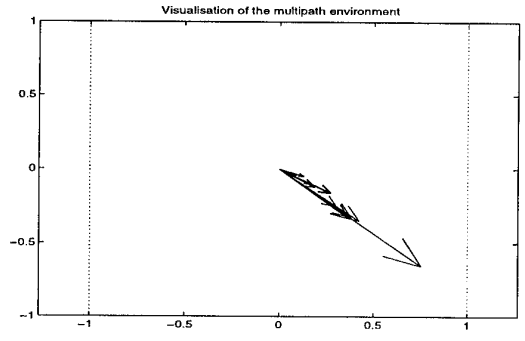


Figure 9: Visualization of power and angle of incidence of the different multipath components (12 in total) corresponding to mobile 1. The length of the arrow is proportional to the power of the component, linear scale.

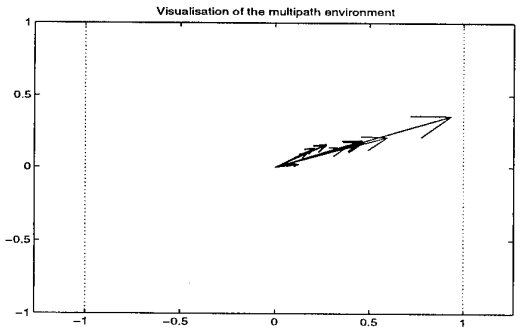


Figure 10: Visualization of power and angle of incidence of the different multipath components corresponding to mobile 2.

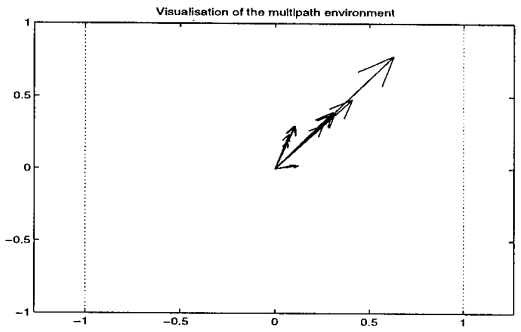


Figure 11: Visualization of power and angle of incidence of the different multipath components corresponding to mobile 3.

Mobile user 1			
Path number	Delay in μs	Power in dB	Angle in degrees
1	0.0	-4.0	-40
2	0.1	-3.0	-40.5
3	0.3	0.0	-41
4	0.5	-2.6	-39.5
5	0.8	-3.0	-42
6	1.1	-5.0	-30
7	1.3	-7.0	-30.5
8	1.7	-5.0	-31
9	2.3	-6.5	-33
10	3.1	-8.6	-20
11	3.2	-11.0	-22
12	5.0	-10.0	-24
Mobile user 2			
1	0.0	-4.0	20
2	0.2	-3.0	20.5
3	0.4	0.0	21
4	0.6	-2.0	19.5
5	0.8	-3.0	22
6	1.2	-5.0	30
7	1.4	-7.0	30.5
8	1.8	-5.0	31
9	2.4	-6.0	33
10	3.0	-9.0	10
11	3.2	-11.0	12
12	5.0	-10.0	14
Mobile user 3			
1	0.0	-4.0	50
2	0.2	-3.0	50.5
3	0.4	0.0	51
4	0.6	-2.0	49.5
5	0.8	-3.0	52
6	1.2	-5.0	70
7	1.4	-7.0	70.5
8	1.8	-5.0	71
9	2.4	-6.0	73
10	3.0	-9.0	10
11	3.2	-11.0	12
12	5.0	-10.0	14

Table 4: Delay, power and angle spread of the corresponding multipaths for all three users.

References

[1] W. C. Jakes, "Microwave mobile communications", John Wiley and Sons NY, 1974.

[2] O. van de Wiel, L. Vandendorpe, "Space Diversity for GSM Channels : Comparison of Performance", *Proc. IEEE Second Symposium on Communications and Vehicular Technology in the Benelux*, Louvain-la-Neuve, Belgium, Nov. 1994.

[3] J. Fernandez, I. R. Corden, M. Barrett, "Adaptive array algorithms for optimal combining in digital mobile communication systems", *IEE/U.R.S.I. eighth international conference on antennas and propagation*, conference publication number 370, pp.983-986, 1993.

[4] P. Zetterberg, *The Spectrum Efficiency of a Basestation Antenna Array System for Spatially Selective Transmission.* Internal Report, Royal Institute of Technology, Department of Signals, Sensors & Systems, Stockholm.

[5] A. F. Naguib, B. Khalaj, A. Paulraj, T. Kailath, "Adaptive channel equalization for TDMA digital cellular communications using antenna arrays", *Proc. ICASSP*, vol. 4, pp. 101-104, 1994.

[6] B. Agee, K. Cohen, J. Reed, T. Hsia, "Simulation Performance of a Blind Adaptive Array for a Realistic Mobile Channel", *Proc. VTC*, vol. I, pp. 97-100, 1993.

[7] G. Raleigh, S. N. Diggavi, A. F. Naguib, A. Paulraj, "Characterisation of Fast Fading Vector Channels for Multi-Antenna Communication Systems", *Proc. of 28th annual Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, Nov. 1994, pp 853-857.

[8] J. Parsons, A. Turkmani, "Characterization of mobile radio signals: model description", *IEE Proc. I.*, 138, (6), pp. 549-556, Dec. 1991.

[9] K. Murota, K. Hirade, "GMSK modulation for digital mobile radio telephony", *IEEE Trans. Communications*, vol. 29, no. 7, pp. 1044-1050, Jul. 1981.

[10] S. Brebels, T. Denotte, *Algoritmen voor adaptieve antennes in cellulaire mobiele telefonienetwerken.* MSc. thesis, Katholieke Universiteit Leuven, Belgium, Jun. 1994.

[11] W. C. Y. Lee, *Mobile Communications Design Fundamentals.* John Wiley & Sons, inc.,New-York, 1986.

[12] C. Ward, P. Hargrave, J. McWhirter, "A Novel Algorithm and Architecture for Adaptive Digital Beamforming", *IEEE Trans. on Antennas and Propagation*, vol. AP-34, no. 3, pp. 338-346, March 1986.

[13] J. F. Böhme, B. Yang, "Parallel and Numerically Robust Algorithms for Adaptive Interference Cancellation", *Proc. of an ESA workshop on Advanced Beamforming Networks for Space Applications*, ESA-ESTEC, Noordwijk, The Netherlands, Nov. 1991, pp. 3.2.1-3.2.8.

[14] A.-J. van der Veen, S. Talwar, A. Paulraj, "Blind identification of FIR channels carrying multiple finite alphabet signals", *IEEE ICASSP '95*, Detroit (MI), May 1995.

[15] *GSM Rec. 05.05 (Draft 3.3.1)*, Belgacom, 1991.

DISCUSSION

Discussor's name: C. Goutelard

Question/Comment:

Votre matrix de corrélation a les dimensions du nombre d'antennes, donc 8 au maximum. Comment faites-vous pour séparer 120 trajets multiples?

Translation:

Your correlation matrix is the size of the number of antenna, that is a maximum of 8. How do you separate the 120 multi-paths?)

Author/Presenter's reply:

Les signaux reçus ne contiennent que les trois informations (des 3 mobiles). Le but de la méthode n'est pas de séparer les 120 trajets mais d'extraire les informations utiles. Pour cela, la connaissance de la séquence d'aprentissage est nécessaire et complémente les données collectées durant le processus de récursion. En fait la matrice a comme nombre de rangées le nombre d'échantillons récoltés jusqu'à la séquence k et comme nombre de colonnes le nombre de coefficients de pondération des antennes.

Translation:

The signals received contain only three items of information (from the three mobiles). The aim of the method is not to separate the 120 paths but to extract the useful information. To do that a knowledge of the learning sequence is necessary to know the learning sequence and complements the data collected during the recursion process. In fact the matrix has the same number of rows as the number of samples collected up to the sequence k, and the same number of columns as the number of the antennas' weighting coefficients.

TESTING HF-MODEMS: DSP'S OPEN NEW POSSIBILITIES AND MAKE NEW DEMANDS ON REAL-TIME HF-SIMULATORS

L. Van der Perre
A. Van de Capelle
KULeuven, Div. ESAT-TELEMIC
Kard. Mercierlaan 94
3001 Heverlee
Belgium
E-mail: Liesbet.Vanderperre@esat.kuleuven.ac.be

SUMMARY

A test set-up for HF-modems with a hardware implemented, mainly analogue simulator based on the Watterson model [1], has been built at the division ESAT-TELEMIC of the K.U.Leuven. It has shown to be a very useful tool to test narrowband HF-modems under stable channel conditions.
In order to improve the transmission quality, more advanced HF-modems are being developed nowadays. The use of DSP's has made it possible to provide large interleaving blocs, intelligent error correcting coding schemes, channel estimators, adaptive equalizers, etc...
The qualities of such modems, that try to adapt to the present channel conditions, can only properly be tested on, provided they are subjected to varying channel conditions. The technological evolution that gave rise to these intelligent modems, has created the need for a simulator that can imitate a varying ionosphere.
This is what is aimed using a digital signal processor for the realization of an HF-channel simulator. Besides the programming capabilities that are automatically present when using DSP's for channel simulation, other advantages come when working digitally. Making adjustments to a design becomes a very simple operation, and some simulation functions are easier to implement when working with digital technology. The design of this digital simulator is done using a digital signal processing design tool called Signal Processing Worksystem (SPW).
The use of spread spectrum modulation techniques on the HF-channel, creates a new challenge for the developers of HF channel simulators. In the first place because there is no wideband model for the HF-channel available. Besides that, the realization of a real-time simulator that can handle signals with bandwidths up to 1 or 2 MHz, will be a complex task, both from a hardware as from a software point of view.

1 INTRODUCTION

The designer of HF-communication equipment faces the following problem: testing of his equipment under real-life conditions requires a distance of hundreds, even thousands of kilometers between transmitter and receiver. He has the additional problem that it is very difficult to know the channel conditions at every instant.
Therefore, a test set-up for HF-modems in the laboratory is very useful in order to facilitate the research, development and fine-tuning of ever faster and more reliable modems for HF fax- and data-communication. The test set-up creates the possibility to know, control and eventually later reconstruct the identical test conditions exactly. This is a power we clearly do not have over the physical ionosphere, which is continuously changing.
The heart of the test set-up for HF-modems is a real-time channel simulator (RTCS) for HF-radiolinks. The settings of the simulator can be varied over a wide range, thus giving us the possibility to subject the modems under test to conditions going from typical to exceptional and worst case.

2 THE PRESENT TEST SET-UP

A. Analogue hardware real-time channel simulator

A mainly analogue hardware simulator, has been developed and built at the K.U.Leuven during the work of several years. The simulator is based on the Watterson model, that is widely accepted as a good model for the ionospheric channel as long as the supposed time and bandwidth periods are not exceeded.
The simulator can provide up to 5 propagation paths, each having their own delay and attenuation circuit. The simulator also includes a white gaussian noise source and a source imitating the behavior of an

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

interferer. The ranges over which the settings of the circuits can be made, largely suffice the CCIR-norm concerning channel simulators for the ionospheric communication.. A scheme of the simulator is given in figure 1.

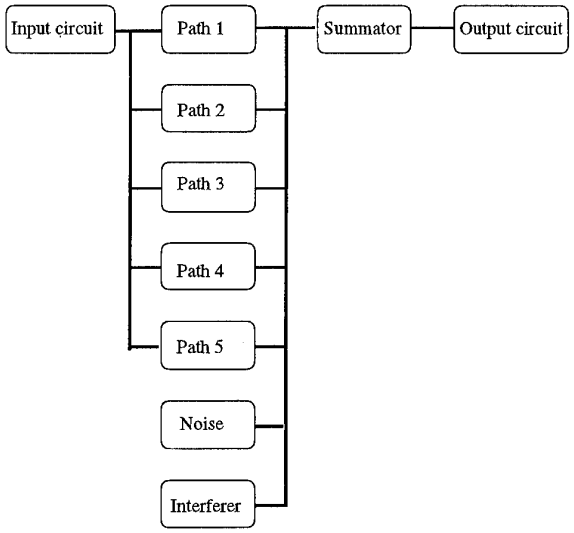
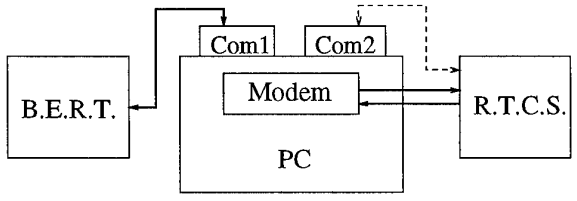


Figure 1: Scheme of the HF-simulator

Through a serial interface with the microprocessor present in the simulator, the settings of the simulator can be sent trough. The microprocessor is on its turn connected to the simulator circuits via a flat cable.

B. Test set-up for HF-modems

Besides the channel simulator and the modem under test, the set-up comprises a bit error rate counter that keeps track of the transmission quality, and a PC that controls all the components of the set-up and the communication between them [2]. The structure of the set-up with the data and control streams are illustrated in figure 2.



——— : data

-------- : control

R.T.C.S. = Real-Time Channel Simulator

B.E.R.T. = Bit Error Rate Tester

Figure 2: Structure of the test set-up

The figure shows the case where the modem is still in a developing stage, implemented as an algorithm on a DSP-board present in the PC. We can also test separate instruments with the set-up, provided they have a baseband output, since the channel simulator can not handle high frequencies.

The simulator is a very useful tool to analyze the behavior of modems under stable channel situations. For e.g. the traditional simple FSK-modem, these tests suffice to predict the complete modem behavior. The test set-up has been used to verify bit error rates generated by a prediction program for the FSK.-modem under different channel conditions [3].

A C++ program has been completed to provide a windows-driven, user friendly interface to the simulator.

Space and connectors have been provided in the simulator to include a second identical channel. This would allow us to do tests on duplex links and communication protocols. Lack of time and energy and the idea that a more advanced simulator is needed, have made that the second channel simulator has not been realized so far.

3 HF-MODEMS USING DSP'S

A. Communication problems

Data and fax and digital communication in general take up an increasing part of the telecommunication traffic. In order to improve the transmission quality, more advanced data-modems for the HF-channel are being developed nowadays. They have the clear advantage over voice modems, that much more transmission delay can be accepted, making the bit error rate the first demand.

The ionosphere is a continuously changing medium, giving rise to specific problems if you want to use it as a communication channel. Both the time and the frequency spreads experienced on the ionospheric channel, are relatively big, causing severe distortion of the transmitted signal.

The time variations can be divided in random fluctuations around an average value due to ionospheric scintillations on the one side, and global variations of that average value due to the structural changes of the ionosphere on the other side.

Only the first kind of changes is included by the Watterson model. The structure of the ionosphere is known to vary daily however. Consequently, multipath components appear and disappear, and their differential time delays can not be considered constant but over very short time spans. Therefore, the Watterson model is valid only for time periods up to 10 msec.

The modeling of the structural variations of the ionospheric communication channel, can be done by

extracting the propagation parameters from ray-tracing programs that use appropriate models for the electron density profile and hence the refractive index structure of the ionosphere.

B. DSP solutions

The use of DSP's has made it possible to design intelligent modems without increasing the cost or the transmission delay dramatically. Indeed, the technological evolutions have made digital computing and memory capacity ever cheaper and more powerful. These benefits are exploited in the HF modem project in cooperation with SAIT-devlonics, with the support of the region of Brussels-Capital through IRSIA.
Using large interleaving blocs reduces the errors that tend to occur in bursts in ionospheric communication, to individual random errors, as illustrated below, in figure 3.

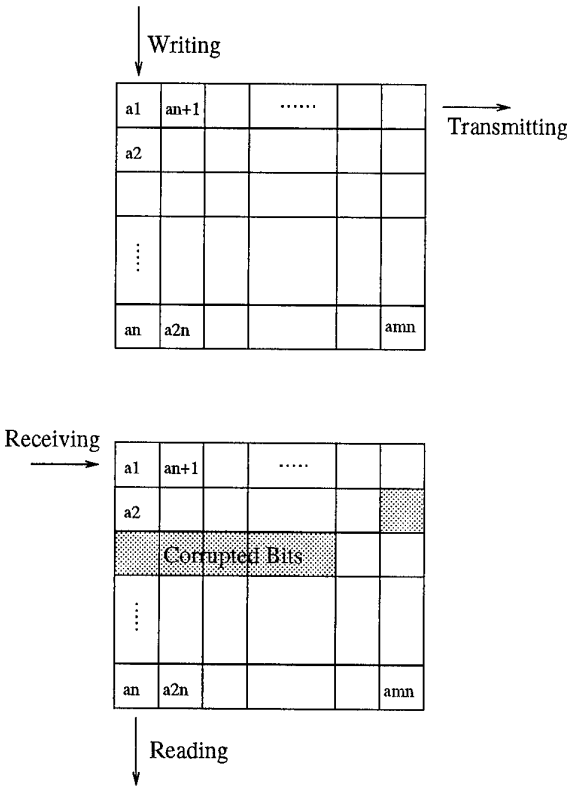


Figure 3: Interleaving process

The larger the interleaving matrices used, the larger the delay introduced and the bigger the memory needed both in the transmitter and the receiver. Unfortunately, the large bursts of errors encountered on the ionospheric channel force us to use relatively large interleaving blocs if they are to perform their job satisfactory.

Detecting and correcting individual errors can be done by intelligent coding schemes, such as a trellis algorithm. Although the Viterbi algorithm was originally designed to decode convolutional encoded messages [4], it has also shown to be very well suited to perform maximum likelihood detection on an uncoded message signal, received after transmission through a channel that may introduce intersymbol interference. In this way, it can be useful to overcome intersymbol interference encountered in HF-communications due to the multipath nature of the channel.
Via channel equalizers, modems can also try to adapt to the present channel conditions. Problems occur however on the HF-channel because of the frequency selective fading, causing deep zeroes in the channel impulse response. As the equalizer tries to compensate this, it will amplify these frequencies, on which mostly noise is received, strongly.
Both the Viterbi algorithm and the equalizer technique suppose the channel conditions and multipath composition are known exactly. This implies that a good channel impulse response estimation algorithm, is needed. As for the coding algorithms, a trade-off between performances at the one side, and complexity and price on the other side, has to be made.
On a higher level, an appropriate communication protocol, like a good ARQ-protocol, can be used to recover blocs of bits that still were lost due to temporarily bad channel conditions. Although a protocol might reduce the net transmission rate significantly, it is the only way out if error free transmission has to be assured. It is important that a very robust algorithm is used on the return channel, as experiments have proven that the transmission rate and quality decreases dramatically if acknowledges are not received correctly.
The improved computing power has also stimulated the use of spread spectrum modulation techniques on the HF-channel to overcome some of the severe communication problems. This subject will be handled separately in paragraph 5.

C. Modem tests

The qualities of modems that try to know and adapt to the present channel conditions, can only properly be tested on, provided they are subjected to a varying channel. The speed at which channel variations take place, is an important factor for modems using repetitive training-sequences for their channel estimators. Heavy ionospheric disturbances and drastic changes of the ionosphere as occur at sunset and sundown, might cause intelligent modems to fail completely. Therefore, it is important that these

conditions can also be simulated by an advanced test set-up for HF-modems.
The test set-up described in paragraph 2, is still useful in determining steady state behavior. For example, critical signal to noise ratios can be determined, and differential time delays between propagation modes causing severe problems with particular modems, have been discovered. In figure 4, the performances of two modems in a channel situation with two propagation paths are displayed.

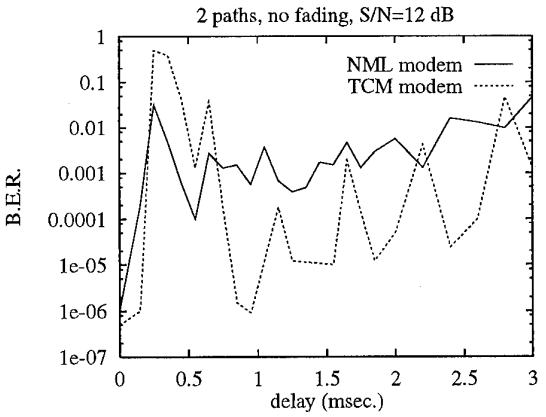


Figure 4: Modem tests

The tests show the bit error rates as a function of the differential delay between the two modes, for a modem using Near Maximum Likelihood detection (NML-modem) and a modem based on Trellis Coded Modulation (TCM-modem). These two modems have been tested and compared in detail in the scope of the above mentioned project.

4 DIGITAL CHANNEL SIMULATOR

A. Advantages of a digital channel simulator

The need to have a real-time channel simulator that can imitate varying propagation conditions, urged us to look out for a channel simulator that can fulfill our demands. The programming capabilities that are automatically present when using DSP's for channel simulation, make them attractive for the task. This issue is further explained in section E of this paragraph.
It goes without saying that general purpose, programmable DSP-boards are very useful in developing the prototype of a real-time channel simulator in particular [5], and as a design tool in a telecommunications research unit in general. Changes to the channel model can be elegantly extended to the simulator by simple software adjustments, whereas the modification of the analogue hardware simulator would imply developing new PCB's and re-assembling the instrument. In this way, designing and developing prototypes has become a much more flexible process.
Besides that, other advantages come when working digitally. Some simulation functions are, by their nature, easier to implement when working with digital technology. For example. delays are simply realizable thanks to the memory that is readily available when using DSP's. The classical presentation by Watterson of the HF-channel as a tapped delay line, has a direct translation in a DSP's architecture via software pointers.

B. Signal Processing Worksystem

The design of a real-time digital simulator was started using a digital signal processing design tool from Comdisco, called SPW. This tool facilitates the design and the simulation of communication systems. The user-friendly interface and the big libraries that are provided, enable the designer to save time and to work in a structured way.

The system comprises three application tools:

1. The Designer/Block Diagram Editor:
   This is the design environment where models can be built on a high level, in a hierarchical manner, and using standard blocs from libraries available in the software package. User-designed blocs and output from the Signal Calculator and the Filter Design System, can also be used.

2. The Signal Calculator:
   As can be understood from its name, signals can be generated through the processing of basic functions available in the system.

3. The Filter Design System:
   This design provides an easy way to design different kind of filters. These filters can then again serve as building blocks in the Block Diagram Editor.

The system also contains several help-tools:

1. The Simulator:
   This tool permits the designed systems to be simulated and analyzed.

2. The Code Generation System:
   Code, to be compiled for a target processor, is generated automatically. Only a limited number of the most common digital signal processors is supported, as it is impossible to keep up with rapidly changing market. For the unfortunate user possessing a non-supported processor, generic C code can also be output.

Figure 5 shows the design cycle that was followed, typically containing a number of repetitive adjustments and simulations.

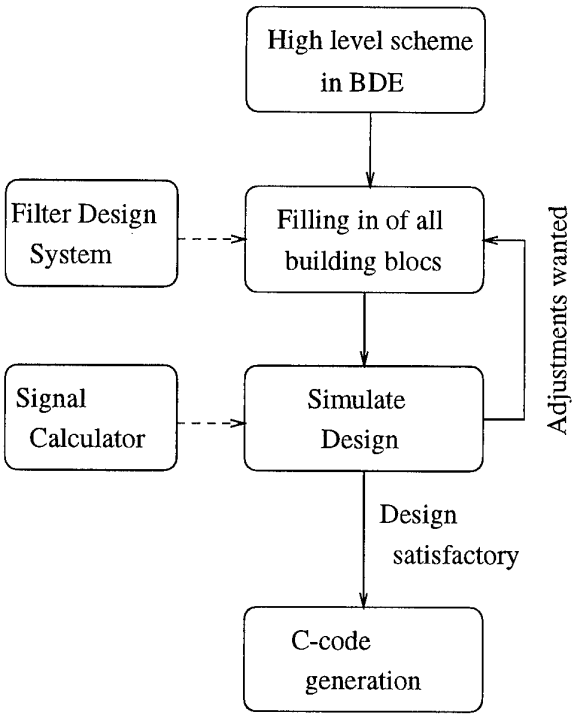


Figure 5: Design cycle followed with SPW

C. Design of the digital channel simulator

In order to facilitate signal processing, the real input signals are converted to analytical signals through the use of a Hilbert filter in the input circuit, as depicted in figure 6.

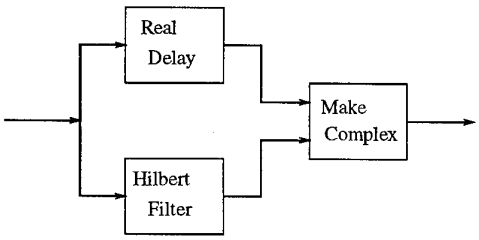


Figure 6: Input circuit of the digital simulator

The real delay in the arm at the top of the circuit is necessary to compensate for the delay that is inevitable introduced by the Hilbert transformation filter.
There has been opted for three different propagation paths, each consisting of two branches to simulate the two magneto-ionic components. The number of paths can easily be augmented as it asks only a very small effort to add an extra identical module, as mentioned above.
The circuit that clearly is the most difficult to implement is the Doppler spreading. A detailed study has been made on the different solution strategies that can be followed for this problem. Also, a well-founded choice for the sample frequency has been made, based on a trade-off between desired resolution and acceptable complexity, and resulting in the option for 40 kHz [6].
Two different noise sources are present in the simulator: one to model the white gaussian noise, the other to take into account the impulse noise that is also encountered on the HF-channel. At last, an adjustable interference source can be switched on.

D. Implementation of the digital simulator

As we are working with C31 DSP-boards which are not supported by the SPW - software, we are using the generic C-code as output from SPW. This code serves us as a guide in programming a real-time channel simulator onto the DSP-boards.
Of course, the scheme of the simulator as designed in SPW for software simulations, contains by itself a lot of useful information. For example, the calculation of the tapcoefficients of the filters is done automatically by the Filter Design System, leaving only a simple programming job to implement them.
Although it is easier to program in a high level language such as C, and all necessary compilers and linkers are provided, it is always interesting to take a look at the assembler code too. In that way, one can verify whether no major efficiency ameliorations can be made. This matter becomes a necessity as more complex structures have to be implemented (for example the realization of a wideband channel simulator, see paragraph 5), and an efficient use of the hardware is a must if we want to keep to the budget.

E. Simulator settings

As for the hardware simulator, an interface has to be written to set the parameters of the simulator. Since we also want to simulate structural changes of the channel in real-time non-stop tests, we have to be able to vary the settings in a continuous way.
One way that has been tried out with the present test set-up described in paragraph 2, is to program the variations on the PC. The PC could change the settings of the simulator during ongoing tests, via the RS-232 connection. However, early experiments to do tests with changing simulator settings in this way have discouraged us to further develop this method. The disturbances caused by the big stabilization time

of some of the circuits made the results not to be trusted.
The method where the changing propagation parameters are generated by the PC, could be used for the simulator implemented on the DSP-board as well, where the PC could communicate the updated settings to the DSP-board via the dual port memory. Although the chance on success of this method is certainly bigger with the DSP's, there is a more elegant solution advisable. Indeed, thanks to the inherent programming possibilities, the evolutions of the propagation parameters can be calculated on the DSP-board itself. In this way, eventual problems due to the interference between tasks of the host PC and the DSP-board and the communication between them, are avoided.

5 WIDEBAND DIGITAL SIMULATOR

A. The use of spread spectrum on the HF-channel

In a second IRSIA-project with SAIT-devlonics in Brussels and the R.U.G. (university of Gent), we are now investigating the use of spread spectrum techniques on the HF-channel. Direct sequence modulation in combination with a RAKE receiver is known to give a solid solution for multipath problems [7]. Even better, the multipath nature has been turned from a problem into an advantage as we can benefit the inherent diversity coming from the different propagation paths.
Although the algorithm for the RAKE receiver has been described in 1958, its realization has only recently become possible, thanks to the evolution of the technology. Even if cheap large memory and powerful computing capacity necessary to built an affordable RAKE receiver is today available, the large delay spreads experienced on the HF-channel will urge us to limit the spreading bandwidth.
If a sufficiently small time resolution is reached with the correlation process in the wideband receiver (asking a wide spreading), it might even be possible to separate the two magneto-ionic components (O - and X - wave), provided an antenna with diversity in polarization is used. This would mean a big improvement over the narrowband modems, as it would not only double the diversity, but in the same time remove the severe problem of polarization fading without needing the complex process of polarization filtering.
It is widely known that spread spectrum techniques have additional advantages in a military context. The signal is inherently coded and buried in the noise, which makes it both undetectable and unreadable for eaves-droppers. The despreading in the receiver of the wanted signal at the same time spreads the unwanted signals, giving a good resistance against jammers.

B. Model for the wideband HF-channel

If we want to use wideband communication on the HF-channel, we need a model for the channel that is valid for bandwidths up to a few MHz. The Watterson model, as mentioned in paragraph 3, is based on assumptions that are no longer valid if frequency dependencies have to be taken into account.
We intend to develop and use an extension of the Watterson model, conserving the delay line and one tap for each physical mode. The tapcoefficients are replaced by time variant FIR-filters in series with time variant path gains, as displayed in figure 7.

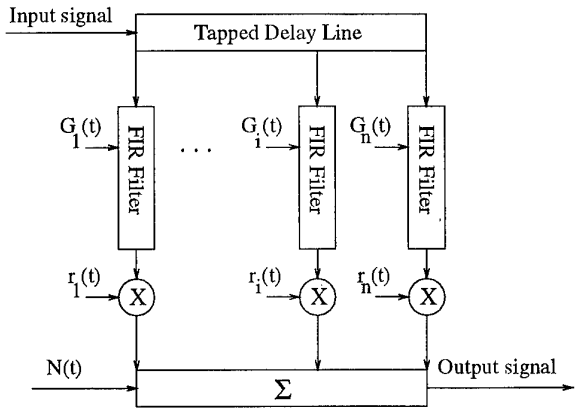


Figure 7: Extended Watterson model

The tapcoefficients of the filters reflecting both the frequency and the time dependencies, could be obtained from the output from prediction programs for ionospheric conditions.
In first instance, a simple parabolic model for the ionospheric density profile is assumed, resulting in analytical and easy to use expressions for the frequency dependencies of the propagation characteristics.
Besides the extension of the model for the propagation phenomena, the expressions for the noise and the interference will have to be adapted [8].

C. Wideband channel simulator

The next step would be the realization of the model for the wideband HF-channel in a wideband real-time channel simulator. The wider bandwidth itself will already complicate the hardware realization, having to handle signals up to a few MHz. The fact that the channel model is also more complex as we want to take frequency dependencies and time evolutions into account, will further reinforce this effect. In first instance, we are looking at a modular construction and the realization of some basic functions.
A splitting in modules that would both be logical from a physical point of view, and practical from a

hardware point of view, would be the division of the total bandwidth in subbands, as depicted in figure 8.

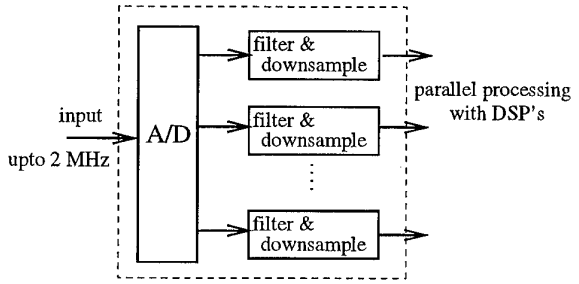


Figure 8: Suggested input circuit for a wideband channel simulator

For the individual subbands, the Watterson model would suffice. The variation of the parameters of the Watterson model over the different subbands (differential time delays of the different modes, gain factors, ...) would again ask a realistic analytical model or a reliable prediction program for the frequency dependencies of the propagation characteristics.
It is clear that using different processors working parallel, will be inevitable. All simulator functions will have to be realized as simple as possible. The efficiency of the code will play a much more important role than for the narrowband channel simulator, and intelligent programming as close to the machine as possible, will be a difficult but necessary task in order to reduce the number of parallel processors and the hardware costs to a minimum.

6 CONCLUSIONS

The use of DSP's has opened new possibilities for real-time channel simulation, making it a very flexible process thanks to the (re-) programmability of fast and nowadays payable signal processors.
The use of these same DSP's in the development of ever more performant modems, makes however new demands on the channel simulator. For intelligent narrowband modems, besides the random fluctuations, the structural changes will have to be imitated by an advanced channel simulator.
In order to properly test spread spectrum modems, a channel simulator that can handle bandwidths up to a few MHz and can also include frequency dependencies, is necessary. Both developing an appropriate model, and realizing the hardware that no doubt will be complex, are jobs in progres.
As a whole, it is clear that we can conclude that modern digital signal processing hardware is very beneficial for HF-communications. It has been illustrated in this paper that a substantial improvement of the transmission quality can be reached by using intelligent modems, either narrowband or wideband. This evolution makes shortwave communication, bearing in mind that it is cheaper than satellite communications, still attractive for data communication .

7 REFERENCES

1. Watterson, C.C., Juroshek, J.R. and Bensema, W.D., "Experimental confirmation of an HF Channel Model", IEEE trans. on communications, December 1970, Vol. COM-18, pp 792-803

2. Van der Perre, L., Van Troyen, D., Van Hove, H. and Van de Capelle, A., "Test set-up for HF-modems", *in* "First symposium on Communications and Vehicular Technology in the Benelux", IEEE, October 1993, Paper 6-2

3. Van Troyen, D., "Propagatiebepaling en simulatie van ionosferische H.F.-radioverbindingen", PhD thesis at the K.U.Leuven, 1995

4. Viterbi, A.J., "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm", IEEE trans. On Information Theory, April 1967, Vol. IT-13

5. Ehrman, L., Bates, L.B., Eschle, J.F. and Kates, J.M., "Real-Time Software Simulation of the HF Radio Channel", IEEE trans. on communications, August 1982, Vol. COM-30, pp 1809-1817

6. Feyaerts, F. and Van Hoof, E., "Simulation of the HF-channel with Digital Signal Processors", may 1994, master thesis at the K.U.Leuven

7. Price, R. and Green, P.E., "A communication technique for multipath channels", Proceedings of the IRE, 1958

8. Van der Perre, L. and Van de Capelle, A., "Study of the HF-channel with the view of the use of wideband communication techniques", *in* "HF 95 - nordic HF-conference", August 1995

DISCUSSION

Discussor's name: J. Harvey

Question/Comment:

Could you provide more detail on the physical causes and the remedy for the polarization fading?

Author/Presenter's reply:

Due to the geomagnetic field, every wave entering the ionosphere is split into two waves having orthogonal polarizations. These waves are known as the O- and X-wave. After their independent propagation through the ionosphere, the wave resulting from the summation of the two magneto-ionic components, has a polarization which most of the time is completely different from that at the transmitter, and rapidly changing, causing polarization fading if a simple antenna is used at the receiver.

To overcome the polarization fading, an antenna having diversity of polarization will be needed. The separation of the O- and the X-wave on the basis of their orthogonal polarization is a very complex process. We are now investigating whether the use of Direct Sequence Spread Spectrum communictions could provide us with a much easier method, based on the differential delay of the two magneto-ionic components. The research is still going on, but more details can be found in the paper I presented in the HF'95 Nordic HF Conference, given as a reference.

DISCUSSION

Discussor's name: P. Cannon

Question/Comment:

1. Comment: The Watterson model is the best model we have but as well as the limitations you have mentioned it is important to remember that it is based on a very limited experiment in both time and geographical area.

2. Question: Could you please explain what validation tests you have performed on your simulator?

Author/Presenter's reply:

1. Indeed, more extensive experiments are necessary to prove the reliability of the Watterson model. However, one of the benefits of the model is that it is widely used and also stated in the CCIR standards. This makes it possible to compare modems developed in different laboratories as they can be tested under similar conditions.

2. The simulator in fact consists of a number of PCB's, where blocks of circuits that perform different simulations of the propagation effects that are encountered on the HF-channel. This gives us the possibility to put a probe connected to a precise oscilloscope, at every stage of the simulation for the 5 propagation paths, the noise and the interference source. In combination with a function generator, the complete behaviour of the simulator has been checked in a qualitative and a quantitative way. A lot of effort and time have been put in the validation of the simulator's performances!

DISCUSSION

Discussor's name: N. J. Farsaris

Comment/Question:

What technique are you going to use in order to simulate both O and X-wave propagation?

Author/Presenter's reply:

We will have to handle the O- and the X-waves as two completely independent propagation paths - which is not in conflict with the physical reality. Of course, this will mean that the delay unit in the tapped delay line model will have to be sufficiently small to provide the necessary time resolution. This is one of the reasons the complexity of the simulator could become a problem.

A Flexibly Configurable Statistical Channel Model for Mobile Radio Systems with Directional Diversity

J. J. Blanz, P. Jung and P. W. Baier
University of Kaiserslautern, Research Group for RF Communications
P.O. Box 3049, 67653 Kaiserslautern, Germany, e-mail: blanz@rhrk.uni-kl.de

1 SUMMARY

A prerequisite for the performance analysis and optimization of future mobile radio systems by simulation is a realistic model of the mobile radio channel. Time variance and frequency selective fading as well as spatial radio spreading have to be considered in such a model to enable the validation of system concepts using directional diversity. In this paper, a flexibly configurable statistical channel model for mobile radio systems using directional diversity is presented. This channel model is based on the superposition of echos which originate from different scattering objects. The parameters of this model can be easily adjusted to various propagation areas as e.g. rural, urban, microcellular and picocellular environments. Therefore, the model which is the basis for a FORTRAN77 program is well suited to perform simulations, evaluations and comparisons of mobile radio systems. Simulation examples are given for typical mobile radio scenarios. The resulting time varying channel impulse responses for both omnidirectional as well as ideal non-overlapping sector antennas are compared. The applicability of the presented channel model to simulations of mobile radio systems is demonstrated for a new mobile radio system concept with joint detection (JD) and directive base station antennas

2 INTRODUCTION

An important issue in mobile radio communications is high capacity, which means that as many users as possible should be accommodated in the allocated frequency bands. Among the various measures to enhance capacity [1], approaches which have not yet been studied extensively are antenna diversity techniques applicable at the base stations (BS) and/or mobile stations (MS), i.e. the use of more than one antenna at BS and/or MS. In such multi-antenna applications, two different approaches can be followed, which may also be combined.

In the one approach termed space diversity, the different antennas are positioned at locations separated by e.g. a few wavelengths. Space diversity has the effect that several transmission channels between BS and MS exist which experience more or less independent fading processes. The effective fading depths can be reduced by space diversity [2].

In the other approach termed directional diversity, the different antennas are directive antennas positioned at virtually the same location, covering however different sectors of space. These sectors may be disjoint or overlapping. The radiated waves follow certain paths, which leave the location of the transmitter in certain directions and which impinge on the receiver location from certain directions. Measurement results obtained in the RACE (Research and Development of Advanced Communications Technologies in Europe) II project TSUNAMI (Technology in Smart Antennas for Universal Advanced Mobile Infrastructure) impressively illustrate the directional anisotropy of mobile radio channels for different environments [3]. In a multi-antenna system using directional diversity one can benefit from this directional anisotropy in several ways. At the receiver, by preferably evaluating the signals emerging from those antennas offering high carrier-to-interference (C/I) ratios, the effective C/I can be enhanced. At the transmitter, by radiating preferably into those directions in which wave propagation to the receiver has the lowest attenuations, the radiated power and the interference to other receivers can be reduced. As another favorable effect, the effective delay spread can be reduced by directional diversity because the subsets of paths addressed by directional antennas usually have lower delay spreads than the aggregate channel that would be measured between omnidirectional antennas [4]. Finally, like in the case of space diversity, a reduction of the effective fading depths can be achieved by directional diversity. To conclude, the application of multiple directive antennas in mobile radio systems

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

may lead to significant capacity benefits. This is also true for SDMA (Space Division Multiple Access) schemes [5, 6], which are closely related to directional diversity considered in this paper.

To quantify the benefits diversity techniques, in particular of the rather novel directional diversity, in mobile radio system concepts simulation is a powerful and inexpensive alternative to building testbeds A prerequisite for simulations is the availability of a realistic model of the mobile radio channel. In particular, the spatially anisotropic wave propagation leading to spatial spreading must be considered in such a model. In this paper, a flexibly configurable statistical channel model for mobile radio systems using directional diversity is presented. The parameters of this model which is available as a FORTRAN77 program can be easily adjusted to various propagation areas. Therefore, the model is well suited to perform simulations, evaluations and comparisons of mobile radio systems.

The paper is organized as follows. In Section 2, the fundamentals necessary to model mobile radio channel impulse responses encountered in systems using directional diversity are presented. Based on these fundamentals, in Section 3 a statistical channel model adjustable to rural, urban, microcellular and picocellular propagation areas is proposed. In Section 4, the validity of the model is illustrated by typical simulation results. and the above-mentioned reduction of the effective delay spread achieved by using directional diversity is shown. Finally, Section 5 presents simulation results concerning a CDMA mobile radio system concept with joint detection (JD) [7] which uses multiple directive BS antennas; it is shown that significant C/I improvements can be achieved as compared to systems using omnidirectional BS antennas.

3 CHANNEL IMPULSE RESPONSES IN MOBILE RADIO SYSTEMS WITH DIRECTIONAL DIVERSITY

As an elementary situation, multi-path transmission between the antenna inputs of a MS and a BS, both using single antennas, is considered. Similar to the approach chosen in [7], in this consideration also the characteristics of the modulator, demodulator, transmitter filter, receiver filter etc. are included. The BS and its antenna are assumed to be at a fixed location, whereas the MS and its antenna may be in motion. In order to formulate the low pass equivalent $\underline{h}(\tau, t)$ of the time varying channel impulse response of said transmission, the space angle variable Ω is used which was introduced in [8]. Ω ranges on a hypothetical sphere S of arbitrary radius centered at the location of the BS antenna and uniquely determines a direction in space relative to the BS antenna. Further, we introduce the following:

$\underline{\vartheta}(\tau, t, \Omega)$: Time variant directional distribution of the channel impulse response seen from the BS antenna. The product $\underline{\vartheta}(\tau, t, \Omega) \cdot \mathrm{d}\Omega$ represents the contribution to the total channel impulse response associated with a differential area $\mathrm{d}\Omega$ on S around Ω. $\underline{\vartheta}(\tau, t, \Omega)$ is time variant mainly due to the motion of the MS and is determined by the location, orientation, and characteristics of the MS antenna, the velocity $\vec{v}$ of the MS and by the topographical and morphographical properties of the propagation area.

$\underline{g}(\tau, \Omega)$: Characteristic of the BS antenna. $\underline{g}(\tau, \Omega)$ is considered to be time invariant. In the case of an infinitely broadband receiving antenna, the dependence of $\underline{g}(\tau, \Omega)$ on τ would be given by the delta function.

$\underline{f}(\tau)$: Impulse response jointly representing the transfer characteristics of modulator, demodulator, transmitter filter, receiver filter etc., see also [2, 7].

With these quantities, the time variant impulse response takes the form

$$\underline{h}(\tau, t) = \int_S \underline{\vartheta}(\tau, t, \Omega) * \underline{g}(\tau, \Omega) * \underline{f}(\tau)\, \mathrm{d}\Omega, \quad (1)$$

where convolution denoted by "$*$" in (1) is to be performed with respect to τ.

Expression (1) can be generalized to cases encountered in mobile radio systems with several co-located directive BS antennas and multi-user operation. With K the number of different users and K_a the number of directive BS antennas, the impulse response

$$\underline{h}^{(k,k_\mathrm{a})}(\tau, t) = \int_S \underline{\vartheta}^{(k)}(\tau, t, \Omega) * \underline{g}^{(k_\mathrm{a})}(\tau, \Omega) * \underline{f}^{(k)}(\tau)\, \mathrm{d}\Omega,$$
$$k = 1 \ldots K, \quad k_\mathrm{a} = 1 \ldots K_\mathrm{a}, \quad (2)$$

is valid for the transmission between MS k and BS antenna k_a. In (2), the upper indices k and k_a identify the individual user or BS antenna, respectively. The $K \cdot K_a$ channel impulse responses of (2) have the discrete-time representations

$$\underline{h}^{(k,k_a)} = \left(\underline{h}_1^{(k,k_a)}, \underline{h}_2^{(k,k_a)} \ldots \underline{h}_W^{(k,k_a)}\right)^T, \quad (3)$$
$$\underline{h}_w^{(k,k_a)} = \underline{h}^{(k,k_a)}((w-1) \cdot T_c, t),$$
$$w = 1 \ldots W,$$

where T_c and W have to be chosen according to the system bandwidth and the delay characteristics of the mobile radio channel, see [2, 7].

When formulating (2) it was assumed that, in the sense of directional diversity, all K_a BS antennas are at the same location. This is e.g. true if the K_a BS antennas are the sectors of a K_a-sectored antenna. If the K_a BS antennas would be at different locations, for each BS antenna a generally different distribution function $\underline{\vartheta}^{(k,k_a)}(\tau, t, \Omega)$ had to be considered.

4 STATISTICAL CHANNEL MODEL

4.1 Basic approach

In what follows, the directional considerations are restricted to the horizontal plane, which does not mean a severe restriction of generality when dealing with mobile radio scenarios. Then, in the above expressions (1) and (2), the quantity Ω has to be replaced by the azimuth angle φ relative to the BS, and (2) takes the form

$$\underline{h}^{(k,k_a)}(\tau, t) =$$
$$\int_0^{2\pi} \underline{\vartheta}^{(k)}(\tau, t, \varphi) * \underline{g}^{(k_a)}(\tau, \varphi) * \underline{f}^{(k)}(\tau)\, \mathrm{d}\varphi,$$
$$k = 1 \ldots K, \quad k_a = 1 \ldots K_a. \quad (4)$$

It is further assumed that the MS antenna is omnidirectional.

In order to model the distribution functions $\underline{\vartheta}^{(k)}(\tau, t, \varphi)$ occurring in (4), a large number of scattering points is placed in the propagation area. These points can e.g. be selected according to the topographical and mormographical situation. The entity of scattering points form a model of the physical propagation area. Fig. 1 shows a BS/MS configuration with one of said scattering points, which has the distances r_1 and r_2 from MS or BS, respectively. The scattering point appears under azimuth angle φ_s from the BS and is assumed to have the scattering coefficient $a \cdot \exp(\mathrm{j}\theta)$. The MS is assumed to have velocity $\vec{v}$ so that r_1 in general varies with time. If the MS in the configuration shown in Fig. 1 is MS k, the contribution of the shown scattering point to the distribution function $\underline{\vartheta}^{(k)}(\tau, t, \varphi)$ in (4) is

$$\underline{\Delta}^{(k)}(\tau, t, \varphi) = (r_1 r_2)^{-\alpha/2} \cdot a \cdot \exp(\mathrm{j}\theta) \quad (5)$$
$$\cdot \delta\left(\tau - \frac{r_1 + r_2}{c_0}\right)$$
$$\cdot \delta(\varphi - \varphi_s)$$
$$\cdot \exp\left(-\mathrm{j}2\pi \frac{\mathrm{d}r_1}{\mathrm{d}t} \cdot \frac{f_0}{c_0} \cdot t\right),$$

where α is the attenuation exponent, c_0 is the velocity of light and f_0 is the carrier frequency. The second exponential function on the right hand side of (6) represents the Doppler shift of the contribution $\underline{\Delta}^{(k)}(\tau, t, \varphi)$. Summing up the contributions of all scattering points yields $\underline{\vartheta}^{(k)}(\tau, t, \varphi)$. In the following, α is set equal to two.

In order to physically model propagation, the total propagation area is hierarchically subdivided as follows, see Fig. 2:

- Regions without scatterers.
- N_a circular scattering areas numbered with $n_a = 1 \ldots N_a$. The diameter of scattering area n_a is termed D_{n_a}. The scattering areas may e.g. represent urban or suburban ensembles of buildings.
- $N_s^{(n_a)}$ circular scatterers numbered with $n_s^{(n_a)} = 1 \ldots N_s^{(n_a)}$ associated with each scattering area n_a, which have the diameters d_{n_a}. The centers of the $N_s^{(n_a)}$ circular scatterers lie within scattering area n_a. The scatterers may e.g. model individual buildings.
- $N_p^{(n_a)}$ scattering points within each circular scatterer of scattering area n_a. The scattering points may e.g. correspond to individual spots on a building from which scattered waves originate.

4.2 Selection of propagation environments

The authors propose variants of the model for five different propagation environments depicted in Fig. 3a-e, namely rural, urban 1, urban 2, microcellular and picocellular. In all five models, scattering area 1 is centered at the MS location. The models urban 1, urban 2, microcellular and picocellular comprise a second scattering area, see Fig. 3b-e. For the geometrical parameters

d, D_1, D_2, d_1 and d_2, see Figs. 2 and 3, and for the numbers ${N_a}^{(1)}$, ${N_a}^{(2)}$, ${N_p}^{(1)}$ and ${N_p}^{(2)}$ default values are given in Table 1. These values were selected based on geometrical considerations. For the model variants urban 1 and urban 2 it was e.g. assumed that the horizontal extensions of buildings and ensembles of buildings are 20m and 1.5km, respectively. In the FORTRAN77 program, the model parameters can be easily adjusted depending on the individual situation. Work has still to be done to determine parameter sets, which, e.g. based on measurements, satisfyingly match with reality.

4.3 Incorporation of statistical properties

The performance of mobile radio systems can e.g. be characterized by mean values and variances of bit error probabilities (BER) which are experienced when the mobiles move through propagation areas of a certain type. In order to reliably obtain such quantities by simulations, which have to take into account practically all possible propagation situations encountered in the considered type of propagation area, a statistical propagation model is required. Such a statistical feature is implemented in the proposed channel model by introducing the possibility of Monte Carlo simulation in the following way:

- Select the model variant for a desired propagation environment, see Fig. 3, and its parameters, see e.g. Table 1.
- Generate a large ensemble of model samples of the selected propagation environment by repeatedly distributing the scatterers over the scattering areas and the fixed number of scattering points within the scatterers in a random way.
- Let MS move a few wavelengths in each model sample and determine e.g. BER say every 10^{th} of a wavelength.
- Take all obtained BERs to determine e.g. mean value and variance of BER.

When generating said random distributions of scattering points, uniform distributions are the most obvious, and scattering points closer to BS or MS than a minimum distance $d_{\text{B min}}$ or $d_{\text{M min}}$, respectively, are omitted because such scattering points would lead to excessive values of $\underline{\Delta}^{(k)}(\tau, t, \varphi)$, see (6). Default values for those minimum distances are contained in Table 1. Concerning the scattering coefficient $a \cdot \exp(\mathrm{j}\theta)$, see Section 3.1, an obvious approach would be to assign all scattering points the same absolute value a and arguments θ selected randomly in $[0, 2\pi)$. This approach was followed when performing the simulation results shown in the following. However, the implemented FORTRAN77 program allows also other selections of $a \cdot \exp(\mathrm{j}\theta)$.

5 EXAMPLES OF $\underline{\vartheta}^{(k)}(\tau, t, \varphi)$ AND $\underline{h}^{(k,k_a)}(\tau, t)$

In what follows, the carrier frequency $f_0 =$ 1800MHz and the default values of Table 1 are presupposed. In Figs. 4a and b, the absolute values of two distributions $\underline{\vartheta}^{(k)}(\tau, t_1, \varphi)$ and $\underline{\vartheta}^{(k)}(\tau, t_2, \varphi)$ normalized to one are shown, which were obtained for one sample of the microcellular model at instants t_1 or $t_2 = t_1 + 2$ms, respectively, with the MS velocity being 25m/s. Time variance and the strong dependence on the azimuth angle φ are evident. Figs. 5a-e show runs of the absolute value $|\underline{h}(\tau, t)|$ of $\underline{h}(\tau, t)$ for cases when MS again moves with the velocity 25m/s in an urban 2 model sample, and the corresponding power delay profiles $P(\tau)$ [9]. Fig. 5a is valid for the case that a single omnidirectional antenna is used at the BS, whereas Figs. 5b-e are valid when four ideal directive antennas are employed at the BS, each antenna having an aperture of 90° and covering disjoint azimuth sectors. It can be seen that the delay spreads experienced by the directive antennas are considerably smaller than the delay spread valid for the omnidirectional antenna. Furthermore, in the case of ideal directive antennas the time variance of the impulse responses seen by the directive antennas is reduced, see Figs. 5b-e.

6 BENEFITS OF DIRECTIONAL DIVERSITY IN A JD-CDMA MOBILE RADIO SYSTEM

Presently, many researchers around the world are engaged in designing multiple access schemes for third generation mobile radio systems. One approach would be to set out from second generation systems like GSM and to arrive at third generation systems by evolution. This is the objective of ongoing work on JD-CDMA [2, 7]. The structure of the uplink of a JD-CDMA mobile radio system is depicted in Fig. 6. Each mobile uses a single transmitter antenna. At the base station, the signals received via K_a receiver antennas are combined by adding the output signals of K_a matched filter banks (MF-bank), each containing K filters matched to the resulting

transmission channel between data source k, $k = 1 \ldots K$, and receiver antenna k_a, $k_a = 1 \ldots K_a$. This combining technique is termed maximal-ratio combining (MRC) [2]. After signal combination, both inter symbol interference (ISI) and multiple access interference (MAI) are cancelled in the ISI & MAI canceller. This unit can be realized by either more elaborate optimal non-linear equalizers or more favorable sub-optimum linear equalizers [2, 7]. Unfortunately the optimum maximum-likelihood JD receiver is prohibitively expensive. Therefore, sub-optimum JD techniques have to be used [2, 7]. By extensive simulations with the above introduced statistical channel model, the authors determined C/I improvements g which could be achieved in JD-CDMA by using multiple directive antennas at BS instead of one omnidirectional BS antenna. As an example, **Fig. 7** shows g versus time t for a particular MS moving with velocity 25m/s through a model sample of the propagation environment urban 2 in the case that an antenna with eight 45°-sectors is deployed at BS. The mean value of g for the considered time interval is 16.2 dB. This C/I improvement allows the reduction of the complexity of the JD-receiver or the enhancement of the transmission quality.

7 CONCLUSIONS

Due to multi-path propagation effects in mobile radio systems, the waves associated with a particular transmitter are not incident completely isotropically at the receiver site. Therefore, besides time variance and frequency selectivity, a basic property of the mobile radio channel is the spatial spreading of incident radio waves. This has to be considered in realistic channel models. In this paper, a flexibly configurable statistical channel model including the effects of spatial spreading is presented in order to enable the validation of system concepts using directional diversity. The parameters of this model which is available as a FORTRAN77 program can be easily adjusted to various propagation areas. Therefore, the model is well suited to perform simulations, evaluations and comparisons of mobile radio systems. Based on this model, quantitative examples for the C/I improvements achievable by directional diversity in a CDMA mobile radio system using joint detection are given. The influence of directional diversity on the system capacity has to be investigated by further simulations based on the proposed channel models. The properties of randomly generated impulse responses have to be compared with measured ones in order to satisfyingly match with reality.

ACKNOWLEDGMENTS

The authors express their thanks to the German Research Society (DFG), which by a research grant within "Schwerpunktprogramm Mobilkommunikation" put them into the position to perform this work. Furthermore, the authors gratefully appreciate the fruitful exchange of ideas with Prof. Dr. Nossek and his coworkers from TU Munich, and are indebted to the supercomputer staff at the Regionales Hochschulrechenzentrum Kaiserslautern (RHRK) for their invaluable support.

References

[1] Calhoun, G.: Digital cellular radio. Norwood: Artech House, 1988.

[2] Jung, P.; Blanz, J. J.: Joint detection with coherent receiver antenna diversity in CDMA mobile radio systems. IEEE Transactions on Vehicular Technology, vol. 44 (1995), pp. 76-88.

[3] Eggers, P. C.: TSUNAMI: Spatial radio spreading as seen by directive antennas. COST231 TD(94)119, Darmstadt/Germany, 1994.

[4] Blanz, J. J.; Jung, P.: Benefits of directional antennas on the performance of third generation mobile radio systems. Proceedings of the International Conference on Telecommunications (ICT'95), Bali/Indonesia (1995), pp. 367-370.

[5] Farsakh, C.; Nossek, J. A.: Application of space division multiple access to mobile radio. Proceedings of the 5th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'94), The Hague/The Netherlands (1994), pp. 736-739.

[6] Ali, M.; Haardt, M.; Nossek, J. A.: 4×S - Single snapshot spatial separation of wavefronts via antenna arrays. Proceedings of the 8th Aachen Symposium on Signal Theory, Aachen/Germany (1994), pp. 369-372.

[7] Blanz, J. J.; Klein, A.; Naßhan, M. M.; Steil, A.: Performance of a cellular hybrid C/TDMA mobile radio system applying joint detection and coherent receiver antenna diversity. IEEE Journal on Selected Areas in Communications, vol. 12 (1994), pp. 568-579.

[8] Fleury, B. H.; Dahlhaus, D.: Investigations on the time variations of the wide-band radio channel for random receiver movements. Proceedings of the 3rd IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'94), Oulu/Finland (1994), pp. 631-636.

[9] Parsons, D.: The mobile radio propagation channel. London: Pentech Press, 1992.

	rural	urban 1	urban 2	micro	pico
d/km	10	3	3	1	0.2
D_1/km	0.2	1.5	1.5	1	0.2
D_2/km	-	1.5	1.5	0.3	0.2
$N_s^{(1)}$	12	12	12	12	12
$N_s^{(2)}$	-	12	12	6	12
$N_p^{(1)}$	50	50	50	50	50
$N_p^{(2)}$	-	50	50	50	50
d_1/m	30	20	20	10	5
d_2/m	-	20	20	20	5
$d_{\mathrm{B\,min}}$/m	-	10	10	5	5
$d_{\mathrm{M\,min}}$/m	10	5	5	5	1

Table 1: Default values of model parameters

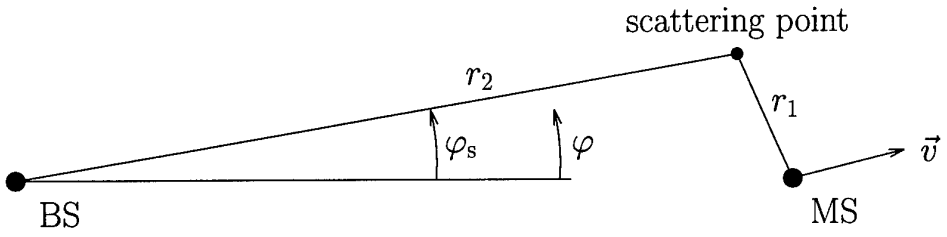


Figure 1: BS/MS configuration with scattering point

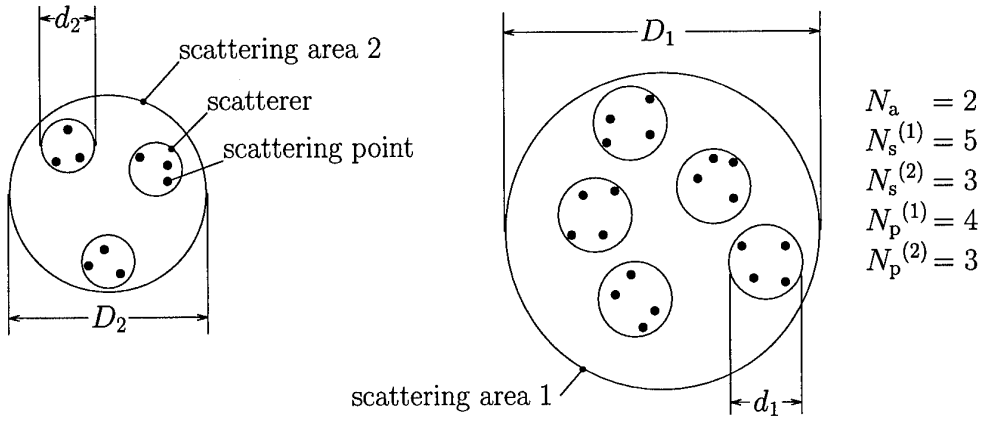


Figure 2: Hierarchical subdivision of propagation area

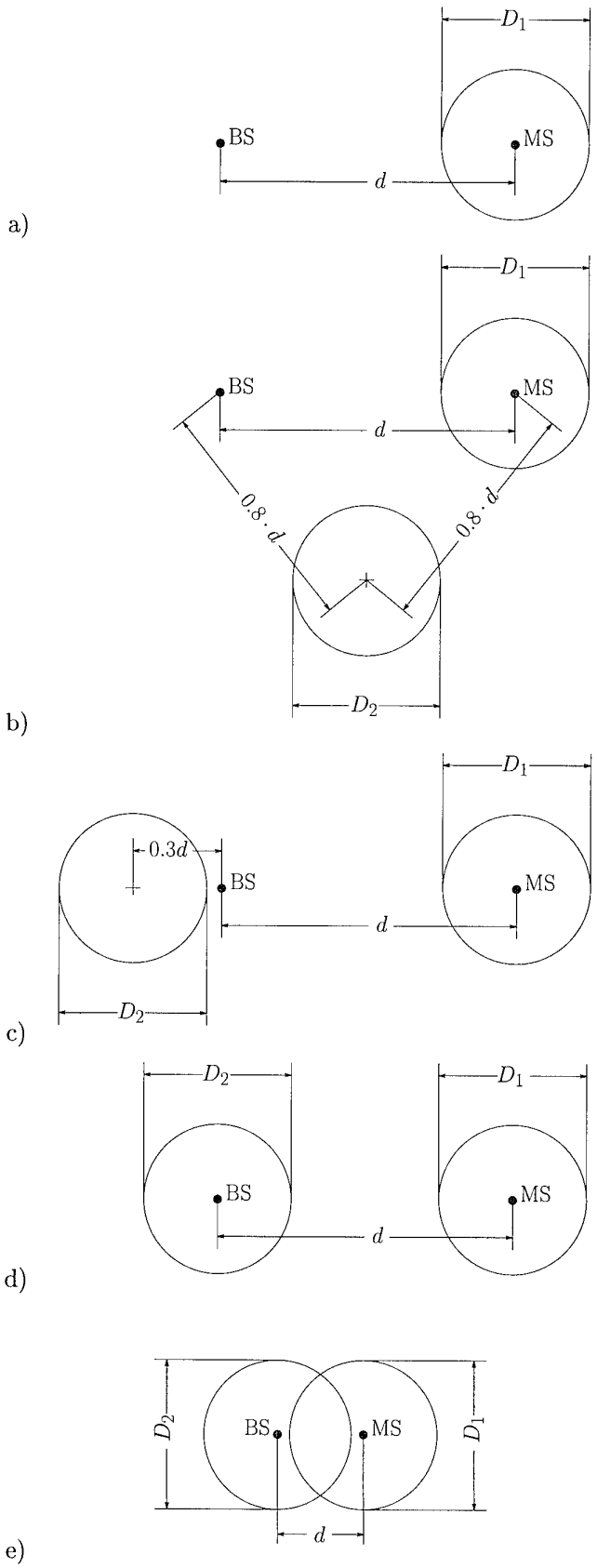


Figure 3: Propagation environments
a) rural; b) urban 1; c) urban 2; d) microcellular; e) picocellular

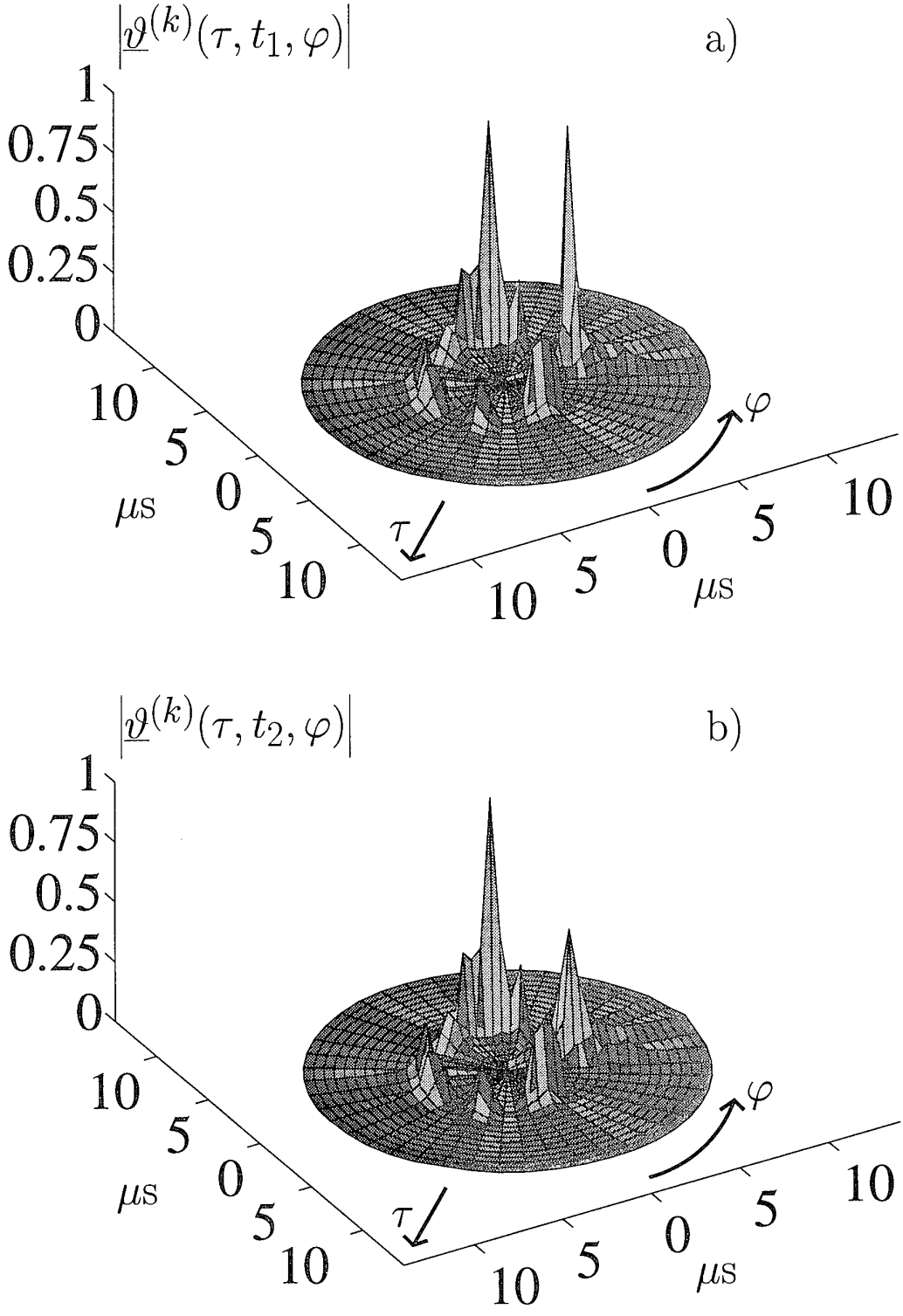


Figure 4: Distribution $\left|\underline{\vartheta}^{(k)}(\tau, t, \varphi)\right|$ for microcellular model a) $t = t_1$; b) $t = t_2 = t_1 + 2\text{ms}$

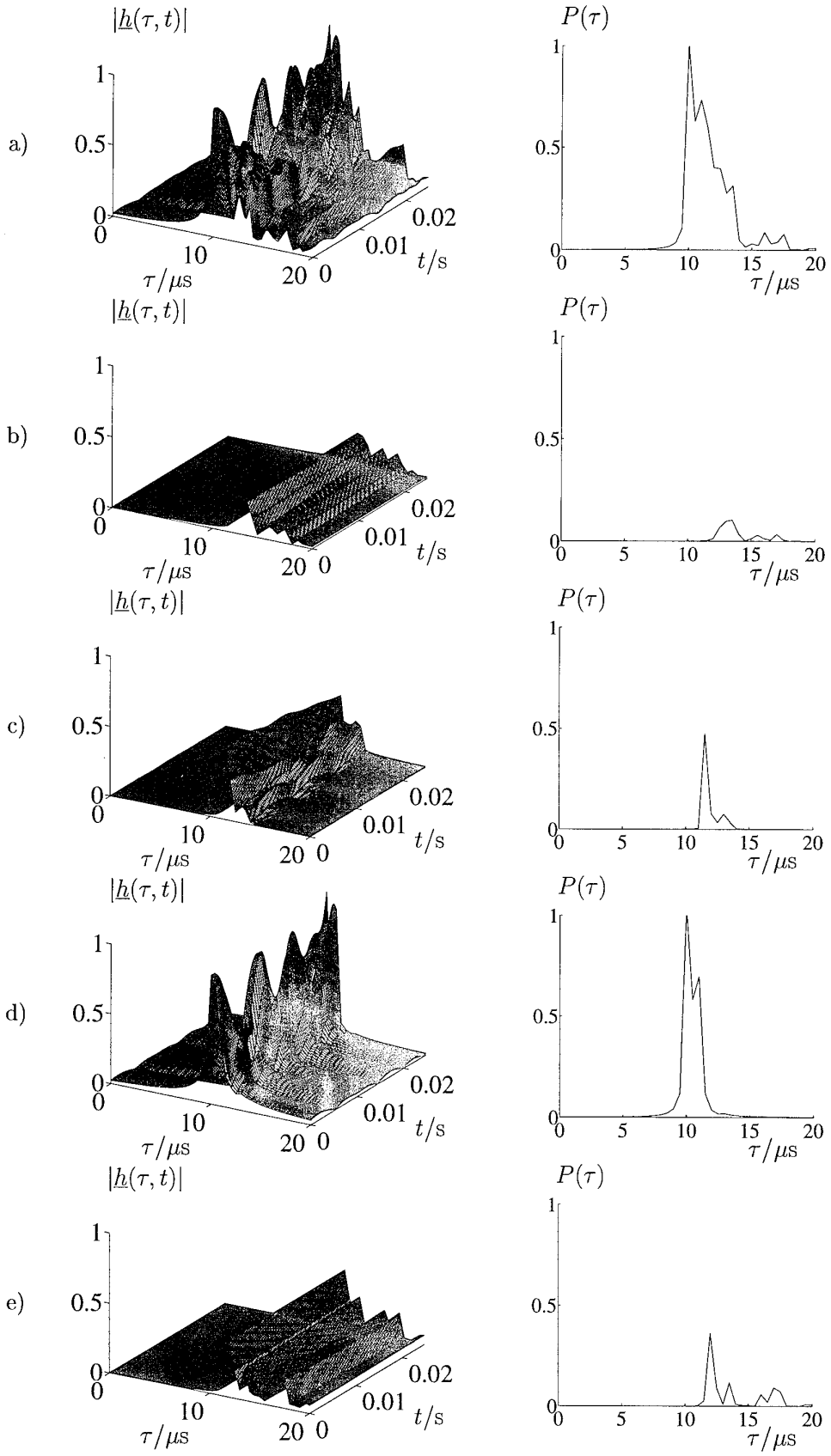


Figure 5: Absolute values $|\underline{h}(\tau,t)|$ of channel impulse responses and power delay profiles $P(\tau)$ for model urban 2; a) omnidirectional BS antenna; b)-e) 90° sectors of 4-sectored BS antenna

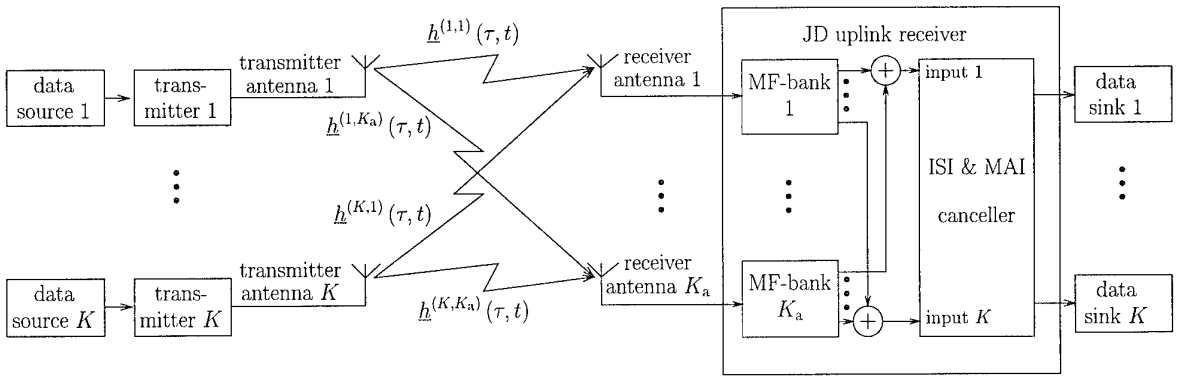


Figure 6: Structure of the uplink in a JD-CDMA system using antenna diversity in the receiver

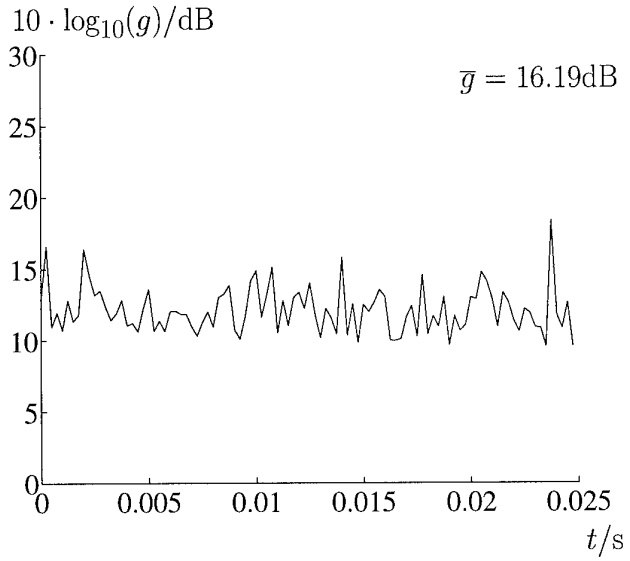


Figure 7: C/I improvement g in a JD-CDMA system versus t; BS antenna with eight 45°-sectors; microcellular model scenario

DISCUSSION

Discussor's name: E. Van Lil

Comment/Question:

As an antenna engineer I am convinced that analog preprocessing will enhance substantially the performance of the system. One of the reasons is that the postprocessing will behave in a digital way. Could you comment on the influence of the discretization error introduced by the A/D conversion process?

Author/Presenter's reply:

Several simulations of the considered JD-CDMA mobile radio system including the effects of finite word length in A/D D/A-conversion and the transfer characteristics of analog filters were performed, see e.g. "Realistic simulation of a JD-CDMA mobile radio system using coherent antenna diversity" in the proceedings of the IEEE International Symposium on Spread Spectrum Techniques and Applications, Oulu, Finland (1994). If a resolution of 8 bits is used in the transmitter and the receiver, the degradation at a BER of 10 -3 is approximately 1dB.

ERROR PROBABILITY COMPUTATION IN OPTICAL FIBER LINKS

M. VANDROOGENBROEK, P. MEGRET, M. BLONDEL

Faculté Polytechnique de Mons - Service d'Electromagnétisme et de Télécommunications
Boulevard Dolez, 31, B-7000 MONS (BELGIUM)
tel : +32 65 374198 - fax : +32 65 374199 - E-mail : Vdgb@fpms.fpms.ac.be

Abstract

Several techniques to compute error probability in optical communication systems exist. Three methods of calculation are analyzed : Gaussian approximation, Gram-Charlier series and Gauss quadrature rule. Some results and comparisons for binary systems are presented.

I. Model

Fig.1 gives a version of the receiver model and allows us to emphasize the noise contributions. The photodetection process (before avalanche) obeys Poisson statistics. The number of photoelectrons generated per second, $\gamma(t)$, is related to the incident optical power by

$$\gamma(t) = \frac{\eta}{h\nu} P_R(t) + \gamma_0 \tag{1}$$

where η is the detector quantum efficiency, γ_0 is the dark current, $P_R(t)$ is the optical signal at the photodetector input and $h\nu$ is the energy of incident photons.

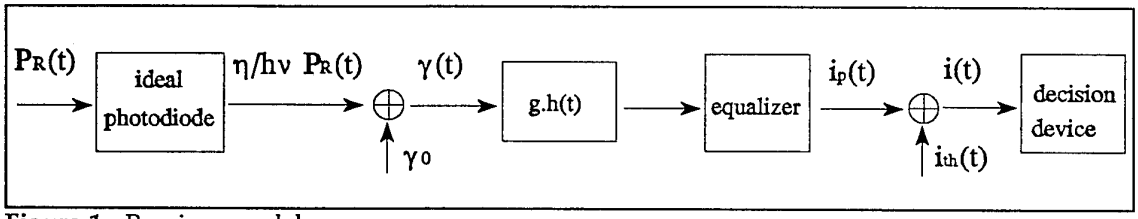


Figure 1 : Receiver model

The output of the receiver is given by the sum of a randomly multiplied filtered Poisson process $i_p(t)$ plus an independent zero-mean Gaussian thermal noise $i_{th}(t)$. The current i(t) at the input of the decision device can be expressed, for $t \geq 0$, as

$$i(t) = \sum_{k=1}^{N(t)} g_k q h_R(t-t_k) + i_{th}(t) \tag{2}$$

where N(t) is an inhomogeneous counting process such that N(t) is the number of photoelectrons generated during [0,t], t_k is the emission time and g_k is the random avalanche gain of the photodiode. Here q is the charge of an electron ($1.6\ 10^{-19}$C) and $h_R(t)$ denotes the photodiode plus equalizer impulse response.

II. Computation

We assume that the pattern process $\{a_\rho\}=\{...,a_{-1},a_0,a_1,...\}$ is strict-sense stationary and that threshold decision on the symbol a_0 is taken from the receiver output at a sampling instant t_0.

Our aim is the evaluation of the average error probability $P_E=E[P_E|a_0]$, where $P_E|a_0$ is the conditional error probability given that a particular a_0 has been transmitted. This requires the knowledge of the conditional distribution

$$F(S|a_0) = P[i_p(t_0)+i_{th}(t_0) \leq S|a_0] \tag{3}$$

where S is the decision threshold.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

Simple a_0-conditioning is not sufficient to assign completely the intensity of i_p. Indeed, conditioning on the whole pattern $\{a_\rho\}$ is necessary. Thus, we introduce the pattern-conditional distribution

$$F(S|\{a_\rho\}) = P[i_p(t_0)+i_{th}(t_0)\leq S|\{a_\rho\}]. \tag{4}$$

From the pattern-conditional distribution so calculated, the desired symbol-conditional distribution is obtained by averaging with respect to $\{a_\rho\}$ with a_0 given :

$$F(S|a_0)=E\left[F(S|\{a_\rho\})|a_0\right] \tag{5}$$

1. Gaussian approximation

The simplest approach to compute the error rate is to approximate $i(t_0)$ as a Gaussian random variable. In this approximation, one only needs to calculate the mean and variance of $i(t_0)$ in each sequence $\{a_\rho\}$. The means $(\mu|\{a_\rho\})$ and variances $(\sigma^2|\{a_\rho\})$ are very simple to obtain [1]. Hence, (4) becomes

$$F(S|\{a_\rho\}) = \Phi\left(\frac{S-\mu|\{a_\rho\}}{\sigma|\{a_\rho\}}\right) \tag{6}$$

where Φ represents the cumulative function of the normalized Gaussian distribution.

2. Gram-Charlier Series

We can use Gram-Charlier series to calculate the pattern-conditional distribution [2]

$$F(S|\{a_\rho\}) = \Phi(S|\{a_\rho\}) + \sum_{n=3}^{N} d_{|\{a_\rho\},n}\ p_0^{(n-1)}(S) \tag{7}$$

The computation of the coefficients $d_{|\{a_\rho\},n}$ only requires knowledge of the cumulants of $i(t_0)$.
If the number of terms N used in the series (equation 7) is large enough, we can consider then we calculate "exact" results.

3. Gauss quadrature rule

With this method, no assumption is made about the statistical distribution of $i(t_0)$.
Let σ^2_{th} denote the variance of zero-mean Gaussian random variable $i_{th}(t_0)$. Then

$$P[i_p(t_0)+i_{th}(t_0)\leq S] = E\left[P[i_{th}(t_0)\leq S-i_p(t_0)|i_p(t_0)]\right] = E[\ \Phi\left(\frac{S-i_p(t_0)}{\sigma_{th}}\right)\ ] \tag{8}$$

As the probability density function of $i_p(t_0)$ is unknown, (8) is evaluated using a Gauss quadrature rule which only requires knowledge of the moments of $i_p(t_0)$, which can be computed from the cumulants. Using the first 2K+1 moments of $i_p(t_0)$ we have [3]

$$P[i_p(t_0)+i_{th}(t_0)\leq S] = \sum_{i=1}^{K} w_i \Phi\left(\frac{S-(E[i_p(t_0)]+x_i)}{\sigma_{th}}\right) \tag{9}$$

where w_i and x_i are the weights and abscissas, respectively, defined by the integration method.
If the number of terms K taken into account (equation 8) is large enough, this method allows the calculation of "exact" error probabilities too.

III. Results and comparisons

Error probability is computed as function of the optical power (Fig.2-7) for a binary system with a wavelength of 1.3 μm. The received optical pulses are Gaussian shaped $s(t)=A\exp(-8B^2t^2)$ where the value of A for a '0' is then percent of the value of A for a '1'. The photodiode impulse response $h_R(t)$ is rectangular time function of duration T=1/B. The thermal noise power of the receiver is assumed to be $\sigma^2_{th}=2.56\ 10^{-15}A^2$. The dark current of the photodetector is neglected and the parameter η is assigned a value of 0.8.

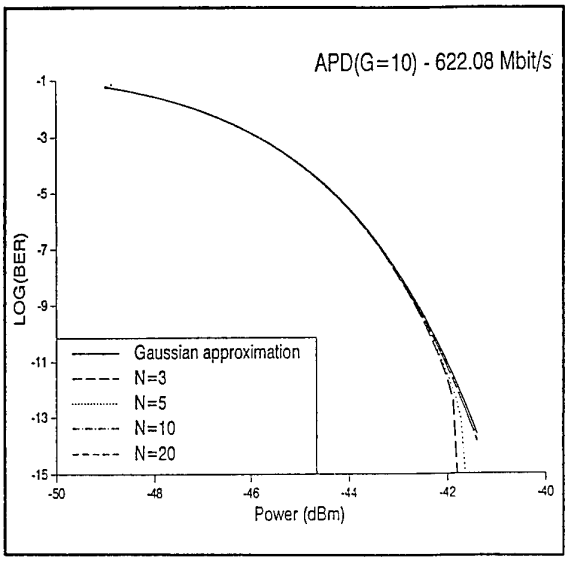


Figure 2 : Comparison between Gaussian approximation and Gram-Charlier series

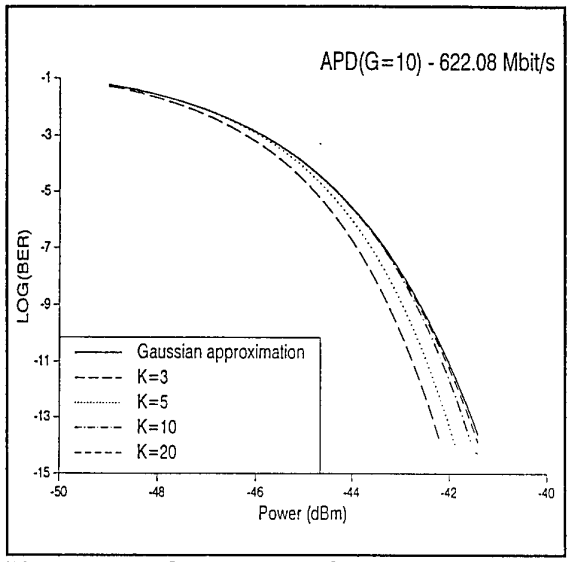


Figure 3 : Comparison between Gaussian approximation and Gauss quadrature rule

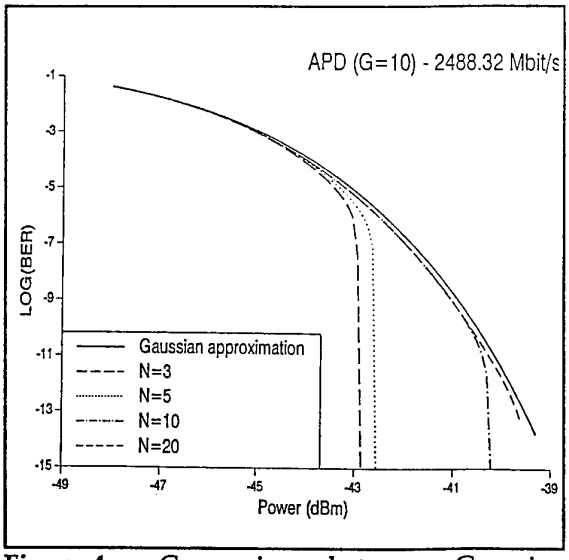


Figure 4 : Comparison between Gaussian approximation and Gram-Charlier series

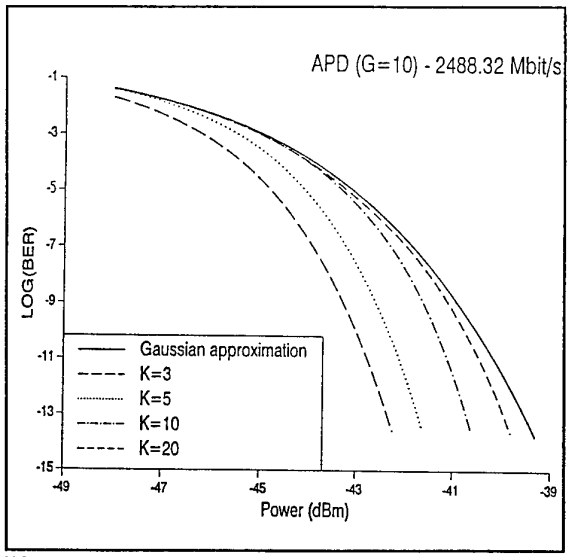


Figure 5 : Comparison between Gaussian approximation and Gauss quadrature rule

First, Fig.2-3 show numerical results for data rate B of 622.08 Mbit/s and an avalanche photodiode with a gain of 10. Fig.2 shows a comparison between Gaussian approximation and Gram-Charlier method with different values of N. For N=3 and N=5, vertical lines mean that for higher power, the compute error probability becomes negative. Fig.3 shows a comparison between Gaussian approximation and Gauss quadrature method with different values of K.

Next, Fig.4-5 show numerical results for data rate of 2488.32 Mbit/s and an avalanche photodiode with a gain of 10.

Finally, Fig.6-7 show numerical results for data rate of 2488.32 Mbit/s and a PIN photodiode.

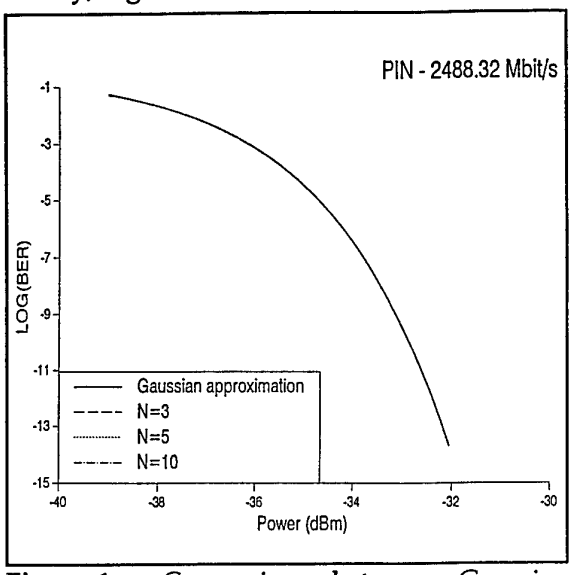


Figure 6 : Comparison between Gaussian approximation and Gram-Charlier series

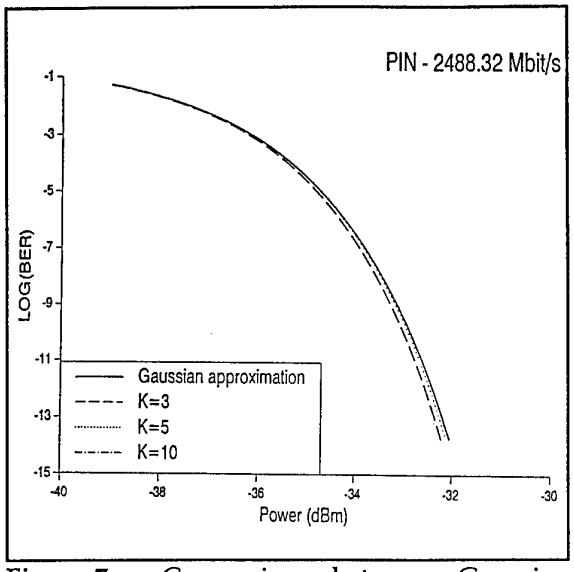


Figure 7 : Comparison between Gaussian approximation and Gauss quadrature rule

Computation duration for different simulations is given in Table I.

Table I : Computation duration for different methods.

	Computation duration
Gaussian approximation	21s
Gram-Charlier (N=3)	59s
Gram-Charlier (N=5)	1min37s
Gram-Charlier (N=10)	4min15s
Gauss quadrature rule (K=3)	1min12s
Gauss quadrature rule (K=5)	1min55s
Gauss quadrature rule (K=10)	7min28s

IV. Conclusions

We have compared three methods for calculating error rate. We can observe that the Gaussian approximation is in good agreement with Gram-Charlier and Gauss quadrature rules methods which allow "exact" calculation.

We can see that, for small values of N, the Gram-Charlier method can fail for small error rates. The higher the bit rate is and the larger the avalanche gain is, the more rapidly simulations fail.

We can also observe that for small values of K, the Gauss quadrature rules method is less accurate than the Gaussian approximation. The higher the bit rate is and the larger the avalanche gain is, the slower the method converge.

Finally, the Gaussian approximation method is much quicker and nearly as accurate as the two others.

Acknowledgment

This work has been funded by Belgacom in the frame of R&D contract RD/TR/UN/1.

REFERENCES

[1] *G.L. Cariolaro,* "Error Probability in Digital Fiber Optic Communication Systems", IEEE Transactions on Information Theory, vol. IT-24, n°2, pp. 213-221, 1978

[2] *M. Mansuripur, J.W. Goodman, E.G. Rawson, R.E. Norton,* "Fiber Optics Receiver Error Rate Prediction Using the Gram-Charlier Series", IEEE Transactions on Communications, vol. COM-28, n°3, pp. 402-407, 1980

[3] *G.H. Golub, J.H. Welsch,* "Calculation of Gauss quadrature rules", Math. Comput., vol. 23, pp. 221-230, 1969

Gain-Switched DFB Lasers as Soliton Sources for High Bit Rate, Long Distance Transmissions

L. Meuleman, X. Wang, M. Blondel

Service d'Electromagnétisme et de Télécommunications, Faculté Polytechnique de Mons, Boulevard Dolez, 31, B-7000 Mons, Belgium
Tel.: +32 65 37 41 95, Fax.: +32 65 37 41 99, E-mail: Meuleman@fpms.fpms.ac.be

Summary

Numerical simulations of soliton transmission are presented. The optical pulses are delivered by a gain-coupled DFB laser using a gain-switching technique. The important source features are reviewed, and a study of the timing jitter is described.

1. Introduction

The use of gain-switched distributed feedback (DFB) lasers as soliton sources for long-haul, high bit rate transmissions is an issue of interest [1]. If this technique is attractive thanks to its simplicity, the quality of the delivered pulses is affected by namely three factors: pulse shape, frequency chirp and pulse-to-pulse timing jitter.
It is well-known that initial departures from the ideal initial situation - i.e. a train of equally spaced, transform-limited, sech-shaped optical pulses - can be evacuated during propagation under the form of dissipative waves, provided that these initial defects are of reasonable magnitude. The optical fibre then acts as a soliton filter, under the assumption that the transmission occurs in the anomalous dispersion regime and that the pulse initial peak power is correctly adapted to the pulse duration. These non-soliton components should anyway be minimized since they disturb soliton propagation.

2. Numerical Simulations

Numerical simulations include two parts: first, a complete model for the gain-coupled DFB laser, then the nonlinear propagation equation.

(1) Semiconductor laser model

The large-signal dynamic behavior of a DFB laser with gain (or loss) coupled grating can be described by a set of generalized rate equations [2], in which the spatial hole burning, standing-wave effect, distributed noises and nonlinear gain suppression have been taken into account.

$$\frac{dS_{Lq}}{dt} = \Delta G(\aleph, u_q^{\pm}) S_{Lq} + \Re_{sp} + \Im_{st} + F_{SLq}(t) \qquad q = 1,2,\cdots,Q \qquad (1)$$

$$\frac{dN_m}{dt} = \frac{J_{bc} + J(t)}{ed} - R(N_m) - R_{st}(N_m, S_{Lq}, u_q^{\pm}) + F_{Nm}(t) \qquad m = 1,2,\cdots,M \qquad (2)$$

Here the mode number is denoted by q and the section number by m. $S_{lq}(t)$ is the variation of optical power emitted from left facet of the laser cavity and $u_q^{\pm}$ are the forward and backward field distribution functions. $\aleph=\{N_1,N_2,...,N_M\}$ where N_m is the average carrier density in the mth section. $\Im_{st}$ is the loss caused by the standing wave effect and the gain coupling coefficient κ_g. $F_{Nm}(t)$ and $F_{SLq}(t)$ are Langevin noise terms, which are related to the local spontaneous carrier recombination as well as the stimulated emission and absorption, respectively. Other parameters were explained in [2].
It is important to note that the output optical pulse shape is greatly affected by the bias current value J_{bc} and the modulation current J(t). Two forms of current pulses are used in our simulations: Gaussian pulses with standard deviation of 96 ps and rectangular pulses with pulse width 90 ps as well as rise and fall times 10 ps. The maximum value of the current pulse is three times the threshold current. In order to obtain the dynamic single mode operation, a 1.55 μm AR-coated λ/4 phase-shifted gain-coupled DFB laser with a 400 μm long cavity is

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

used as the soliton source. The gain coupling coefficient ($\kappa_g L$) is -0.2 and the index coupling is 0.8, 1.6, 2 and 2.8. The laser is biased just below the threshold ($J_{bc} = 0.95\ J_{th}$) and modulated by a train of electrical pulses. The modulation frequency is 2.5 GHz.

(2) Soliton transmission

Provided that the amplifier spacing is short when compared to the soliton period Z_0, one may describe the periodically amplified link by the following averaged Nonlinear Schrödinger Equation (NLSE) [3]. This equation is solved by the split-step Fourier method [4]:

$$\frac{\partial u}{\partial z} = \frac{j}{2}\frac{\partial^2 u}{\partial \tau^2} + j\gamma |u|^2 u \qquad (3)$$

In this equation, u is the normalized complex field amplitude, z is the normalized distance and τ is the normalized time in a framework moving at the group velocity [4]. The link parameters are the following : Group Velocity Dispersion D = 2 ps/(nm.km), Nonlinear Refractive Index $n_2 = 3\ 10^{-20}$ m²/W, total length: 10,000 km. The EDFA noise is not included in these simulations in order to isolate the influence of the source on the link performances.

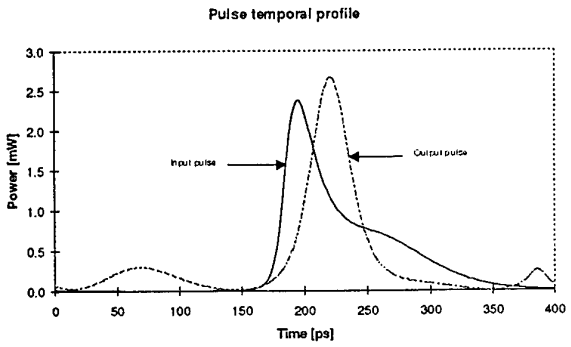


Figure 1: Pulse initial and final temporal profiles

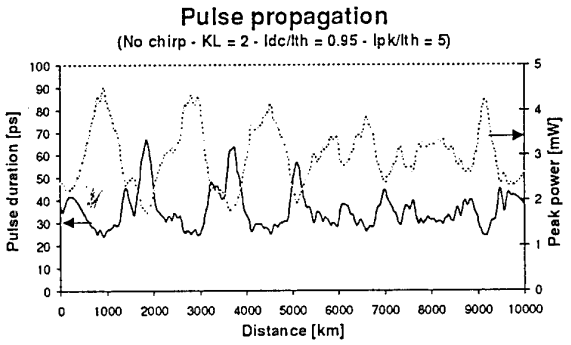


Figure 2: Pulse parameters during propagation

The effect of pulse dissymmetry is shown hereabove: the strongly asymmetric temporal profile in fig.1 results in strong fluctuations of the pulse parameters (peak power, FWHM temporal width) along the transmission (cf. fig. 2). These fluctuations are out of phase and more or less periodic, their period is given by $8Z_0$ where Z_0 is the soliton period. This spatial resonance behaviour is a common feature of any perturbed soliton propagation [5]. At the end of the link, the temporal profile is well fitted to a sech one. The pulse frequency chirp is not included in this simulation. Such strong temporal asymmetry can be avoided thanks to the use of a less important peak current.

One way to solve the chirp problem is the so-called narrow-band spectral filtering technique [6]. In this technique, a narrow-band Fabry-Pérot filter is used to select the transform-limited part of the pulse. We assume a lorentzian profile for the filter. The power transfer function is then:

$$T(f) = \frac{1}{1+\left(\frac{f-f_0}{B/2}\right)^2} \qquad (4)$$

where f is the optical frequency, f_0 the filter center frequency and B is the FWHM bandwidth of the filter. Prior to propagation simulations, the filter parameters (B and f_0) were carefully chosen. In the case presented, B = 0.08 nm and $f_0 = -0.4$ nm.
The resulting pulse is then propagated on the same link than previously, with its initial peak power adapted to its duration. The output pulse profile is shown at fig.3 with the ideal output pulse (no frequency chirp). These results are pretty close to each other.

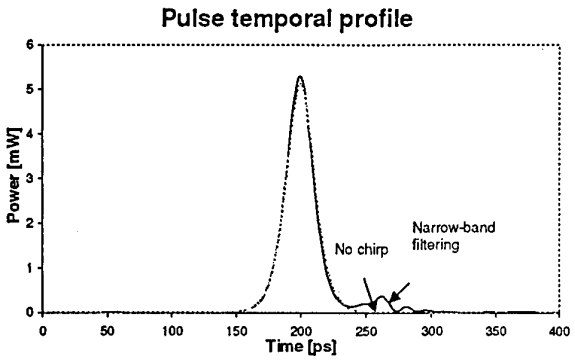


Figure 3: Pulse final teporal profiles

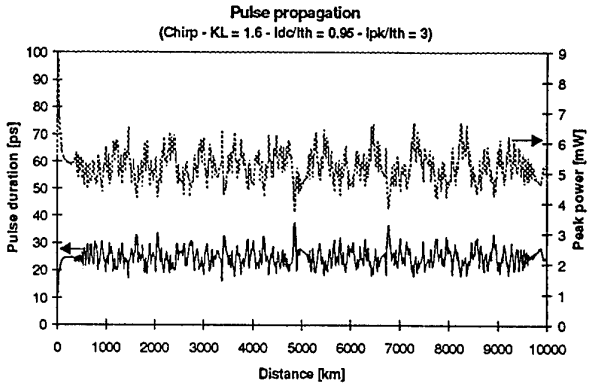


Figure 4: Pulse parameters during propagation

Nevertheless, even though the pulse spectrum fulfils the transform-limited condition, phase noise can affect the pulse train at the output of the laser. The spectral signature of this noise is very narrow, so that the transform-limited condition is not sufficient to characterize the quality of a pulse source. Experiments lead by Mollenauer *et al.* [7] show that pulse-to-pulse jitter is a major defect of the gain-switching technique for soliton propagation over long distances.

The pulse-to-pulse jitter is certainly the most demanding effect as far as computation effort is concerned. It can be analysed following two ways. The first way consists in a computation of the timing of individual pulses along their propagation. The jitter is then characterised by the standard deviation of this timing. One alternative way consists in drawing the eye diagram at the end of the link. This analysis is thus qualitative and only valid for comparative studies.

Such comparative analysis is presented. A pseudo-random 16-bit pulse sequence with intensity and phase noises is propagated along the link. The bit slot is 400 ps, corresponding to a 2.5 Gbit/s bit rate. The jitter effects were studied for different κL parameters. A value of $\kappa L = 1.6$ was chosen as it results in a better eye diagram opening at the laser output. The eye diagram of fig. 5b is strongly degraded by the pulse-to-pulse jitter. This problem can be overcome by using soliton-control techniques [9-11], which have first been used to reduce the so-called Gordon-Haus effect [8]. Equation (3) is then perturbed in the following way [11]:

$$\frac{\partial u}{\partial z} = \frac{j}{2}\frac{\partial^2 u}{\partial \tau^2} + j|u|^2 u + \delta u - \beta(j\frac{\partial}{\partial \tau} - \omega_f)^2 u \qquad (5)$$

The filtering action is described by three parameters : the filter strength β, the sliding term $\omega_f = \omega_{f0} + d\omega_f/dz$ and the excess gain coefficient δ. The results are shown in fig. 5c, with $\delta = 0.01$, $\beta = 0.03$. One can easily see the dramatic improvement obtained thanks to this technique (fig. 5c). The use of sliding filters allows still more complete recovery of the initial eye diagram aperture (fig. 5d). In this method, the central frequency of the filters is gradually changed along the link. The soliton then follows the filter displacement whereas linear components (noise, CW components) are not able to readjust their frequency spectrum and are eventually damped out [11].The filter parameters were in this case: $\delta = 0.05$, $\beta = 0.1$, sliding rate $d\omega_f/dz = 0.03$. One can see that this technique allows the use of stronger filters (higher β values) and thus a better suppression of non-soliton components.

3. Conclusions

Numerical simulations of the propagation of picosecond optical pulses delivered by a gain-coupled DFB laser have been presented. These simulations show the influence of different features of gain-switched pulses: dissymmetry, chirp and timing jitter. If solutions exists for the first two problems, the latter is more delicate. However, this problem can be solved in the same way that the so-called Gordon-Haus effect, thanks to the use of fixed or sliding filters placed along the link.

4. Acknowledgements

This work has been partly funded by BELGACOM in the frame of R&D contract RD/TR/UN/ 12.

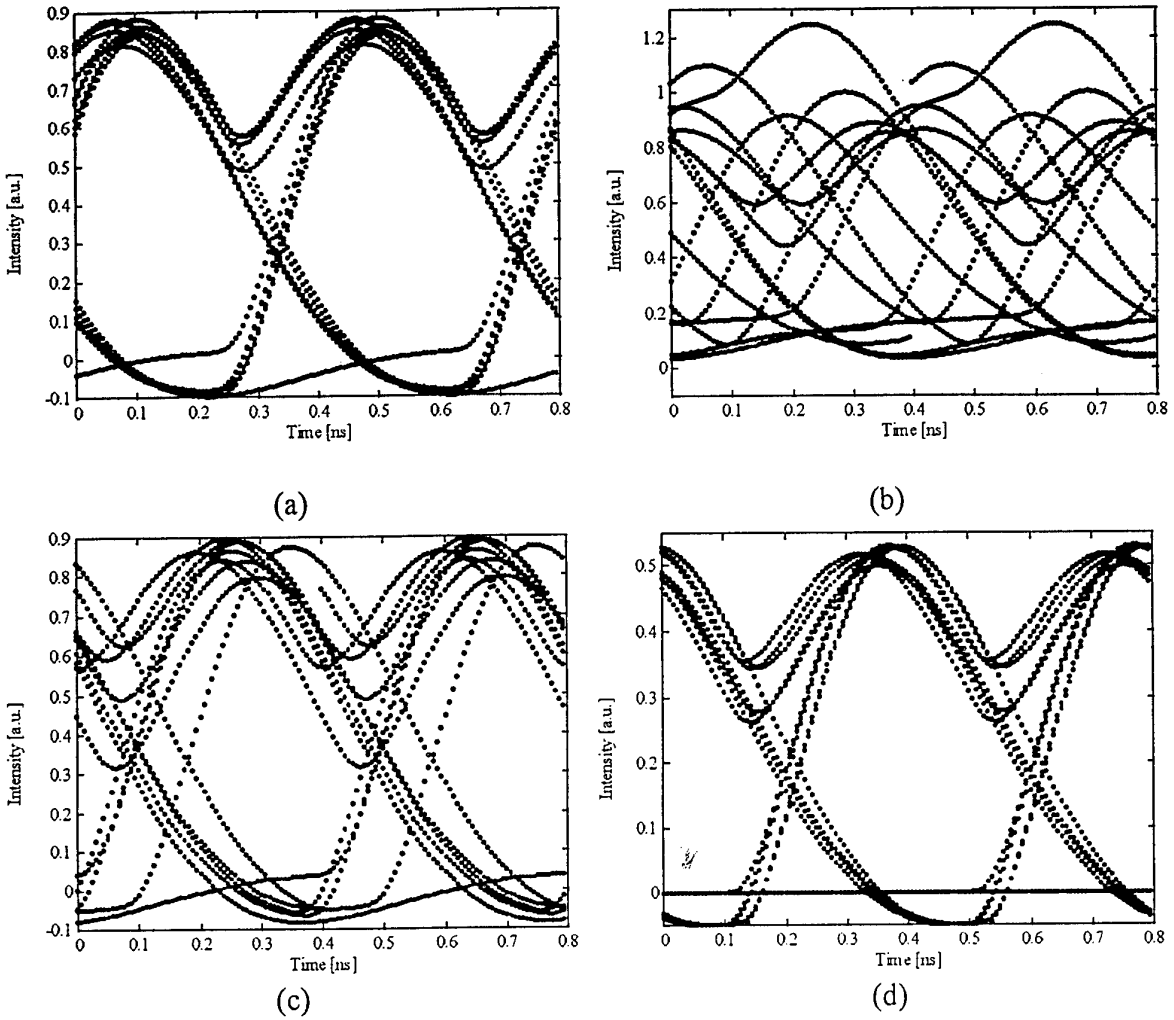


Figure 5: Eye diagrams : (a) at the laser output, (b) at the link output, (c) at the link output for fixed filters, (d) at the fibre output for sliding filters

5. References

[1] M. Nakazawa, K. Suzuki, E. Yamada, H. Kubota, Y. Kimura, M. Takaya "Experimental demonstration of soliton data transmission over unlimited distances with soliton control in time and frequency domains", Elect. Lett, vol. 29, n°9, pp. 729-30, April 1993.

[2] X. Wang, M. Blondel "An accurate large-signal dynamic model of distributed feedback lasers," *Technical Digest, Conf. on Lasers and Electro-Optics Europe*, paper CTuS4, pp. 181, Aug. 1994.

[3] A. Hasegawa, Y. Kodama "Guiding-center soliton in optical fibres", Opt.Lett., vol. 15, n°24, pp. 1443-5, December 1990.

[4] G.P. Agrawal, *Nonlinear Fibre Optics*, San Diego, CA: Academic Press, 1989.

[5] J.P. Gordon "Dispersive perturbations of solitons in the nonlinear Schrödinger equation", J. Opt. Soc. Am. B, vol. 9, pp. 91-7, 1992.

[6] M. Nakazawa, K. Suzuki, Y. Kimura "Transform-limited pulse generation in the gigahertz region from a gain-switched distributed-feedback laser diode using spectral windowing", Opt. Lett., vol. 15, n°12, pp. 715-7, June 1990.

[7] L.F. Mollenauer, B.M. Nyman, M.J. Neubelt, G. Raybon, S.G. Evangelides "Demonstration of soliton transmission at 2.4 Gbit/s over 12,200 km", Elect. Lett., vol. 27, n°2, pp. 178-9, January 1991.

[8] J.P. Gordon, H.A. Haus "Random walk of coherently amplified solitons in optical fibre transmission", Opt. Lett., vol. 11, n°10, pp. 665-7, October 1986.

[9] A. Mecozzi, J.D. Moores, H.A. Haus, Y. Lai "Soliton transmission control", Opt. Lett, vol. 16, n°23, pp. 184-3, December 1991.

[10] Y. Kodama, A. Hasegawa "Generation of asymptotically stable optical solitons and suppression of the Gordon-Haus effect", Opt. Lett., vol. 17,n°1, pp. 31-3, January 1992.

[11] L.F. Mollenauer, J.P. Gordon, S.G. Evangelides "The sliding-frequency guiding filter: an improved form of soliton jitter control", Opt. Lett., vol. 17, n°22, pp. 1575-7, November 1992.

DISCUSSION

Discussor's name: N. Farsaris

Comment/Question:

Did you take into account fibre attenuation and what was the distance between two optical filters in the link?

Author/Presenter's reply:

The fibre attenuation was taken into account by assuming periodical optical amplification by means of Erbium-Doped Fibre Amplifiers (EDFA's). The spacing between the filters was assumed to be the same as the amplifiers' spacing (typically 30-50 km).

SITE-SPECIFIC RADIO PROPAGATION PREDICTION METHODS FOR URBAN ENVIRONMENTS

R. Luebbers
J. Schuster
Department of Electrical Engineering
The Pennsylvania State University
University Park, PA 16802, USA

SUMMARY

With high frequency portable communication systems becoming more common, the need for fast and accurate predictions of high frequency radio propagation in urban environments has rapidly increased. Until recently such predictions were usually based on empirical methods, especially if relatively large numbers of predictions were required. But advances in both computational methods and computer speeds now allow for fast, site-specific predictions of radio propagation even in complex urban areas. In this paper we are primarily concerned with narrow band signal attenuation, but the ray methods to be discussed are also applicable to determination of signal delay and delay spread. Various approaches being used for such predictions are discussed in this paper. The most popular current approaches involve some combination of shooting multiple rays and evaluating their amplitudes using diffraction theory. For this reason we will concentrate our paper on this approach. Basic considerations in the application of ray methods will be discussed, and some typical results are presented.

1 INTRODUCTION

In this paper we are concerned with predicting radio propagation path loss for high frequency communication in urban environments. By high frequency we mean frequencies such that the building dimensions are very large when measured in wavelengths. This is often considered to include frequencies above 900 MHz, but lower frequencies may also be considered.

Until relatively recently, site-specific predictions were not possible given realistic limits on the amount of both human and computer time that could be expended. Instead of site-specific methods, empirical methods were applied. These are not specific to a particular site, but classify sites in categories, such as dense urban, suburban, rural, etc. Path loss is estimated based on these classification and on the distance between the transmitter and receiver antennas. These methods obviously cannot provide predictions for specific paths. They are useful for providing area coverage statistics, and have the advantage of being very fast to apply even on computers of limited capability.

Before considering site-specific predictions in any detail, let us first discuss what it is we hope to predict, and what we can't predict. Variations in signal level due to interference between energy traveling via different paths usually cannot be predicted. This "fast" fading occurs over distances of about 1/2 wavelength, and can be many dB. Site-specific calculation of this fast fading would require knowledge of the spatial positions of the buildings and antennas to within a fraction of the wavelength, and for the frequencies considered here this is not feasible.

An exception to this is the signal variation caused by rays whose relative path lengths are know accurately. For example, consider the pair of rays that travel the same path except that one is reflected from the ground. If the transmitting and receiving antennas are relatively close to the ground, these two rays are traveling paths that differ only slightly. Thus the fading caused by their interference can be calculated for a specific location.

Therefore our site-specific predictions are primarily concerned with predicting "slow" fading for specific sites. Slow fading is not determined by the relative phase of energy arriving via different paths, but rather by the net energy that propagates from the transmitter to the receiver and is available to be received. The relative phases, other than for special cases like the direct and ground reflected ray, are neglected since they contribute to "fast" rather than "slow" fading.

In the remaining sections of this paper the problem of predicting site-specific propagation will be approached as an electromagnetic scattering problem. We will first briefly consider different possible methods. After this survey an

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

approach based on a combination of two of them will be described. Finally, some example results will be presented.

2. HIGH FREQUENCY METHODS

Since the problem we have set for ourselves involves predicting the scattering from structures (buildings in our case) that are very large compared to the wavelengths of interest, we need to consider high frequency electromagnetic computation methods. Candidate methods include Geometrical Optics, Physical Optics, the Physical Theory of Diffraction, the Geometrical Theory of Diffraction (GTD), and Shooting and Bouncing Rays (SBR).

Geometrical Optics is a ray tracing method that is simple and straightforward. With this method the propagation of the radio waves is approximated as following straight ray paths. These rays may reflect from building surfaces, but diffraction around building corners and over roofs is not included. Geometrical optics may be used in situations where direct and reflected rays dominate. But when energy diffracted around building corners or over building rooftops is important, Geometrical Optics cannot be used.

Physical Optics approximates the surface currents flowing on the scattering surfaces (the buildings), and then calculates the fields radiated by these currents. The major difficulty with physical optics for our application is that it is extremely slow for large structures. It is also not accurate for computing the fields diffracted into shadow regions, behind buildings for example.

The Physical Theory of Diffraction is an improvement on Physical Optics. It makes simple approximations for most of the currents on a scattering surface, with special terms for the currents flowing near the edges. It is more accurate at computing the fields diffracted into shadow regions than physical optics, but is still relatively slow for the electrically large structures we are considering.

The Geometrical Theory of Diffraction (GTD) method is accurate, and has been successfully applied to propagation prediction where the geometry to be analyzed is relatively simple. For example, site-specific propagation prediction over hilly rural terrain has been calculated using GTD for over a decade [1]. For this situation the terrain would be approximated by 10 or 20 connected flat "plates" forming diffracting wedges. Then wedge diffraction coefficients [2,3] would be applied to determine the energy reflected and diffracted by the terrain. The drawback for the GTD is that for each pair of transmitter and receiver locations the specific GTD ray paths must be determined. Even for the simple two-dimensional (2D) terrain problem the GTD requires considerable computer time to search out all the necessary rays. For the urban propagation situation considered here, with hundreds or thousands of building faces, the GTD method would require very large amounts of computer time.

The Shooting and Bouncing Ray (SBR) method is a more recent development. There are different implementations of the method, varying in how the rays are combined to determined the field strength at the receiver and in how the amplitudes of the diffracted rays are evaluated. The main advantage of the SBR is that it is extremely fast at propagating large numbers of direct and multiply-reflected rays which are interacting with large numbers of reflecting/diffracting surfaces.

Since the GTD and SBR methods are the two most widely used high frequency methods, the next two sections of this paper will describe them. This will be followed by a description of a method which combines them to produce both a 2D "canyon" model and a fully three-dimensional (3D) urban propagation model.

3. Shooting and Bouncing Ray Method

The SBR method was developed by Professor Lee and coworkers of the University of Illinois [4]. Prior to this development the GTD was the preferred ray method for calculating high frequency radar scattering from aircraft and other complicated structures. But as the amount of detailed geometry information available from Computer Aided Design data bases increased, the algorithms for finding the GTD rays were not efficient enough to find the ray trajectories. In the GTD approach, the rays traveling from the transmitter to the receiver must follow precise paths determined by the extended Fermat's principle. This prescribes that each ray path must have a length that is a local minimum, maximum, or inflection. When thousands of flat reflecting surfaces are involved, finding these particular

rays takes a large amount of computer time. Furthermore, for each receiver location the process must be repeated, since the ray paths will change.

The SBR method avoids this by "shooting" large numbers of rays from the transmitter location. These rays are shot, at least initially, without regard for the receiver location. They then "bounce" off the reflecting surfaces following the laws of reflection. Since no particular ray paths are being searched for, each ray travels through its trajectory extremely quickly. Many more rays are considered than with GTD, but each SBR ray is much faster to compute.

In addition to the relative speed for computing individual SBR rays, there is another advantage. With one "shooting" the SBR rays for all the receiver locations can be found. That is, the rays are "shot" without regard to receiver location, and then the rays going through a collection region surrounding a particular receiver location are combined to determine the field strength at that location. To consider another receiver location, one merely changes the collection region and repeats the process on the rays going through the new region.

The shooting and bouncing process is very straightforward. Consider one of the many rays shot from the transmitter location. It is shot in a particular direction, and continues until it either intersects a building face or exits the computation region. If more than one intersection occurs, the one closest to the source is selected since this one intercepts the ray before it can strike any more distant ones.

Once the closest reflection point is found, the ray is then reflected. This process continues until the maximum number of reflections (as set by the user) is reached, or until the ray leaves the computation region. If we let point 0 denote the location of the transmitter, point 1 the first reflection point, N the last reflection point, R_n the reflection coefficient for the nth reflection, r_{0F} the total distance traversed by the ray between in traveling from the source to the field point, and E_0 the peak amplitude of the electric field 1 meter from a unity gain transmitting antenna, then the complex amplitude of the qth SBR ray after undergoing the N reflections is given by

$$E_q = E_0(\theta,\phi)\,\frac{e^{j\beta r_{0F}}}{r_{0F}} \cdot \prod_{n=1}^{N} R_n \qquad (1)$$

where in general the R_n reflection coefficients will be angle and polarization dependent. For the 3D case the electric fields are 2D vectors (one term for each polarization), the reflection coefficients are dyads, and the polarization vectors must be converted at each reflection into the ray-fixed coordinated suitable for application of the reflection coefficients.

In (1) E_0 is given an angular dependence indicating that any reduction in the radiated electric field due to directionality of the transmitting (or receiving) antenna must be included. This is easily done since the direction of transmission or reception of each ray is known.

Equation (1) describes the basic SBR approach. However, there can be variations in how the method is applied even before diffracted energy is considered. Let us consider how the rays reflected from the ground might be included. For each SBR ray that travels directly to the receiver point as described in equation (1), another will reach the receiver after being reflected from the ground surface. One approach to include the ground reflected rays is to locate the transmitter antenna at point 0′, the image of point 0 in the ground surface. Then the reflected SBR ray is calculated by the same process as the direct SBR ray. It is evaluated using an equation similar to (1), but modified to include the ground reflection coefficient R_G. The equation for the qth reflected SBR ray E'_q after the nth building reflection is

$$E'_q = E_0(\theta,\phi)\,R_G\,\frac{e^{j\beta r_{0'F}}}{r_{0'F}} \cdot \prod_{n=1}^{N} R_n \qquad (2)$$

The advantage of treating ground reflected rays as a special case is that these two rays can then be combined coherently. That is, the direct ray and corresponding

ground reflected ray, as a pair, can be combined as complex fields with phase included. For 3D calculations, where it is assumed that the transmitter is significantly higher than the receiver, the ray path lengths for this pair of rays will differ by much larger amount than for the 2D calculations. For this reason in a 3D calculation the ground reflected rays usually are combined incoherently with the corresponding non-ground-reflected rays. However, different propagation models may combine rays coherently or incoherently in different circumstances than these.

For a fully 3D calculation the polarization of the electric fields should be included in the equations since the reflection coefficients are polarization dependent. For 2D calculations (with low transmitter and receiver heights and high buildings) the polarization dependence is often approximated to simplify the calculations. For example, if the transmitting antenna is vertically polarized, then it may be assumed that all building face reflections are calculated using perpendicular reflection coefficients, and all ground reflections are calculated using parallel reflection coefficients.

Using the SBR method, one cannot initially evaluate the field at a particular point, since one cannot be guaranteed that all of the important SBR rays will pass through that point. One must determine a collection area enclosing the point, and then use the rays that pass through this collection area to determine the predicted field strength.

The collection area must not be made too small relative to the number of SBR rays, or not enough SBR rays will be intercepted. For calculations made at frequencies of 900 MHz and 1.9 GHz we have had good success with 5 meter squares (cubes for 3D calculations) centered on the receiver location. However, collection areas as small as one meter square have yielded good results. Smaller collection areas require more SBR rays and therefore more computer time and memory.

The density of the SBR rays must be adjusted for the collection aperture size so that at least two or three rays for each unique ray path are intercepted. In order to set the number of rays to be shot the following approximate rule can be applied. Let d be the longest distance traveled by a ray (approximately 2 to 3 times the distance across the area of buildings being considered), let L be the length of one side of the collection area or volume in the same units as d, and let n be the minimum number of rays for each distinct path desired to be intercepted by the collection area. Then an estimate of the number N of rays to be shot from the source is

$$N \approx \frac{2\pi nd}{L} \tag{3}$$

All of the rays passing through the collection area are sorted and combined to determine the predicted field strength. The sorting is done to select one SBR ray for each unique path. For example, perhaps several line-of-sight rays are intercepted by the collection area. The sorting process finds all of these, and the one line-of sight ray closest to the center of the collection area will be selected. Using the approach described above, the ground reflected ray may be automatically included with it for a 2D calculation with low antenna heights, while for 3D it may be determined separately. The sorting process may also find several rays that were collected after reflecting from the same set of building faces. Again, from these rays the one closest to the center of the collection area, and its corresponding ground reflected ray, will be selected. Thus for Q rays, one for each distinct path, passing through the collection area, the predicted complex field amplitude E_p at the center of the collection area is given for a 2D calculation by

$$E_p = \sum_{q=1}^{Q} E_q + E_q' \tag{4}$$

In (4) each pair of rays is evaluated coherently for a 2D calculation. Then for 2D each pair of rays is usually combined incoherently (in magnitude) with all the other pairs of rays using

$$E_{Tot} = \sqrt{\sum_P | E_p |^2} \tag{5}$$

to determine the magnitude of the total electric field E_{Tot}. For 3D each separate ray is combined incoherently as in (5)

with no distinction between rays reflected from the ground and other rays except for using the appropriate ground reflection coefficient. The result of (5) is readily converted to path loss once the gain, receiving pattern, and polarization of the receiving antenna are used to convert the total electric field to received power.

If the field strengths at all receiver points were determined by line-of-sight and reflected energy, then the approach described above would be sufficient to determine the field levels. However, for many receiver locations diffracted fields make an important contribution. For these receiver locations the SBR method must be augmented by adding diffraction. While different diffraction coefficients exist, a reasonable approach is to use the Geometrical Theory of Diffraction.

4. Geometrical Theory of Diffraction

The Geometrical Theory of Diffraction (GTD) was originated by Joseph Keller [5]. As with the SBR method, it uses rays to predict electromagnetic scattering from conducting objects. But while SBR includes line-of-sight and multiple bounces (reflections), the GTD also includes edge diffracted rays. In the approach described here, the SBR is used to determine the ray paths, and GTD to evaluate the amplitudes of the diffracted rays. Within the framework of the GTD, the Uniform Theory of Diffraction wedge diffraction coefficient developed by Kouyoumjian and Pathak [2] and modified for imperfect reflection due to finite conductivity and local surface roughness [3] is widely used. For multiple diffraction the spreading factors given in [1] are used, although the UTD is not strictly accurate for multiple diffraction.

As outlined above, the hybrid SBR/GTD approach considered here first shoots many SBR rays. Each of these rays is then bounced through the building geometry. After the rays are determined they are sorted by building interaction. In order to find the diffracting edges (formed by building corners), this sorting process finds adjacent rays that interact differently with the building geometry. For example, let us assume that of two adjacent SBR rays one is reflected from a certain building face while the other is not. This means that a diffracting edge of a building face lies between these two rays. It is then quite simple to locate a diffraction point on this edge.

For a 2D calculation each diffracting building corner will have one disk of diffracted rays. But for 3D, different rays shot from the transmitting antenna will have diffraction points at different locations along the same diffracting edge. From the theory of the GTD, each diffraction point is then the source of a cone of diffracted rays. For a 2D calculation, this cone degenerates to a disk of rays. Usually when GTD is applied only a singe ray of the cone or disk is used, since the ray trajectory from the diffraction point to the field point is known. However, applying SBR at each diffraction point involves "shooting" a partial cone (a disk in two dimensions) of diffracted rays in the region external to the building. These rays then bounce through the building geometry just as if they were "shot" from the original source location, except that their amplitude and phase are determined by the GTD wedge diffraction coefficient. For 2D there is also a companion ray that follows the same ray path except that it comes from the image of the diffraction point in the ground and is reflected once by the ground. For 3D calculations the ground reflected rays are treated separately, as for the non-diffracted rays. For double diffraction the same process is continued to locate the edges for the second diffraction and to shoot the resulting cones (2D disks) of diffracted rays.

Thus in the Hybrid SBR/GTD approach as described here, each building edge that is a source of diffracted rays is considered to be another SBR ray "source". The amplitude of these pseudo-sources is determined using the GTD. But once this amplitude is determined, the SBR method is used as described above to find the ray paths and evaluate the ray amplitudes. The exception to this is that certain combinations of direct, reflected and diffracted rays must be combined coherently for maximum accuracy.

5. Making Calculations using the Hybrid SBR/GTD Method

Let us extend the SBR equations given above to a case of single diffraction. Consider that a particular building corner has been identified as the source of a diffracted ray. Further, let this building corner be illuminated by a ray that originally came from the actual source, and that has been reflected N times, with the notation as given in conjunction with (1) above. Then assume that his ray is further reflected M times as it bounces from the diffracting building edge to the collection area near

the field point. Then for the qth such ray the electric field at the field point will be given by

$$E_q = \left(E_0(\theta,\phi)\ \frac{e^{j\beta r_{oD}}}{r_{oD}} \cdot \prod_{n=1}^{N} R_n \cdot D \cdot \sqrt{\frac{r_{DF}\cdot r_{oD}}{(r_{oD}+r_{DF})}} \right) \cdot \frac{e^{j\beta r_{DF}}}{r_{DF}} \cdot \prod_{m=1}^{M} R_m \qquad (6)$$

This equation includes three separate aspects of propagation. First, the ray propagates from the original source with amplitude E_0 (adjusted for the antenna radiation pattern if necessary) and undergoes N reflections until it reaches the diffracting edge with r_{0D} the distance from the source to the diffracting edge. Then it is diffracted by the edge with GTD diffraction coefficient D, and with its energy spreading changed by the spreading factor given in the square root. Then it propagates from the diffracting edge to the field point F, undergoing M additional reflections along the way.

The hybrid SBR/GTD approach treats this as a ray which is shot from the diffracting building corner, but with amplitude as given by the expression within the large brackets rather than just E_0. But other than this amplitude change, the SBR calculation for this diffracted ray proceeds from the diffracting edge as if the original transmitter source was located there, except that no rays are shot into the interior of the building. For a 2D calculation there is a corresponding equation for the ground reflected ray which can be combined coherently if desired. Equation [6] can be extended to provide a reasonable approximation to double diffraction by using the spreading factors given in [1].

The diffraction coefficient used in (6) is given in [1,2,3]. Thus it also includes the angle-dependent (and polarization-dependent for 3D calculations) reflection coefficients for the particular building face. This is necessary for continuous fields to exist at reflection boundaries. For 3D calculations the electric fields are vectors and the reflection and diffraction coefficients are dyads so that the polarization information is included.

We know that the building locations cannot be known accurately enough so that all rays can be combined coherently.

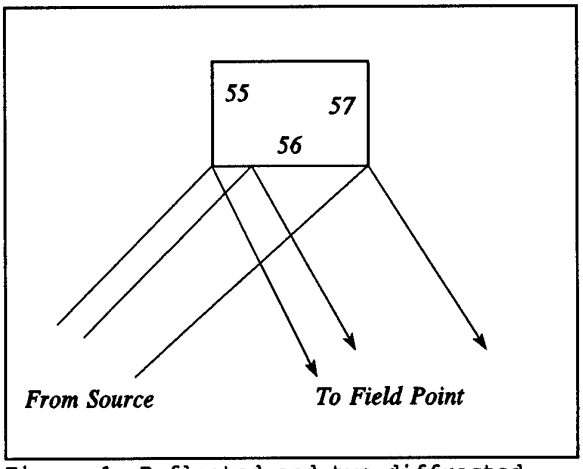


Figure 1 Reflected and two diffracted rays for a single building face

However, the rays reflecting from the same building surface must be combined coherently in order to obtain the full accuracy of the GTD. For example, consider three rays that interact with the same building face as shown in Figure 1. If we imagine a distant field point in Figure 1 moving in a left-right direction, the reflection point of the reflected ray will move on the reflecting surface. It may "walk" off one edge or the other if the field point moves far enough. The GTD maintains continuous fields at these "reflection boundaries" by introducing a jump in the diffracted ray to compensate for the presence or absence of the reflected ray. For this mechanism to work and produce continuous fields as the field point is moved through a reflection boundary, these three rays must be combined coherently. This is possible since they travel nearly identical paths. Even though the building locations are not precisely known, since these rays only interact with the SAME buildings, they can be combined coherently.

A similar situation exists at shadow boundaries. A typical situation is shown in Figure 2. Here a movement of the field point in a direction transverse to

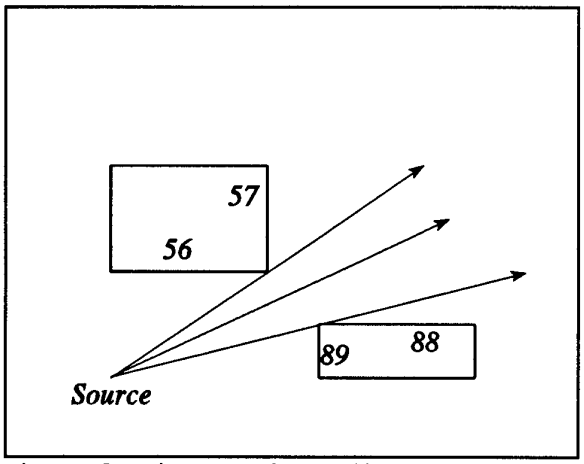


Figure 2 Direct and two diffracted rays which must be added coherently.

the ray paths will cause the ray going between the two buildings to eventually be blocked by one or the other. As the field point moves through these "shadow boundaries," the GTD diffraction coefficient will jump so as to maintain continuous fields at the field point. For proper calculation of the total field produced by the combination of these three rays, they must be combined coherently. This is not an attempt to predict "fast" fading, but rather the way the GTD determines how much energy passes between the two buildings. This complicates the ray combining process somewhat. The Hybrid SBR/GTD approach determines which rays to combine coherently during the sorting process. As a part of the SBR method, all rays passing through a rectangular collection area centered on each field point are collected. These rays are then sorted by their building multipath interactions. The first part of the sorting process is to determine duplicate rays. Only the one passing closest to the center of the collection area is retained.

The second part of the sorting process is to determine which rays to add coherently. To understand how this can be done, we assume that each building face has a unique number assigned to it, and each building corner is labeled using the numbers of the two faces which form it. The sorting process produces a list of rays that are identified by a sequence of numbers which correspond to the path traveled by the ray in going from the original transmitter location to the field point. For example, a ray might have a sequence of [0, 4, 54, 33, 56, 12, 44, 66, F]. This ray would have traveled from the transmitter (0) to the receiver by reflecting from building faces 4, 54, and 33, then reflecting from building face 56, and finally reflecting from faces 12, 44, and 66 to receiving field point F. This will be the only such ray being collected at the field point, since any duplicate SBR ray traveling the same path will already have been discarded. This ray (and its ground-reflected mate for a 2D calculation) will be added incoherently to all other rays EXCEPT for the diffracted rays involved with its shadow or reflection boundaries.

Rays traveling the same path but diffracting from one of the edges of reflecting face 56 must be combined coherently with the reflected ray. We extend our ray notation to include diffractions involving an edge formed by building faces numbered x and y to include the term {x-y}. For the rays shown in Figure 1, the diffracted rays described by the sequences [0, 4, 54, 33, {55-56}, 12, 44, 66, F] and [0, 4, 54, 33, {56-57}, 12, 44, 66, F] must be combined coherently with the reflected ray.

The other rays that must be combined coherently are the shadow boundary rays. For the example shown in Figure 2 let us assume that the ray passing between the two buildings has traveled path [0, 4, 54, 33, 56, 12, 44, 66, F] as before. The two shadow boundary rays will have sequences which differ from this only by a diffraction. For the geometry of Figure 2, the shadow boundary rays will have the sequences of [0, 4, 54, 33, {56-57}, 12, 44, 66, F] and [0, 4, 54, 33, {88-89}, 12, 44, 66, F].

For the example of Figure 2 the ray path does involve different buildings, and thus it first appears that coherent combination is not possible due to the uncertainty in the building locations. But for these paths the rays are very nearly parallel, and thus the phase differences will be small. It is this relative phase information that is used by GTD to evaluate how much energy passes between the two buildings. As the building get closer together less energy can pass between them. The accurate evaluation of this requires coherent addition of the three rays shown in Figure 2.

We believe that this combination of coherent and incoherent combination of rays is a better approach than either all coherent or all incoherent. If all rays are combined incoherently, the GTD cannot provide the full accuracy of which it is capable. But if all rays are combined

coherently, the rapid fading that is predicted tends to overestimate the path loss. The combination of both methods, as appropriate, seems to us to provide an optimum approach and the most accuracy.

6. Practical Considerations

For our example calculations, the first step was taking the terrain and building data and converting them to a form suitable for propagation prediction. The area considered is Rosslyn, VA, just across the Potomac river from Washington, DC. Rosslyn is a dense urban area, with buildings of height varying from approximately 30 stories to just a few stories. The original building data base was obtained from aerial photography and manually converted to digital points in three dimensions by GDE Systems, Inc. Before applying the propagation model this data was used to form approximations to the buildings. For 2D calculations the vertical variations in the buildings were ignored. The buildings were approximated as being infinitely high with a cross section determined by the footprint of the building at ground level. The ground was approximated as a flat plane passed through the average terrain height. The 2D designation refers to the building being approximated as having no vertical variation, but the vertical heights of the transmitter and receiver antennas are included in the calculations since they effect the coherent combination of the direct and ground-reflected rays. Thus the 2D model really involves 3D calculations in a 2D approximation to the buildings.

For the full 3D calculations the building surfaces were approximated in three dimensions using flat polygons. These could be of arbitrary shape with an arbitrary number of corners, but with the constraint that they must be convex, that is, no polygon interior corner could be greater than 180 degrees. The polygonal edges were not constrained to be horizontal or vertical since the model was implemented to include diffraction from slanted edges. The ground was approximated using polygonal plates as well. For the Rosslyn geometry the diffractions from the edges where the ground plates connected was not included, since these edges are very near to 180 degrees. But this diffraction could be added for other situations where diffraction over terrain hills was significant.

For the reflection coefficient calculations angular-dependent plane wave reflection coefficients for an infinite dielectric half-space were used. For the 2D calculations the relative permittivity of both the ground and building surfaces was taken as 4.0, while for the 3D calculations 5.0 was used. The effect of this change is very slight.

For the 2D calculations in Rosslyn both single and double diffraction were included, with a maximum of 5 total reflections, but with no reflections after the second diffraction. Thus a ray could reflect N times, diffract, reflect M times, and diffract again, with $N+M \leq 5$. Of course all rays reaching the collection area for a particular field point with fewer interactions would also be included. And for 2D the corresponding ground-reflected ray was automatically included and combined coherently.

For the 3D calculations only single diffraction was included. For paths with no diffraction up to 5 reflections were included. Paths involving diffraction could have 0 or 1 reflection, then a diffraction, then 0, 1 or 2 reflections. Ground-reflected rays interacted with the polygonal surfaces approximating the ground without distinction from the rays reflecting from buildings.

All antennas were vertically polarized. For the 3D calculations the antenna pattern of the transmitting antenna was highly directional, and was included in the predictions. The 3D receiving antenna had a vertical dipole ($\sin\theta$) pattern. For the 2D calculations the vertical antenna patterns were not included, and the horizontal patterns were omnidirectional. The path loss L_p defined using

$$L_p = P_r - P_t - G_t - G_r - L_c \qquad (7)$$

where P_r is the received power, P_t is the transmitted power, G_t is the maximum transmit antenna gain, G_r the maximum receive antenna gain, L_c is the cable loss, all given in dB. Thus the path los depends only on the building and terrain geometry and the directional gains of the antennas.

7. Typical Results Obtained using the Hybrid SBR/GTD Propagation Model

In this section of the paper some typical 2D and 3D predictions made using the Hybrid SBR/GTD model described above will

be presented. All predictions were made totally blind, that is, the predictions were made and provided to the group at AT&T Bell Laboratories making the measurements before any measured path loss data was supplied. Additional information about how the measurements were made, and some other comparisons with different propagation models, is included in [6]

The approximate building plan used for the predictions is shown in Figure 3. For the 2D predictions the buildings do not vary in the vertical direction. For the 3D calculations the vertical variation of the buildings was included in the calculations. The receiving antenna was located on a van, with the antenna approximately 2.3 m above local ground. The paths driven by the receiving van are shown in Figure 3. Measurements were made for 900 MHz and 1.9 GHz.

For the 2D example the transmitting antenna was located at site 2B on Figure 3, supported by a tower at a height of 10.0 m above local ground. For the 3D examples the transmitting antennas were located on buildings at sites 5 and 6 on Figure 3. These antennas were located 2 m above the building roofs and approximately 3 meters from the roof edge. Diffraction from the edge of the roof supporting the transmitting antenna was not included in the calculations. The antenna height at site 5 was approximately 42 m above local ground, while at site 6 approximately 43 meters above local ground. These latter two heights are above many of the nearby buildings, so that propagation over the buildings is important.

A typical 2D prediction result for the 10 m high transmitting antenna at site 2B at 900 MHz is shown in Figure 4 for the receiving van driven along the 19th Street/Nash Street path. This path involves some line of sight regions, and some that must be reached by multiple reflection/diffraction. A 3D prediction for transmitter site 5 at 1.9 GHz is compared with measurements in Figure 5 for the Ft. Myer Dr. van path. This path also involves both line of sight and reflection/diffraction regions. Finally, a pair of predictions for transmitter site 6 at 900 MHz is shown in Figures 6 and 7 for Lynn St. and Kent St. respectively. Going from Lynn St. to Kent St. introduces an additional block of buildings between the transmitting and receiving antennas. The 3D prediction is quite accurate in showing the additional attenuation.

8. Acknowledgements

The building data base was supplied by GDE Systems, Inc. The measurements were made by a team from AT&T Bell Laboratories led by J. D. Moore, T. Willis, and B. Guarino supported by R. Rinaldi, D. Jacobs, G. McArthur, J. Johnson, and B. Witkowski.

9. References

1 Luebbers, R., "Propagation Prediction for Hilly Terrain Using GTD Wedge Diffraction", IEEE Transactions on Antennas and Propagation, vol 32, no 9, pp. 951-955, September 1984.

2 Kouyoumjian, R. G., and P. H. Pathak, "A Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," Proc. IEEE, Vol. 62, No. 11, pp. 1448-1461, November 1974.

3 Luebbers, R., "Finite Conductivity Uniform GTD Versus Knife Edge Diffraction in Prediction of Propagation Path Loss", IEEE Transactions on Antennas and Propagation, vol 32, no 1, pp. 70-76, January 1984.

4 Ling, H., R. Chou, and S. Lee, "Shooting and Bouncing Rays: Calculating the RCS of an Arbitrarily Shaped Cavity," IEEE Transactions on Antennas and Propagation, Vol. 37, No. 2, pp 194-205, February 1989.

5 Keller, J. B., "Geometrical Theory of Diffraction," J. Opt. Soc. Amer., Vol. 52, No. 2, pp 116-130, February 1962.

6. T. M. Willis et al, "UHF and Microwave Propagation Prediction in an Urban Environment," Paper 8, AGARD Meeting on Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design, Athens, Greece, September 18-21, 1995.

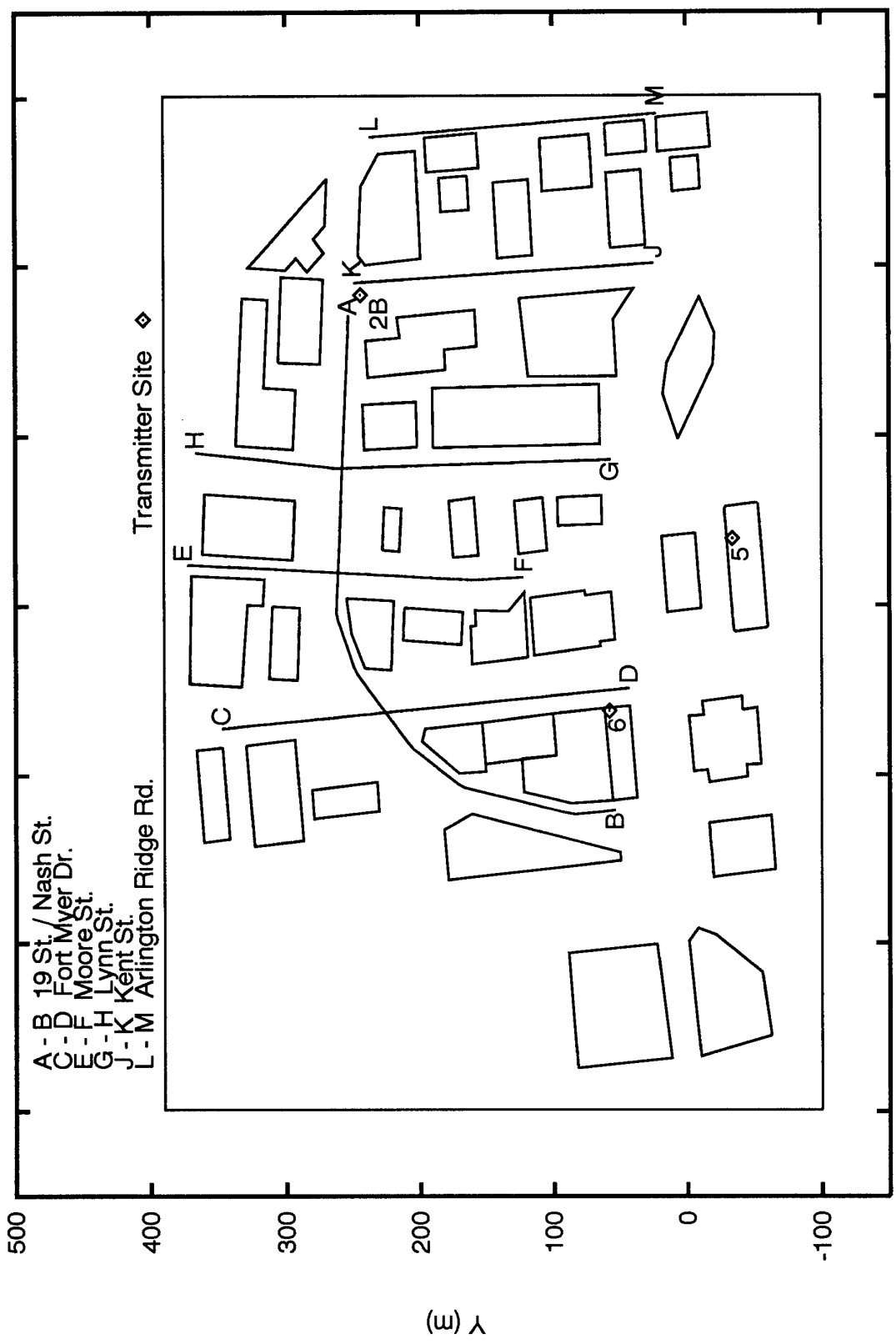


Figure 3: Plan view of Rosslyn VA showing the building locations, transmitter sites, and paths driven by the receiving van

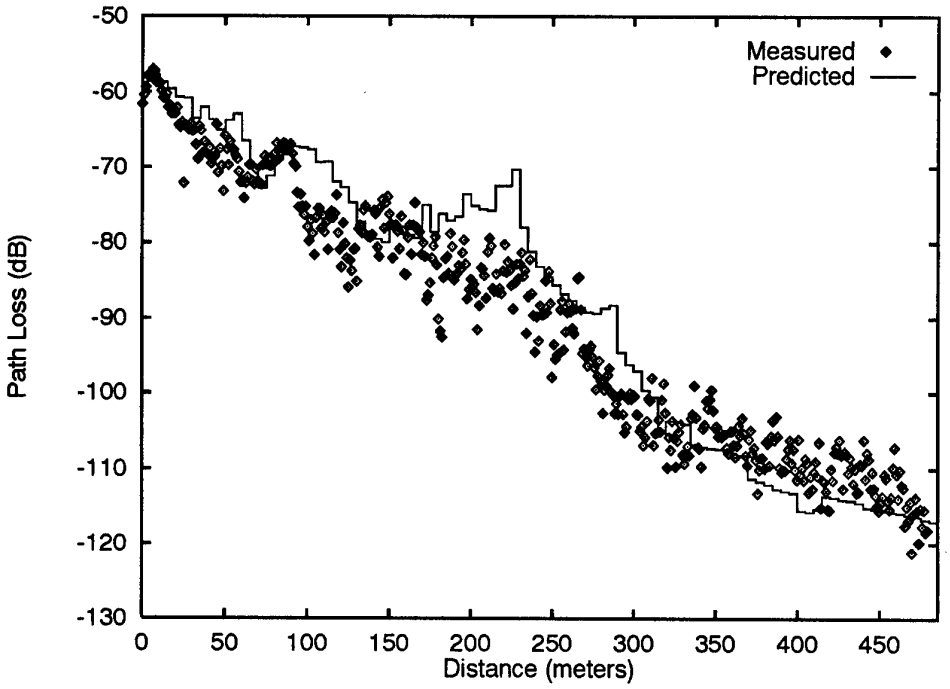


Figure 4: Prediction by 2D SBR/GTD model at 900 MHz for transmitting antenna at site 2B located at 10 m above local ground for the 19th St./Nash St. receiving van path

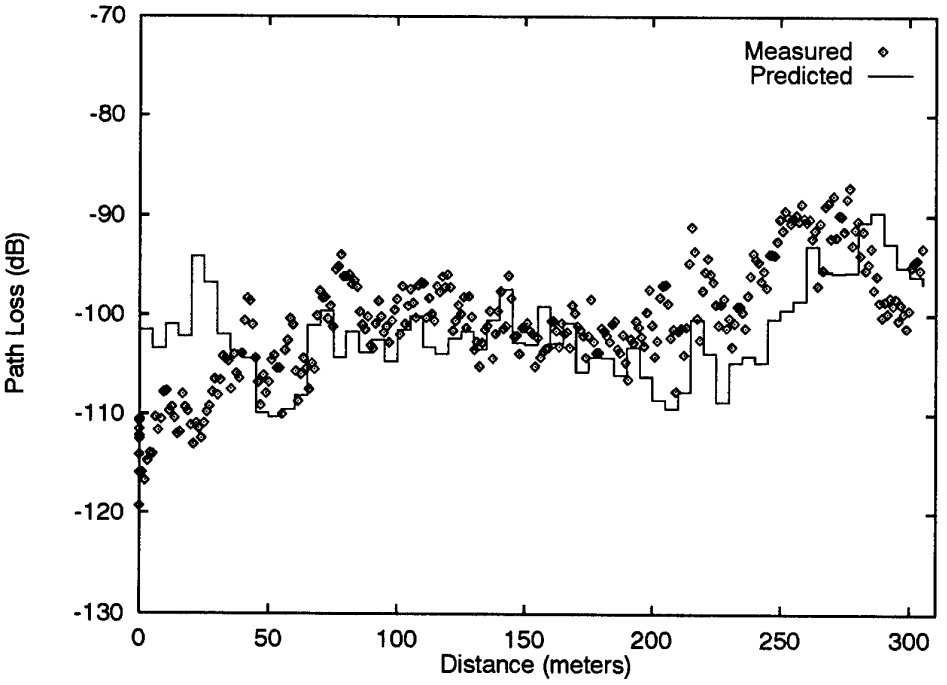


Figure 5: Prediction by 3D SBR/GTD model at 1.9 GHz for transmitting antenna at site 5 located on a roof 42 m above local ground for the Ft. Myer Dr. receiving van path

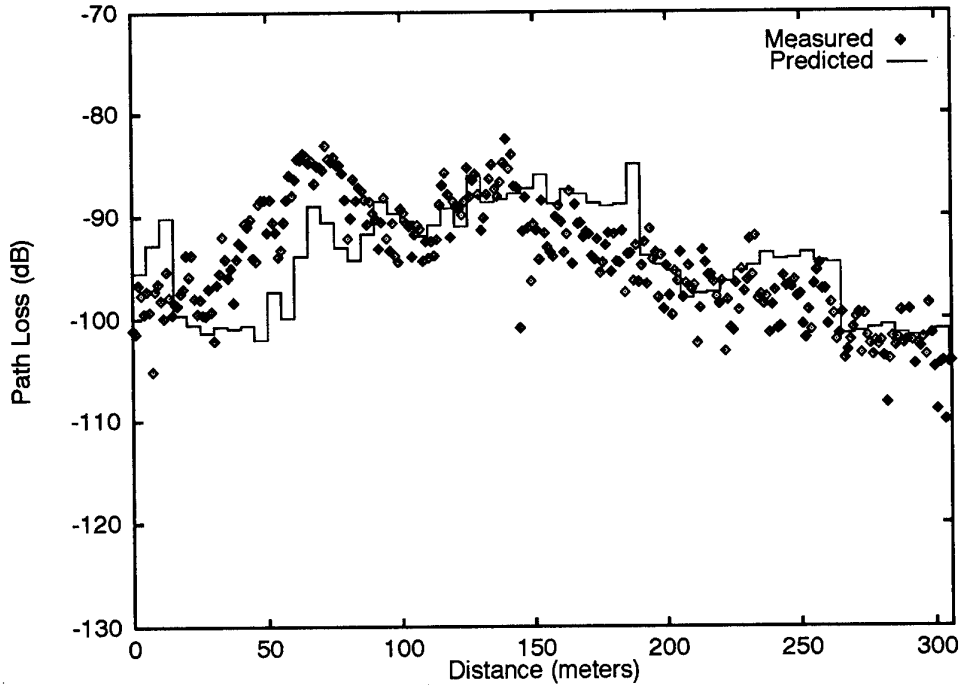


Figure 6: Prediction by 3D SBR/GTD model at 900 MHz for transmitting antenna at site 6 located on a roof 43 m above local ground for the Lynn St. receiving van path

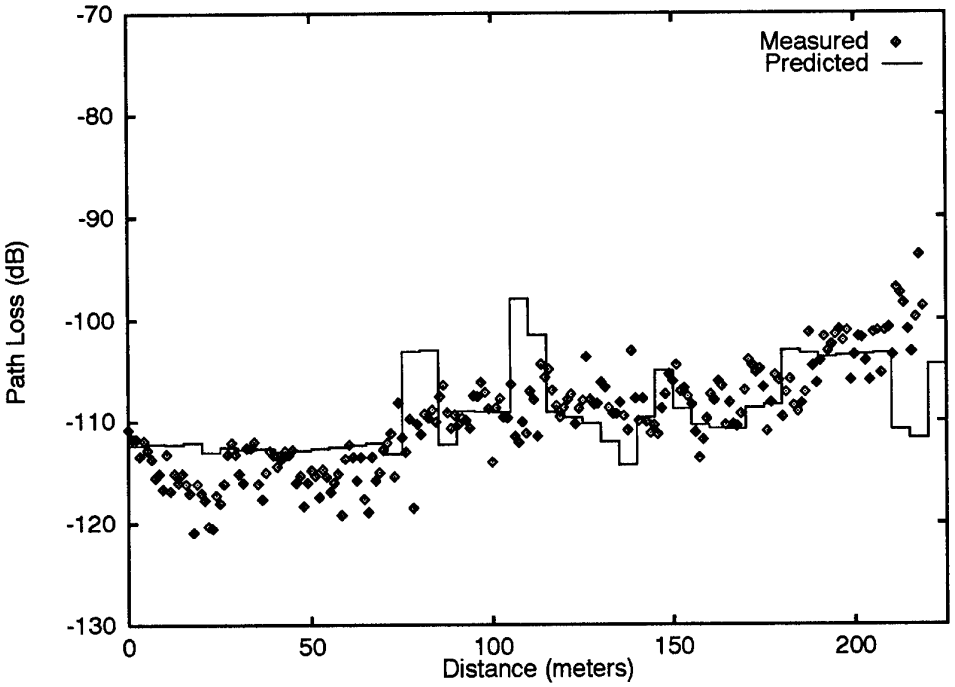


Figure 7: Prediction by 3D SBR/GTD model at 900 MHz for transmitting antenna at site 6 located on a roof 43 m above local ground for the Kent St. receiving van path

DISCUSSION

Discussors's name: G. Brown

Comment/Question:

Do you include the ray that bounces off the ground surrounding the transmitting antenna? Did you account for the topographic variation in your model?

Author/Presenter's reply:

We include the ground reflected ray whenever it exists. In our 2-D model we neglect terrain variation. In our 3-D model we approximate terrain variation with connected triangular flat "plates".

Discussor's name: B. Audone

Question/Comment:

You have said that the difference between measured and predicted values depends on the missing rays. Don't you think that it depends on the reflection coefficients?

Author/Presenter's reply:

This is a good comment. There is certainly some error due to reflection coefficients being approximated, but missing or incorrect rays due to errors/approximations in the building geometry do contribute to errors in predictions.

Discussor's name: F. Davarian

Question/Comment:

1. How are incoherent rays combined? How do you combine rays that are 180 degrees out of phase?

2. What are the applications of the method you have described? Transmitter-site selections?

Author/Presenter's reply:

1. Most rays are combined incoherently, neglecting phase. In special situations we combine coherently, by adding computed fields. A phase difference of 180 degrees would not be treated in any special way.

2. We can use the model to evaluate different transmitter sites.

DISCUSSION

Discussor's name: W. Keydel

Comment/Question:

Each target acts as a polarization transformer. Did you ever introduce polarization considerations? Did you measure with different polarization? Did you consider the loss due to windows, and measure inside the buildings.

Author/Presenter's reply:

All measurements have been outdoors with only vertical polarization. We neglect polarization effects in our 2-D model but include them in our 3-D model.

Discussor's name: K. Craig

Question/comment:

a) One common problem with "shooting" methods of ray tracing is determining the uniqueness and completeness of the rays. Could you comment on how your algorithm deals with this, please? Your comparisons with measurements seem very good, and this does not seem to be a problem in practice.

b) What is a typical size of your receiver "collection area"?

Author/presenter's reply:

a) We check completeness by adding more rays and observing convergence. We assume that each ray follows a unique path by storing the path history and sorting to remove duplicate rays.

b) Our collection areas are typically 3m to 5m diameter spheres.

Name of Discussor: H. J. Strangeways

Question/Comment:

Some of our questioners seem a little sceptical that you have found all the ray paths to a given receiver position accurately. Have you considered doing a superresolution analysis (best using an MLE based method because of the coherent sources) of the received field to compare your predicted DOAs for each ray with experimentally determined values?

Author/Presenter's reply:

This may be a good suggestion. We have not tried this, but rely on simple convergence tests, adding more rays until the predictions no longer change.

An Impulse Response Measurement System and Some Experimental Results from a Forward Scatter Meteor Burst Link

K.J. Ellis[1], A.R. Webster[2], J. Jones[3], S. Chow[1]

[1] Communications Research Centre, 3701 Carling Ave., Box 11490, Station H, Ottawa, Ontario, K2H 8S2, Canada

[2] Department of Electrical Engineering, The University of Western Ontario, London, Ontario, N6A 3K7, Canada

[3] Department of Physics, The University of Western Ontario, London, Ontario, N6A 3K7, Canada

1. INTRODUCTION

Meteor burst is a unique mode of communication which relies on the trails of ionization that are created when meteoric particles enter and burn up in the Earth's atmosphere. Due to the serendipitous nature of these events, these meteor burst channels are inherently intermittent but do possess characteristics which are attractive in military applications. Among these features are the potential for long-range communication (up to distances of 2000 km), low probability of jamming or intercept, and invulnerability to attack by either nuclear or conventional means.

An important factor limiting the overall performance, and consequently the usefulness, of a meteor burst communication system is the maximum data rate that can be supported by the channel. This maximum rate is constrained by bandwidth limitations arising from channel impairments such as multipath and fading. In the majority of cases, these limitations are deduced from knowledge of the channel impulse response, which is usually empirically derived. To date little research, with the notable exception of that conducted by Akram [1], Weitzen [2,3] and Eriksson [4], has been undertaken to evaluate systematically the channel impulse response. An endeavor such as this is justified by the need to analyze and develop communication strategies that fully exploit the limited resources of the meteor burst channel. In pursuit of this knowledge, a novel system incorporating chirp radar techniques was developed to obtain measurements of the channel impulse response, path loss, flight-times and path-lengths. Instrumentation details and experimental results from a 500 km meteor burst link between the cities of London and Ottawa (Ontario, Canada) will be presented in this paper.

2. PRINCIPLES OF OPERATION

The impulse response measurement system relies on the transmission and subsequent demodulation of a wideband linear FM (chirp) waveform. When a meteor trail of suitable ionization and orientation is present, the signal from the transmitter will be scattered or reflected from one or possibly several regions along the trail and will arrive at the receiver as illustrated in Figure 1. At the receiver the signal is demodulated by mixing it with a locally generated chirp whose starting time has been delayed to coincide with the expected arrival time of the signal from the transmitter. Because of the intermittent nature of the channel, the measurement system was designed to collect data only when trails having suitable ionization and orientation are present. This is accomplished by transmitting a brief carrier, or "probe" signal in advance of the chirp transmission. Whenever this probe signal is encountered, a meteor trail is known to exist and the demodulated signal $d(t)$ is digitized and stored in a convenient manner for analysis. Information such as the channel impulse response, path gain and flight-times (path length) can then be determined from the spectral characteristics of the receiver output.

The chirp signal emitted by the transmitter is a sinusoid whose instantaneous frequency is linearly increasing with time and is described by the equation:

$$s(t) = p(t)\cos\left(2\pi f_o t + \pi\mu t^2\right) \qquad (1)$$

where f_o is the starting frequency of the chirp (Hz), μ is the chirp rate (Hz/s) and $p(t)$ is a unipolar rectangular switching waveform which controls the duration and repetition rate of the transmitted signal. At the receiver, a similar chirp signal is generated for demodulation purposes and is described by the equation:

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

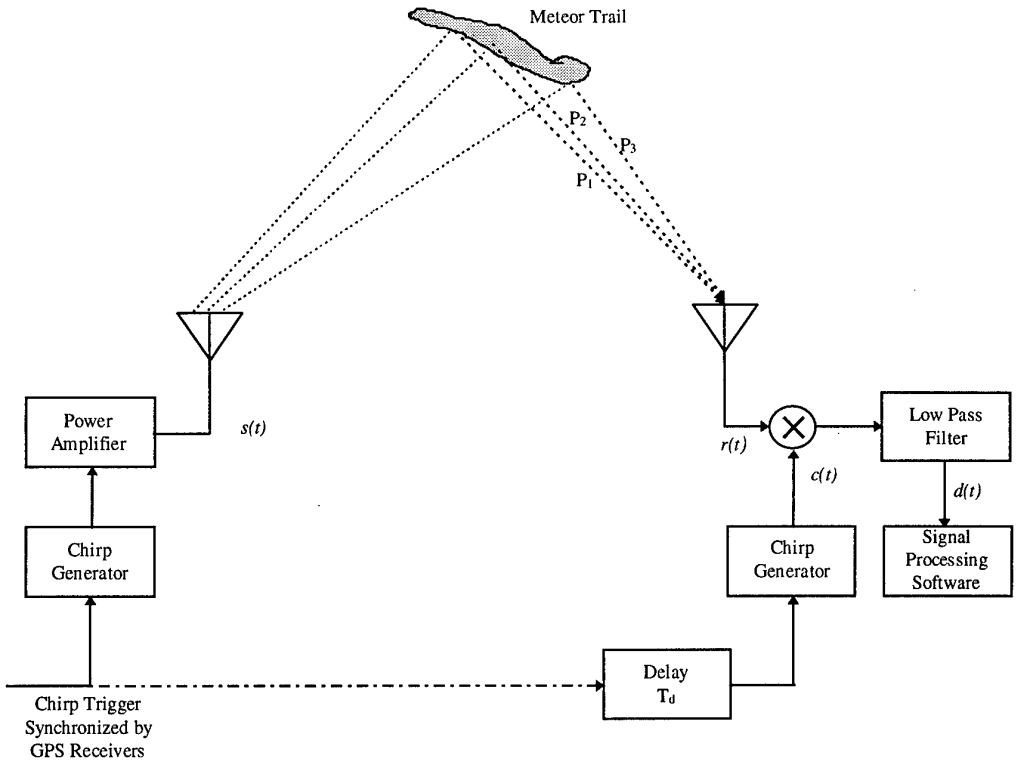


Figure 1 Simplified block diagram of the measurement system

$$c(t) = p(t - T_d)\cos\left(2\pi(f_o + f_d - \mu T_d)t + \pi\mu t^2 + \theta\right) \quad (2)$$

where f_d is a fixed frequency offset (Hz), T_d (s) is the delay between the starting times of the chirp waveforms generated at the transmitter and receiver, θ is a phase term (rad.) and all other symbols have their previously defined meanings. When the signal emitted by the transmitter arrives at the receiver, after taking one or more paths, it is mixed with the locally generated chirp, *c(t)*, and filtered to produce the receiver output signal, *d(t)*, given by:

$$d(t) = \sum_{k=1}^{N} \beta_k p\left(t - T_{pk}\right)\cos\left(2\pi\left(f_d + \mu\left(T_{pk} - T_d\right)\right)t + \theta_k\right) \quad (3)$$

where N is the number of paths between the transmitter and receiver, β_k (real-valued) and T_{pk} represent the gain and flight-times for the k^{th} path, θ_k is a phase term which accounts for phase changes resulting from the reflection/scattering process and all other parameters have their usual meanings. All measurements of interest (time-of-flight, path gain, etc.) are derived from the spectrum of the output signal, *d(t)*. A sketch of an ideal spectrum is shown in Figure 2 for the case in which three paths exist between transmitter and receiver as illustrated in Figure 1. In this case, the flight-times for each path can be estimated from knowledge of the frequency components, f_k, which from equation (3) are given by:

$$f_k = f_d + \mu\left(T_{pk} - T_d\right) . \quad (4)$$

Since the parameters μ , f_d and T_d are known constants which are fixed at the outset of the experiment and f_k is measurable, the preceding equation can be solved in terms of the flight-time giving:

$$T_{pk} = T_d + \frac{f_k - f_d}{\mu} . \quad (5)$$

The spectrum also indicates the relative strengths of the multipath components. It should be noted that the delay time, T_d, must remain constant throughout the measurements and this is achieved in practice by synchronizing the transmitter and receiver timing through the use of Global Positioning System (GPS) receivers located at both ends of the link. Having obtained the flight-times, the corresponding path lengths, P_k, can then be computed using:

$$P_k = cT_{pk} \quad (6)$$

where c is the velocity of light in a vacuum. In addition to providing time-of-flight information, the channel impulse response can be obtained by computing the spectrum of the receiver output signal *d(t)* and rescaling the frequency axis according to equation (5) (see Figure 2). Furthermore, when the receiver is properly calibrated to measure the power of each of the frequency components, f_k, the path losses between transmitter and receiver can also be measured.

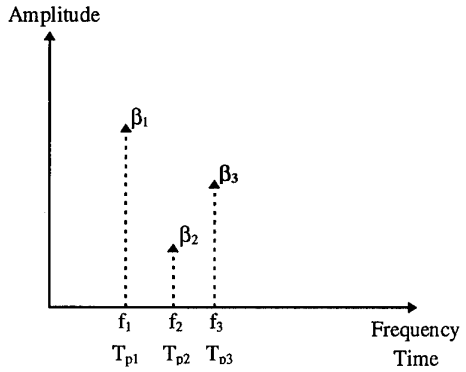


Figure 2 Ideal receiver output spectrum and corresponding time scale.

3. THE EQUIPMENT

A block diagram showing the constituent components of the transmitter system is shown in Figure 3. The signal emitted by the transmitter is initially generated at a low intermediate frequency (IF) with the chirp signal occupying the frequency range from 1.0 to 3.5 MHz at the chirp generator output. This signal is synthesized digitally with the signal samples being generated on a personal computer and uploaded into a large memory array contained within the chirp generator unit. Once the samples have been stored in the memory array, they are then clocked out to a digital-to-analog (D/A) converter and filtered to obtain the desired analog signal. At this point the low IF chirp signal is sequenced with a 2.0 MHz signal, the low IF equivalent of the probe signal, giving the transmission cycle indicated in the diagram. This composite signal is then upconverted to the desired frequency range in a dual conversion process producing a chirp signal with a time-bandwidth product ($\tau\Delta f$) of 125000 which sweeps over a 2.5 MHz bandwidth in 50 ms beginning at 38.75 MHz and ending at 41.25 MHz as well as a 10 ms probe signal at 39.75 MHz. This signal is then amplified and applied to the antenna.

At the receiver, the signal from the antenna is split into two branches (see Figure 4). The first branch is connected to the probe receiver which detects the presence of the probe signal and triggers the data recording system. The second branch is applied to the chirp receiver which translates the spectrum of the received signal to a low IF and demodulates the signal by mixing it with a locally generated version of the chirp. It should be noted that the chirp signal generated at the receiver is identical to that generated at the transmitter, but has a frequency offset, f_d, of 12 kHz so that a signal having a propagation time equal to the delay time, T_d , will generate a beat note of this frequency. After demodulation, the signal is amplified, filtered and stored for subsequent analysis by the data recording system shown in Figure 5.

Signals present at the receiver output are sampled by the analog-to-digital (A/D) converter at a rate of 200×10^3 samples per second when triggered by reception of the probe signal. Because of the relatively high sampling rate involved, the samples representing the receiver output are transferred into the personal computer (PC) using direct memory access (DMA) techniques. At the end of a predetermined number of sweeps the information stored in computer memory is transferred to hard disk and is eventually archived on tape.

4. ANALYSIS AND EXPERIMENTAL RESULTS

As all the information of interest is contained in the spectrum of the receiver output signal, each sweep recorded by the instrument is processed by first weighting the time domain signal samples with a Dolph-Chebyshev window function having a peak sidelobe level 40 dB below that of the main lobe (Harris [5], Nuttal [6]) and then computing its power spectrum using the Fast Fourier Transform (FFT) algorithm. This particular weighting function was selected on the basis of simulation results which showed that it offered

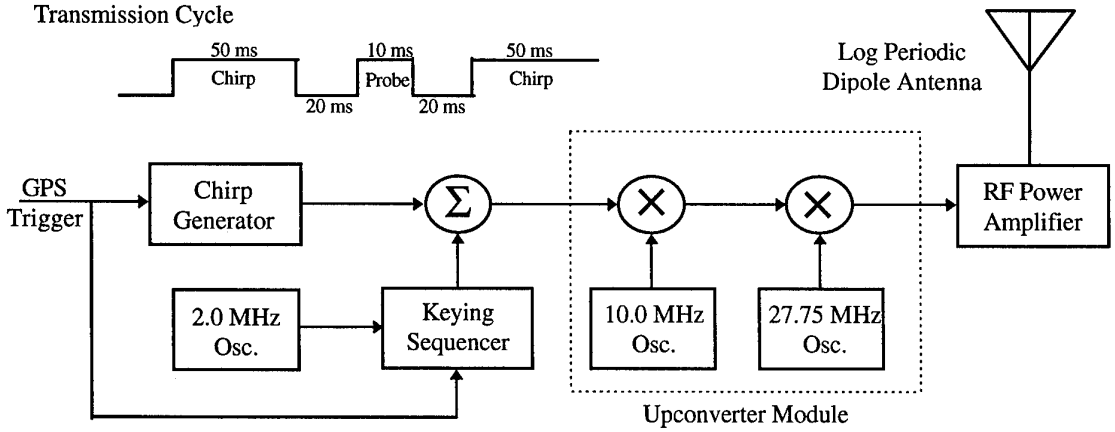


Figure 3 Transmitter block diagram

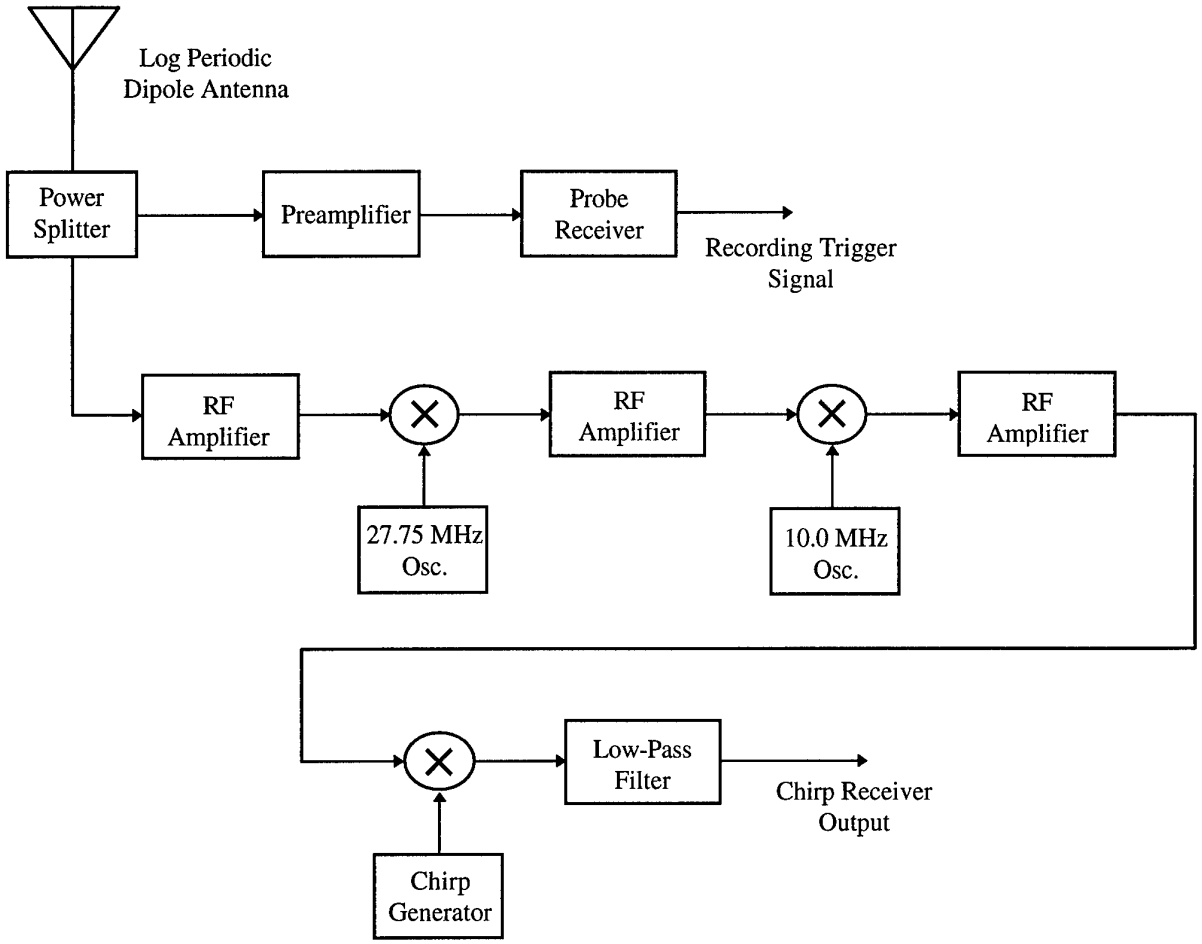


Figure 4 Block diagram of the receiver

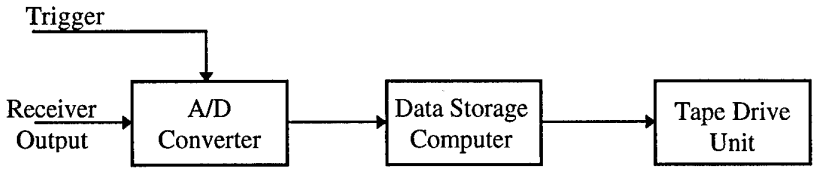


Figure 5 The data recording system

marginally better resolution than that achievable using the Hamming window function. These simulation results showed that the system employing Dolph-Chebyshev windowing was capable of resolving a flight-time differential of 0.8 μs, which is equivalent to a path length differential of 240 metres. After the power spectrum of an individual sweep has been computed, the power levels and frequencies associated with each spectral peak are then used, in conjunction with the receiver calibration data, to determine the gains and flight-times of each of the resolved paths. During the receiver calibration process, the dynamic range of the receiver was found to be limited to approximately 35 dB.

Applying this analysis procedure to the data collected on a 500 km meteor burst link between London and Ottawa, the series of impulse response profiles shown in Figures 6 and 7 were obtained. In each of these profiles, the sweeps recorded during the passage of a single meteor echo have been sequentially arranged and plotted as a grayscale image. The values plotted along the vertical axis represent the flight-times (T_p , ms), while those plotted along the horizontal axis identify the individual sweeps which were recorded at intervals of 100 ms. Path gain, or equivalently signal strength, is conveyed by means of a gray-scale whose varying shades represent the path gain variability between the transmitter and receiver. One of the advantages of this particular presentation format is that when multipath conditions exist, distinct bands are formed (in the sweep axis direction) with each band representing a particular flight-time and corresponding path length. In the case of the impulse response profiles shown in Figure 6, it is apparent that only a single path exists between the transmitter and receiver, while for those profiles shown in Figure 7, there is clear evidence of multipath propagation. The largest flight-time differentials for each of these profiles (Figure 7 (a-c)) are approximately 1.0, 2.7 and 6.2 μs respectively.

In the impulse response profiles of Figure 6 (a) and (b), both echoes are relatively short lived and have durations of 0.8 and 0.7 s, respectively. During this time there is some path gain variation due to fading before the ionized trail finally disperses and becomes unusable. Because of the relatively short durations of these recordings, it is highly probable that these echoes were produced by trails having underdense ionizations (i.e. an electron line density less than 10^{14} electrons/metre). Although the impulse response of Figure 6 (c) is much longer in duration (2.0 s) than the first two profiles (most likely the result of an overdense trail) it exhibits similar path gain variations. Neglecting fading effects, these three impulse response profiles are relatively well-behaved throughout the duration of the respective echoes and are unlikely to have any serious detrimental effects on communication. In contrast, those shown in Figure 7 illustrate the vacillating conditions that can be experienced during the lifetimes of some ionized trails. It should be noted that, while in many instances the path gains associated with the secondary paths are on the order of 10 dB or more below that of the dominant path, cases have been observed where the gains of several paths are nearly equal, as is the case with the profiles of Figure 7 (b) and (c). Under conditions such as these, the channel transfer function will become frequency selective and will impair communications.

4.1 Multipath Characterization

An easy way to assess the multipath nature of the channel is to observe the frequencies with which multiple paths are resolved by the instrument. Applying this analysis to the impulse response measurements collected on the 17th and 18th of December 1993 yields the results shown in Table 1. These figures show that in 97% of all sweeps made during this two-day period there were no more than two resolved paths between the transmitter and receiver. More importantly, however, is the fact that in roughly 70% of these measurements no multipath was observed at all. This finding suggests that adverse effects due to multipath interference may be encountered only on an occasional basis. Although none of the sweeps considered in this analysis indicated the presence of more than 4 paths, the impulse response profile of Figure 7 (c), measured during a separate campaign in 1995, possesses as many as 8 paths and has been included to illustrate the severity of conditions that can be encountered in meteor communications.

Date	1 Path	2 Paths	3 Paths	4 Paths	>4 Paths	Total
1993/12/17	113	38	2	1	0	154
1993/12/18	168	69	8	1	0	246
Combined	281	107	10	2	0	400

Table 1 Summary of Path Count Measurements

The multipath nature of a communications channel is frequently described in terms of its root-mean-square (rms) delay spread (Eriksson [4], Cox [7]) which is, in essence, a measure of the width of a power-delay profile. In the context of the work presented here, a power delay profile is the power spectrum of the receiver output signal and is intimately related to the individual cross-sections of the impulse response profiles shown in Figures 6 and 7. The delay spread, S (s), is defined as the square root of the second central

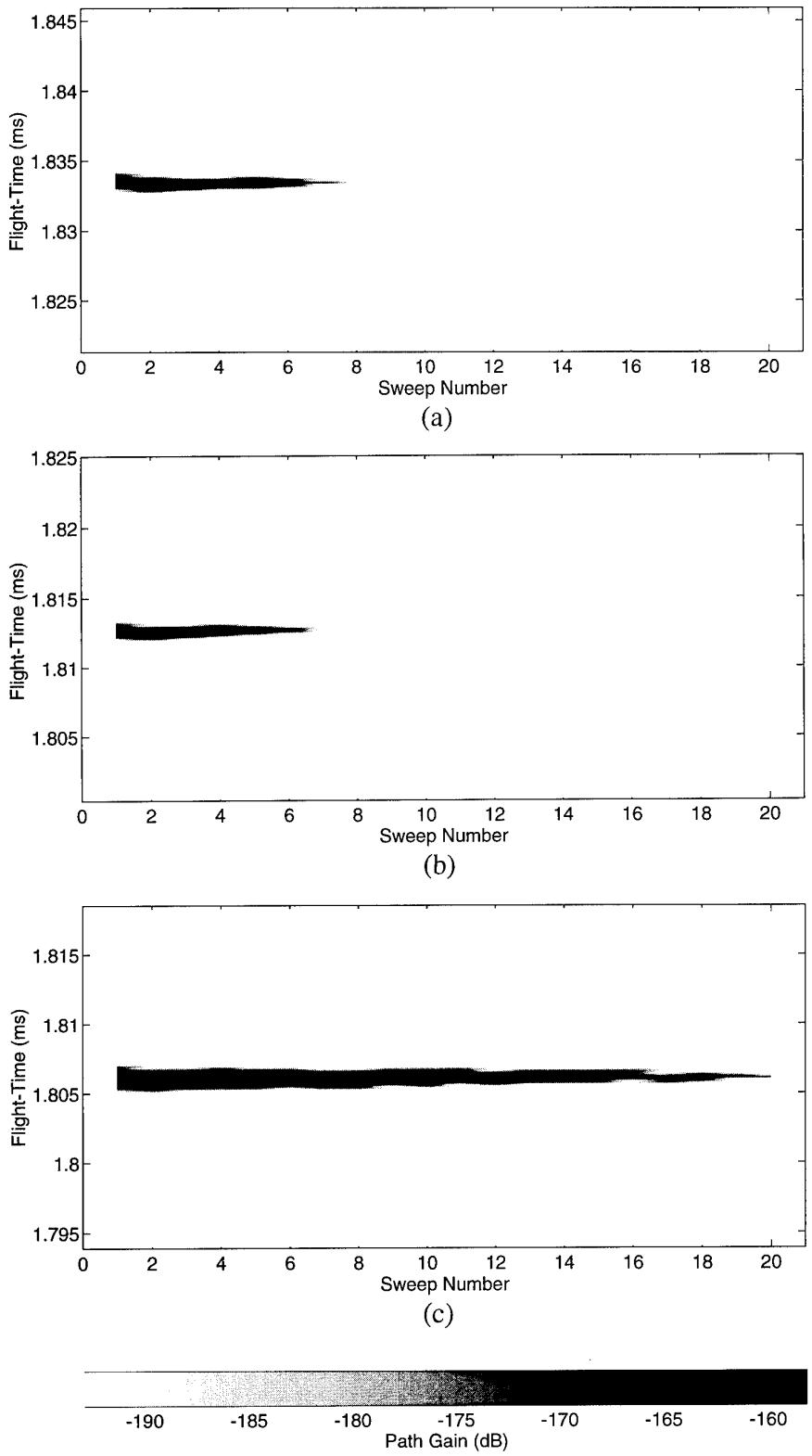


Figure 6 Impulse response profiles showing no evidence of multipath. Profiles such as these were the most frequently observed.

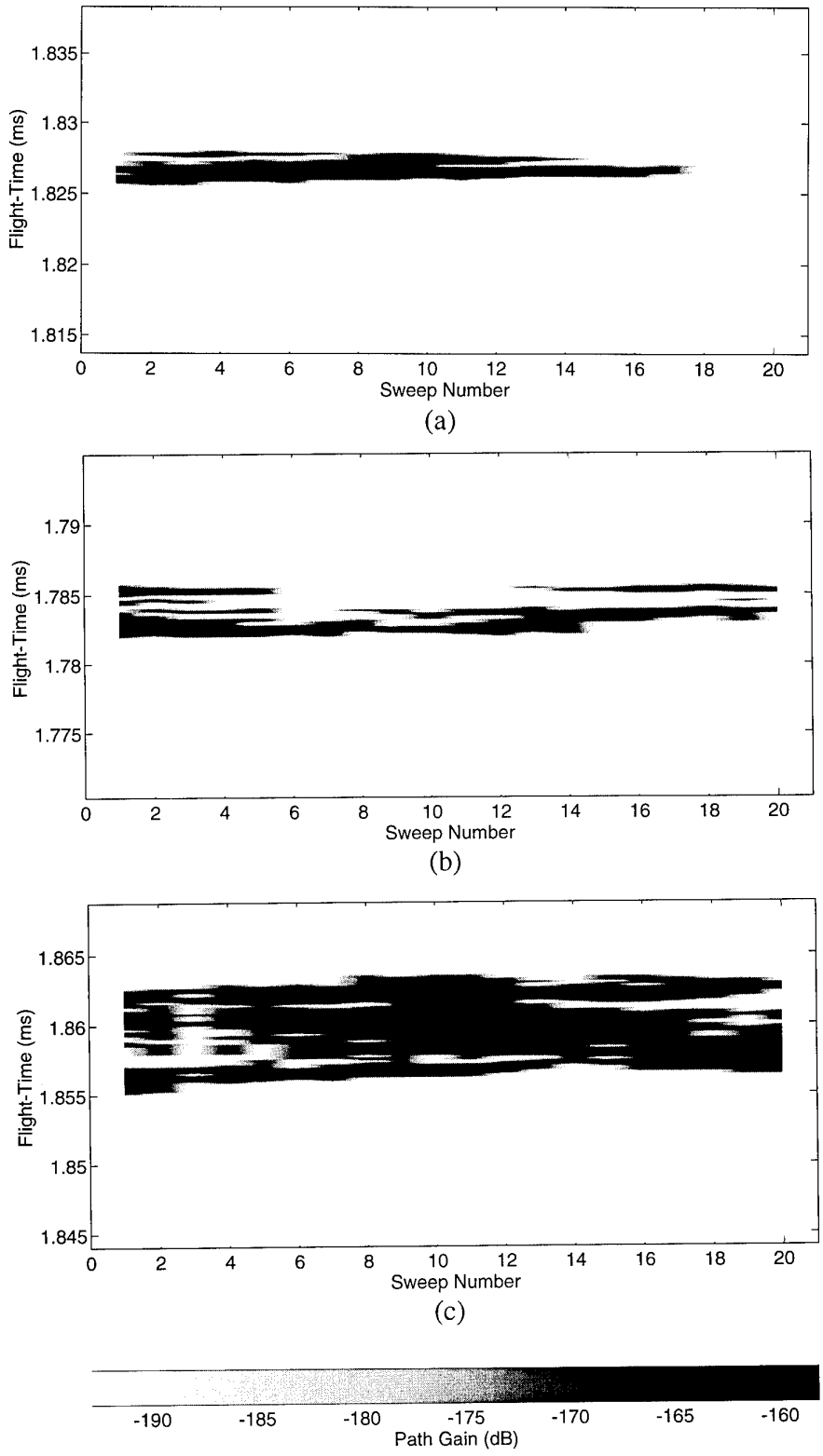


Figure 7 Impulse response profiles showing clear evidence of multipath. Profiles of this type occurred relatively infrequently.

moment of such a power delay profile and is computed according to the equation (Eriksson [4], Cox [7]):

$$S = \left(\frac{\sum_{k=1}^{N} (\tau_k - D)^2 P(\tau_k)}{\sum_{k=1}^{N} P(\tau_k)} \right)^{1/2} \quad (7)$$

where $P(\tau_k)$ is the power level at the k^{th} delay sampling time and D is the first moment of the profile, or average delay time (s), and is given by:

$$D = \frac{\sum_{k=1}^{N} \tau_k P(\tau_k)}{\sum_{k=1}^{N} P(\tau_k)} \quad . \quad (8)$$

Applying equations (7) and (8) to the data set collected in the 1993 propagation experiment, the rms delay spread distribution shown in Figure 8 was obtained. According to these statistics, in 90% of these measurements the rms delay spread was found to be less than 100 ns, while in 99% of all measurements the rms delay spread was found to be less than 400 ns. In several rare instances rms delay spreads ranging from 1.0 to 7.0 μs have been observed. Statistics indicating similar findings have been reported by Eriksson [4]. In light of this evidence, it would appear that trail conditions causing large delay spreads are relatively infrequent.

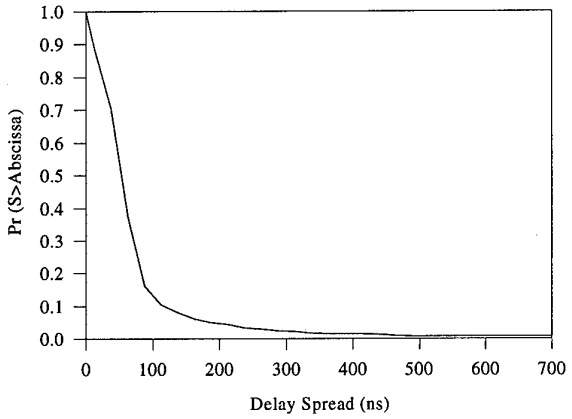


Figure 8 Delay spread statistics

5. CONCLUSIONS

The impulse response measurements analyzed to date clearly indicate the presence of multipath propagation. Although the data studied thus far suggest that the channel impulse response is for the most part well-behaved in terms of its multipath characteristics, there are occasions when severe multipath propagation is encountered. As these severe multipath events are likely to impose signaling rate limitations, additional research is currently underway to confirm the frequency of these events, characterize their nature and evaluate their impact on meteor burst communication systems.

6. ACKNOWLEDGMENTS

The authors would like to thank Mr. T. Ohno and Dr. N. Serinken of the Communications Research Centre for their assistance in the various phases of this project.

7. REFERENCES

[1] F. Akram, N.M. Sheikh, A. Javed, M.D. Grossi, "Impulse response of a meteor burst communications channel determined by ray-tracing techniques", IEEE Trans. on Commun., pp. 467-470, Apr. 1977.

[2] J.A. Weitzen, M.D. Grossi, W.P. Birkemeier, "High-resolution multipath measurements of the meteor scatter channel", Radio Science, vol. 19, No. 1, pp. 375-381, Jan.-Feb. 1984.

[3] J.A. Weitzen, M.J. Sowa, R.A. Scofidio, J. Quinn, "Characterizing the multipath and Doppler spreads of the high-latitude meteor burst communication channel", IEEE Trans. on Commun., vol. COM-35, No. 10, pp. 1050-1058, Oct. 1987.

[4] G.Eriksson, "An investigation of the meteor channel for adaptive communication systems", IEEE MILCOM 90, pp. 873-877, Vol. 2, 1990.

[5] F.J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform", IEEE Proceedings, Vol. 66, No. 1, pp. 51-83, Jan. 1978.

[6] A.H. Nuttal, "Generation of Dolph-Chebyshev weights via a fast Fourier transform", IEEE Proceedings, Vol. 62, p. 1396, 1974.

[7] D.C. Cox, "Delay Doppler characteristics of multipath propagation at 910 MHz in a suburban mobile radio environment", IEEE Transactions on Antennas and Propagation, Vol. AP-20, No. 5, pp. 625-635, Sept. 1972.

DISCUSSION

Discussor's name: D. Yavuz

Comment/Question:

As you know, there are many factors that enter into MB trail characterisation (diurnal, seasonal, solar cycle, distance, location .....). As mentioned in my talk yesterday, link-level protocols are, in my view, the best manner to obtain robust and efficient links. Can you comment on the possible validity of the 70% to 90% figures for 1 & 2 paths under other conditions.

Author/Presenter's reply:

As you pointed out there is a great deal of variability associated with meteor burst communications. The reliability of the figures presented is currently unknown and for this reason we have repeated this experiment on the same 500 km path and will conduct additional measurements for both long range (~1300 km) and short range (<200 km) paths. A comparison of these experimental results should provide some indication as to the reliability of the original measurements.

DISCUSSION

Discussor's name: F. Davarian

Comment/Question:

You have shown the delay spread of the channel to be in the order of about 200 ns. This implies a channel bandwidth of about 5 MHz indicating a maximum throughput of about 1 Mb/s. Are these assumptions correct? Please comment on the bandwidth size of the channel.

Author/Presenter's reply:

For delay spreads on the order of 100-400 ns one would expect to be able to signal at a data rate that is inversely proportional to the delay spread. In reality such an estimate should be scaled by a constant (usually less than unity) as the exact relationship between the maximum transmission rate and the delay spread is unknown. A transmission rate of 1.0 Mb/s as you have suggested is reasonable, however it should be remembered that the meteor burst channel is very lossy and consequently a large transmitter power would be required to achieve a suitable signal-to-noise ratio at the receiver. Realistic signalling rates are more likely to range from 10's of Kb/s to perhaps an upper maximum of 500 Kb/s depending on the available link budget and system requirements.

DISCUSSION

Discussor's name: P. Cannon

Question/Comment:

I think you should be congratulated on a very nice study.

Have you looked at modes other than meteor scatter with your instrumentation (i.e. ion scatter or sporadic-E)?

Author/Presenter's reply:

Thank you ... no, not individually.

Discussor's name: S. Karp

Question/Comment:

Was there any doppler spread?

Author/Presenter's reply:

We did not attempt to measure any above 20 Hz.

DISCUSSION

Discussor's name: L. Bertel

Comment/Question:

La présence du champ magnétique terrestre rend anisotrope le plasme associé aux trainées météoriques.

1. Ne pensez-vous pas que les doubles trajets que vous obtenez très frequemment sont en partie dus à cet effet d'anisotropie observé de la même façon en propagation ionosphérique (modes dénommés O et X)?

2. L'utilisation d'antennes adaptées à la polarisation associée à cet effet d'anisotropie (polarisation elliptique en général) ne vous permettrait-elle pas de sélectionner l'un des deux modes de propagation?

(Translation:

The presence of the earth's magnetic field renders the plasma associated with the meteor trails anisotropic.

1. Do you not think that the double paths that you frequently obtain are partly due to this anisotropic effect observed in the same way in ionospheric propagation (modes known as O and X)?

2. *Would not the use of antennas adapted to the polarisation associated with this anisotropic effect (general elliptic polarisation) allow you to select one of the two propagation modes?*

Author/Presenter's reply:

It is possible that these double paths are in fact created much in the same manner as observed in ionospheric propagation as you have suggested. As we have not yet begun to identify the physical processes responsible for the features observed in our impulse response profiles, it would be premature to attribute them to any particular mechanism.

While the use of adaptive antennas may permit you to select one of the propagation modes (O or X), the implementation of such a scheme may be be advantageous since it would appear that for meteor burst channels the signalling rate limitations due to multipaths are secondary to the problem of achieving an adequate signal-to-noise ratio.

Description expérimentale et modélisations du canal météorique

O. RAVARD *, D. SORAIS **
* Laboratoire Radiocommunications
URA CNRS 834, Campus de Beaulieu
Université de Rennes 1
35042 Rennes Cedex, France
** THOMSON-CSF/RGS
66, rue du Fossé Blanc
92231 Gennevilliers Cedex, France

Abstract:

The meteoric channel can be described, over a given period of time, by the duty cycle and the mean waiting time. The objective of this paper is to give an analysis of the channels statistical parameters and to describe different kinds of channel modelling.

The statistical analysis is based on experimental data. We show the effects of fading, observed on the received signal, on the channel parameters.

We present different kinds of meteoric channel modelling which take account for underdense sporadic meteor trails.

These modelling allow the consideration of the most important problems:

- *the astronomical problem to evaluate the meteor distribution on the sky,*
- *the ionisation problem to determine the electronic density created by the meteor,*
- *the propagation problems.*

The performances of these modelling are evaluated.

1 INTRODUCTION

On décrit dans ce papier notre expérience de l'étude du canal météorique sous ses aspects théorique et expérimental. Les météores pénétrant dans l'atmosphère terrestre créent des trainées ionisées qui diffusent suffisamment ($f < 100$ MHz) les ondes radioélectriques pour établir des liaisons intermittentes. Afin d'estimer les potentialités de ce canal nous avons réalisé des expérimentations sur le territoire français et développé des modélisations dans le but de prolonger les résultats obtenus à des configurations de paramètres quelconques (antennes, distances, ...). Les champs de validité de ces différents modèles seront précisés par comparaison avec nos résultats expérimentaux ainsi qu'avec les données de la littérature.

2 PARAMÈTRES CARACTÉRISANT LE CANAL MÉTÉORIQUE

Les météores incidents sur la terre peuvent être séparés en deux classes : les essaims de météores et les météores sporadiques. Les essaims sont formés d'un ensemble de particules se déplaçant sur des orbites précises autour du soleil. Lorsque la terre rencontre ces orbites pour des périodes données de l'année (mois d'août pour les Perséïdes par exemple) on obtient un flux de météores qui semblent provenir d'un même point de la sphère terrestre (radiant). Pour un essaim donné l'intensité du phénomène se reproduira identiquement chaque année si les particules de cet essaim sont uniformément distribuées sur la totalité de son orbite. Les essaims représentent une faible partie du flux total de météores incidents sur la terre et ne sont observables que durant certaines périodes de l'année.

Les météores sporadiques ont des orbites dont les paramètres sont distribués aléatoirement et sont observables quelle que soit la période de l'année. Les performances des liaisons radioélectriques sont donc essentiellement déterminées par cette classe de météores à laquelle nous ferons implicitement référence dans la suite de ce papier.

Les traînées ionisées par les météores produisent des ouvertures de quelques centaines de millisecondes séparées par des fermetures dont les durées sont liées au nombre de météores incidents sur la terre. Pour une position géographique donnée de la liaison, ce flux de météores est fonction de la position de la terre sur son orbite autour du soleil (mois de l'année) et de la rotation de la terre (heure de la journée). Ces mouvements de la terre forment la composante déterministe des variations temporelles annuelle et diurne de l'activité météorique.

La variation de cette composante déterministe étant négligeable pour une heure d'observation, les durées moyennes d'ouverture et de fermeture peuvent être estimées pour cet intervalle de temps. Pour une période

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

d'observation donnée, la durée de service d_s représente le pourcentage du temps durant lequel le canal est ouvert :

$$d_s = \frac{\bar{\tau}}{\bar{\tau} + \bar{\delta}} \qquad (1)$$

La composante aléatoire de l'activité météorique est représentée par les fluctuations de jour en jour de ces trois paramètres, pour une heure h fixée de la journée. Les résultats expérimentaux de la figure 1 ont ainsi été obtenus en réalisant, pour chaque heure, une moyenne sur une dizaine de jours (voir ¶ 6). Ce filtrage permet de faire apparaître la signature diurne du canal météorique caractérisée par un maximum vers 6 heures et un minimum vers 18 heures. La valeur moyenne de cette signature diurne sur les 24 heures de la journée (4.1 % sur la figure 1) sera donc la durée de service moyenne liée aux dix jours de l'expérimentation. Les variations déterministes annuelles seraient représentées par l'évolution de cette denière moyenne de dix jours en dix jours. On peut évidemment réaliser des traitements semblables pour les durées d'ouverture et de fermeture.

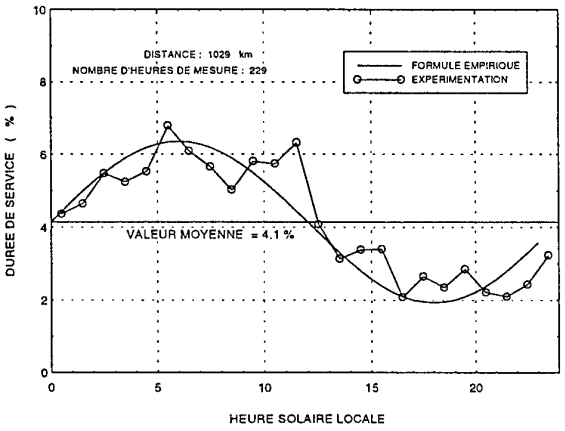


FIG. 1 - *Variation diurne de la durée de service (liaison de 1029 km).*

Classiquement, les caractéristiques optimales des antennes pour réaliser une liaison donnée sont liées à des directions particulières pour lesquelles il est nécessaire de fournir le maximum de puissance possible. Pour le canal météorique, l'ensemble du volume commun étant a priori utile, l'efficacité d'une antenne dépend des gains de l'antenne dans les directions balayant le volume V_c, pesés par l'activité du canal dans ces directions. Le choix des antennes adaptées à un service donné (couverture radioélectrique) est donc particulièrement délicat pour ce canal.

Pour chaque ouverture du canal, la variation temporelle du signal reçu complète la connaisssance du canal en permettant d'évaluer les durées des ouvertures ainsi que leur distribution. On peut également exploiter ces réponses temporelles pour identifier les modes de propagation non météoriques présents sur les liaisons réalisées (voir ¶ 7).

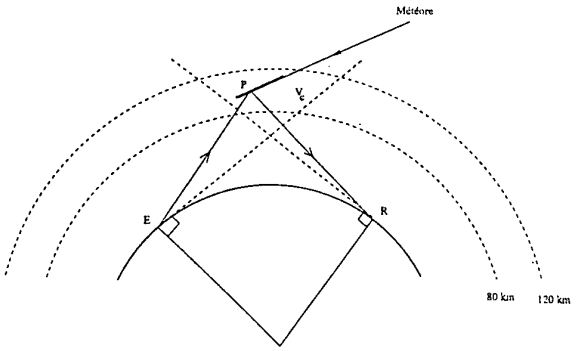


FIG. 2 - *Représentation du volume commun.*

3 INTRODUCTION À LA MODÉLISATION DU CANAL MÉTÉORIQUE

Les traînées ionisées a priori observables se trouvent dans le volume commun V_c compris entre 80 et 120 km d'altitude, simutanément vu de l'émetteur E et du récepteur R (figure 2).

Pour un point P du volume commun, on souhaite estimer la densité électronique de la trainée ionisée induite par un météore de masse et de vecteur vitesse fixés. Les météores radioélectriquement détectables doivent avoir des trajectoires tangentes au point P à l'ellipsoïde de révolution de foyers E et R. Par ailleurs, la zone fortement ionisée de la traînée doit recouvrir au moins la moitié de la première zone de Fresnel centrée sur le point P. Les météores passant par le point P seront radioélectriquement exploitables si la densité électronique créée est supérieure à un seuil lié à la configuration de la liaison (puissance, distance, E/N_o minimum, ...).

Il est donc nécessaire d'analyser différents problèmes de physique pour réaliser une modélisation du canal météorique :

- L'astronomie : Les astronomes ont réalisé de nombreuses observations optiques et radioélectriques (RADAR) des météores incidents sur la terre. On peut obtenir à partir de ces données expérimentales les distributions statistiques des masses et des vecteurs vitesse de ces météores. Ces éléments sont exploités dans les modèles pour évaluer les variations temporelles des paramètres du canal.

- La physique de l'atmosphère : Les météores se vaporisent dans la haute atmosphère en formant une traînée ionisée qui possède une densité électronique dont on souhaite estimer la valeur maximum et l'altitude de formation.

- La propagation radioélectrique : Pour chaque point P du volume commun, on souhaite déterminer les bilans de liaison associés aux météores potentiellement détectables.

Bien que relativement complexe à mettre en œuvre, la modélisation du flux de météores incidents sur la terre peut être correctement réalisée et décrit relativement bien les

phénomènes observés. Il est par contre difficile de modéliser la totalité des phénomènes qui contribuent à déterminer la densité électronique induite par les météores. L'évaluation du nombre moyen de météores détectés par unité de temps est cependant en bon accord avec les mesures.

Lorsque les électrons de la traînée ionisée diffusent l'onde incidente d'une manière indépendante les uns des autres (traînées sous-denses), la surface efficace de diffusion peut être évaluée analytiquement. L'examen des réponses temporelles mesurées permet effectivement d'identifier des formes prévues par le modèle théorique. Un grand nombre de traînées sous-denses sont cependant affectées d'évanouissements, notamment dus aux distorsions de la traînée par les vents présents dans la haute atmosphère. Pour les densités électroniques élevées, tous les modèles développés utilisent l'approximation du cylindre métallique qui semble correcte pour la rétrodiffusion et les courtes distances. *Jones* [8] a montré que pour toutes les situations pratiques de prodiffusion, on obtient des écarts importants (par rapport à une résolution WKB) sur l'amplitude maximale de la réponse temporelle et sur sa durée. Compte tenu de ces éléments, il apparait que la durée de service et la distribution des ouvertures du canal seront des paramètres difficiles à modéliser sans introduire dans la simulation un certain nombre de données empiriques.

Nous avons décrit au début de ce paragraphe les contraintes géométriques que les traînées ionisées doivent satisfaire pour être potentiellement détectables. *Eshleman* et *Manning* [6] ont montré que ces contraintes imposent une distribution des traînées ionisées utiles non uniforme sur le volume commun de la liaison. Les "points chauds" de l'activité météorique sont situés de part et d'autre du grand cercle tandis que cette activité est quasi nulle au voisinage du point milieu de la liaison. Les auteurs ont supposé que le flux de météores incidents sur la terre est spatialement uniforme et que les densités électroniques maximales se produisent à une altitude fixe au dessus du sol (100 km par exemple). Pour chaque point de cette surface, l'évaluation de la densité de probabilité des météores détectables a permis de réaliser des calculs des durées de service sur différentes liaisons [11].

La modélisation développée par *Hines* et *Pugh* [12], [7] est liée à une géométrie plus analytique que celle précédemment exposée et permet d'introduire avec rigueur les différentes hypothèses physiques du problème. Pour chaque point de la surface d'intégration, ce modèle permet d'évaluer le nombre de météores effectivement détectés. Bien que dans leurs travaux Hines et Pugh aient choisi une distribution spatiale uniforme des météores, il est possible d'exploiter ces travaux en introduisant des distributions plus élaborées.

Cette potentialité a été utilisée par *Meeks* et *James* [10] qui ont admis que les météores sporadiques proviennent essentiellement du plan de l'écliptique. Contrairement aux précédents travaux, la simulation ainsi construite rendait compte des variations diurne et annuelle de l'activité météorique observée.

Les modélisations actuelles utilisent des distributions de météores nettement plus précises que celles de Meeks et James et réalisent une intégration dans le volume où se vaporisent les météores [15], [2], [4]. Ces modélisations fournissent des prévisions précises des variations temporelles du nombre de météores détectés.

4 MODÈLES « INTÉGRATION EN SURFACE »

Le modèle décrit est basé sur les travaux de *Pugh* [12] qui détermina la distribution des météores observables dans le ciel. *Bain* [1] utilisa la méthode de Pugh pour déterminer la distribution azimutale des météores observables, et en déduisit une distribution héliocentrique de ceux-ci. La méthode de Pugh, que nous dénommons "intégration en surface", a été améliorée par *Sorais* [13] pour tenir compte notamment des pertes de polarisation. Nous avons repris cette méthode, pour évaluer les performances d'un système, en y introduisant la distribution des radiants des météores donnée par Bain.

De la relation théorique donnant la puissance reçue associée à une trainée météorique sous dense, on déduit le nombre de météores observables passant à travers une surface élémentaire dS de normale $\vec{n}$ et provenant d'un angle solide $d\Omega$ autour d'une direction $\vec{u}_m$.

$$N_{det}(P) = K dS \vec{n}\vec{u}_m \cos\xi D(\vec{u}_m)\sqrt{\frac{W(P,\beta)}{P_{min}}}d\Omega \qquad (2)$$

Ce nombre dépend de la distribution des directions d'arrivées des météores $D(\vec{u}_m)$ dans la direction $\vec{u}_m$, de la puissance minimale détectable P_{min}, de l'angle β que fait la trainée avec la trace du plan de propagation sur le plan tangent et du point P du volume commun considéré.

On travaille dans le repère de Pugh (V, x, y, z) défini comme suit : V est le milieu du segment émetteur-récepteur porté par l'axe V_x. L'axe V_z est porté par la verticale en V et l'axe V_y est orthogonal à V_x et V_z. On considère un élément de surface $dxdy$ du plan (V_x, V_y).

Pugh [12] détermina alors l'élément de surface dS, qui s'étend sur toute l'épaisseur du volume commun, en fonction de dx et l'angle solide $d\Omega$ en fonction de $dyd\beta$. Le choix d'une distribution des directions d'arrivées des météores $D(\vec{u}_m)$ permet d'obtenir le nombre de trainées ionisées observables par unité de temps attribuées à l'élément de surface $dxdy$ du plan de Pugh.

$$N_{det}(P) = g(P)\int_{\beta_{min}}^{\beta_{max}} \xi(P,\beta)D(\vec{u}_m)d\beta dxdy \qquad (3)$$

Les angles limites β_{min} et β_{max} sont déterminés de telle sorte que les météores soient vus dans des directions formant un angle zénithal inférieur à $\frac{\pi}{2}$. Le nombre total de météores observables est alors obtenu en intégrant la contribution de tous les éléments $dxdy$ du plan de Pugh. Pugh détermina une expression analytique de la relation 3 en supposant une distribution isotrope des directions d'arrivée des météores dans le ciel ($D(\vec{u}_m) = \frac{1}{4\pi}$). Nous avons choisi d'utiliser la distribution donnée par *Bain* [1], afin de pouvoir retrouver les variations diurne et annuelle

de la durée de service. La relation 3 est, par conséquent, évaluée numériquement en utilisant cette distribution.

5 MODÈLES « INTÉGRATION EN VOLUME »

Contrairement aux modèles "intégration en surface" (décrits dans le paragraphe précédent), les modèles "intégration en volume" peuvent être utilisés pour prendre en compte la formation des trainées ionisées et la distribution spatiale des météores détectés.

Le modèle que nous décrivons est basé sur le même principe que celui élaboré par *Desourdis et al.* [5]. Pour chaque volume élémentaire du volume commun de la liaison, on détermine le taux de météore $P_{m,\theta,\phi,V_H}(m,\theta,\phi,V_H)dmd\theta d\phi dV_H$ dont la vitesse héliocentrique est V_H à dV_H près, dont la masse est m à dm près, dont le vecteur vitesse se trouve dans l'angle solide $d\sigma$ autour des deux coordonnées sphériques (θ,ϕ). La distribution des masses des météores incidents est supposée être en m^{-2}, et la densité de probabilité P_{θ,ϕ,V_H} est déduite de distributions expérimentales (énergie orbitale et inclinaison) données par *Delcourt* [3].

Les météores potentiellement observables, définis par les conditions de tangence classiques, ont des vecteurs vitesse définis par (θ,ϕ,V_H) qui forment un cône de demi angle au sommet ϕ_0. Le nombre total de météores potentiellement observables est donc donné par :

$$N_{pot}(P) = \iiint P_{\theta,\phi,V_H}(\theta,\phi_0,V_H)m^{-2} \times \sin\phi_0 dm dV_H d\theta \quad (4)$$

Il s'agit maintenant de déterminer le nombre de météores observables pour une configuration donnée. Pour cela, il suffit de connaître les météores potentiellement observables qui produiront une densité linéique d'ionisation q supérieure à une valeur minimum q_{min}. On introduit alors une modélisation de l'ionisation reliant la masse et le vecteur vitesse du météore à la densité linéique d'ionisation produite, ainsi que sa variation avec l'altitude.

La modélisation classique, décrite par *McKinley* [9], n'est pas suffisamment précise et donne souvent des altitudes d'ionisations trop élevées et des longueurs de trainée ionisée trop importantes. On propose donc d'utiliser le modèle d'ionisation proposé par *Delcourt* [3] qui consiste à résoudre un système d'équations différentielles. Ce modèle rend bien compte des observations, qu'elles soient visuelles ou radioélectriques, en traitant les problèmes suivants :

- l'effet d'écran produit par l'interaction entre le corps céleste et les molécules d'air avant la vaporisation intense,
- l'échauffement du corps et le rayonnement thermique de celui-ci, pendant la vaporisation intense.

Lorsque la densité linéique q, au point de tangence, est déterminée, le problème radioélectrique intervient alors et on considère les deux types de réponses temporelles associées aux traînées météoriques sous dense et sur dense.

Les durées d'ouvertures associées à ces deux types sont déduites de la théorie classique et respectivement données par :

$$t_{0_{SD}}(q) = \tau \ln\left[\frac{q}{q_{min}}\right] \quad \text{et} \quad \tau = \frac{\lambda^2}{16\pi^2 D\cos^2\Phi} \quad (5)$$

et l'approximation :

$$t_{0_{SSD}}(q) = \frac{1}{4D}\frac{r_e\lambda^2 q}{\pi^2\cos^2\Phi}\left[\frac{1}{e^{\frac{q_{min}}{eq}}} - \frac{1}{2e - e^{\frac{q_{min}}{q}}}\right] \quad (6)$$

La durée d'ouverture du canal météorique associée au point P du volume commun est donc :

$$d_S(P) = \iiint t_o(q)m^{-2}P_{V_H,\theta,\phi}(V_H,\theta,\phi_0) \times \sin\phi_0 dV_H dm d\theta \quad (7)$$

où $t_0(q)$ correspond à la relation 5 ou 6 selon que la trainée observée est de type sous dense ou sur dense. La densité linéique de transition entre les deux types de trainées est choisie de telle sorte qu'il y ait continuité entre les puissances maxima fournies par les théories associées. Les durées de fermeture moyenne et la durée de service sont alors évaluées de la même manière que pour les modèles "intégration en surface". L'intégration de la relation 7 à travers le volume commun de la liaison permet de caractériser le canal.

6 DESCRIPTION DES EXPÉRIMENTATIONS RÉALISÉES

On rappelle ci-après les principales données d'une campagne de mesures réalisée en France sur la fréquence de 42 MHz (voir [14] pour une description plus complète). Les antennes d'émission et de réception sont des antennes Log-périodiques horizontales placées à une hauteur h_{ant} au-dessus du sol. L'expérimentation a été réalisée pour 5 liaisons de distances D distinctes, pour lesquelles le nombre total d'heures de mesure H_m est compris entre 50 et plus de 200 heures. Par ailleurs ces expérimentations se sont déroulées entre les mois de mai et septembre 1989 pour des dates médianes notées Da_m.

Les autres caractéristiques nécessaires à l'évaluation du bilan de liaison sont fournies dans le tableau 2.

Liaison	h_{ant}	H_m	Da_m
107 km	3	52	15-05
310 km	3	223	20-04
363 km	4	165	15-06
706 km	6.5	84	29-06
1029 km	11.5	229	26-09

TAB. 1 - *Configuration des liaisons*
h_{ant} : hauteur des antennes en mètres, H_m : Durée des mesures en heures,

Puissance rayonnée	1200 W
Antennes	Log-Périodiques
Gain	6.5 dBi
Facteur de bruit du récepteur	3.4 dB
Rythme binaire	16 kb/s
E/N_0 min nécessaire	6 dB

TAB. 2 - *Caractéristiques radioélecctriques des transmissions*

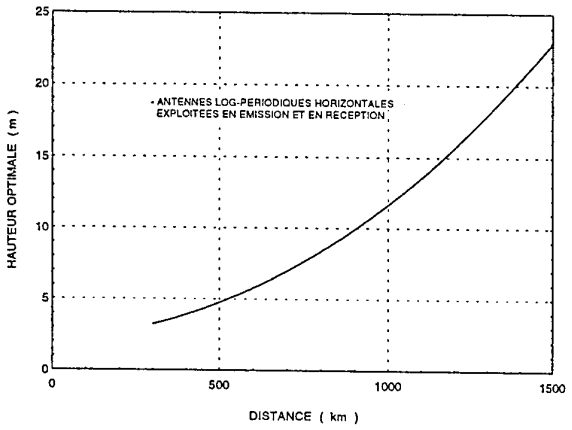


FIG. 3 - *Variation de la hauteur optimale de l'antenne placée au-dessus du sol.*

La forme du diagramme de l'antenne étant très sensible à sa hauteur au-dessus du sol, il existe, pour une fréquence et une distance données, une hauteur qui maximise la durée de service produite par les météores. Les hauteurs utilisées pour les expérimentations précédemment décrites sont très proches des hauteurs optimales obtenues par simulation (figure 3).

Pour les mêmes liaisons, d'autres mesures ont été réalisées en faisant varier les hauteurs des aériens et la fréquence (70 MHz).

7 FILTRAGE DES MÉCANISMES DE PROPAGATION PRÉSENTS

La propagation en vue directe ou par diffraction sur le sol peut facilement être identifiée par des durées d'ouverture très supérieures à celles du canal météorique. La région E-sporadique produit des réflexions notamment caractérisées par une amplitude du signal reçu très supérieure à celles des météores. Pour des fréquences comprises entre 40 et 100 MHz et des latitudes moyennes, l'occurrence de ce mécanisme de propagation est cependant rare. Il est également possible d'obtenir des échos dus à la diffusion ionosphérique pour des fréquences suffisamment basses (autour de 40 MHz). Ces deux derniers phénomènes sont plus probables pour des distances supérieures à 500 km que pour des courtes distances. Enfin, la diffusion troposphérique peut contribuer d'une manière importante à la durée de service totale pour des distances inférieures à 400 km.

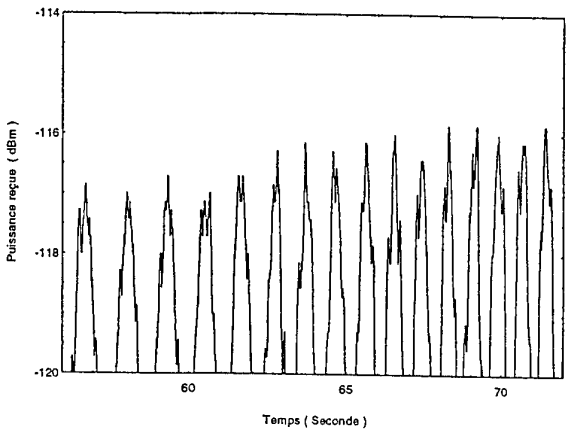


FIG. 4 - *Exemple de formation des ouvertures liées à la diffusion troposphérique.*

Si l'on connait les réponses temporelles des ouvertures élémentaires du canal il est possible de classer automatiquement ces ouvertures en réalisant un logiciel qui identifie des signatures radioélectriques aux différents mécanismes de propagation. Nous avons utilisé cette technique pour une partie des résultats de nos expérimentations.

Parmi tous ces mécanismes de propagation, le plus intense est la diffusion troposphérique qui masque fréquemment l'activité du canal météorique aux courtes distances. La présence de cette composante parasite (pour l'expérimentation) peut être détectée en analysant les distributions statistiques des fermetures. Les durées moyennes des ouvertures dues à la diffusion troposphérique sont semblables à celles des météores (quelques centaines de ms). Par contre les instants d'arrivée des météores produisant les ouvertures suivent une distribution Poissonnienne contrairement à ceux associés à la diffusion troposphérique qui ne sont pas indépendants. Ces derniers résultent en effet du découpage d'un évanouissement dû à quelques diffuseurs (voir figure 4).

7.1 Filtrage des évènements Poissonniens

Les durées moyennes de fermeture du canal météorique étant grandes devant les durées moyennes d'ouverture, une distribution Poissonnienne des instants d'arrivée des météores déterminera une distribution exponentielle des durées de fermeture. L'analyse graphique des résultats permet d'obtenir la durée de fermeture moyenne "Poissonnienne" $\bar{\delta}_p$ ainsi que le pourcentage d 'événements x_p effectivement Poissonniens.

Pour le canal météorique la hauteur optimale de l'antenne au-dessus du sol est inférieure à 4 m pour les courtes distances (< 400 km). Lorsque l'on élève progressivement l'antenne au-dessus de cette hauteur, on constate que la durée de service s'accroît tandis que le nombre d'événements Poissonniens diminue. Ce résultat est dû à la présence de la diffusion troposphérique qui est favorisée par l'élévation des aériens, dans la mesure où cette élévation permet d'accroître les gains des antennes pour les angles d'élévation faible (< 5 degrés).

Lors de l'expérimentation No 2 (310 km) on a mesuré une durée de service de 1 % pour une antenne Log-périodique placée à 3 m au-dessus du sol. Pour la même liaison et la même période de l'année (début avril) cette durée de service devenait égale à 2 % lorsque la hauteur de l'antenne était élevée à 4 m de hauteur. Bien que durant la première partie du mois d'avril il n'y ait pas d'essaims météoriques présents, cette variation pourrait être due à des variations de l'activité météorique. La méthode d'analyse précédemment décrite permet de décider que cet accroissement est essentiellement dû à la diffusion troposphérique puisque le nombre d'événements Poissonniens diminue de 65 à 30 % (h_{ant} : 3 m $\rightarrow$ 4 m).

Cette diffusion troposphérique peut donc être détectée aux courtes distances par une mesure de quelques heures.

On peut cependant noter que nos résultats expérimentaux fournissent un pourcentage d'événements Poissonniens toujours inférieur à 70 % alors qu'en présence de la seule composante météorique on pourrait s'attendre à obtenir un pourcentage supérieur à 90 %. La réponse à ce problème peut être obtenue en examinant un résultat du filtrage des réponses temporelles réalisé sur la liaison No 5 (1029 km) pour une durée d'observation de 2 heures (figure 5). La fonction de répartition complémentaire des durées de fermeture pour les données brutes permet d'estimer le pourcentage d'événements Poissonniens à 50 % environ. L'examen des réponses temporelles mesurées fait apparaitre un évanouissement produisant des coupures sur les durées d'ouverture du canal. Après suppression de ces coupures et d'autres modes de propagation reconnus par le logiciel de filtrage on obtient une seconde droite associée à un pourcentage d'événements poissonniens de l'ordre de 90 %. Pour cet exemple, on n'obtient pas 100 % d'événements Poissonniens à cause, de la mauvaise ou la non-identification d'un certain nombre d'ouvertures.

Une partie notable des événements non Poissonniens est donc due aux coupures des ouvertures du canal météorique. Ce phénomène réduit simultanément les durées moyennes d'ouverture et de fermeture mesurées (données brutes). La durée de fermeture "Poissonnienne" $\bar{\delta}_p$ qui est égale aux pentes des droites de la figure 5 est donc la durée de fermeture moyenne obtenue en ne prenant pas en compte les coupures des ouvertures.

Le pourcentage d'évènements Poissonniens x_p a été estimé pour les liaisons dont les caractéristiques sont décrites dans les tableaux 1 et 2. Sachant que la durée des ouvertures sous-denses croît avec la distance, il semble logique d'observer une décroissance du paramètre x_p avec cette distance puisque la probabilité d'évanouissement sur une trainée ionisée sera d'autant plus forte que sa durée est importante (figure 6).

Compte tenu des éléments précédents il semble raisonnable de supposer que pour les faibles hauteurs d'antenne utilisées aux courtes distances, la saturation du paramètre x_p est principalement produite par les évanouissements des ouvertures du canal météorique (hauteurs optimales d'antenne). Ce résultat peut être confirmé en examinant la signature diurne du canal météorique : pour la liaison No 2 (310 km) cette signature est nettement déformée pour une hauteur de 4 m alors que pour une hauteur de 3 m elle est correctement formée.

Lorsque les antennes sont placées à leur hauteur optimale pour le canal météorique (configurations du tableau 1) les durées de service mesurées peuvent donc être principalement attribuées a celui-ci.

7.2 Filtrage des réponses temporelles

Nous avons également utilisé les enregistrements des réponses temporelles pour déterminer les proportions que l'on peut attribuer à chaque mécanisme de propagation.

Le filtrage des réponses temporelles se décompose en deux étapes. La première étape consiste à classer chaque ouverture à partir d'un certain nombre de tests effectués sur cette ouverture. La deuxième étape modifie le classement effectué à l'issue de la première étape en fonction du classement des ouvertures précédentes et suivantes. Les tests effectués sur chaque ouverture sont de

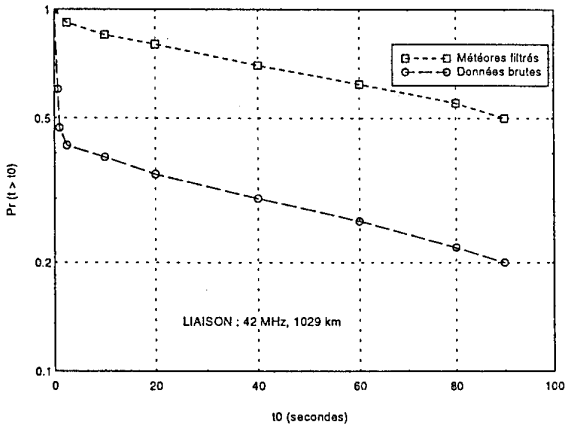


FIG. 5 - *Filtrage des réponses temporelles du canal. Exemple de droite Poissonniene avant et après filtrage.*

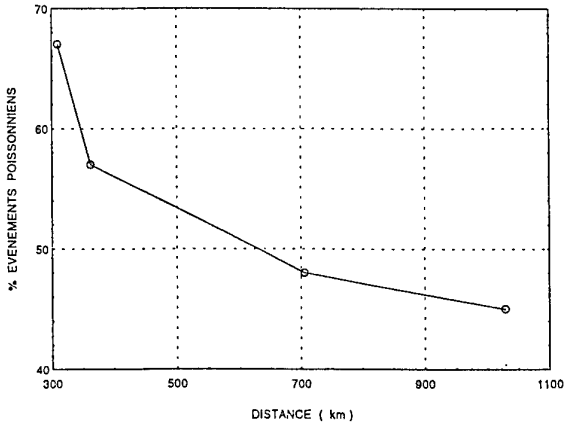


FIG. 6 - *Variations du nombre d'événements Poissonniens avec la distance.*

différents types :

- les tests de type « géométriques » où l'on tente de reconnaître la réponse temporelle associée à l'ouverture commme une forme géométrique simple. Par exemple, la réponse temporelle associée à une traînée météorique sous-dense sera reconnue géométriquement comme un triangle grâce à la décroissance exponentielle propre à ces traînées.
- les tests associés à la densité spectrale de la réponse temporelle (utilisation de la FFT, nombre de passage par la valeur médiane)
- les tests liés à la densité de probabilité de la réponse temporelle.

A l'issu des deux étapes, les ouvertures du canal sont classées en plusieurs catégories :

- la classe « sous-dense » regroupant les ouvertures associées aux trainées sous-denses,
- la classe « sur-dense » regroupant les ouvertures associées aux trainées sur-denses,
- la classe « fading » associée aux ouvertures du canal produites par des météores dont l'origine (sous dense ou sur dense) est indéterminée. La réponse temporelle est affectée d'évanouissements,
- la classe « réflexion » associée à des ouvertures dont les réponses temporelles évanescentes et dont les origines peuvent être multiples (diffusion troposphérique, trainées météoriques non spéculaires),
- la classe « diffusion » dont les ouvertures correspondent à des réponses temporelles suivant une loi de Rayleigh.

La figure 7 donne les résultats de classement pour la liaison No 2. On remarque, pour cette configuration de liaison, la contribution majoritaire du canal météorique (classes « sous-denses », « sur-denses » et « fading »).

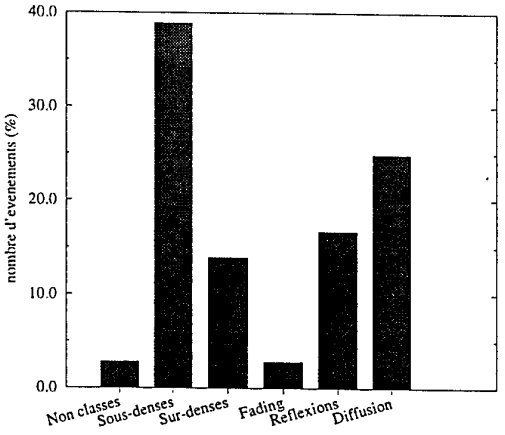


FIG. 7 - *Exemple de filtrage des modes de propagation. Liaison No 2 du tableau 2.*

8 ANALYSE GLOBALE DES PERFORMANCES DU CANAL

8.1 Durées de fermetures du canal

La durée moyenne de fermeture étant inversement proportionnelle au nombre moyen de météores observés durant une seconde ($\bar{\delta} << \bar{\tau}$), on peut espérer prévoir correctement ce paramètre (voir ¶ 3). Les fermetures "Poissonniennes" $\bar{\delta}_p$ mesurées pour les cinq liaisons du tableau 1 peuvent être comparées aux estimations de la modélisation en volume (météores sous-denses) qui prend notamment en compte les variations diurne et annuelle de l'activité météorique (figure 8).

Les résultats théoriques et expérimentaux sont raisonnablement proches compte tenu des incertitudes présentes dans certains éléments de la simulation.

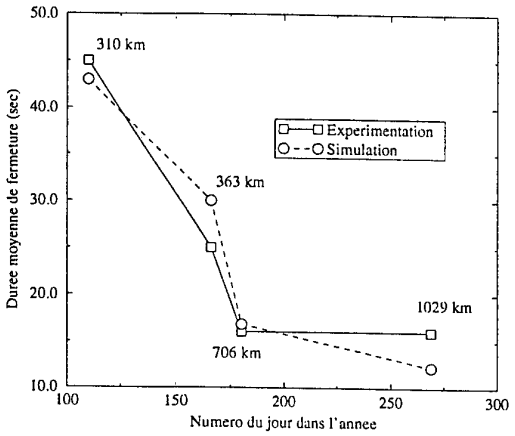


FIG. 8 - *Durées de fermetures moyennes pour les liaisons reportées dans le tableau 1. Les simulations ont été effectuées pour les dates médianes associées aux mesures.*

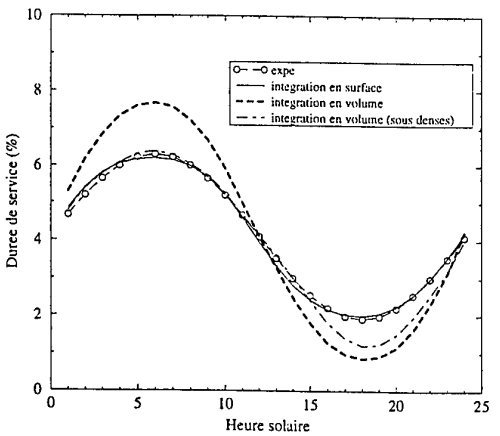


FIG. 9 - *Variation diurne de la durée de service du canal météorique pour la liaison No 5 du tableau 1 (1029 km).*

8.2 Paramètres liés aux durées d'ouverture du canal

Les performances du canal sont essentiellement déterminées par la durée de service qui dépend des durées d'ouverture du canal. Bien que ce dernier paramètre soit mal évalué par notre modèle, la variation diurne de la durée de service est correctement estimée par la modélisation en volume (figure 9). Ce résultat est notamment dû aux faibles variations de la durée d'ouverture dans une journée.

Les durées d'ouverture moyenne mesurées, fournies dans le tableau 3, montrent que ce paramètre évolue peu avec la distance. On peut également estimer le nombre moyen de coupures N_c par météore à partir du pourcentage d'évènements Poissonniens x_p:

$$N_c = \frac{100 - x_p}{x_p} \tag{8}$$

La faible variation de la durée d'ouverture avec la distance est en partie expliquée par l'accroissement du paramètre N_c. Pour les courtes distances, la durée d'ouverture est nettement supérieure à celle fournie par les modèles. La présence des météores sur-denses, modélisés par l'approximation du cylindre metallique, ne semble pas suffisante pour justifier ce résultat expérimental.

9 Conclusion

Le filtrage des mécanismes de propagation présents, pour des liaisons VHF transhorizons, a permis d'identifier les phénomènes physiques liés au canal météorique que les modélisations doivent décrire le plus finement possible.

Liaison	D (km)	$\bar{\tau}$ (ms)	N_c
2	310	253	0.49
3	363	248	0.75
4	706	298	1.08
5	1029	280	1.22

TAB. 3 - *Durées d'ouverture moyenne du canal et nombre de coupures moyen par météore.*

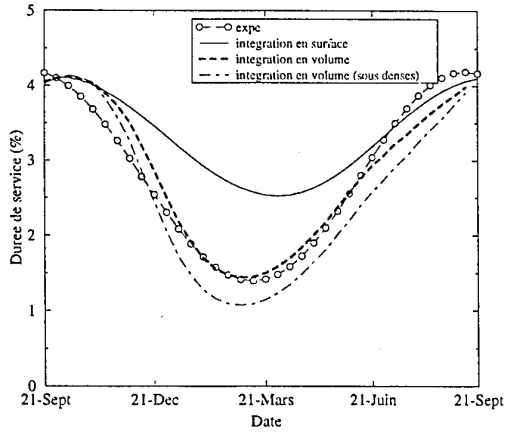


FIG. 10 - *Variation annuelle de la durée de service du canal météorique pour la liaison No 5 du tableau 1.*

La modélisation du canal météorique fait intervenir des phénomènes physiques complexes que nous n'avons pas tous complètement décrits. En particulier, la distribution des ouvertures du canal semble difficile à modéliser dans la mesure où elle résulte, d'une part, d'un phénomène déterministe (diffusion par la traînée) qui ne peut être simplement analysé que pour les traînées rectilignes strictement sous-denses, et d'autre part, des distorsions de la forme de la traînée qui produit des évanouissements.

Les travaux théoriques et expérimentaux que nous avons réalisés fournissent cependant un ensemble de données qui permettent de définir les principaux paramètres d'un système de transmissions. Le choix des antennes optimales peut être facilement déterminé en utilisant une modélisation avec intégration de surface tandis que les variations temporelles des paramètres du canal seront mieux évaluées avec une intégration de volume.

Références

[1] W. Bain. The azimuth distribution of oblique reflections from meteor trails and its relation to meteor radiant distributions. *J. Atm. and Terr. Phys.*, 17:188–204, 1960.

[2] D. Brown. A physical meteor burst propagation model and some significant results for communication systems design. *IEEE J. Select Areas Commun.*, SAC-3(5), 1986.

[3] J. Delcourt. Observations radioélectriques des petits corps météoriques. etude du mouvement héliocentrique d'un corps de faible masse soumis aux perturbations jovienne et de poynting-robertson. Thèse de Doctorat d'état, 1992.

[4] R. Desourdis. *Meteor Burst Communications: Theory and Practice.* Wiley. Chap. 3.

[5] R. Desourdis, J. Wojtaszek, S. Merrill, and K. Hernandez. Meteor burst link performance sensivity to antenna pattern, power margin and range. In *Military Comm. Conf.* IEEE, Oct. 1988.

[6] R. Eshleman and Manning. Radio communication by scattering from meteoric ionization. *Proc. of the IRE*, 42, Mar. 1954.

[7] O. Hines and R. Pugh. The spatial distribution of signal sources in meteoric forward scattering. *Can. J. of Phys.*, 34, Oct 1956.

[8] W. Jones. The application of the WKB approximation to the calculation of the scattering of radio waves from overdense meteor trails. *Planet. Space Science*, 40(11):1487–1497, 1992.

[9] D. McKinley. *Meteor Science and Engineering.* Mc Graw Hill, 1961.

[10] L. Meeks and C. James. On the influence of meteor radiant distribution on meteor scatter communication. *Proc. IRE*, 45, Dec. 1957.

[11] H. Nes. Dimensioning technique for meteor-burst communications systems. *IEE Proceedings*, 136(6):505–510, 1985.

[12] R. Pugh. The number density of meteor trails observable by the forward scattering of radio waves. *Can. J. of Phys.*, 34(10):997–1004, 1956.

[13] D. Sorais. Modélisation des performances du canal météorique. Technical report, Thomson CSF-RGS, 1991.

[14] D. Sorais, O. Ravard, and L. Bertel. Analyse des résultats expérimentaux de liaisons météoriques réalisées pour différentes distances et fréquences. In *CP-486*. AGARD, 1990.

[15] J. Weitzen. Predicting the arrival of meteors useful for meteor burst communication. *Radio Science*, 21(6), 1986.

DISCUSSION

Discussor's name: P. Cannon

Comment/Question:

As you know it is critical in meteor burst modelling to use a realistic model for the distribution of meteor radiants. Have you investigated the use of models other than that due to Pugh? Have you included shower meteors in your model?

Author/Presenter's reply:

Nous avons modifié le modèle de Pugh pour tenir compte des radiants des météores. Dans les modèles "intégration en volume", nous tenons compte également de la distribution des radiants. Nous utilisons les distributions expérimentales des orbites des météores (énergie orbitale $\frac{1}{a}$ et inclinaison i) déterminées par J. Delacourt).

Nous sommes intéressés uniquement aux météores sporadiques, mais les distributions associées aux essaims météoriques peuvent être introduites dans les modèles.

(Translation:

We have modified Pugh's model to take into account meteor showers. In the "integration in volume" models we also take into account the distribution of radiants. We use the experimental distribution of meteor orbits (orbital energy $\frac{1}{a}$ and inclination i) determined by J. Delacourt.

We are only interested in sporadic meteors, but the distributions associated with meteor showers can be introduced into the models).

Beam Forming Techniques in Meteor Scatter Communications Systems

Ammad Akram
Paul S Cannon

Radio Science and Propagation Group
Tactical Communications Dept
Defence Research Agency
Malvern, Worcs, WR14 3PS
United Kingdom

1.0 Summary

Experimental and modelling results from a programme to improve the data throughput of meteor scatter communications systems are presented.

The method of passive beam formation using a 4-element Butler matrix to improve the signal availability of meteor scatter communication systems is investigated. Butler matrix signal availability is compared with the performance of a single Yagi reference system using ~6.5 hours of data from a 720 km North-South temperate latitude link. The signal availability gain of the Butler matrix is found to range between 1.6-1.8 over the SNR threshold range 20-30 dB. The signal availability can be further improved by ~10%-20% in a system employing two 4-element Butler matrices with squinted beams so as to illuminate the sky with eight high gain beams. Space diversity is found to increase the signal availability of a single antenna system by ~10%-15% but the technique has very little advantage in a system already employing passive beam formation.

At low SNR, the Butler matrix also compares favourably with a system employing adaptive beam formation driven by direction finding. The overhead time needed to continually direction find with the gain of a single antenna mitigates against the adaptive scheme at low SNR. For high SNR signals, however, the adaptive system outperforms the Butler matrix.

A detailed computer model designed to predict the duty cycle of a general forward meteor scatter communications link has also been developed. The model incorporates effects such as major shower streams, a non-uniform radiant distribution and antenna polarisation coupling, including Faraday rotation. A particularly useful aspect of the model is its capacity to predict the passage of hotspot regions across the sky. Experimental results are presented to validate the accuracy of these predictions. A computer controlled single high gain beam meteor scatter system is eventually envisaged.

2.0 Introduction

A recurring barrier to the widespread application of meteor scatter communication (MSC) is its low data throughput, typically of the order of a few tens of bits/sec in simple simplex systems [e.g., *Cannon and Reed*, 1987]. The low throughput is due to the intermittent nature of meteor trails, the low received signal power and the decay of the signal-to-noise ratio (SNR) with time.

One strategy to overcome the low data throughput focuses on improved error coding. Fixed rate forward error correction (FEC) coding offers an improvement of ~25% in the data throughput over un-coded systems [*Miller and Milstein*, 1990]. Further slight improvements can be made by using more complex variable rate coding schemes [*Pursley and Sandberg*, 1989] but significant advances are constrained by the limited bandwidth of most MSC systems.

An alternative technique is to employ variable data transmission rates whereby the data rate profile mimics, in some fashion, the decay profile of the trail SNR. Such a scheme takes advantage of the relatively high SNR at the beginning of the trail [e.g., *Weitzen et al.*, 1984].

Shukla et al. [1992] and *Cannon et al.* [1993] investigated increasing the trail duration, and thereby the data throughput, by using antenna space diversity. This technique takes advantage of the fact that towards the end of a trail, and particularly in the case of long duration trails, the incoming wavefront is no longer planar due to interference between decorrelated signals. *Cannon et al.* [1993] recommended an antenna spacing of at least 20λ, preferably greater, to take advantage of the diversity gain of all types of trails, typically around 15% [*Shukla and Cannon*, 1993].

There are also a variety of high gain antenna approaches to increasing the data throughput [e.g., *Mawrey and Weitzen*, 1995] which can be sectioned into two categories, namely passive and adaptive beam forming techniques. In these schemes those regions of the sky geometrically favoured for meteor signal return (hotspots) are illuminated with high gain. Adaptive beam formation entails the synthesis of a high gain beam based on directional information about the hotspots or the trail itself whilst the method of passive beam formation synthesises a number of high gain beams looking at fixed areas of the sky.

The primary advantage of the adaptive scheme is its ability to synthesise a high gain beam in the direction of incoming individual trails. Unfortunately, only the single array element gain is available during the trail detection and beam synthesis process and consequently additional compensatory coding gain is required at the start of transmission. Systems using passive fixed direction beams are not burdened by the trail detection overhead and require no extra coding. *Kilpatrick et al.* [1991], in a theoretical study, concluded that the passive method of beam formation by a Butler matrix is preferred over the adaptive method. Ease of hardware implementation and applicability to either transmit or receive stations were further advantages cited in favour of passive synthesis of multiple beams. *Kilpatrick et al.* [1991] also pointed out a shortcoming of the Butler matrix due to the ~-4 dB beam crossover points. Minimisation of this shortfall is an issue of importance in this paper.

The Butler matrix was the preferred method of *Mawrey and Weitzen* [1995] to passively synthesise multiple beams for reception purposes. A 4-element hardware radio frequency (RF) Butler matrix was used for reception. An approximate doubling of the average hourly data throughput relative to a single antenna reference system was observed.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

In this paper, we move forward from the conventional RF implementation of the Butler matrix to a baseband DSP realisation. This advance provides for more flexibility, together with lower hardware and implementation costs. Our results confirm the advantages of fixed beam MSC systems observed by *Mawrey and Weitzen* [1995]. The parallel operation of a number of 4-element Butler matrices each producing a set of overlapping squinted beams is also investigated. Such an approach can provide, at the expense of a number of parallel DSPs, the full theoretical gain of a single 4-element Butler matrix regardless of the azimuthal orientation of incoming meteor trails. This investigative exercise highlights the greater flexibility of the digital implementation of the Butler Matrix over the RF approach.

Finally, the paper details a MSC computer prediction model that may be used to help steer a single high gain beam. Such a system offers the prospect of an adaptive system which is not hampered by the direction finding and beam synthesis overhead whilst retaining the high gain premiums. Experimental evidence is presented to qualitatively show that the model predictions are accurate enough to allow the realisation of a computer model driven adaptive MSC system.

3.0 Butler matrix

3.1 Experimental details

Data for this experiment were collected during November 1994 from a 720 km temperate latitude link between Wick, Scotland (58.56^{o}N, 3.28^{o}W) and Pershore, England (52.14^{o}N, 2.04^{o}W). Measurements were collected in 30 minute periods with start times from 1035 UT to 1510 UT. The Wick station transmitted 400 W continuous wave at 40 MHz using a 4-element horizontally polarised Yagi antenna positioned at a height of ~1λ above the ground. The reception system at Pershore consisted of five 4-element, horizontally polarised Yagi antennas mounted ~1λ above the ground along an East-West line pointing towards the transmitter at Wick. The first four antennas were equally spaced at 1.25λ (~10.2 m) apart to form a linear array with a fifth diversity antenna located ~30λ (~240 m) from the fourth antenna.

Fuller details about the hardware and the digital signal processing steps can be found in *Akram and Cannon* [1995]. It is sufficient to note here that five channels of in-phase and quadrature components were logged to a PC at ~300 samples/s in each data run. Thirteen data runs were conducted to provide a total of ~6.5 hours of data which were subsequently post processed.

3.2 Butler matrix method of passive beam forming

The Butler matrix is a passive method of synthesising N beams from N receivers where each beam has the full gain of the receiver array and N is an integer power of two. The azimuthal gain profile of a single 4-element Yagi antenna at ~40 MHz is such that the 3 dB points are approximately at $\phi=\pm30^{o}$. The directional property of the 4-element Yagi antenna thus dictates that the four synthesised matrix beams be constrained to lie approximately within the azimuthal sector $-30^{o}<\phi<30^{o}$. A 4-element Butler matrix receiving inputs from a linear array with inter-element spacing 1.25λ synthesises four beams at azimuthal angles $\pm6^{o}$ and $\pm17^{o}$. The equal inter-element spacing of 1.25λ was derived from an optimisation of the duty cycle using the computer model of *Akram and Cannon* [1994].

The Butler matrix provides an improved SNR figure by coherently combining the wanted signals whilst incoherently combining the noise. A 4-element Butler matrix, taking its inputs from identical antennas, can provide a maximum increase in SNR of 6 dB relative to a single input antenna when the noise sources are uniformly distributed in space and the incoming wave is perfectly planar. The latter condition is an important assumption behind the use of the Butler matrix. There is no advantage in using the Butler matrix if the input signals are not phase coherent, e.g., if they are the result of scatter from a multitude of dynamic sources. Space diversity is more appropriate in increasing the SNR in such situations.

Another limitation of the Butler matrix, even with planar wavefronts, is the drop in the SNR gain to ~2 dB at the beam crossover points. Thus the use of a Butler matrix in the reception system, is equivalent to increasing the transmitter power by a factor lying between 2 and 6 dB depending on the azimuthal arrival angles of the meteor trails.

3.3 Single antenna and Butler matrix results

One method of comparing the relative merits of MSC systems is to examine the improvement in signal availability above some fixed SNR threshold whereby a common bit error rate is ensured in the systems being compared. The signal availability is defined as the integrated duration (ms) a signal spends above some SNR threshold (dB). Data from the thirteen blocks, each of 30 minutes duration, were individually processed and analysed. No attempt was made to categorise the data into meteoric or non-meteoric events.

The phased array data were processed as a 4-element Butler matrix with the output beams at azimuthal angles $\pm6^{o}$ and $\pm17^{o}$. The four beams were continuously monitored to pick that beam with the highest instantaneous SNR. The single antenna reference system signal was taken to be the mean of the four phased signals feeding the Butler matrix. Figure 1 shows the signal availability improvement factor as a function of operational SNR for the Butler matrix system F(BM) over the reference system. The standard error on the F(BM) ratio values have been calculated from the thirteen separate data files and show a clear decrease with improving statistics at the lower SNR thresholds. The large experimental errors have resulted from the limited observation period of ~0.5 hours preventing the detection of similar number of high SNR trails during the thirteen separate data acquisition runs. These errors can be stabilised by extending the observation period or the number of the data runs.

Also shown in Figure 1 is the improvement in signal availability to be expected if all incoming trails follow the classical underdense distribution. According to *Sugar* [1964], a 6 dB increase in transmitter power leads to roughly a two-fold increase in the number of trails. Hence a doubling of signal availability is to be expected through the use of a 4-element Butler matrix. The expected improvement is actually around ~1.7 if we take into account the mean improvement in the SNR of the Butler matrix over the active azimuthal range.

The experimental improvement values are found to be threshold dependent and lie just below, or match, the classical improvement values over the SNR range 20 to 25 dB before rising towards the upper line over the SNR range 25 to 30 dB. Thereafter, the experimental values are seen to increase dramatically for SNR>31 dB. The Butler matrix appears to be under performing for SNR<25 dB even though the improvement factors are within acceptable bounds overall. The experimental results of Figure 1 are consistent with the

RF based 4-element Butler matrix [*Mawrey and Weitzen*, 1995] where the average hourly meteor arrival rate was observed to be approximately double that of a single antenna system.

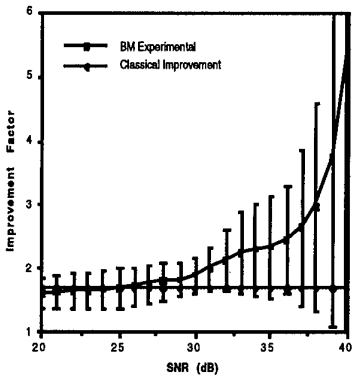


Fig. 1. Signal availability improvement of the Butler matrix.

The theoretical analysis of *Akram and Cannon* [1995] predicts the dramatic improvement increase seen in the BM system over the reference system for SNR>31 dB.

3.4 Multi-Butler matrices and space diversity results

We now address the issue of the shortfall in the Butler matrix performance owing to the ~-4 dB beam crossover points as highlighted by *Kilpatrick* [1991]. The full gain of the Butler matrix is not utilised if the incident wavefront is not coincident with one of the four main beam directions and indeed in the worst case the SNR gain of a 4-element Butler matrix can fall from 6 dB to only 2 dB. The obvious solution to minimise the shortfall is to increase the number of beams (e.g., from 4 to 8) that are used to illuminate the common volume. This will increase the likelihood of main beam excitation but the beam crossover points still occur at ~-4 dB. The implementation of an 8-element Butler matrix will double the hardware requirements in terms of the number of antennas and receivers and also significantly increase the level of DSP complexity. It should be noted that the implementation of such a Butler matrix solely in hardware will increase the number of directional couplers to 12 and the number of phase shifters to 8.

In our DSP implementation, we find that the number of illumination beams can be increased from 4 to 8 more simply by running two 4-element Butler matrices in parallel. There is no increase in the number of antennas and receivers with this method and the DSP processing overhead is lower than in the case of an orthodox 8-element Butler matrix. In our implementation the second 4-element Butler matrix has its main beams coincident with the beam crossover points of the first resulting in eight 'look' directions. This technique can be likened to diversity in the azimuthal domain as we are attempting to counter the 'fades' in the output of one Butler matrix by switching to the second. Whilst more squinted Butler matrices could be implemented a trade of between complexity and throughput would probably place a limit at two or three pseudo arrays.

In synthesising eight 'look' directions, the second 4-element Butler matrix can be derived from the original matrix by progressively phase shifting the inputs to steer the directions of the four main output beams to the required azimuthal angles. This method of superimposing beams has the additional advantage of raising the beam crossover points to ~-2 dB, an increase of 2 dB on both the single 4 and 8-element Butler matrix.

It is instructive to extend the signal availability analysis to quantify the improvement offered by the new technique of overlapping Butler matrices. The advantages of a MSC system that is able to switch between passively synthesised beams and space diversity depending on which of these two techniques gives the highest instantaneous SNR are also investigated.

Four configurations of improved MSC reception systems are considered: a dual Yagi selection diversity combiner (DVT); a single 4-element Butler matrix with the output beams at azimuthal angles $\pm 6^{o}$ and $\pm 17^{o}$ (BMA); a phase shifted 4-element Butler matrix (BMB) derived from BMA; parallel operation of the 4-element Butler matrices BMA and BMB to produce eight 'look' directions (BMC); and finally, parallel operation of the diversity combiner and the eight 'fingered' Butler matrix BMC (DVT+BMC). The signal availability improvement factors, relative to the reference system (REF), as a function of SNR threshold for these four systems are summarised in Figure 2.

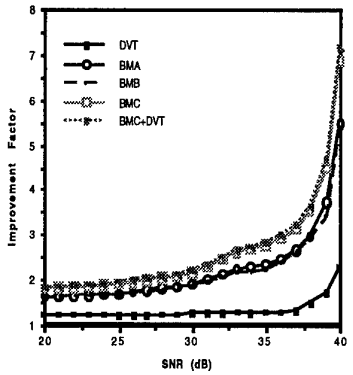


Fig. 2. Comparison of the signal availability improvement factors for various MSC systems.

The diversity system DVT shows the most modest improvement, typically around ~10%-15%, a figure that is almost threshold independent over the SNR range 20-35 dB. The rise in the improvement seen for SNR>35 dB could be an artefact of poor statistics (small sample size) rather than an actual increase. It was verified that the diversity improvement was higher, rising to almost 50%, for low amplitude signals with SNRs between 10 dB to 20 dB. This substantiated the view that diversity effects play a more prominent role towards the end of a trail lifetime. Practical MSC systems are, however, unable to exploit these fairly significant 50% improvements as the SNR is below usable thresholds.

The improvement factors of the two 4-element Butler matrices BMA and BMB is expected to be very similar if a few general assumptions are satisfied. An important assumption is that all 'look' directions lie within the active azimuthal profile of the single Yagi. If so, the envelope of the maximum instantaneous SNR of the beams of each matrix should be very similar when averaged over a large number of randomly arriving sporadic meteors. It will be noticed that the BMA improvement always exceeds the BMB value where there is a difference. This is probably a direct result of the alignment of one of the beams of the BMB matrix with the great circle at $\phi=0^{o}$, a region of the sky with the lowest probability of signal return. Over the

SNR range 20 to 30 dB where the meteor numbers are large and arrival angles fairly random, the performance of the two Butler matrices is, as expected, very similar.

The BMC signal availability improvement factor in Figure 2 shows the added advantage of employing two 4-element Butler matrices in parallel. Signal availability has increased by ~10% over BMA or BMB at the lower thresholds rising to >20% at the highest thresholds. This improvement can be viewed as arising from the approximate doubling of the coverage of the sky with the full gain of the original 4-element Butler matrix. Alternatively, and perhaps more accurately, the improvement results from raising the beam crossover points from ~-4 dB to ~-2 dB.

Turning finally to the DVT+BMC improvement factor in Figure 2, we find that the addition of diversity increases the signal availability by an almost threshold independent figure of ~2.5% over BMC. This strikingly confirms the qualitative conclusion reached earlier that diversity has little to offer in beam synthesised systems.

4.0 Direction finding driven adaptive beam forming scheme

We have shown above that the parallel operation of Butler matrices offers an effective method of monitoring almost the entire sky with the combined gain of the reception array. The method of parallel Butler matrices approaches the ideal reception solution of targeting incoming individual trails with the full available gain. We now investigate the relative merits of such an adaptive beam forming scheme and compare its performance against the single 4-element Butler matrix.

For this exercise, the 4-element array at the Pershore station was modified such that the inter-element spacing was reduced to ~0.625λ. This prepared the use of the array for direction finding purposes. Data was taken for roughly the same length of time as that for the Butler matrix study and once again post processing was applied to the data.

The mean phase difference between adjacent antennas was computed for two consecutive blocks of ten sample points. A new beam was synthesised if the arrival direction associated with the second block was outside the 3 dB drop-off points of the 4-element array factor centred on the arrival direction of the first block. Beam forming with the old phase weights took place every sample point even though variations in the phase differences (to detect changes in arrival directions) were only monitored every ten points. As with the Butler matrix, the performance of the adaptive beam system was referenced to a single antenna system. This was taken to be the mean of the four signals input into the beam forming network.

Figure 3 shows the comparison of the improvement factor of the 4-element adaptive scheme with the results of the Butler matrix study already shown in Figure 2. The performance of the adaptive scheme is below that of the Butler matrix over the SNR range 20-32 dB but is superior thereafter. It appears that there is no advantage in changing the beam direction to keep up with the high frequency of direction changes that occurs at low SNR. The adaptive process is also limited by the poor accuracy of the direction finding algorithm at low SNR. Therefore, it is better to use the simpler passive scheme of fixed direction beams to monitor the more numerous low SNR meteors. However, for the less frequent high SNR meteors with well defined arrival directions, it is advantageous to form a beam in the correct direction.

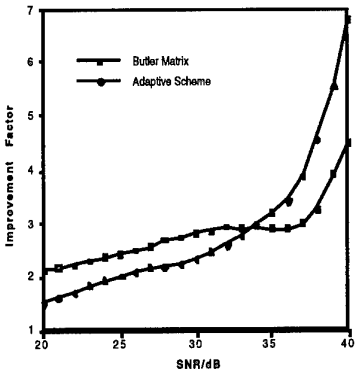


Fig. 3. Comparison of passive and adaptive beam forming schemes.

5.0 Computer prediction model driven adaptive scheme

We now consider an adaptive scheme where the beam direction is controlled by a prediction model such that a beam is synthesised towards the hot spots, i.e. regions of significant meteor scatter. This approach is simple to implement, the gain is likely to be comparable to the 4-element Butler Matrix and it is not subject to the disadvantages of the real-time direction finding adaptive scheme. The theoretical framework behind the computer prediction model presented here is fully described in *Akram and Cannon* [1994]. We describe here the graphical output of the model and its usefulness in providing directional information to steer a high gain beam.

We first discuss model results from a link where the general characteristics are well understood. Consider a link running West to East through 10^{o} of longitude on the equator. Figure 4 shows a typical example of the graphical output from the model at various local times when the non-uniform Davies-Pupyshev hybrid radiant distribution [*Akram and Cannon*, 1994] is specified. The contour plots are displayed over identical (x, y) planar regions at an altitude of 100 km. The (x, y) values range over the domain (-2D<x<2D, -2D<y<2D) where 2D is the ground range. It should be noted that the transmitter-receiver separation along the x-axis is approximately equal to the ground range. More rigorously, the conversion of (x, y) distances to distances along and perpendicular to the great circle joining the transmitter-receiver results in errors typically less than ~0.5%, e.g., the transformation of the maximum ground range of 2000 km leads to a distance along the x-axis which differs by only ~10 km. Thus as a simplifying step all results are presented in link co-ordinates rather than in distances along and perpendicular to the great circle joining the transmitter-receiver.

Panel (I) describes the meteor detection probability levels at local time (LT) 0600, the higher the contour level, the higher the probability of locating a usable meteor trail. Immediately obvious is the asymmetry resulting from the tilt of the earth's spin axis to the ecliptic normal which causes the hot spots to differ in size and strength. The tilt has the effect of increasing the probability of meteor detection north of the great circle (x-axis), albeit over a smaller area in relation to the southern dispersed hotspot.
The remaining three panels of Figure 4 show the subsequent evolution of the hotspots at six hourly intervals. In panel (II) (local noon) an overall decrease in the probability of meteor scatter with a particularly sharp decline northwards of the transmitter-receiver great circle is observed. At this time the

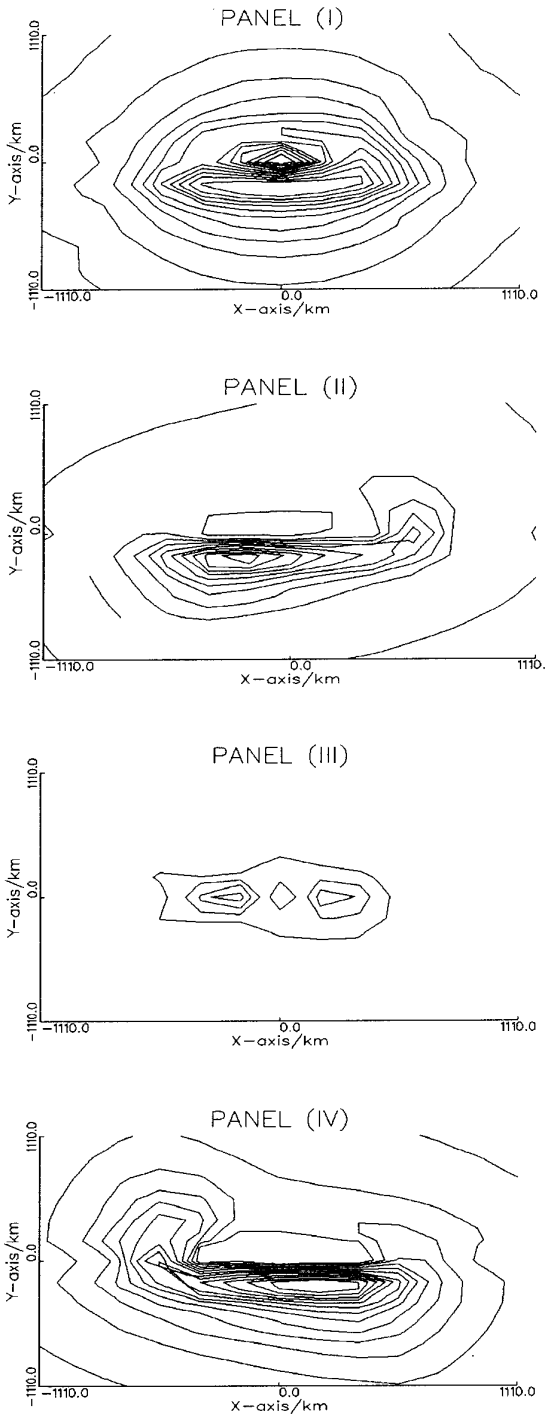


Fig. 4. Evolution of hotspots at six hourly intervals.

concentrated meteor flux originating from radiants centred at $\lambda g = 295°$ (λg being the geocentric longitude) and the apex direction in the Davies distribution have set over the link. Thus most of the sporadic meteoric activity over the link at noon can be attributed to radiants centred at $\lambda g = 65°$. Even this single prong of the Davies distribution sets over the link at 1800 LT, panel (III), and we find scant evidence of meteor scatter. What little activity there is results from the very low constant intensity of radiants specified in the non-uniform radiant distribution in this local time sector.

Not shown are pictures leading up to 1800 LT where the only signs of meteoric activity are located in ever diminishing hotspots near overhead of the receiver. The spatial extent of the link ensures that the receiver end of the link is able to retain 'sight' of the $\lambda g = 65°$ leg of the Davies distribution even though it may have set over the transmitter. The situation is reversed once local dusk has been crossed whereby the hotspot over the transmitter increases in strength at a faster rate than that over the receiver. Turning to panel (IV) (local midnight), the last snapshot in this time series, we notice the overall increase in the probability of meteor scatter southward of the great circle with a small pronounced hotspot over the transmitter. It is clear that the concentrations of meteoric flux at $\lambda g = 0°$ and $\lambda g = 295°$ are beginning to rise over the link.

The sequence of these four snapshots has allowed the user to follow the time evolution of major regions of meteor scatter. Carried out at shortened time intervals the process would have revealed detailed passage of the hotspot regions across the sky. It is the monitoring of these regions that is critical to the design of improved meteor scatter communication systems.

6.0 Verification of model against experimental data

In the context of this paper detailed knowledge about the location of the hotspots is most important since if it can be achieved it would be possible to preferentially illuminate such regions with increased gain and, hence provide improved system performance. Complete verification of the model requires a wide range of experimental links distributed in geographic area, orientation and operating for substantial periods.

The model was first verified using data from the Pershore-Wick link with the reception 4-element array spacing at ~0.625λ. The data set consisted of three contiguous periods totalling approximately 187 hours or nearly eight days which were post-processed. We focus on a sub-set lasting 24 hours to show the diurnal variation in the arrival distribution of meteors.

It is not possible to resolve either azimuth or elevation angle of an incoming meteor trail with a simple one dimensional linear array. Rather, it is the phase difference $\Psi = 2\pi(d/\lambda)\sin\phi\cos\Delta$, where ϕ is the azimuth angle, Δ is the elevation angle and d is the array inter-element spacing, between consecutive elements of the array that can be measured. In order to validate our model we consequently compared observed values of the phase difference with model predicted values. Data from the link were transformed to phase differences Ψ in the following manner.

The data were first averaged over 35 ms (ten 3.5 ms sample points) and then searched for sub-sections where the mean of the five signals exceeded 20 dB SNR. The sub-sectioning process was halted once either the mean signal dropped below 20 dB SNR or the length of the sub-section exceeded 1000 points. A further more complex condition for termination of the sub-section data block was based on consideration of the arrival of a second trail before a first trail had finished decaying. Changes in the amplitude and phase variation between consecutive 35 ms points were monitored to detect the arrival of a second trail. Two parameters, besides the start time, were used to characterise the sub-section data block. These were the average value of Ψ and the correlation factor

R. The latter determined how closely the amplitude drop off within the data block approximated a simple underdense trail exponential decay. Setting R to a high acceptance value provided a simple method of filtering out the majority of non-meteoric (e.g., sporadic-E) contaminant signals. The less frequent overdense trails were also filtered by this process but it was felt that the resulting reduction in the meteor count, dominated in any case by the underdense trails, would not unduly affect the analysis.

Figure 5 depicts the experimental variation of Ψ as a function of local time. Individual values of Ψ were placed into bins 25^{o} degrees wide whilst the associated start times were converted to bins half an hour wide. The plot is composed of 48 local time bins running midnight to midnight and 18 Ψ bins spanning the range $\pm 225^{o}$. Every value of (Ψ, LT) is associated with R > 0.9, i.e. we only consider trails that appear to be underdense at confidence level greater than 90%. It should be noted that the figures show the number of meteors that arrived within a certain range of Ψ over an interval of half an hour rather than the integrated trail duration which is sensitive to a small number of long lived trails.

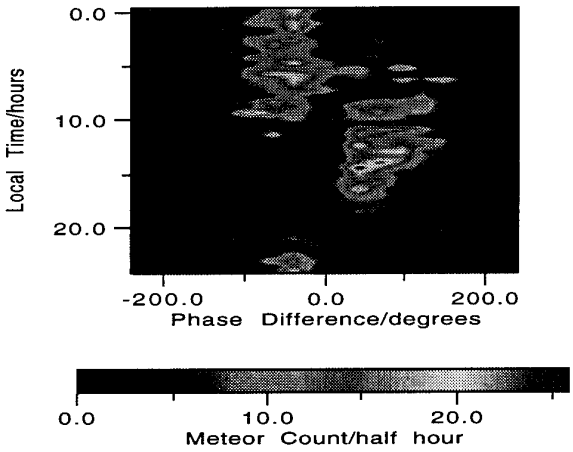


Fig. 5. Observed meteor arrival distribution.

Ψ is zero for signals originating from the great circle path whatever the elevation angle. We find that the zero phase difference bins are generally sparsely populated at all local times, thus confirming our expectations that the great circle is a region almost devoid of meteor scatter.

Bearing in mind the north-south orientation of the link, trails with positive(negative) Ψ have arrived east(west) of the array normal. It is difficult to be more precise in specifying the direction of incoming trails as all elevation angles are possible except for those constrained by the Yagi antenna vertical polar diagram. Despite this inability to resolve the ambiguity between azimuth and elevation angles, the asymmetric nature of Figure 5 does provide valuable insight into the diurnal variation in the meteor arrival rate.

Most of the meteor scatter activity at midnight is centred to the west of the array normal with almost no activity to the east. Meteor trails arrive from the east starting at ~0200 and increase in number reaching a constant level at ~0500 which is sustained almost to dusk. The western hotspots come to an abrupt end at ~1100 after which time there is negligible activity on this side until ~1900. A common expectation of meteor links is the maximisation and minimisation of the meteor rate at dawn and dusk respectively owing to the orbital motion of the earth. We certainly find a minimum in the meteor rate at ~1800 when no hotspots of any significance are present over the link. However, the situation regarding the time of maximum meteoric activity is less clear and almost no activity is observed at ~0600 on the eastern side. Rather than a localised morning maximum, the meteoric activity seems to maximise at a fairly constant level over the hours 0700 to 0900.

Figure 5 exposes the limitations of the standard configuration of single Yagi antennas for transmit and receive purposes illuminating a common volume centred at the midpoint. The time dependent distribution of arrival angles clearly demonstrates the advantage of some form of beam forming, (e.g. through a linear array of N elements), especially in the afternoon and night hours. Uniform illumination either side of the great circle is likely to be satisfactory only in the early morning. Antenna gain at other times could be employed more usefully by preferentially illuminating a smaller volume. At its simplest, just two changes in the direction of a high gain beam at ~0100 and ~0900 would offer considerable advantage over a fixed beam system.

The model was executed with the geographic parameters of the Pershore-Wick link to determine how closely predicted values of Ψ matched those observed experimentally. Four element horizontally polarised Yagi antennas positioned 1 λ above the ground were specified at both link ends and polarisation coupling was included. The meteor arrival rate was computed as a function of (x, y, z) link centred co-ordinates at a height of 100 km every 10 minutes for 24 hours starting at 2300(0000) UT(BST). Each (x, y, z) point was transformed to equivalent elevation and azimuth angles (Δ,ϕ) from which the phase difference, given by $\Psi = 225\sin\phi\cos\Delta$, as would be measured by a 0.625λ array, was calculated. In like manner to the experimental data the phase differences were placed into one of 18 bins each 25^{o} wide giving a range of Ψ from $\pm 225^{o}$. The Ψ bin values of three 10 minutes model predictions were integrated to provide half hourly Ψ values to facilitate comparison with Figure 5.

Figure 6 depicts the predicted half hourly values of Ψ for the Pershore-Wick link. It is fortuitous that the absolute binned values have come out so close to the observed values as no effort was made to scale the predicted values in any way. The predicted values are highly dependent on the chosen link geometry and antenna configuration. Diurnal changes in the values of Ψ are more important than the absolute values in determining the validity of the model.

As a prerequisite for any MSC model, the great circle running down the centre of Figure 6 duly shows up as a region of negligible meteor scatter. Overall, there is a remarkable degree of similarity between the two plots of Figures 5 and 6 regarding both the position and time of the hotspots. The model correctly predicts the western quadrant as the predominant location of meteor scatter from around 1800 hours onwards followed by the subsequent switch over to the eastern quadrant around 0100 hours. There are, however, minor differences which suggest that the Davies non-uniform radiant distribution as currently used in the model requires fine tuning. One such difference is the temporal extent of the afternoon western blackspot. The observed blackout of meteoric activity lasts about an hour or so longer than that predicted and there is no longer a clean minimum of activity around dusk. It is tempting to correct such timing offsets by

altering the longitudinal positions of the $\lambda g = 65^{o}$ and $\lambda g = 295^{o}$ legs of the Davies distribution. But the day to day variability in Ψ (not shown here but clearly apparent in the 8 days of available data) indicates the dangers of making such empirical changes to the model based on a single day of data. It is more prudent to delay these modifications until the present data set can be extended to at least match the sample size of the *Davies* [1957] experiment which ran for one day every month for a year and observed 2400 meteor radiants and included more long term effects such as seasonal variations of the meteor arrival rate. Any changes to the Davies distribution then become statistically justifiable.

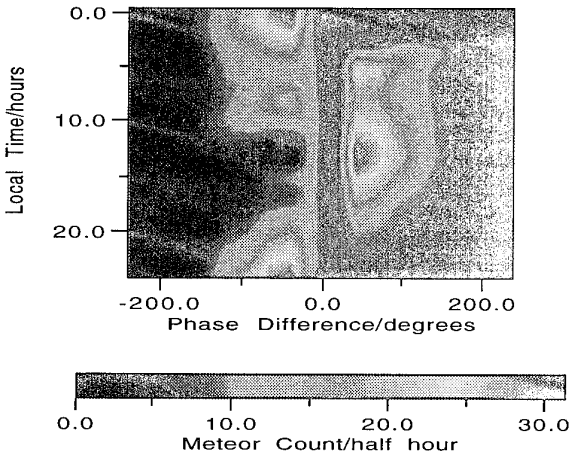


Fig. 6. Predicted meteor arrival distribution.

7.0 Conclusions

The techniques of passive beam forming with and without space diversity have been investigated as a means of increasing the signal availability of meteor scatter communications systems. A 4-element Butler matrix was implemented using a DSP approach to synthesise four beams in fixed azimuthal 'look' directions. The signal availability improvement factor of the Butler matrix relative to a single antenna system ranged between 1.6-1.8 in a 300 Hz bandwidth over the SNR interval 20-30 dB. Beyond SNR>30 dB, the improvement increased with increasing SNR threshold rising to a value of >5 at 40 dB. It was shown that the signal availability could be further increased by 10%-20% by employing two overlapping 4-element Butler matrices in parallel. Extension of the overlapping Butler matrix system to include space diversity resulted in only a minor improvement in signal availability.

The performance of a real-time radio direction finding based adaptive beam forming scheme was compared with the Butler matrix. It was found to be superior only at high SNR (>33 dB in a 300 Hz bandwidth) where the direction of the incoming signal can be accurately determined. The simpler Butler matrix was more effective in a high noise environment.

The predicted diurnal variation of the arrival distribution of underdense trails with our representation of the non-uniform radiant distribution has yielded encouraging results and gives impetus to our speculation that a computer controlled adaptive beam forming system is a viable proposition for improving MSC systems.

References

Akram, A., and P. S. Cannon, A meteor scatter communications system data throughput model, *Proc. IEE HF94 - HF Radio systems and techniques Conf.*, York University, England, 392:343-347, 1994.

Akram, A., and P. S. Cannon, Passive beam forming and spatial diversity in meteor scatter communication systems, Accepted for publication in *Rad. Sci.* 1995.

Cannon, P. S., and A. P. C. Reed, The evolution of meteor burst communications systems, *J. Inst. Electr. Rad. Eng.*, 57,101-112, 1987.

Cannon, P. S., A. K. Shukla and M. Lester, Spaced antenna diversity in temperate latitude meteor burst systems operating near 40 MHz: Variation of signal cross-correlation coefficients with antenna separation, *Radio Sci.*, 28(2), 203-212, 1993.

Davies, J. G., Radio observations of meteors, *Advances in Electronics and Electron Phys.*, Vol.9, pp.95-128, 1957.

Kilpatrick, J. A., J. A. Weitzen and S. A. Parl, Adaptive meteor burst antenna study, RL-TR-91-234, Rome Laboratory, September 1991.

Mawrey, R. S. and J. A. Weitzen, Measured performance of meteor burst systems using antenna beam steering, *IEEE Trans. Comm.*, 43:1467-1476, 1995.

Miller S. L. and L. B. Milstein, Error correction coding for a meteor burst channel, *IEEE Trans. Comm.*, 38:1520-1529, 1990.

Pursley, M. B. and S. D. Sandberg, Variable rate coding for meteor burst communications, *IEEE Trans. Comm.*, 37:1105-1112, 1989.

Shukla, A. K., P. S. Cannon and M. Lester, Spaced antenna diversity in temperate latitude meteor burst systems operating near 40 MHz: Variation of signal cross-correlation coefficient with time, *Radio Sci.*, 27(6), 841-849, 1992.

Shukla, A. K. and P. S. Cannon, Seasonal variation of diversity gain in a broadcast meteor scatter communications system, *Proc. Ionospheric Effects Symposium*, 150-156, 1993.

Sugar, G. R., Radio propagation by reflection from meteor trails, *Proc. IEEE*, 52: 116-136, 1964.

Weitzen, J. A., W. P. Birkemeier and M. D. Grossi, An estimate of the capacity of the meteor burst channel, *IEEE Trans. Comm.*, 32:972-976, 1984.

A. Akram and P. S. Cannon, Radio Science and Propagation Group, Defence Research Agency, Malvern, Worcs WR14 3PS, England (UK).

© British Crown Copyright 1995/DERA
Published with the permission of the controller of Her Britannic Majesty's Stationery Office

DISCUSSION

Discussor's name: R. Bauman

Question/Comment:

1. Can the presenters comment on the following reasoning regarding the concern about multipath in the meteor scatter channel; fading aoccurs in approximately 26-30% of meteor trails. When fading occurs, it tends to occur towards the end of the trail, i.e. the last 20-30%. Therefore, concern about multipath in MB channels would appear to be a concern about a phenomenon that occurs approximately 4% - 9% of the time. Does this justify complicating the MB system with multipath and robust techniques?

2. Do the presenters see user need or demand for MB communication that might in some way drive their research?

Author/Presenter's Reply:

1. The participants who took part in the discussion following this question agreed with the thrust of the comments that preceded the question, and generally felt that the added complexity (and, hence, cost) associated with providing robust techniques to mitigate such multipath effects in MB systems was not warranted.

2. The discussion following this question did not indicate that there was strong "user" demand for MB communication systems at this time, but that perhaps some "potential users" might not be fully aware of some of the attractive features of MB systems, such as low-probability-of-intercept, anti-jam, low power, transportability and relatively low cost.

Coverage area computations: an approach with diffuse reflections

E. Van Hoof
E. Van Lil*
K.U.Leuven
Div. ESAT-TELEMIC
Kard. Mercierlaan 94
B-3001 Heverlee
Belgium
F.J. Rodriguez Blanco
F. Perez Fontan
ETSI de Telecommunicacion
Universidad de Vigo
E-36200 Vigo
Spain

Summary

More and more computations of coverage areas of a transmitter are now deterministic, as opposed to the empirical or semi-empirical methods that don't give the precision nowadays needed. The deterministic methods mostly use direct, specular reflected and diffracted rays.
It has been shown recently, that specially for indoor propagation, the diffuse reflection cannot be neglected. Diffuse reflection is caused by the roughness of the terrain and can occur outside the incidence plane. Formulas treating this subject can be found in Beckmann [1] .
In a digital terrain model, triangles can be used to accurately describe the environment. Therefore the classical formulas for diffuse reflection have to be adapted to make the use of triangles possible. Some mathematical problems that arise, caused by the division of the rectangle in two triangles, with two different center points, are discussed in this paper.
As diffuse reflection is caused by the roughness of the terrain, this is an important parameter. This roughness is, just as the specular reflection coefficient, determined by the sort of terrain that reflects. Research has still to be conducted on the determination of this parameter, as well as on the optimal distribution.
Finally this model is used in a program for computing the coverage area optical. It is especially useful in cases where the digital terrain model has a resolution of a few $\lambda's$ and when computer time is not unlimited.

*National Fund for Scientific Research (NFWO), Belgium.

1 Introduction

With the growing use of terrestrial wireless communications for different purposes: cellular telephones, military telecomsystems, ..., there is more than ever need for an accurate description of the coverage area of a transmitter. Problems as absence of signal in certain areas, blocked by hills for example, and interference of an other base station, can occur. The empirical or semi–empirical methods of Okomura–Hata and of Ibrahim&Parsons [2] no longer give the precision needed. Therefore deterministic models are necessary.
Most deterministic models only take into account the direct, specular and diffracted rays. Diffuse reflections can give important contributions to the total field [3]. This method of diffuse reflection has the advantage that the accuracy can be enhanced while still working with the existing resolutions of most digital terrain models. In Belgium, for example, the resolution of the digital terrain model provided by the NGI (National Geographical Institute) is 2" in latitude and 1" in longitude. This gives, knowing that Belgium lies at around 50 degrees latitude, a resolution of 30,8 m in north–south direction and 38.8 m in east–west direction. So this method does not need costly measurements of the terrain. The amount of computer time is off course more than the empirical formulas but it is much less than would be needed if a resolution of 0.1λ would be taken to compute the coverage area only using specular reflections. This resolution is a minimum if a total precise solution is wanted with only specular reflection.
In this article we look at we have to do to adjust the formulas of diffuse reflection to be usable with

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

a digital terrain model. In section two the theory of diffuse reflection is briefly explained. The changes that have to be made to the formulas are described in section three. The next section deals with the mathematical problems that can arise when using the formulas for the triangles. The parameters influencing the diffuse reflection depend on the kind of terrain used. In section six conclusions and remarks are given.

2 Diffuse reflections

Apart from the direct ray and the specular ray, there also exists a contribution to the total field by diffuse reflection. Here the reflection between two points does not take place in the specular reflection point, but can occur at any point of the terrain(see figure 1), as long as there is a line-of-sight .

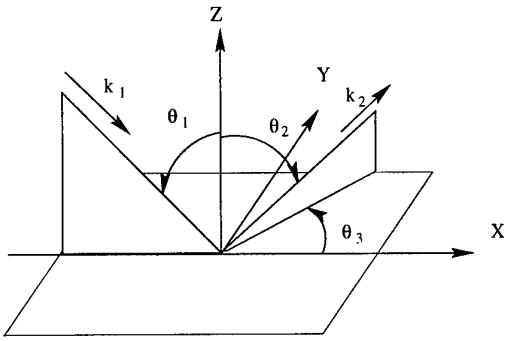


Figure 1: *Configuration of a diffuse reflection*

Seen from another viewpoint, with a fixed transmitter and reflection point, this means that all the points receive some reflection, and only the point that lies in the plane of incidence and whose incidence angle equals the reflected angle, receives the specular reflection, attenuated due to the non–flatness of the surface. It is easily concluded from this viewpoint that specular reflection is nothing else than a special case of diffuse reflection. Therefore, the theory of diffuse reflection must equal the theory of specular reflection at this point. This subject will be discussed later on.

Now, the theory of diffuse reflection [1] will be briefly summarized.

Let E_1 and E_2 respectively be the incident and scattered field, with wave vectors $\vec{k_1}$ and $\vec{k_2}$ respectively. For simplicity reasons, we take only a horizontal or vertical polarization. These polarizations are defined with the incidence plane (for E_1) and the scattering plane (for E_2) as references. This means that E is horizontally polarized when normal on the reference plane and vertically polarized when in the reference plane.

Let the rough surface be given by

$$z = \zeta(x, y) \tag{1}$$

Then, with P the observation point and R' the distance from P to a point (x,y,ζ(x,y)) on the surface S, the scattered field is given by the Helmholtz integral:

$$E_2(P) = \frac{1}{4\pi}\int_S\int (E\frac{\partial\Psi}{\partial n} - \Psi\frac{\partial E}{\partial n})dS \tag{2}$$

and Ψ is the free space Green's function:

$$\Psi = \frac{e^{jk_2R'}}{R'} \tag{3}$$

Now a few approximations are made. The distance R' is moved to infinity : $R' \rightarrow \infty$ and the radius of curvature of the irregularities is considered to be small in comparison with the wavelength, so that the field on any point of the surface can be approximated by the field that would be on the tangent plane at that point.

The first approximation allows to assume that the distance to the observation point can be replaced by the distance to a reference point R_0. The second approximation, only valid if the surface does not include sharp edges or sharp points, gives the classical equations for the fields at an infinite boundary with reflection coefficient R:

$$(E)_s = (1 + R)E_1 \tag{4}$$

$$(\frac{\partial E}{\partial n})_S = (1 - R)E_1\vec{k_1}\cdot\vec{n} \tag{5}$$

Now we can substitute these equations into the Helmholtz integral, see equation 2.

$$E_2(P) = \frac{je^{-jkR_0}}{4\pi R_0}\int\int_S (R\vec{\nu} - \vec{p})e^{j\vec{\nu}\cdot\vec{r}}\cdot\vec{n}dS \tag{6}$$

with:

$$\vec{\nu} = \vec{k_2} - \vec{k_1} \tag{7}$$

$$\vec{p} = \vec{k_2} + \vec{k_1} \tag{8}$$

With all these equations, we are able to compute the scattered field at every point. In [1], this is done for a rectangle. An extra supposition is added, namely the length and width of the rectangle have to be considerably larger than the wavelength.

Finally this yields, with the configuration as in figure 1 the next equation, where $\vec{\nu}$ is perpendicular to $\vec{p}$ ($|\vec{k_1}| = |\vec{k_2}|$).

$$E_2(P) = \frac{-jkRe^{-jkR_0}}{2\pi R_0} \cdot \frac{1+\cos\theta_1\cos\theta_2-\sin\theta_1\sin\theta_2\cos\theta_3}{\cos\theta_1+\cos\theta_2} \cdot \int_{-X}^{X}\int_{-Y}^{Y} e^{j\vec{\nu}\cdot\vec{r}}dxdyE_1 \quad (9)$$

Now if we assume that $\zeta(x,y)$ is a random variable, then the remaining integral in equation 9 is the characteristic function of the probability density function.
For a rectangular surface with a Gaussian probability density function, we get:

$$E_2(P) = \frac{-j2kXYe^{-jkR_0}}{\pi R_0}R. \frac{1+\cos\theta_1\cos\theta_2-\sin\theta_1\sin\theta_2\cos\theta_3}{\cos\theta_1+\cos\theta_2} \cdot \mathrm{sinc[k}(\sin\theta_1-\sin\theta_2\cos\theta_3)\mathrm{X}] \cdot \mathrm{sinc[k}(-\sin\theta_2\sin\theta_3)\mathrm{Y}] \cdot e^{-\frac{1}{2}\sigma^2(\cos\theta_1+\cos\theta_2)^2}E_1 \quad (10)$$

with $\mathrm{sinc} = \frac{\sin(x)}{x}$

A few remarks can be made about this equation.
In the first factor you can see that the field in P is proportional to $\frac{e^{-jkR_0}}{4\pi R_0}$ where R_0 is the distance from the reference point to the observer and not of the observer and transmitter like in the specular case. But it's also obvious from this factor that the received field at P is proportional to the surface considered, that equals $4XY$. Great care must be taken when adding those fields to the direct field of the transmitter.
The fact that the integral is proportional to the surface is very important if comparisons are to be made with the specular reflection. This specular reflection just uses an infinite surface, while in equation 10 only a small piece of the surface is considered.
The R factor accounts for the reflection coefficient. Given that E is horizontally or vertically polarized, this coefficient can be computed if the material of the surface, e.g. grass, wood, etc., and the angle of incidence are known. Only at higher frequencies, above 20 GHz, the properties are not yet well been documented.
The third factor is a factor that is responsible for the geometry in the case considered. The sinc factors originate from the rectangle. The exponential factor is caused by the surface that has a certain Gaussian distribution of the heights.

3 Diffuse reflection for a digital terrain model

A digital terrain model is a database that contains the heights of several points. These points have a fixed spacing , for example 1 second in latitude and 1 second in longitude. Off course this is not a fixed distance on a large (world) scale, due to the curvature of the earth, but in an area of about 20km x 20km considered here, this will not make a great difference.
4 points, that form a rectangle in longitude and latitude, are rarely in one plane. To describe the terrain without extra effort in a smoother way, we divide the terrain into triangles. The criterion we used, was:
If $(z2-z1)*(z3-z1) > 0$ the rectangle is divided by the line 23. Else by 14.
The principle is, if z3 and z2 are both higher or lower than z1, then 1 leads either to a mountain or a valley. Therefore we use the line 23 to divide the rectangle and to follow the terrain more closely, see figures 2 and 3.

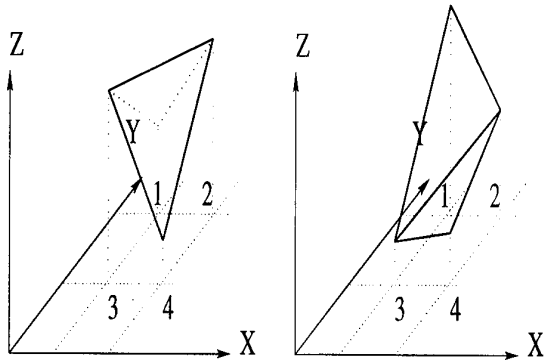


Figure 2: *Division of rectangle when* $(z2-z1)*(z3-z1) > 0$

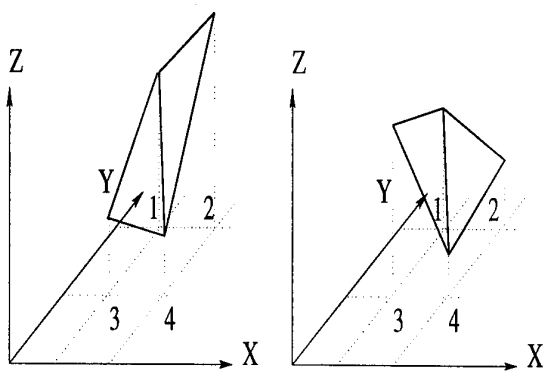


Figure 3: *Division of rectangle when* $(z2-z1)*(z3-z1) \leq 0$

This also means that the formula for diffuse reflection has to be adjusted. We have to find an analytical expression for the integral in equation 6. This integral is now over a surface in the form of a triangle that is not horizontal, requiring a coordinate transformation. To denote this we have used the subscript 0.
This leads to:

$$E_2(P) = \frac{je^{-jkR_0}}{4\pi R_0} R \frac{|\vec{\nu_0}|^2}{\vec{\nu_0}\cdot\vec{n_0}} \cdot e^{-\frac{1}{2}\sigma^2\nu_z^2} \int_T \int e^{j\nu_x x} e^{j\nu_y y} dS \quad (11)$$

valid for all orientations of the triangles, including perpendicular to the horizontal XY-plane. Because we have 2 triangles and 2 possible division of the rectangle, there are 4 formulas to be derived. Also when these integrals are used in a computer program, it is important for the accuracy that a difference is made when $|\nu_x|$ is greater than $|\nu_y|$ and vice versa. So a total of 8 integrals has to be computed. Take for example $|\nu_x| < |\nu_y|$ then the integral for triangle 123 is:

$$I = \frac{D_{12} e^{j\nu_x \frac{x_1+x_2}{2}}}{j\nu_y} \cdot [e^{j\nu_y y_1} \mathrm{sinc}(\nu_x \frac{D_{12}}{2}) - e^{j\nu_y \frac{y_2+y_3}{2}} \mathrm{sinc}(\nu_x \frac{D_{12}}{2} + \nu_y \frac{D_{13}}{2})] \quad (12)$$

with D_{13} = distance between points1&3
with D_{12} = distance between points1&2

And the integral for triangle 234 is:

$$I = \frac{D_{34} e^{j\nu_x \frac{x_3+x_4}{2}}}{j\nu_y} \cdot [e^{j\nu_y \frac{y_2+y_3}{2}} \mathrm{sinc}(\nu_x \frac{D_{34}}{2} + \nu_y \frac{D_{24}}{2}) - e^{j\nu_y y_4} \mathrm{sinc}(\nu_x \frac{DH_{34}}{2})] \quad (13)$$

If the two triangles are in the same plane, the sum of these 2 integrals yields back the original sinc–factors of equation 10. Again the assumption is made that the horizontal and vertical distances are a lot greater than the wavelength.

4 Calculation of a diffuse reflection

The calculation of the field diffusely reflected by a rectangle or two triangles in one plane consist of different steps we have to take. Let T be the transmitter, R the receiver and 1,2,3,4 the corner points of the rectangle (see figure 4).

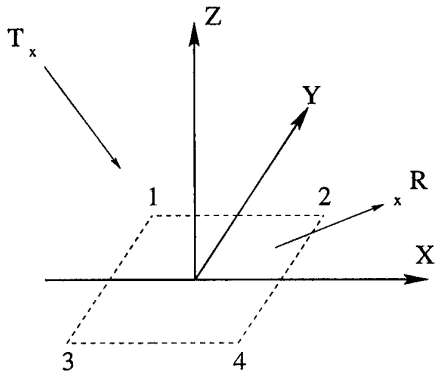


Figure 4: *Configuration of diffuse reflection.*

Firstly we have to determine the field strength of the antenna in the direction of the reference (step a — see figure 5). For the moment we only use vertical or horizontal dipoles, allowing to simulate most kinds of antennas. Then the distance from T to the reflection point is calculated and used to determine the field at the surface (b). Given the direction of the incident and reflected wave, the different vectors of horizontal and vertical polarization, respectively $\vec{E_l}$ and $\vec{E_p}$, can be calculated both before and after reflection (c). Then the adjusted reflection coefficient taken into account diffuse reflection (d) for horizontal and vertical polarization must be multiplied with the corresponding field. Next the distance from the reflection point to R is determined and used to calculate the attenuation (e). At last the polarization of the receiving antenna gives the received signal strength (f).

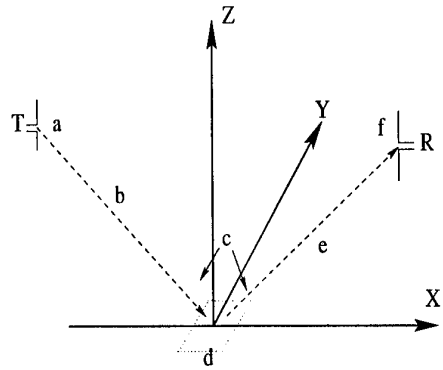


Figure 5: *Different components to calculate the diffuse reflection*

With the formulas for the diffuse reflection of a rectangle, this is easily done.But with the triangle formulas, there are some difficulties.
First of all, because you mostly have two different planes you must compute twice those vectors $\vec{e_l}$ and

$\vec{e_p}$. One has to be careful to use the correct sign for the fields, in order to avoid cancelation of the fields of the 2 triangles by each other. This has been done by checking whether the internal product of the unity vectors $\vec{e_p}$ before and after reflection is not negative. So we ensure that the vector after reflection lies in the same half plane as the vector before reflection. The second problem has really to do with the formulas of the triangles. When you divide the rectangle in two triangles, then you have two barycenters of the triangles that you would use to compute the distance to T and R. Let these distances be Dtbx (x equals I or II) and Dbrx respectively and let the distance from T, R to the middle point of the rectangle be Dtm, Dmr respectively (see figure 6).

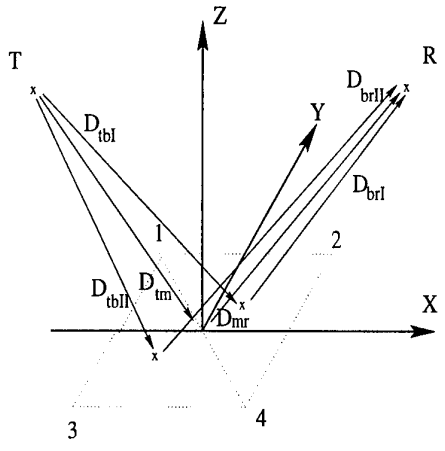


Figure 6: *Definition of different distances.*

Now if we use the distances to the barycenters of the triangles (Dtbx and Dbrx), then the solution is not the one that we get with the equation for a rectangle.

We use a rectangle that lies in a plane. So the 2 triangles are situated in the same plane. We do not have to rotate our coordinate system, we only move the center point of the coordinate system to the barycenters. We have taken Dtm and Dmr to be both 100 m, and the sides of the rectangle to be both 10 m. The zenith angle of the incident ray is 60 degrees and the frequency used is 417 MHz. This gives a wavelength of 0,72 m. This is more than ten times smaller than the length or width of the rectangle that are both 10 m. So this satisfies the supposition made to get equations 12 and 13, but it is smaller than the far–field distance $2D^2/\lambda$. In the figures 7 and 8, the full line is the normal rectangle solution. The trace made by the '+' is made by moving the center of the coordinate system to the center point of the rectangle . The 'o' gives the trace for the two barycenters in the triangle.

Also the total diffusely reflected power can be computed and compared with the power that reflects

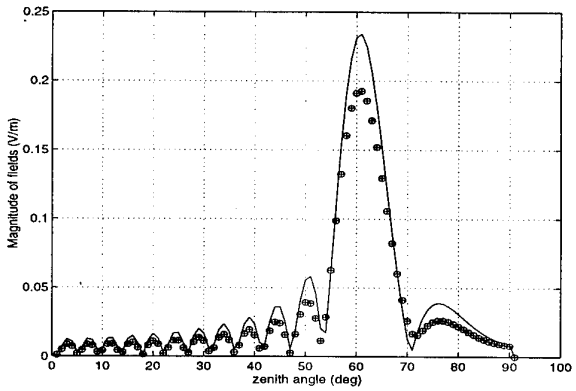


Figure 7: *Magnitude of fields with different calculation methods.*

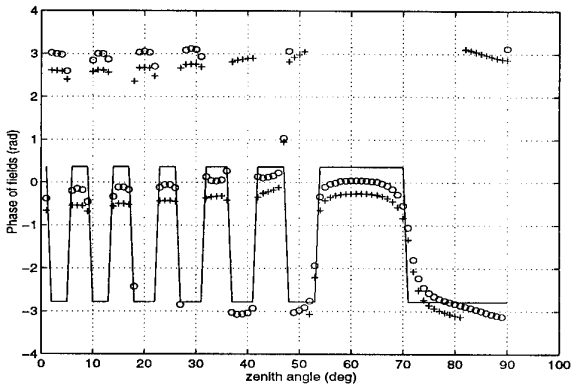


Figure 8: *Phase of fields with different calculation methods.*

from the rectangle. If the rectangle is completely reflecting then this power is also equal to the incident power. In our configuration the power equals $7.51E-3$ W with the rectangular equation. With the method of using the center point of the rectangle, this is $7.615E-3$ W. On the other hand, when using the barycenters we get $7.623E-3$ W . This gives a slight difference of 0.1% between the solutions using the triangles.

This difference becomes more and more apparent with smaller distances Dtm and Dmr. When the axes are moved over a distance (X_s, Y_s), then the integral in equation (11) gets an extra factor, equal to $e^{j\nu_x x_s + j\nu_y y_s}$. When summed over the 2 triangles this gives a difference. Also the total phase factors are different. Because we are talking about phases small differences can be significant. These phase factors are for the center point method ($\psi 1$) and for the barycenter method (ψx with x equal to I or II) equal to:

$$\psi 1 = -\beta(Dtm + Dmr) \qquad (14)$$

$$\begin{aligned}\psi x &= -\beta(Dtbx + Dbrx) + \nu_x x_s + \nu_y y_s \\ &= -\beta(Dtbx + Dbrx) + (k_{2x} - k_{1x})x_s \\ &\quad +(k_{2y} - k_{1y})y_s \end{aligned} \tag{15}$$

and these 2 factors have a second order difference. To correct for this, formulas 12 and 13 can be replaced by formulas containing Fresnel Integrals.

5 Parameters concerning the diffuse reflection

There are several points that influence the calculations of the diffuse reflection. We will discuss successively the width and length of the rectangle, the heights of the corner points of the rectangle and the different parameters for the different surfaces.
Firstly, the resolution of the digital terrain model plays an important role. As can be seen from the assumptions made to derive the equations 12 and 13, this resolution is linked to a minimal frequency that can be calculated accurately. On the other hand, it is also the case that when you use higher frequencies the digital terrain model must have higher resolution, otherwise you could miss some important features of the terrain [4]. Also you can get into trouble with the distance between the two triangles becoming to great, so that the direction of the vectors $\vec{e_p}$ and $\vec{e_l}$ becomes opposite.

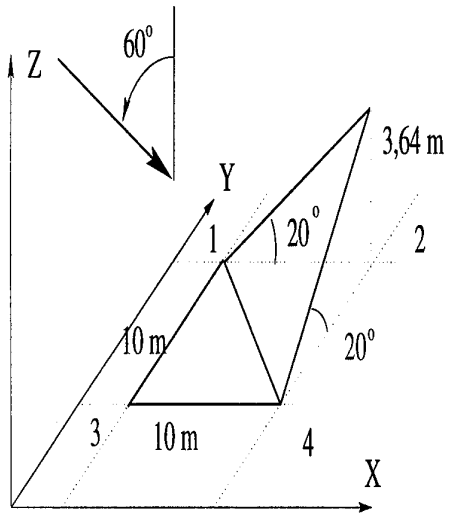


Figure 9: *Configuration used for figure 10*

A second parameter that is monitored by the digital terrain model, are the heights of the corner points. It usually happens that the rectangle does not lie in one plane. Take for example the configuration of figure 9. One corner point has a height of 3,64 m so that the angle in X and Y direction is 20^o.
With this configuration we can expect two areas of high reflection, one from the flat triangle in the 0^o azimuth, and one that will be turned in azimuth direction (fig. 10).

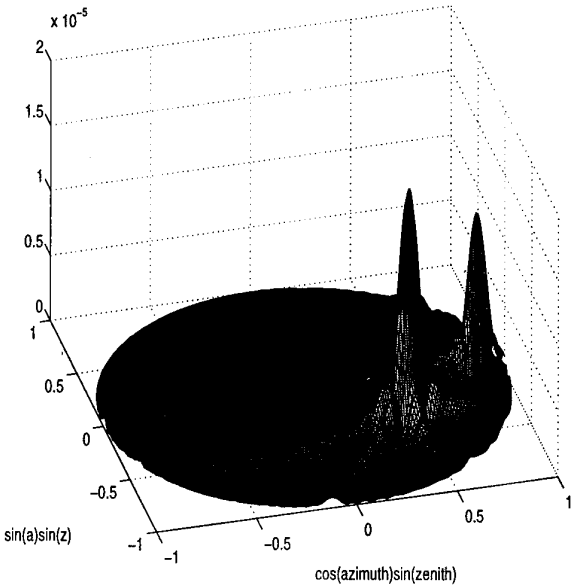


Figure 10: *Magnitude of the fields in the configuration of figure 9*

The different strength of the peaks can be explained by the fact that the the triangle that is not in the XY–plane, has a very low contribution to the 0^o azimuth direction. So the heights are a very important parameter to influence the outcome of the model. Normally these heights are given in one or the other digital terrain model. It may be possible in future also to use GIS (Geographic Information Systems) to take into account significant buildings or big squares in towns.
A third factor that directly influences the diffuse reflections, are the different parameters concerning the type of surface we are dealing with. There is the reflection coefficient that already has been studied well. So most of the different terrain types can easily be linked to some value of the reflection coefficient for horizontal and vertical polarization. These coefficients depend on the incidence angle. Another parameter that depends on the sort of the terrain that is reflecting, is the statistical distribution and the roughness of the terrain, in equation 10 expressed by σ. This means that this roughness is equal to the standard deviation for this case. Because there are not yet many publications on this subject, and mostly only backscattering is considered, this is an interesting field for further research. From the few publications ([5], [6]) that are available, you can get some experimental data about roughness for grass plains and woods.

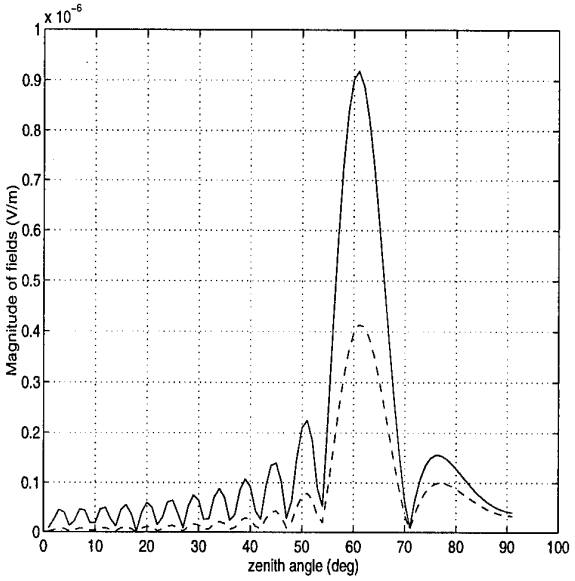


Figure 11: *Magnitude of the field with σ = 4cm (straight line) and σ = 15cm (dashed line)*

In figure 11 we show the influence of this parameter on the diffuse reflection. In the first case the roughness is 4 cm (the straight line) and in the other case 10 cm (the dashed line). Again we use a 2–dimensional plot: the rectangle lies in a plane and in the figure only the incidence plane is plotted.
The most dominant effect is a diminishing of the field with greater roughness. This is explained by the negative interference of the waves reflected by different parts of the surface, mathematically expressed by the exponential factor in equation 10. A second remarkable point is the shift towards higher zenith angles. This is caused by the factor ν_z in the exponent of equation 11.

6 Conclusions and Remarks

In this article, we have shown that the theory of diffuse reflection as described in [1] can be employed in the calculation of a coverage area. It's therefore necessary to change the shape of the area considered from a rectangle to two triangles. In that case one must be careful for a few mathematical problems. This diffuse reflection is influenced by the resolution of the digital terrain model, by the heights of the corner points and by the sort of terrain. Especially in this last fields more research has to be conducted. First of all the different values of roughness have to be established. Therefore it will be necessary to conduct some measurements. Also other probability density functions than Gaussian can be considered, because it's not certain that this function reflects optimally the most used surfaces experimentally. Finally this verified model can then be used in a program that totally computes the coverage area for indoor communications. This method is useful if the resolution of the surroundings is a few λ. It allows to reduce the very high amount of data as well as computer time needed if a total deterministic model with a resolution of 0.1 λ would be used.

Acknowledgments

I wish to thank prof. Dr. Ir. A. Van de Capelle, who made it possible for me to do this work, and my colleagues at ESAT-TELEMIC for the constructive discussions, especially over lunch.

References

[1] P. Beckmann, A. Spizzichino,"The scattering of electromagnetic waves from rough surfaces", Norwood, USA,Artech House Inc., 1987 (ISBN 0 890 06238 2), pp 9 — 99.

[2] M. F. Ibrahim and J. D. Parsons,"Signals Strength Prediction in built–up Areas Part1: Median Signal Strength", IEE Proceedings, Part F, 130, 5, pp. 377 - 384, august 1983.

[3] K. R. Schaubach and N. J. Davis IV,"Microcellular Radio–Channel Propagation Prediction.", IEEE AP Magazine, 36, 4, August 1994, pp 25 — 34.

[4] R. Grosskopf,"Prediction of Urban Propagation Loss", IEEE AP,42 ,5 , May 1994, pp 658 — 666.

[5] G. Van der Plas, E. Van de Schoor, P. Van Nieuwenhove,"Meting van de speculiere reflecties op grondoppervlakken", Revue HF, Vol. XV, nr.2, May 1986, pp 55 — 66.

[6] W. P. Lonc and J. H. Whitteker,"Over-Tree Propagation Meausurements-Phase 3",Communications Research Centre Ottawa, March 1990.

Computer Aided Frequency Planning for the Radio and TV Broadcasts

A. Altıntaş, O. Ocalı, S. Topçu, S.G. Tanyer, and H. Köymen
Bilkent University, Department of Electrical and Electronics Eng.,
Bilkent, 06533, Ankara, Turkey.

1 Summary

The frequency planning of the VHF and UHF broadcasts in Turkey is described. This planning is done with the aid of computer databases and digital terrain map. The frequency offset is applied whenever applicable to increase the channel capacity. The offset assignment is done through Simulated Annealing algorithm. The international rules and regulations concerning Turkey are also considered.

2 Introduction

The introduction of terrestrial VHF and UHF broadcast is not new. However, especially in many developing countries, this broadcast has been under the monopoly of the government. Only recently, the private broadcasting became a reality, even though in a *de facto*, and unorganized manner in many occasions. This brought up the need to have a nationwide re-planning, allocation, and management of the broadcast frequencies. The pressure from the public as well as the international agreements require that the planning has to be done in the shortest possible time with realistic and applicable goals. Therefore, an integrated approach with maximum utilization of computers and relevant databases has to be followed.

For the planning of Turkish broadcast spectrum, we have used five main computer tools and databases. The first one is a propagation simulation software which models the propagation environment and calculates the attenuation of the field strength. The second one is the geographical database of the planning region. This includes a digital terrain map of the country. The third is the demographical database. The forth is the list of all present transmitters and their electronic and geographical parameters. The last one is the international rules and agreements concerning the region. All of these tools are either acquired or produced and they have been properly integrated to have an optimum frequency allocation.

The main purpose is to maximize the interference free coverage area for each transmitter. To this end, for each transmitter propagation simulation studies are obtained. From these studies, channel assignment can be done. However, a better option is to use frequency offsets to get the maximum number of available channels. This is done using a Simulated Annealing algorithm. In [1], the Simulated Annealing algorithm is used for the efficient frequency assignment, but for a different problem. After specifying the offsets, channel assignment is done automatically by considering international agreements.

In addition to the channel assignment plan, the population in the coverage area and the services provided to any town in the country is determined. This information is used for the licencing fees to be charged to the stations and also for the classification of broadcasters into the national, regional, or local categories.

3 Tools of Planning

The necessary ingredient of a computer aided frequency planning is the prediction of field strength variation over the terrain due to a given transmitter. This requires a propagation simulation software which utilizes the topography of the terrain. Fortunately, these are available from different sources. The next step is to decide on the propagation models to be used. At present, quite a few number of different propagation models are available and the comparative advantages of these models are still under investigation [2]. The most widely known models are prepared by the CCIR of ITU and FCC of US Government. These models are based on intensive experimental measurements in the VHF and UHF bands. The terrain effect on the propagation is included in these models only statistically. The actual topografic structure may cause strong reflections and diffractions which can also be added to these models. For the coverage area studies, we have accepted and used such a model based on FCC measurements and augmented by reflection and diffraction contributions of the terrain. As is well known, CCIR attenuation curves begin at 10 km. away from the transmitter, making it unsuitable for coverage estimates. We have made some measuments in Ankara and in İstanbul, and the measured field strengths were within 10 dB of the computer generated data obtained through the adopted propagation model.

Around the transmitter, we can define an area, in which the field strength is strong enough to give a satisfactory grade of picture quality. We can call this area as "the coverage area". Similarly, "co-channel and interference areas" are defined such that in these areas either the same or the adjacent channels may

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

not be used by any other transmitter. To be able to transmit the same or adjacent channels, the relevant protection ratios between the desired and undesired (interference) field strengths have to be satisfied.

These protection ratios are listed in the ITU and EBU publications. A sample TV coverage area contour plot is given in Figure 1 for a TV transmission station in the VHF band in İstanbul. A better insight is gained through a gray scale plotting of the field strength around the transmitter. Figure 2 shows such a plot for the same station around İstanbul. Here it is easier to identify the effect of the topography on the field strength distributions. It is noted that the field strength values are calculated over radial lines centered at the transmitter location. This analysis is valid when the reflection and diffractions from the vertical discontinuities are small. Fortunately, most natural topographic boundaries are not vertical as opposed to man made buildings. So, the limitation to radial lines (2-D) is expected to work well especially outside highly populated urban areas [3].

The next step is to decide the transmitter locations, transmitter powers and antenna characteristics. In practice, this is usually done by the demand of the public. Therefore, the coverage starts from densely to sparsely populated areas. The same logic is applied here and the aim is set to cover all cities and towns over a threshold population. Also, all the information about the transmitter locations of the existing broadcasters are collected into a database, and this database is consulted whenever necessary for the proper selection of the transmitter locations. Transmitter powers are set according to the intended viewer population. Antenna pattern is set to be omnidirectional in the horizontal plane unless otherwise specified.

4 Processing of Study Files

After the specification of the transmitter, the results of electromagnetic field simulations are obtained as study files, which contain field strengths as dB $\mu V/m$, over radial lines centered at the transmitting locations. For TV planning, simulations are performed at frequencies that are chosen as representatives for each band. Angular and radial resolution of these study files vary for different transmitting cites depending on demographic requirements. Performing the electromagnetic simulation studies with the terrain information is the most time consuming part of the planning.

Observe that it is sufficient to obtain these study files for one transmission power only. For powers other than the simulated value, the field strengths are easily computed as

$$\Delta F = \Delta P \tag{1}$$

where ΔP, and ΔF stand for dB changes in the transmission power and the field strengths, respectively. The simulations are performed using omnidirectional antennas, and angular dependences of the emitted RF power from the actual planned antennas are easily incorporated similarly.

In storing these study files, field strengths at the simulation points are hardlimited between the values -128 and 127 dB, and quantized to integers, so that they can be stored in byte sized variables. The large number of transmitter cites makes such a data reduction necessary. Using these files, the field strengths at arbitrary coordinates are obtained by interpolation.

In this project about 1000 TV transmitting cites are planned, whose study files required a storage space of approximately 0.6 Gigabytes after such a reduction described above.

5 Interference Graph

The nodes on the interference graph $G(\alpha)$ correspond to transmitters (transmitting cites). The interference graph is a directional graph. Let $F_i(\vec{x})$ denote the field strength in decibels that is caused by the i^{th} transmitter at the world coordinate $\vec{x}$. An edge E_{ij} from node i to node j is introduced if transmitter i interferes with transmitter j. So, an edge E_{ij} indicates that there exist at least one significant point $\vec{x}$ in the coverage area of j^{th} transmitter such that

$$F_j(x) - F_i(x) < \alpha \tag{2}$$

where α is the protection ratio. Hence, if α is the co-channel protection ratio and in the interference graph $G(\alpha)$ there is an edge between two nodes, the same channel should not be used on these two nodes. Significance of a world coordinate $\vec{x}$ is determined by using demographic information about that point, e.g. the population around that point, the proximity of that point to a road, lake, or a military area etc.. Mountain tops are the points that interference occurs most likely. Fortunately, they usually are not demographically critical. The interference graph is generated for various protection ratios using the study files and a demographic database of Turkey.

Repeating the same channel is highly desirable for efficient utilization of the spectrum. It is possible to lower the co-channel protection ratio and reduce the number of edges in the interference graph by using frequency offset between transmitters. For practical reasons, three nonprecision offset values (8M, O, 8P) are used. If different offsets can be assigned to two

different transmitting cites, the co-channel protection ratio drops by 12 dB.

Using the study files, the interference graph is generated for two values of co-channel protection ratio α_1, α_2, corresponding to co-channel protection values with and without offsets, respectively. Naturally $\alpha_1 < \alpha_2$, and, if there exists an edge E_{ij} in $G(\alpha_1)$ the same edge exists also in $G(\alpha_2)$, but the converse does not necessarily hold. Let $\mathcal{E}$ denote the set of edges that exist in $G(\alpha_2)$ but not in $G(\alpha_1)$. After assigning offsets to transmitters, the resultant interference graph $\mathcal{G}$ is formed by adding edges $E_{ij} \in \mathcal{E}$ to $G(\alpha_1)$ under the condition that the same offset is assigned to i^{th} and j^{th} transmitters. Let $M(\mathcal{G})$ define the total number of edges in graph $\mathcal{G}$. The offsets must be assigned such that $M(\mathcal{G})$ is as small as possible. There are also some constraints in assigning offsets to transmitters. These constraints are determined by rights and restrictions that are specified by international agreements (Stockholm 1961) and by specifications of some of the transmitters that are already operational in Turkey.

The task of minimizing $M(\mathcal{G})$ under the above mentioned constraints by assigning offsets to transmitters is a combinatorial optimization problem. This problem belongs to the class of NP-complete problems and there is no known algorithm that can find the exact minimizing offset combination for 1000 transmitters in a reasonable time. Observe that the total search space consists of 3^{1000} combinations. Simulated Annealing is employed for finding acceptable solutions to this problem. There is a vast amount of literature on the method of Simulated Annealing [4], so we suffice to present its application to the offset assignment problem.

6 Offset Assignment by Using Simulated Annealing

Initially, a random offset combination which complies with the constraints is generated and $M(\mathcal{G})$ is evaluated. Then the following steps are repeated for a gradually decreasing parameter T(temperature) until it reaches a final predetermined value.

i) Choose a transmitter randomly and generate a random offset value for that transmitter that complies with the constraints.

ii) Compute the change in cost function

$$\Delta M(\mathcal{G}) = M(\mathcal{G}_{\text{new}}) - M(\mathcal{G}_{\text{old}}). \qquad (3)$$

iii) Accept the new configuration with the probability P computed as

$$P = \begin{cases} 1 & \text{if } \Delta M(\mathcal{G}) \leq 0 \\ exp(\frac{-\Delta M(\mathcal{G})}{T}) & \text{if} \Delta M(\mathcal{G}) > 0. \end{cases} \qquad (4)$$

The above algorithm generates a sequence of configurations with statistically decreasing costs.

Observe that during iterations a new configuration is immediately accepted if it decreases the cost, otherwise it is accepted with a probability that decreases exponentially by an increase in the cost. It is this nonzero probability that enables the algorithm to escape from the local minima. As T decreases, it becomes less likely for the algorithm to accept a configuration that increases the cost. The process of decreasing T over iterations is called "cooling schedule" in the Simulated anealing terminology. Decreasing the temperature too fast may give rise the algorithm to be stuck at a local minimum that is too far from the absolute minimum. The algorithm on the other hand takes a long time if the temperature is decreased too slowly.

Various methods are proposed in the literature for determining the initial value of T and the cooling schedule [5]. Simulated Annealing takes a reasonable amount of time (a few minutes compared to several hours for finding $G(\alpha)$ and weeks to complete the electromagnetic simulations in a Sparc 20). Hence, the easiest way of determining the cooling schedule and the initialvalue of T is trial and error, and monitoring the value of the cost function versus T over the iterations. The algorithm eventually reaches a statistical steady state if the iterations are performed at a fixed T. A good idea is to decrease the temperature just before the statistical steady state is reached. If the temperature is decreased sufficiently slowly then the final cost should be very close to the minimum cost that occurs until then, and different runs with the same cooling schedule should result in very close optimal values as the solution.

The final (nearly optimal) offset assignment yields the desired interference graph with minimal number of edges. The resultant interference graph is then passed to another program that assigns channels to transmitting cites.

7 Channel Assignment For Transmitters

A program has been developed for channel assignment for all transmitters in a network. This program uses three different kind of input data. One of them is the interference graph. The second type of data fed to the program is a list of forbidden chan-

nels for each transmitter. Due to the international agreements like Stockholm 1961 Plan, some channels cannot be used in the transmitters throughout the border area of a country. For this reason, the program assigns channels from a set of allowed channels to each transmitter. The third type of data used by the program is the demographic database including the population of cities, towns, and villages. By means of this database, the program assigns more channels to those transmitters whose coverage area includes bigger population.

Using these input data, firstly, the program tries to assign a minimum number of channels to every transmitter in the network. This minimum number of channel is given by the user. Then, the program starts to assign additional channels to each transmitter by increasing the number of channels one by one. At this point, if there are several alternatives, a new channel is assigned to that transmitter whose coverage area has the biggest population.

8 Generated Outputs

The frequency plan is a list of transmitter sites with geographic and technical properties. The geographical properties consist of the coordinates, the height above the sea level, and the antenna height. The technical properties consist of the transmitter power, maximum effective radiated power in all directions, assigned frequency, polarization, offset, and the vertical tilt angle. For the Turkish TV network, a total of 21 652 frequencies at 880 transmitter sites are listed.

In addition to the plan, two databases are generated for licensing purposes. In the first one, for each transmitting station a list of places down to smallest inhabited districts in the coverage area are given together with total population. For each provincial district, a list of transmitting stations and channels that can be viewed are listed in the second one. Various statistics like average number of channels per transmitting cite that can be viewed, etc. can be extracted from those databases.

9 Conclusions

A computer aided frequency planning is presented for the VHF and UHF broadcasts. A similar approach has been followed for the FM broadcasting in Turkey. The approach has utilized the Simulated Annealing algorithm for the optimization of number of broadcast channels. It is seen that, this has resulted in a very fast and efficient way of channel assignment. The approach can be generalized to similar mobile and fixed networks.

10 Acknowledgements

This work is sponsored by the Supreme Council of Radio and TV of Turkey.

References

[1] A. Quellmalz, A. Knällmann, B. Müller, "Efficient Frequency Assignment with Simulated Annealing" *IEE ICAP'95*, Eindhoven, The Netherlands, 4-7 April 1995.

[2] R. Grosskopf, "Comparison of Field-Strength Prediction Methods in UHF Range", *EBU Review-Technical*, No.229, June 1988.

[3] R. Grosskopf, "Prediction of Urban Propagation Loss", *IEEE Trans. Antennas and Propagation*, Vol.42, No.5, May 1994.

[4] E.H.L. Aarts, and P.J.M. van Laarhoven, *Simulated Annealing: Theory and Applications*, Reidel, Dordrecht, 1987.

[5] B. Hajek, "Cooling Schedules for Optimal Annealing", *Mathematics of Operations Research*, Vol.13, 1988.

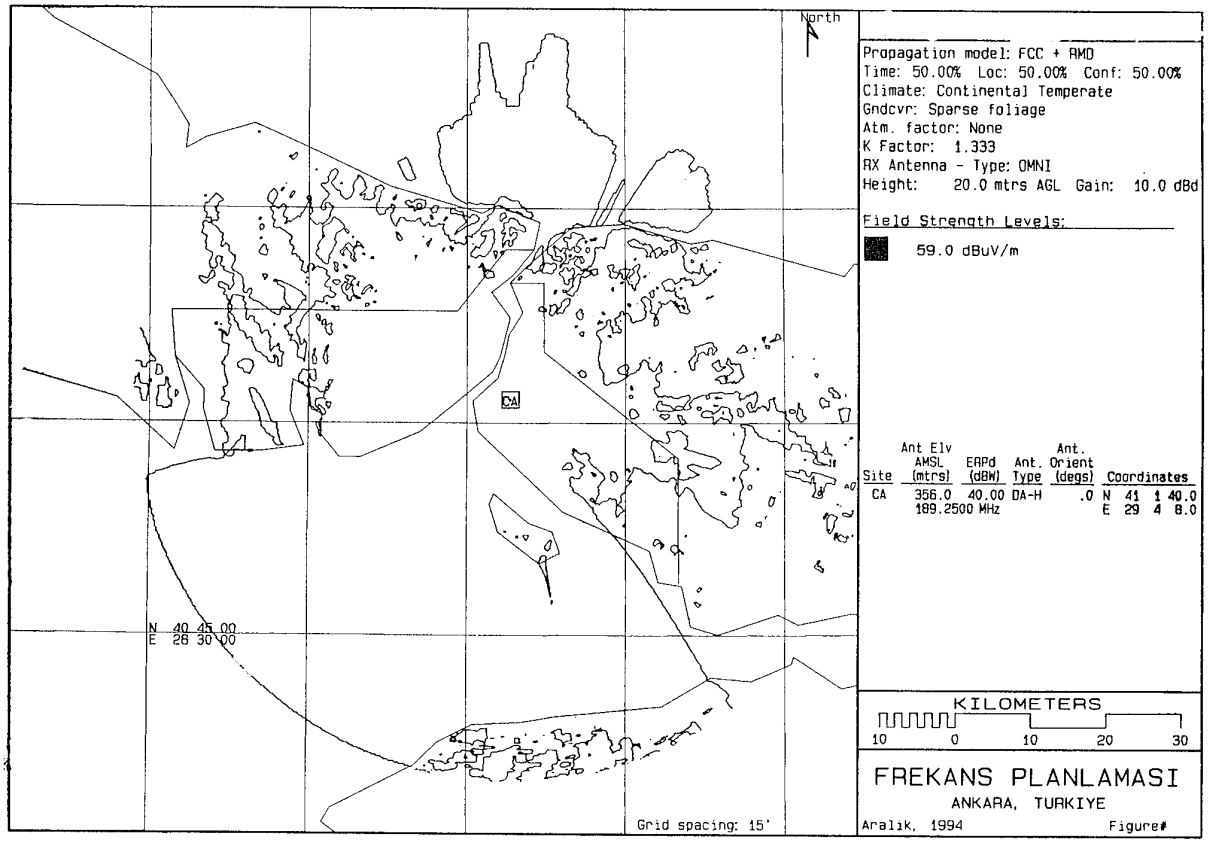


Fig.1. Coverage area contour plot of the Çamlıca, İstanbul station in the VHF.

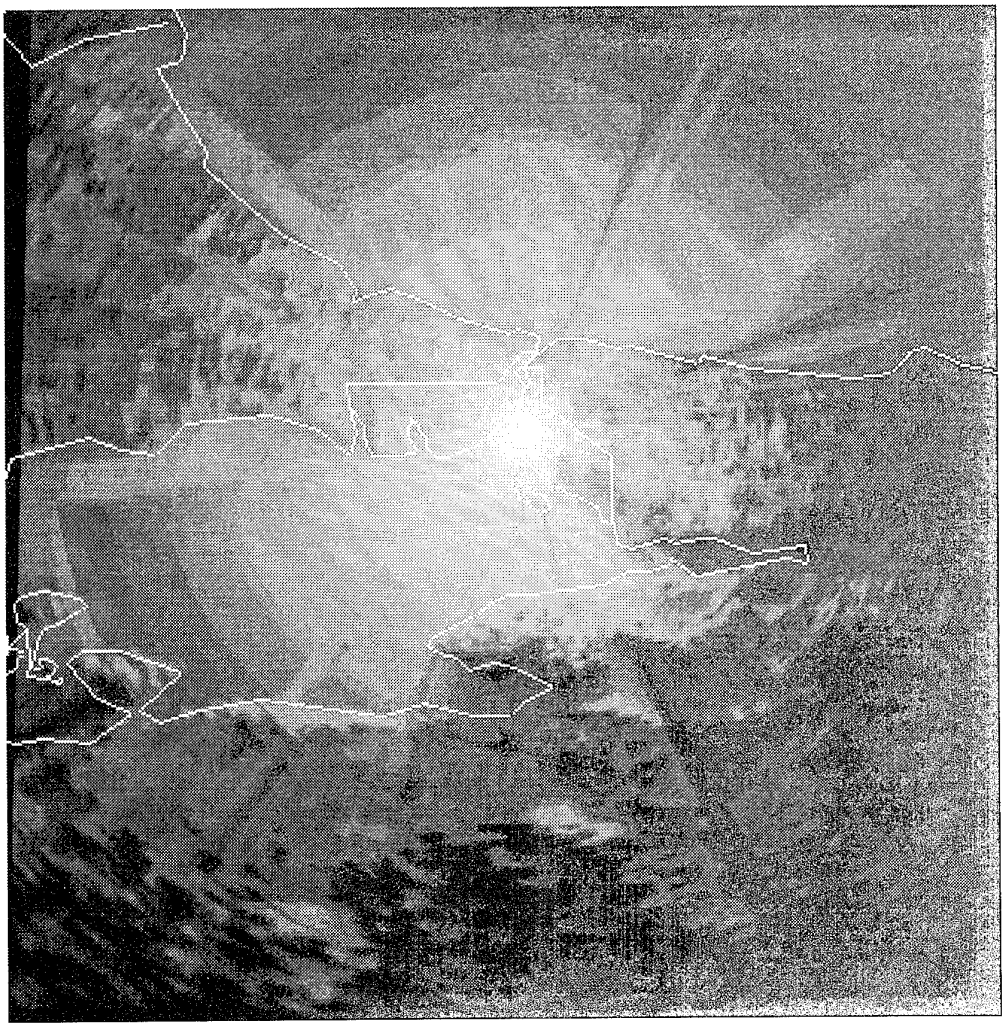


Fig.2. Gray scale plot of the field strength due to the Çamlıca, İstanbul station in the VHF.

INTRODUCTION TO RURAL AREAS TELECOMMUNICATIONS AND DEVELOPMENT OF A PILOT TERMINAL UNIT

Rodoula Makri
Michalis Gargalakos
Nikolaos K. Uzunoglu
Institute of Communications and Computer Systems
Department of Electrical and Computer Engineering
National Technical University of Athens
42 Patission Str. 106 82 Athens, Greece

SUMMARY

Rural areas present special characteristics which must be taken under consideration when designing their telecommunication networks. These characteristics and their effects on systems design are examined in detail. The design of a pilot terminal unit in a block diagram form is presented with a brief explanation of its specifications and subunits. Various applications of this system concerning rural areas are viewed. Modern applications such as telemedicine and teleteaching can be realised with the use of this system. Military use is also a promising perspective of the system. Alternative solutions that could increase systems capacity and range are also examined.

LIST OF SYMBOLS

ISDN: Integrated Services Digital Network
TDMA: Time Division Multiple Access
FDMA: Frequency Division Multiple Access
FSK: Frequency Shift Keying
IF: Intermediate Frequency
RF: Radio Frequency
QPSK: Quadrature Phafe Shift Keying
QAM: Quadrature Amplitude Modulation
ICCS: Institute of Communications and Computer Systems
CCITT: International Organization
PCM: Pulse Coded Modulation
VSWR: Voltage Standing Wave Ratio
MMIC: Monolithic Microwave Integrated Circuit
MUX: Multiplexer

1. INTRODUCTION

As it is well known rural areas which are geographically isolated but usually rich in natural resources play a significant role in the evolution of a country.This role is even greater when the country is dominated by high mountains and small isolated islands.

So, the absence of well developed telecommunications in these areas results in a low level of education and economy which increases the internal immigration towards the urban zones. The side effects of this phenomeno is the increase of unemployment and the decay in the quality of life. In the same time rural areas become deserted and underdeveloped. In conclusion, a modern telecommunication network can attract investements and encourage the local population to remain to his local area.

2. CHARACTERISTICS OF RURAL AREAS

A rural area generally consists of scattered settlements, villages and small towns which can be found either in high mountaineer regions or in small groups of islands. These areas can be characterised from the following aspects.

a) Partial or total lack of public services such as reliable electricity or water supply

b) Access and transport difficulties

c) Rarity of locally available qualified personell (technicians, doctors)

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

d) Unusual climatic conditions that constrain the requirements of equipment life and maintenance.

e) Scattered population which some times has temporary housing.

f) Underdeveloped health and educational services.

g) Primitive economy activities such as agricultute, fishing, animal raising and complete lack of idustry.

3. NEEDS AND PRIORITIES

The diversity of rural areas demands a variety of services especially adapted to the characteristics of each area. For example, there are areas in which the first step will be to provide a simple telephone service and others more developed which need modern telecommunication services.

If we take under consideration a small island there are several factors that should be taken into account. The most important thing is that the population changes dramatically each season because of the tourism. In addition, the presence of high quality tourism in some places increases the demand for more reliable services. This means that a rural telecommunication network should be designed in such a way in order to be able to deal with the periodic changes in the services demand and there must be a prediction for quality servicies such as fax, telex e.t.c.

On the other hand, in mountaineer regions the population changes again periodically and especially in holidays when people leave the big cities to visit their home villages. In these areas, some times, the telecommunication network does not exist at all and the prior need is for single telephone lines.

These two examples prove what was already mentioned in the beginning that rural areas present great differencies between them so there must be a special design for the implementation of a telecommunication network in each of them.

In general, in rural areas special study should be made for emergency services. They must be easily accesssible and free of charge for the user. These services should have dedicated channels with abreviated dialing and a smart reception system that will be able to recognize the service required.

Apart from the services mentioned previously, the telecommunication system must take into consideration the future development of the rural area. That means, that apart from telephone and telegraph other special services must be easily applied in the existing network in the future. Some of these applications are for example, telemedicine, distant teaching, image and data transfer and other ISDN applications. It is clear, that these services will enable the rural area to be developed rapidly along with the rest of the country.

In conclusion, the most important services in application priority for a general rural network development plan are:

1.Community telephony with an automatic service between communities.

2.Public rural telephony.

3.Private rural telephony.

4.Special services, data transmission, telex, fascimile.

4. TECHNOLOGY USED

Before introducing our own system it would be useful to present all the common technologies that exist and have been proposed for rural areas and examine the advantages and disadvantages of them.

a) Open wire lines and cables.

This traditional method is widely used all over the world for the linking of a remote subscriber to the nearest local exchange center. Experience has shown that this method is

sufficient and low cost when the distance involved is relatively short. When we are refferring to mountaineer regions or isolated islands distant from the shore, the cost of the installation becomes extremely high. In some cases, it is impossible to install cables because of the terrain morphology and the depth of the sea bottom. Even if modern technics permit the installation of underwater or underground cables the cost effective is not acceptable.

b) Optical fibre systems

Optical fibre systems which have been developed in recent years, have many advantages over cables because they present greater reliability, higher capacity and a variety of applications. Nevertheless when we are refferring to a small number of subscribers optical fibres are not yet a cost effective solution. In addition, the installation procedure can some times be extremely long or even impossible in case of rugged terrain.

c) Sattelite systems

These systems are used for very long-distance telecommunications (over a thousand miles). So, for small countries with dispersed population this solution is extremely expensive. However, they can be used in very special circumastances where every other possible solution can not be applied.

d) Mobile systems

These systems can be very attractive to subscribers due to the facilities they provide and the simplicity of their use. Unfortunately, the full development of such a system requires a large number of base stations. Especially in mountaineer regions, this number increases rapidly because of the propagation problems that exist. In general the use of mobile systems demands highly sophisticated and expensive infrastructures.

Apart from the technologies mentioned previously, we also have microwave radio systems. These systems can provide low medium and high capacities in a variety of ranges. They are well suited, effective and in most of the cases low cost. Modern radio systems use complex TDMA or FDMA technics for point to multi point communications.

To what it concern rural areas, the examination of their special characteristics has shown that radio systems tend to be in general the most sufficient for these cases. Based on these conclusions the ICCS developes a wireless link terminal unit. This unit is especially designed in order to meet the following conditions

a) low power consumption
b) modular design
c) easy installation
d) high reliability
e) easy and economic expansion to new subscribers.

When these conditions are met the system becomes ideal for small isolated areas with disperse population. In conclusion it is expected that such small flexible systems will make possible the overcoming of the physical and economic isolation of rural areas and their development along with the rest of the country.

5. SYSTEM DESIGN

After examining all the concepts that have been mentioned above the first step was to derive the basic specifications of the novel rural area telecommunication unit. These specifications are the following:

i) frequency of operation: 1.7-1.9 GHz (L-band)

ii) Type of modulation: FSK

iii) system rate: 2.112 Mbits/sec

iv) output power: > 2Watt

v) range: > 20 Km

vi) bit error rate: < 10exp(-3)

The block diagram of the system is shown in figure 1.
The system is full-duplex which means that both the transmitter

and the receiver are located in each side of the link. So, when we are refferring to the block diagram of figure 1 it means that an identical system is placed at the other side of the link. From now on, this block diagram will be refferred as tranceiver.

From this figure it is obvious that the system consists of the following main subunits:

a) the baseband subunit
b) the IF-RF subunit
c) the front-end subunit

Each of these main subunits will be analysed separately.

a) the baseband subunit

This subunit mainly consists of a multiplexer - demultiplexer. The multiplexer has two different input interfaces. The first one can be connected directly to a telephone line and receive either analog voice signals or digital signals from the telephone modem. The other one can be connected directly to the telephone exchange center. The difference is that the second interface can identify a call in the transmitter side and regenerated it in the receiver side.

Both of them convert the analog or digital signal to a PCM signal corresponding to a 64 Kbit/s digital bit sequence. The digital signal is multiplexed on time axis and a 2.112 Mbit/s digital tributary is derived.

The systems capacity is 30 channels. Each of them includes the original 64 Kbit/s signal and the corresponding signaling. So, the output rate of each channel is expected to be a little greater than 64 Kbit/s. Another option of the system is a co-directional interface based on the CCITT G703 Recommendation. This interface can directly receive data without the interference of a modem.

b) The IF - RF subunit

First of all, it must be made clear that the system operates in two different IF-RF bands because it is full-duplex. So, when it transmits in the first band it can simultaneously receive in the second one and vice versa in the other side of the link.

As it has already been explained in the output of the baseband subunit we have the generation of a 2.112 Mbit/s digital tributary. This sequence is inserted to the Modulator 1 (MO 1 ref.fig.1) where the baseband signal directly modulates the IF carrier. Practical MO 1 is a phase-locked Voltage Control Oscillator (VCO) with an operating frequency of 478.5 MHz for system A (one side of the link) or 370 MHz for system B (opposite side of the link). The phase locking is realised with the use of a frequency synthesizer with appropriate dual modulus prescaler with a reference frequency input of 10 MHz. The output power of MO 1 is greater than 10 dbm and the spurius and harmonics less than -60 dbc. Finally, the phase noise will be less than -90 dbc/1 KHz offset.

The FSK scheme has been chosen because it demands greater simplicity in the circuits design than other schemes (QPSK, QAM e.t.c.). In addition, the characteristic of linearity is not of such great importance because of the the constant envelope of the modulation scheme.

The next unit that is shown in fig. 1 is the RF Synthesizer. This component creates an output frequency band of 1343.5 - 1424 MHz with a frequency step of 125 KHz and a reference frequency of 10 MHz which will be the same with the one of MO 1.

The signals from the MO 1 and the RF synthesizer are fed to the RF mixer. So, the transmitted signal is up converted in the transmission frequency band which is 1822 - 1902.5 MHz for system A and 1713.5 - 1794 MHz for system B. The output of the mixer is at least 0 dbm and the lower side band and carrier supression are greater than 40 dbc.

Then the RF signal must be filtered in order to achieve the

image rejection. So, it has been designed a bandpass filter with the following characteristics:

Second order maximally flat response
center frequency 1.8 GHz
fractional bandwidth 11%
cutoff 20 db at 1.4 GHz and 2.2 GHz

This filter has been realized with coupled lines in a stripline substrate. Alternatively, instead of striplines, ceramic - ready modules can be used.
In the end of the transmitter unit the signal is being amplified by the power amplifier. This amplifier will be operating at 1.7 - 1.9 GHz with 2 Watts output and 35 db gain. This choice makes the use of preamplifiers unnecessary because the power level at the output of the filter is sufficient for driving the power amplifier.

As it has already been mentioned the range of the system will be at least 20 Km but this distance can get even greater according to the application (isolated islands). So, an alternative solution of 10 Watts output has been examined. In this case, one or two preamplifier stages will be needed to drive the power amplifier.

In the receiver stage after the front end subunit the signal is amplified by a low noise amplifier. The gain of this unit will be 25 db and the noise figure less than 1.5 db.

The received signal will be in the frequency band of 1713.5 - 1794 MHz for system A or 1822 - 1902.5 MHz for system B.

After the low noise amplifier (LNA) the signal is then filtered using a filter identical with the RF filter of the transmitter. Notice that the characteristics of the filter were chosen in order to comply with both used bands.

Next, it is possible to have one or two intermediate amplifer stages depending on the chosen range of the system. Then we meet the first down conversion. Spesifically, we have a mixer which is driven by the RF received signal and the same local oscillator signal used in the transmitter. Both the local oscillator signals used in the transmitter and the receiver are obtained by the RF frequency synthesizer with the use of a power splitter.

The output of the mixer will be at 370 MHz for system A and 478.5 MHz for system B. These frequencies will be the center frequencies for the filter following the mixer in order to rejects the image and the intermodulation products.

At this point instead of proceeding with the demodulation of the signal it is necessary to have a second down conversion at 70 MHz. This is suggested because only at these low frequencies it is easier to obtain surface acoustic wave filters which are needed for the correct demodulation of the signal. So, for the second down conversion we use a circuit identical with the MO 1 in the transmitter. The only difference is the frequency of the VCO used. So, for system A (370 MHz) 440 MHz is used and for system B (478.5 MHz) 408.5 MHz is used at 10 dbm output.

The final step is the demodulation proccess which is realised by a very simple circuit because of the scheme used. The input frequency is 70 MHz and the input power level is -10 dbm ± 1 db. The frequency response is -3 db/2 MHz and the output is ± 1 V at 75 Ohm.

The 2.112 MBit/s baseband sequence which is extracted by the modulator is fed into the demultiplexer which creates 30 channels of 64 Kbit/s.

c) front - end subunit

This subunit consists of Tx (for transmitter) and Rx (for receiver) branching filters, the diplexer (circulator) and the antenna.
The Tx nad Rx filters are used for the final filtering of the signal before the transmission and just after the reception in order to reject any noise associated with the signal. The characteristics of these filters

are extremely strict and in the first approach it was decided to use ready modules.

The diplexer is necessary because the system is full duplex and is used for isolating the receiver from the transmitter. It consists of a common circulator with the specific demand of great isolation between the ports at which the transmitter and the receiver are connected.

To what it concerns the antenna the choise of its type depended on the range of the system. For ranges less than 15 Km panel antennas would be sufficient for our applications. In our case in order to achieve a range of 20 Km or even greater with a high gain we decided to use parabolic antennas. The characteristics of these antennas were

frequency range 1.7 - 2.11 GHz
reflector diameter 1.2 m
gain 25 db
half power beamwidth 10.7 - 8.7 degrees
VSWR 1.5:1
front to back ratio 30 db

These characteristics were extremely adequate for our application and gave our system great flexibility with an operational range much higher than 20 Km.

6. APPLICATIONS

As it was already mentioned the applications of this system are plenty. Specifically, the Institute of Communications and Computer Systems which is directly connected to National Technical University of Athens in cooperation with the Polytechnic University of Tirana is developing a novel pilot terminal unit based on the previous design. This unit will be installed and operated in Albania.

A similar system will be developed with the cooperation of the State Engineering University of Armenia at Yerevan and the Georgia Technical University. Both projects are funded by the European Community in terms of programs which reinforce the technilogical cooperation between Europe and Eastern countries.
Another system based on the same block diagram design is being developed in Greece with a change in the modulation scheme. Specifically, the modulation scheme used is QPSK. This results in a change in the modulation and demodulation unit. Furthermore, the design of the filters and amplifiers has to be more strict because it has to take under consideration the demand for a better linearity in the response of these modules.

In general, it is obvious that the design of the system is extremely flexible because we can easily alter its range and its modulation. The change in range results in a greater variety of applications especially when we are considering Greece which presents a great geographical dispersion.

As it is well known, more than three hundred islands are inhabited both in Aegean and Ionian sea. These islands are found in distances which vary from few to several hundreds of miles from shore. So, with the proper design of the antennas and the increase of the output power the whole of the sea territory of Greece will be covered.

Apart from the islands, Greece in the mainland is characterised from high mountaineer regions. These regions are usually full of small isolated villages a fact which means a great population dispersion. So, the system under development will be the most suitable one for these cases.

The main idea is to develop a network whose draft architecture is shown in figure 2. Sites A, B and C represent small villages or small islands with some hundreds of residences. Site D is the exchange center of the area either in the mountains or in a bigger island. So, each of these small regions could be equiped with this system and then connected to the main communication network of the country through the local exchange center.

The design and the development of our system was based on this idea. That is, to develop a small, easy to use and to install system suitable for such cases. The systems capacity is 30 channels which is enough for small areas with few inhabitants but it can easily be increased with the use of sub multiplexer units if the demand for teleplone services becomes greater. Apart from this, the modulation scheme can also be changed in order to make better use of the bandwidth. This would result again in an increase of the capacity.In conclusion the capacity and the range are two flexible concepts of the system that can be easily altered in order to serve the specials needs of a region.

To what it concerns the special applications of the system many services can be offered apart from simple telephone services. As it has already been explained the multoplexer can be connected to a telephone modem or it can directly receive data from a PC. This means, that the system has the capability to transfer images or data packages. This feature is extremely useful for transferring medical data from small isolated emergency medical units (electrocardiogram, X-rays tomography, e.t.c.).

Additionally, the feature for data transfer can be useful for tele-consulting, tele-teaching, meteorology and tele-agriculture.

Another special application is the thansfer of radar images and data from small airforce or navy bases situated in high mountains or small islands to the army headquarters. An expansion of this idea would be to develop a network of radars which would be installed in the islands. Each of them will be equiped with this system and it will be able to transfer immediately any useful radar information. By this way the whole of the sea teritory of Greece and especially the borders would be multicovered.

This system is expected to be operational within the next few months when it is planned to be installed in Albania. When the proper use of the system has been sufficiently tested there will be an attempt of descreasing its size and cost by integrating some RF and IF stages. The final structure of the system is predicted to consist of some modules (MUX, antenna, Tx-Rx filters e.t.c.) and RF MMIC chips.

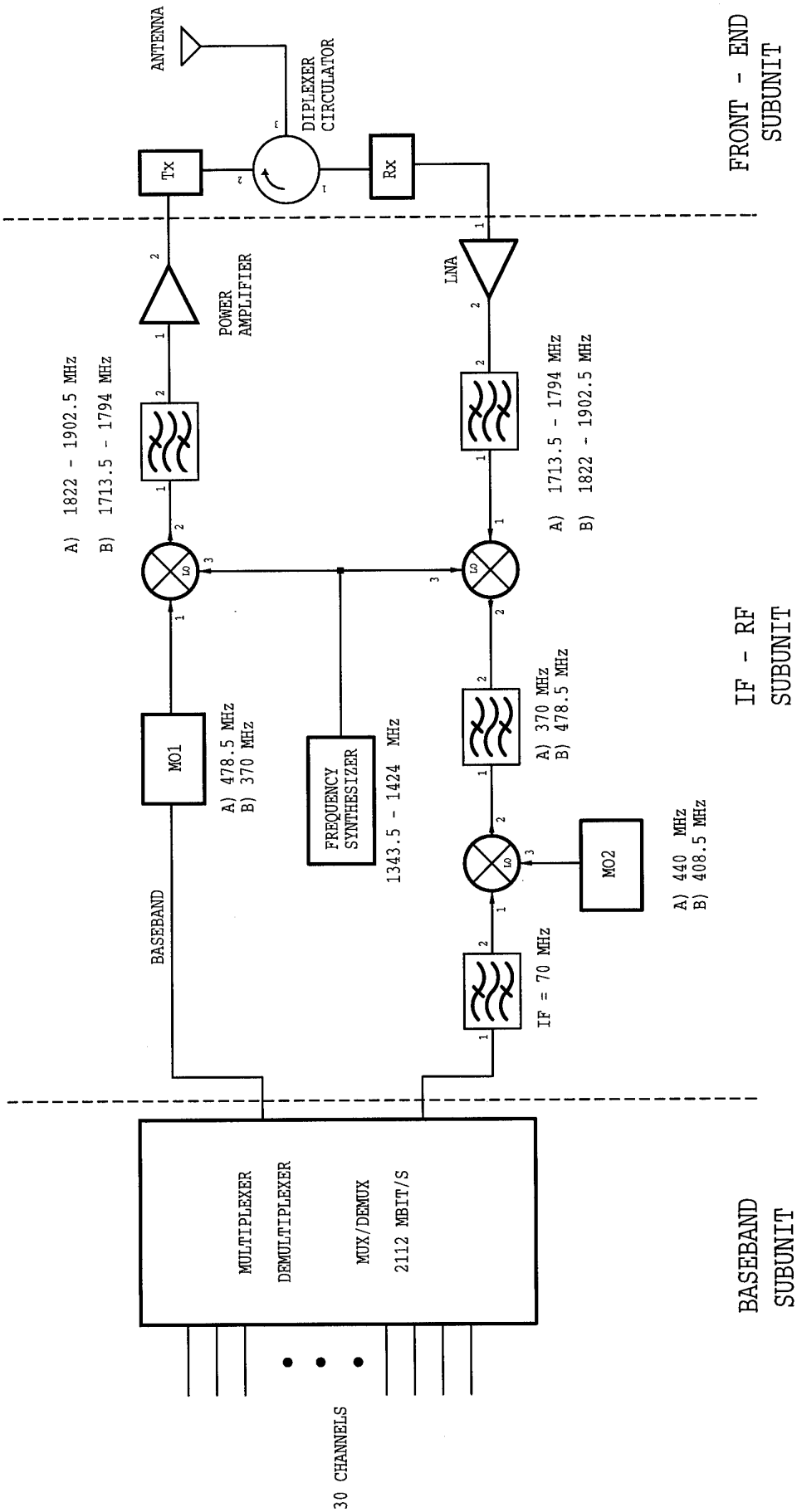

ANTENNA
DIPLEXER CIRCULATOR
Tx
Rx
FRONT - END SUBUNIT
POWER AMPLIFIER
LNA
A) 1822 - 1902.5 MHz
B) 1713.5 - 1794 MHz
A) 1713.5 - 1794 MHz
B) 1822 - 1902.5 MHz
MO1
A) 478.5 MHz
B) 370 MHz
FREQUENCY SYNTHESIZER
1343.5 - 1424 MHz
A) 370 MHz
B) 478.5 MHz
MO2
A) 440 MHz
B) 408.5 MHz
IF = 70 MHz
IF - RF SUBUNIT
BASEBAND
MULTIPLEXER
DEMULTIPLEXER
MUX/DEMUX
2112 MBIT/S
30 CHANNELS
BASEBAND SUBUNIT

FIGURE 1

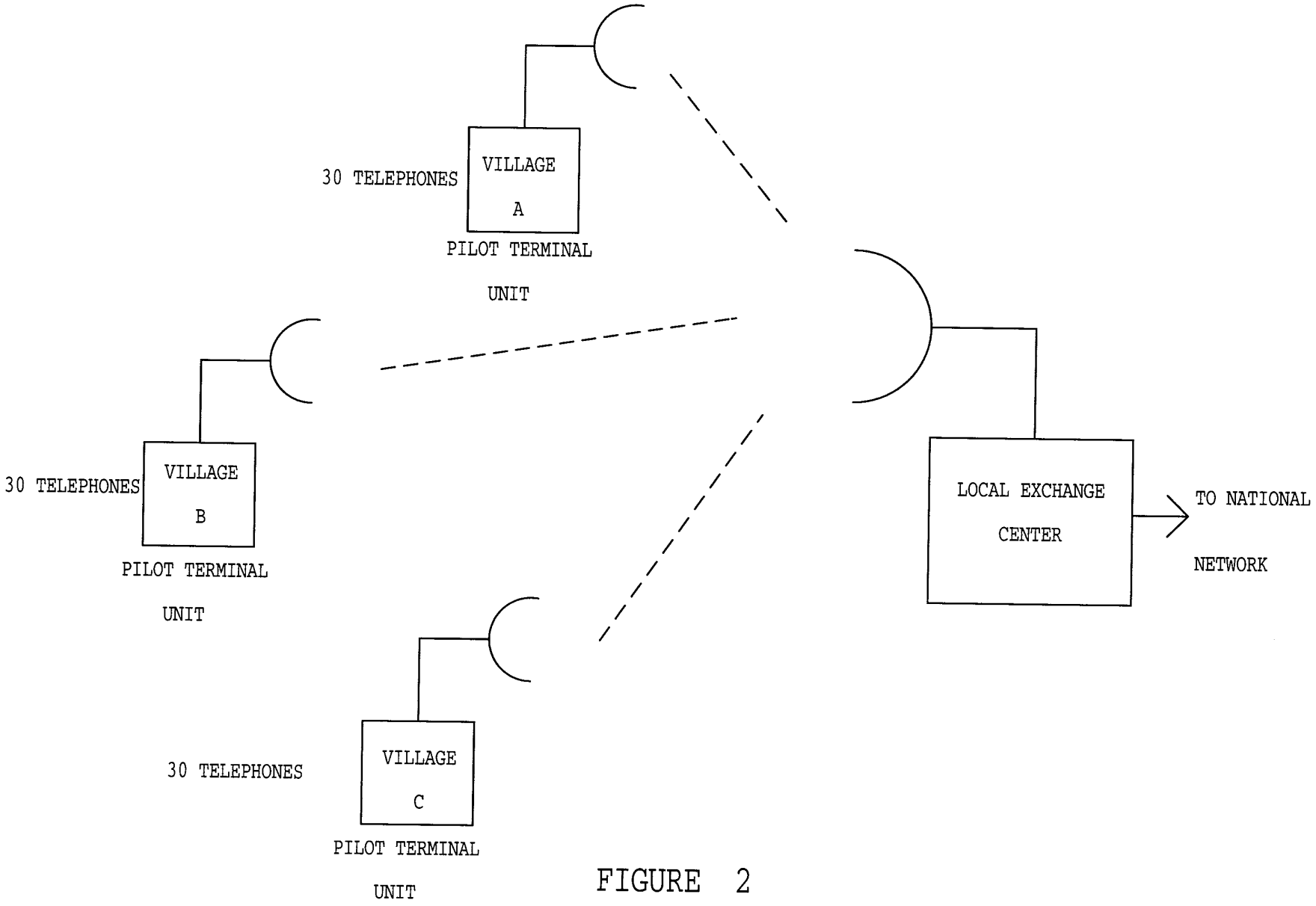

30 TELEPHONES
VILLAGE A
PILOT TERMINAL UNIT
30 TELEPHONES
VILLAGE B
PILOT TERMINAL UNIT
30 TELEPHONES
VILLAGE C
PILOT TERMINAL UNIT
LOCAL EXCHANGE CENTER
TO NATIONAL NETWORK

FIGURE 2

DISCUSSION

Discussor's name: L. Bertel

Comment/Question:

What are the range limitations of your system?

Author/Presenter's reply:

The system was designed for a 30 km range. In the USA the range is approximately 20Km.

COMMUNICATIONS AND SITUATION AWARENESS ON HOSTILE BORDERS

R. W. Kocher, Lt. Colonel, US Army
Advanced Research Projects Agency
3701 N. Fairfax Drive
Arlington, Virginia 22203

J. G. Allen, Ph.D.
Science Applications International Corporation
4001 N. Fairfax Drive, Suite 300
Arlington, Virginia 22203

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government

SUMMARY

In military operations, it is essential that commanders maintain an awareness of the location of subordinate units. The U.S. and other nations have deployed troops under United Nations' (UN) auspices to observe activity along potentially hostile borders. In these operations, there is always a danger that troops may inadvertently stray into a restricted area or cross a border. A system is needed to provide situation awareness by monitoring the position of soldiers (or the vehicles they are mounted in) and to warn when a soldier is entering a restricted area. Such warning should be provided to the soldier himself and also to higher echelons. If a higher echelon is aware of an emergency situation it may attempt to warn the soldier by independent means and also initiate the preparation of a reaction team.

The Advanced Research Projects Agency (ARPA) has undertaken to deliver and demonstrate such a system. This is accomplished by the use of a very compact package of Global Positioning Satellite (GPS) receivers and microprocessors to provide the individual with an autonomous capability. In addition, communications are included so that the position and status of soldiers can be automatically reported to higher echelons. Emergency messages, such as a call for medical evacuation, can be passed up or down the links. This paper describes the design and development of such equipment to support U.S. Army personnel deployed in the Republic of Korea (ROK) and in the Former Yugoslav Republic of Macedonia (FYROM).

I. INTRODUCTION

ARPA is developing fundamental technology as well as providing field demonstration systems which will maintain the track and status of soldiers or units. The system is based on the use of the GPS satellites to provide the geographical location of any soldier, mounted or dismounted. A small device (a monitor unit) is carried by an individual soldier or a vehicle. It measures the soldier's position and transmits the coordinates by radio to a base unit. The

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

tracks of each soldier can then be relayed to other base units.

In addition to maintaining the track of soldiers, the system also includes a status reporting capability. Each monitoring unit can transmit a set of pre-defined messages, such as a request for medical evacuation, which are automatically routed and relayed to all base units. In this way, all echelons would be warned simultaneously of any critical incidents or dangerous activity. A set of response messages are available for any base unit to respond to the requests by a monitor unit.

In this paper, the term situation awareness is used to mean the functions just described and the system is called "Soldier 911." In the United States, the telephone number 911 can be used automatically to directly access all emergency services: fire, police, rescue, and medical.

In order for the situation awareness system to perform adequately, the monitor unit must have regular, unmasked access to multiple GPS satellites and must periodically have a clear propagation path to a base unit. Terrain masking and foliage masking can both interfere with the GPS and radio link propagation paths thereby limiting the performance of the system.

Because of the potential need to extract military units from danger (i.e., individual soldiers, vehicles, pilots or aircraft), it is desirable that the system be 100 percent compatible with current international search and rescue (SAR) capabilities. The importance of this capability was recently highlighted by the dramatic rescue of the F-16 pilot, Lt. O'Grady, in Bosnia. Maintaining this compatibility is an essential requirement and has driven the design of the protocols for the deployed systems to be discussed.

Many nations have troops deployed for the purpose of monitoring activity on international borders where hostilities are possible. There are three general types of operations where troops may cross into a restricted area: foot patrols, convoys, and airborne patrols. The applications described in this paper relate to these three types of operations as they affect U.S. Army personnel deployed in the ROK and in the FYROM. Each of these areas has unique requirements which require a custom solution.

II. TECHNICAL REQUIREMENTS - ACTUAL SITUATIONS

In the FYROM, the U.S. provides a battalion sized task force to monitor a portion of the border between Serbia and the FYROM. A sketch of this situation is shown in Figure 1. The task force operates with two companies operating forward. Each company controls four to five Observation Posts (OPs) located near the border through a forward deployed Command Post (CP) and the companies are controlled by task force headquarters located near the Skopji airport. The OPs are typical of most other UN OPs set up in the Balkan region and support a squad of about nine men. Both mounted and dismounted patrols are conducted from the OPs. Reconnaissance units from the battalion task force may also conduct patrols.

For the FYROM, base stations are required at each OP, at each company CP, and at the task force headquarters. The location of a patrol would then be available at its parent OP, the controlling CP, and at headquarters. Monitor units would be carried by

individuals on foot patrol or would be employed inside of vehicles for mounted patrol.

In the ROK, the U.S. performs many UN missions to include:

- Conducting aerial surveillance of the De-Militarized Zone (DMZ)

- Providing aerial delivery of personnel and equipment to OPs located close to and along the DMZ

- Providing ground based patrols to monitor activity in the Joint Security Area (JSA).

A sketch of this situation is shown in Figure 2.

The JSA is similar to any of the OPs in FYROM and shares the same requirements as stated above. Helicopter operations present a completely different set of requirements. Of particular interest are the flights of Black Hawk UH-60 transports which fly VIPs and delivers sustainment to units operating near the DMZ. These helicopters are especially vulnerable to crossing into denied areas and they were selected for demonstration of the situation awareness system.

Most helicopter operations are controlled from the Guardian flight following facility located in Yongson, Seoul. Flights are flown at relatively low altitude and at relatively long ranges from Guardian. To implement communication with helicopters, radio relay stations are employed to ensure a nearly continuous communication over most of the Northwest portion of the ROK. To ensure complete communication, thus complete situation awareness, it would require that satellite communications be employed.

Currently, the system is being deployed for UH-60 model helicopters only; subsequent phases may expand this to other models. Initially, five helicopters will be equipped so that the on-board monitor unit can communicate with Guardian via UHF satellite communications (SATCOM). The SATCOM equipment is presently installed on UH-60s in the ROK and it will be modified to accommodate the Soldier 911 monitor unit. These five helicopters, as well as an additional five, will also be equipped so that the monitor unit can communicate with Guardian via VHF/UHF radio communications. For this latter purpose, radio transceivers will be located at the existing relay sites and remotely controlled from Guardian. In addition, an airborne radio relay system is also to be provided.

Between ROK and FYROM, one can see that there are a variety of missions with somewhat different requirements. The system to provide situation awareness must accommodate these different needs as well as maintain 100 percent compatibility with SAR systems.

In the FYROM, the OPs are generally located on high ground and there is line of sight between most OPs, the two company CPs and the Task Force HQ. Link distances are less than 40 km. Communications between OPs, CPs, and HQ are not generally a problem. On the other hand, the patrol area around many of the OPs is extremely rugged with many steep slopes and high peaks. Communications between patrols and their OP may often be limited by line of sight. In the ROK, there is a wider variation in conditions. Aerial patrols are limited to low flight altitudes - below air defense

fighter operating altitudes. The terrain is extremely rugged with very steep slopes and high peaks. Line of sight conditions are rare.

As a proof of principle, ARPA conducted a functional demonstration in the ROK this past February. Prototype equipment was installed in an OH-58 helicopter and at the aircraft's operating base. Figure 3 shows the results of a test flight. Both views depict the path of the helicopter as it flew up to a "no fly" line (about 10 km below the DMZ). The left side shows the position history of the helicopter as measured and displayed inside the helicopter (i.e., the complete track). The right side shows the situation as seen at the base station. The difference in the two views is due to communication line of sight conditions, or better stated, lack of line of sight.

It is apparent that the situation in the ROK necessitates a satellite communications solution. The FYROM on the other hand is generally supportable by use of terrestrial communications.

A final set of requirements result from the need to maintain compatibility with current SAR equipment. Air crews carry a SAR radio which enables them to transmit a distress beacon signal at standard international frequencies, and communicate by voice with SAR teams via radio in the 225 to 300 MHz band. In addition, some SAR radios can also generate special signals to allow SAR aircraft to use distance measuring equipment to fly quickly to the downed air crew.

III. BUILDING BLOCKS

ARPA has invested in advancing the technology for miniaturizing GPS receiver and processing hardware for military applications. This has resulted in a chip set, manufactured by Motorola, which allows all of the GPS position location functions to be performed in a small amount of real estate - about 25 mm by 25 mm. Motorola also manufactures a line of SAR radios, type designated the PRC-112. (This in fact was the equipment used by Lt. O'Grady.)

One shortfall in existing SAR radios is that anyone (such as an adversary) may intercept and exploit the distress beacon or voice communications. Motorola has invested in eliminating part of this problem and improving the functions of the SAR system by embedding a GPS capability into their PRC-112 radio, resulting in an integrated unit designated the GPS-112 or the Air Force HOOK-112. The device is shown in Figure 4 with modifications made for the ARPA program. Motorola also has developed an interrogation system which allows a SAR team to directly interrogate a GPS-112/HOOK-112 unit and receive the unit's actual geographic location. Rescues can then take place with a minimum amount of communication. In addition, communication privacy codes are available to prevent anyone from exploiting the data communications. The interrogation unit consists of a VHF radio (e.g., a Motorola URC-200), a special Motorola proprietary smart modem/interface, called a Six Gun™, and a PC computer.

The Integrated Systems Research Corporation (ISRC) produces a line of position tracking and display systems which fuses GPS type position reports and displays tracks on computer generated maps. ISRC has participated in previous ARPA location tracking programs and has software available that is suitable for the program described here which operates on a PC computer.

Motorola and ISRC were the two contractors selected to participate in the ARPA program and are providing all of the equipment necessary to implement the concept and install the system in both the FYROM and the ROK.

IV. THE EVOLVING SYSTEM ARCHITECTURE

To summarize, the Soldier 911 system consists of monitor units (based on the GPS-112/HOOK-112 radio/GPS set) and Base Stations comprised of a VHF radio, the Motorola Six Gun™, a PC computer and tracking/display software provided by ISRC. Different configurations (for ROK and FYROM) are implemented by the communication protocols. A block diagram of the system for ROK is shown in Figure 5.

The protocols for the system were developed in stages and will continue to evolve as other deployment requirements are added. The initial deployment was oriented on the situation in the FYROM; it was desirable to deploy a basic capability as soon as possible. Later, the need for a capability in the ROK prompted the re-examination of the protocol requirements and the broadening of the capabilities. Consequently, a more efficient protocol was developed and applied to both FYROM and ROK.

For the initial protocol, a simple asynchronous scheme was employed where each Base Station simply interrogated the monitor units under its control, at a prescribed but variable rate. Because of frequency allocation restrictions, it was desirable to minimize the number of frequencies needed. Thus, the same frequency was used by all base stations operating at OPs and CPs.

With the expanded requirements for ROK, the protocol had to be designed to accommodate mixed communication modes: UHF satellite and VHF terrestrial. Again, because of frequency allocation restrictions, a single frequency is employed for all terrestrial communications and, of course, a separate frequency is needed for SATCOM. Figure 6 describes the timing basis for the protocol used for VHF terrestrial communications.

The general requirements for the protocol are as follows. The system can handle up to 48 monitor units which report at least every 60 seconds. Any monitor unit should be able to transmit an emergency "911" call, and a Base Station should receive the call within 10 seconds; acknowledgment of a 911 call should be provided within 30 seconds.

The timing of the protocol begins with the same sequence of data blocks for each radio relay (three terrestrial: RR 1, RR 2, RR 3, and an airborne relay - RR Air). These blocks contain differential GPS correction data relative to the respective relay point, space for a canned message to be transmitted through the respective relay, and space for a script message. Emergency 911 calls, for example, are acknowledged via canned messages in the relay data blocks. As shown in Figure 6, a 30 second half cycle is used to ensure that the differential correction is accurate to 10 meters and that the 911 acknowledgment requirement can be satisfied. In each half frame, there are 24 time slots (labeled 1, 2, etc. in Figure 6) used by monitor units to report their position reports (time and 3-D position). Special slots are provided as follows. The Remote slot is used for scripted messages sent from monitor units to the Base Station and the log-in slot is used by a monitor unit to inform the Base Station when it enters the

network. Collisions in these time slots are possible, but are expected to be extremely rare because of pre-scheduling.

If less than 24 monitor units are active, the second half frame can be used the same as the first and the overall timing then equates to a period of only 30 seconds.

V. STATUS AND PLANS

The system is operational in the FYROM and will be delivered and installed in the ROK in early October of this year. It is expected that continuous feedback will be received from the users and periodic system improvements will be made. In addition, ARPA will pursue the development of the enabling technologies for Soldier 911 systems of the future.

ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of various individuals as well as Motorola and Integrated System Research Corporation. Dr. Larry E. Stotts of ARPA provided the motivation and direction necessary for establishing the basic program. Dr. Sherman Karp, a private consultant, was responsible for defining the communication protocols which form the basis for the system. Lastly, Ms. Jeni Porter and Mr. David Stenger are solely responsible for the editorial and artistic content of this paper.

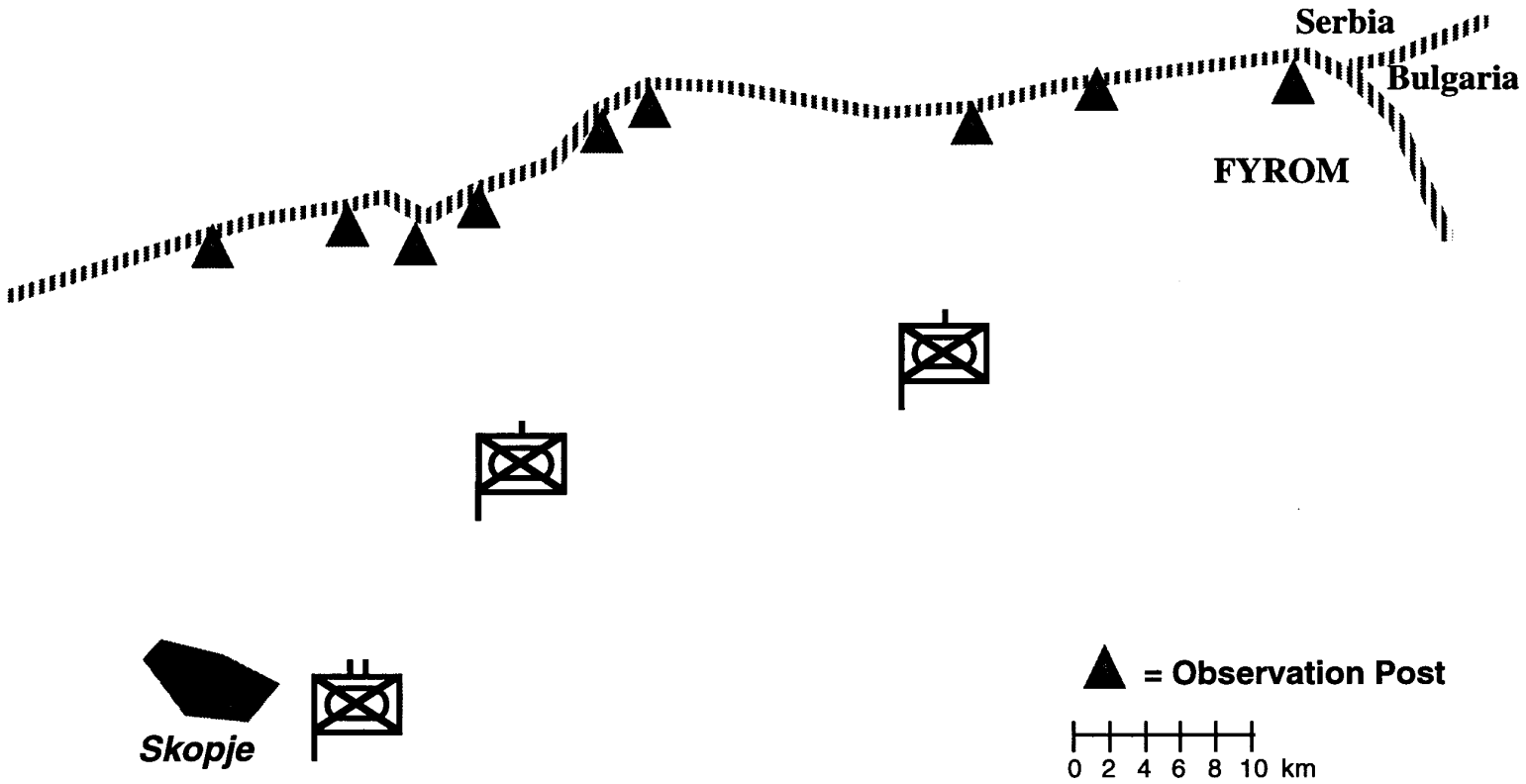


Figure 1 - FYROM Situation

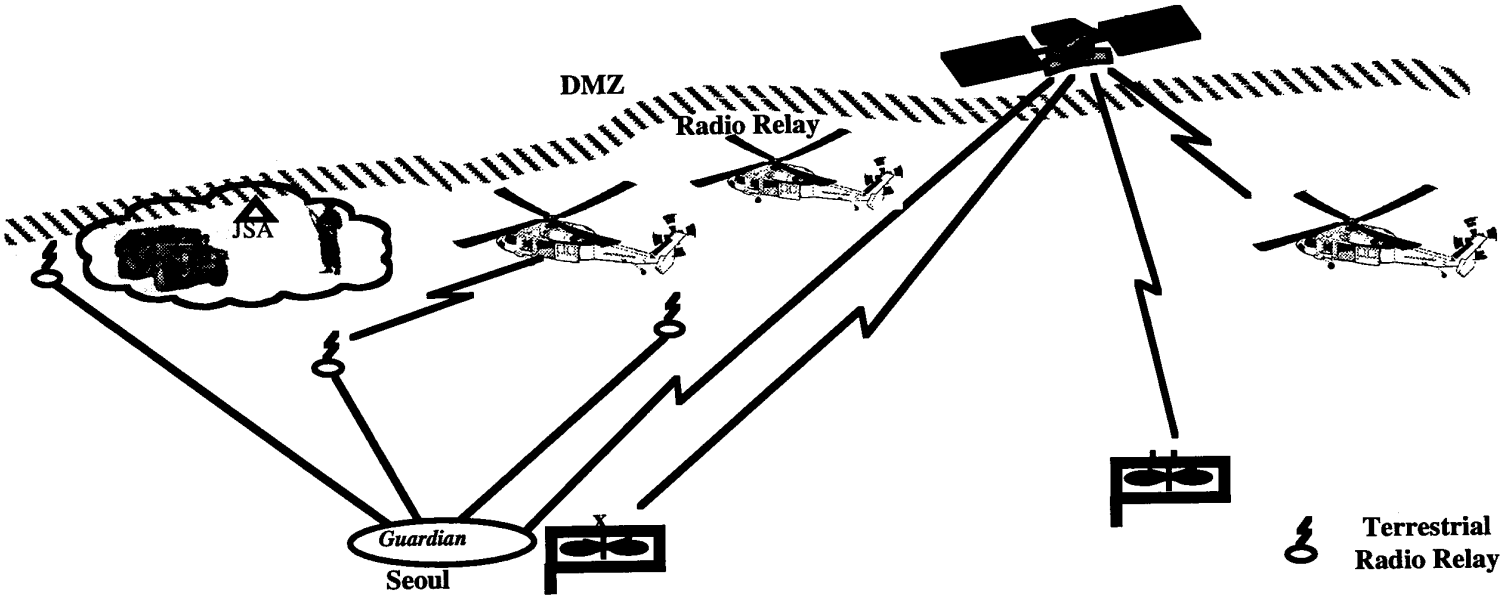


Figure 2 - ROK Situation

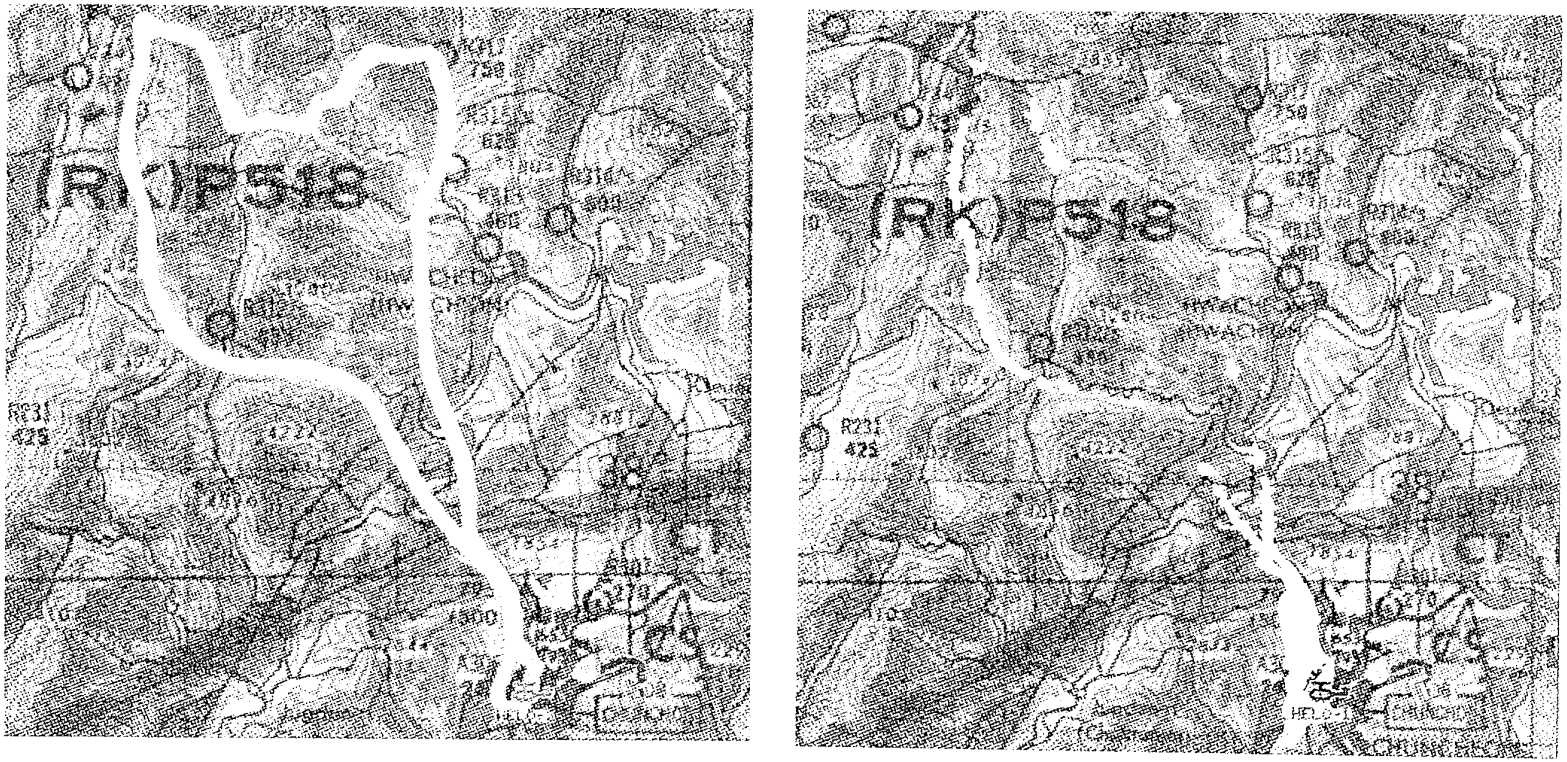


Figure 3 - Demonstration Flight in ROK

Figure 4 - Soldier 911 Monitor Unit

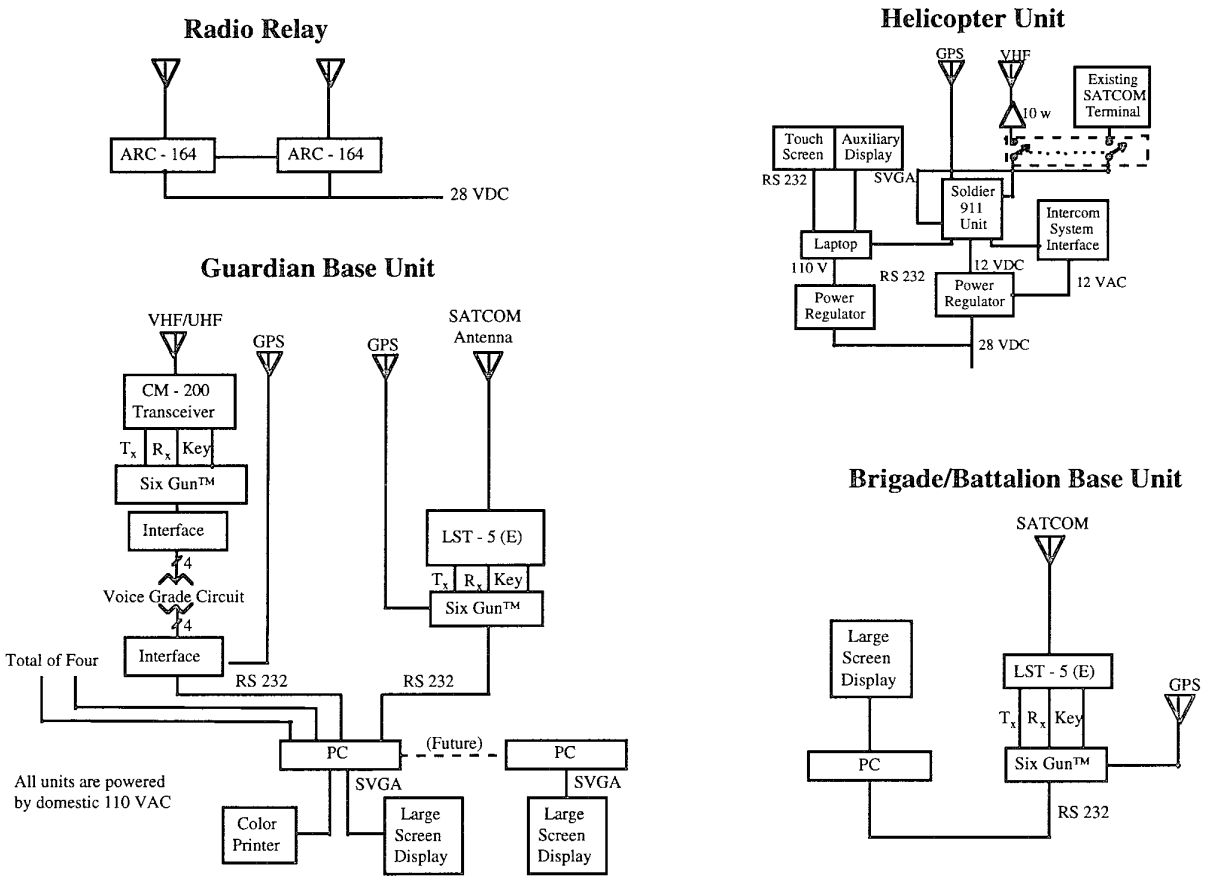


Figure 5 - System Components for ROK

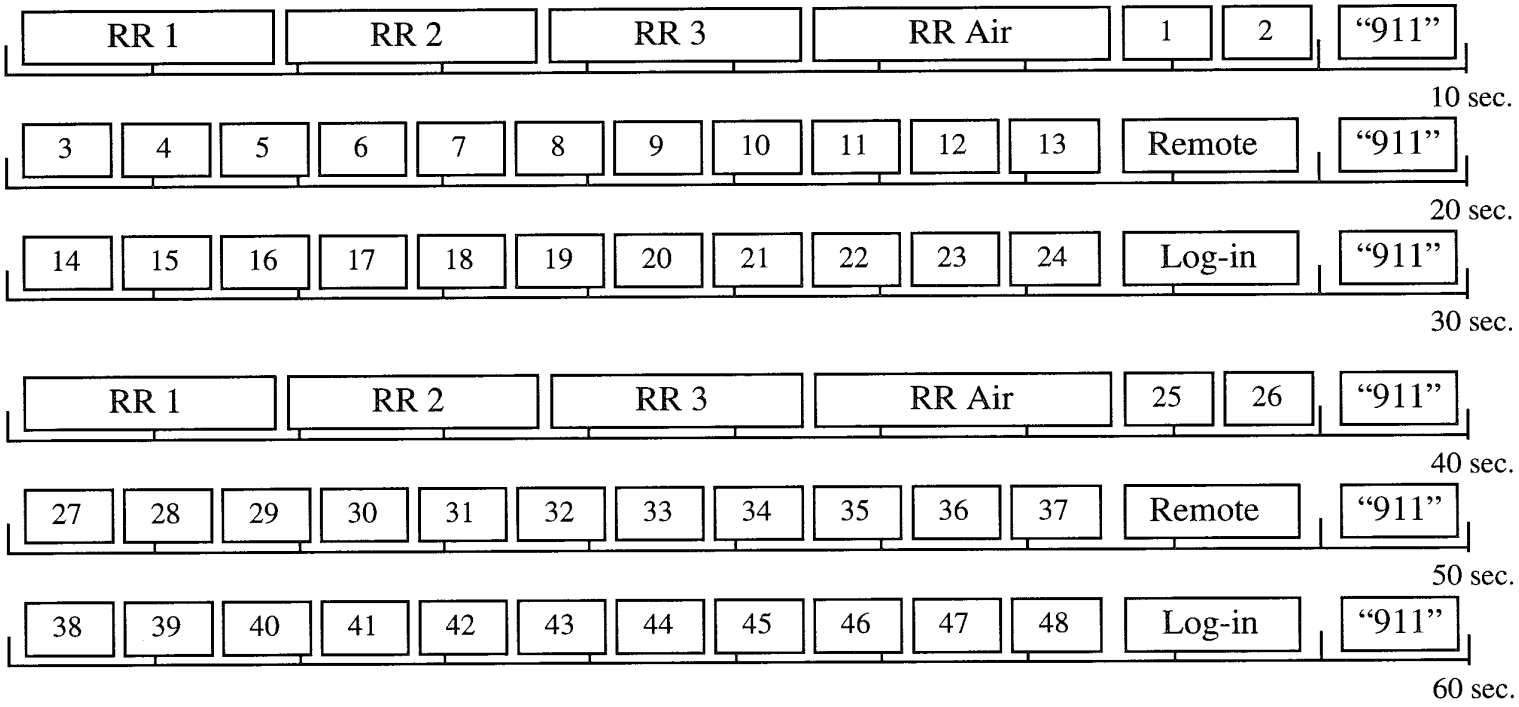


Figure 6 - System Components for ROK

COMMUNICATIONS REALISM WITH INTEGRATED TERRAIN-ENVIRONMENT-MULTIPATH MODEL

Chandrahant Sheth
Space and Terrestrial Communications Directorate
U.S. Army Communications-Electronics Command
Directorate, AMSEL-RD-ST-CE-M
Fort Monmouth, New Jersey 07703, USA

Salvatore Barone
Command and Control Systems Integration Directorate
U.S. Army Communications-Electronics Command
Fort Monmouth, New Jersey 07703, USA

Mark R. Lambert
Maden Tech Consulting, Inc.
Shrewsbury, New Jersey

Lonnie R. Kennedy
Telos Corporation
Shrewsbury, New Jersey

Abstract

Modeling and simulation (M&S) that assists the system developer or warfighter in defining, prototyping, testing and training should be executed with a high level of fidelity—this includes communications simulations. Current virtual, constructive, and systems performance simulations do not provide the dynamic irregularities at the lower link/physical layers that stress on-the-move (OTM) communications, either assuming perfect communications or including statistical averages of error effects. The U.S. Army communications-Electronics Command Research, Development and Engineering Center (CECOM RDEC) has been developing models employing "communications realism." Communications realism is a simulation feature through which a synthetic environment reflects the same communications problems that exist in a real environment. The purpose is to provide realistic tools for developing on-the-move (OTM) communications systems that ensure reliable connectivity on the digitized battlefield.

The Combined Arms Command and Control (CAC2) Advanced Technology demonstration (ATD) is a key program for supplying digital capabilities for the United States Army's concept of the digitized battlefield. The CAC2 system performance model (SPM) was developed under the CAC2 ATD as an economical means of evaluating proposed system designs and modifications prior to costly field demonstrations and tests.

The Integrated Terrain-Environment-Multipath Model (ITEMM) provides the basis for the Communications-Realism Submodel (CRS) of the CAC2 SPM. ITEMM provides communications parameters for real-time physical and link-layer models to implement communications realism. ITEMM dynamically simulates the communications channel problems that exist in real environment. These problems include signal fades/error bursts, signal propagation interference due to terrain shadowing and destructive multipath effects, moving platforms, and other factors affecting received signal strength as a function of time. True communications impacts can only be obtained through dynamic simulation of the environment rather than statistical averages of stationary communications link errors.

Introduction

CAC2 is the leading ATD in the effort to support digitizing the battlefield. Its objective is to develop an operational and technical approach that will provide effective and efficient information distribution across the battlefield at echelons brigade and below. In order to achieve this goal, M&S will play a substantial role in the development of an architecture, its refinement, and its evolution in terms of both technical performance and operational effectiveness. Because this effort includes hundreds of platforms at and below the brigade level, M&S is the only possible way to conduct economical evaluations of proposed system designs and

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

modifications prior to costly field demonstrations and tests. Integrating performance simulation with realistic scenarios allows designers and users to quantify aspects of these systems that could not previously be envisioned because of the system's structural complexity.

ITEMM prototype software, originally developed as a generic radio model, led to the development of the CRS which has been integrated with the CAC2 SPM to provide a realistic means of simulating the physical propagation environment on the tactical battlefield. ITEMM dynamically simulates the communications channel problems that exist in real environments. These problems include signal propagation interference due to terrain shadowing and destructive multipath effects, moving platforms, and other factors affecting received signal strength (RSS) as a function of time.

This paper provides an overview of the CAC2 SPM, the communications environment on the tactical battlefield, the modeling approach used by CECOM in support of the digitized battlefield, and ITEMM.

Combined Arms Command and Control Systems Performance Model

The objective of the CAC2 ATD is to automate and digitally integrate all friendly forces on the battlefield by providing an information based digital system architecture at brigade and below to support horizontal and vertical integration capabilities that improve shared situation awareness, battlefield synchronization, and near-real-time target handover capabilities.

The CAC2 SPM is used to determine how proposed C3 systems architectures impact communications systems, and how the communications systems will impact the CAC2 system.

The CAC2 SPM utilizes TRADOC-generated force laydown information for the Task Force XXI Advanced Warfighting Experiment (AWE) to simulate brigade and below Operational Facilities (OPFACS) (Figure 1). The high-resolution CAC2 SPM consists of modular submodels that interact at run-time to simulate the overall CAC2 system (Figure 2).

The Command and Control (C2) submodel initializes the simulation with set-up data on OPFAC deployment, net structures, message descriptions, mission threads, and scenario data. The mission thread data was developed by subject-matter experts (SMEs), representing all battlefield functional areas, during an Operational Requirements Analysis (ORA). The ORA is a process through which SMEs, using a scripted scenario, develop information exchange requirements. A sample call-for-fire mission thread is shown in Figure 3.

The Combat Net Radio (CNR) protocol model simulates a variety of CNR protocols including TACFIRE, AFATDS, IVIS, MIL-STD-188-220, and MIL-STD-188-220A. The SINCGARS ICOM submodel emulates the Single-Channel Ground and Airborne Radio System (SINCGARS) Integrated COMSEC (ICOM) Received Transmitter (R/T). The SINCGARS SIP Submodel emulates the SINCGARS System Improvement Program (SIP) R/T.

Other Army legacy communications systems represented in the CAC2 SPM include the Mobile Subscriber Equipment (MSE) Tactical Packet Network (TPN) and the Enhanced Position Location Reporting System (EPLRS) R/T. Connectivity between the various communications submodels is provided by the Tactical Multi-Net Gateway (TMG) Submodel which provides a gateway between the SINCGARS, EPLRS and MSE Submodels, and the Internetwork Controller (INC) Submodel which provides a gateway between the SINCGARS SIP and EPLRS Submodels.

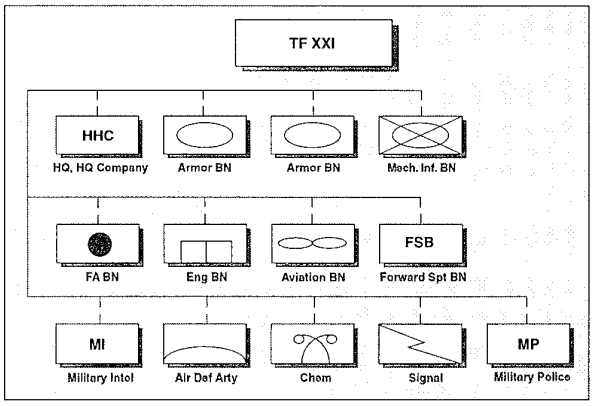


Figure 1. Task Force XXI Organization

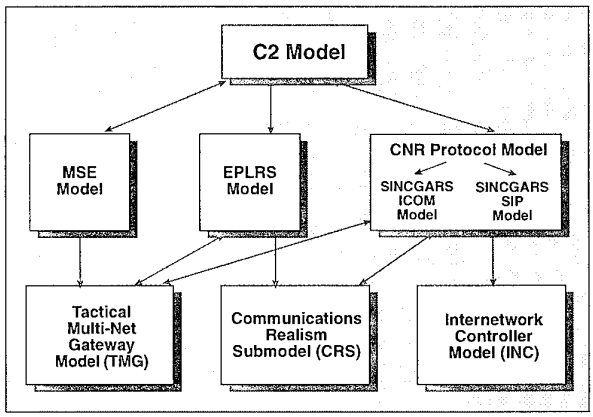


Figure 2. CAC2 SPM Architecture

The CRS, based on ITEMM (see Figure 4), provides realistic communications efforts including dynamic terrain loss and multipath effects using actual terrain data (see Figure 4).

The CAC2 SPM has been used to conduct comparative analyses of the performance of existing transmission protocols-with and without voice traffic. The model collects output statistics to evaluate the performance of the networks (utilization, throughput, collisions), OPFACs (messages received/transmitted, queue size/build up, retries), and mission threads (average delay and delays for individual mission threads).

The CAC2 SPM is used as a systems engineering tool to develop and evaluate the systems and technical architectures

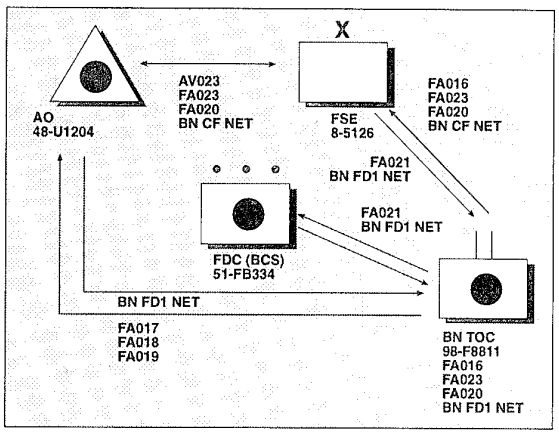


Figure 3. Call-For-Fire Mission Thread

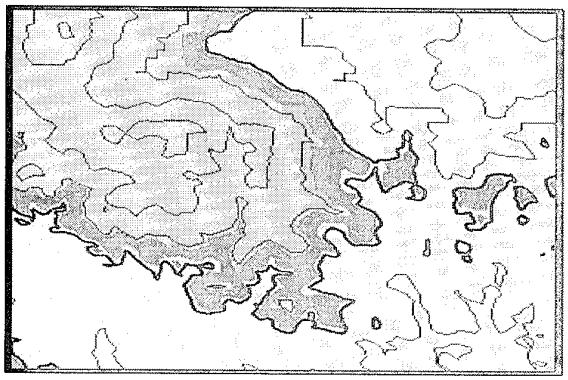


Figure 4. CAC2 SPM CRS Uses Actual Terrain Data to Provide Communications Realism

that support battlefield digitization for AWEs and Task Force XXI. The systems architecture is the physical connectivity of an information system. The technical architecture specifies how those systems will be implemented, e.g., protocols.

The CAC2 SPM allows variations in the interrelationship of the operational architecture (force structure), with the technical architecture to achieve the optimum systems architecture for the CAC2 system.

The Tactical Environment

A critical issue addressed by the CAC2 SPM is how to represent the communications environment found on the tactical battlefield. Communications problems on a digitized battlefield are often attributable to radio sources. As forces move across the battlefield, the quality of radio connectivity between them varies dramatically and continuously as a function of terrain, propagation, climate, radio-link capabilities, and other factors. Platform movement continuously varies the RSS because the receiver is always exposed to multiple replicas of the signal that interact destructively (Figure 5). These replicas are signal returns from structures, hills, foliage, and atmospheric layers.

The time rate of the RSS variations increases with vehicle speed. There can also be RSS variations when vehicles are stationary because phase balances of the interacting rays change as other vehicles move, as foliage moves in the wind, and, on longer paths, because atmospheric layers are dynamic. Typical fading of a 900-Mhz signal received by a mobile unit traveling at 15 mph is shown in Figure 6.

These communications signal impairments cause time-varying error bursts in the digital bit stream. As shown in Figure 7, every time the received signal strength falls below a radio system's performance threshold, the digital bit stream will have an unacceptable burst of errors of varying duration and occurrence time. Modeling of variations in radio connection quality is accomplished by error burst/multipath effects algorithms and look-up tables consisting of such factors as variable obstructions, climate, thermal noise, and foliage.

Modeling Approach

Communications on a digitized battlefield are principally data communications. The Army has developed a data communications standard, MIL-STD-188-220A, which has been selected as the objective protocol for digitizing the battlefield. MIL-STD-188-220A will be used in conjunction with commercial Internet protocols to provide seamless connectivity on the digital battlefield. The communications functions of MIL-STD-188-220A fall within five of the seven layers of the open system interconnect (OSI) model (Figure 8).

In addition to detailed representations of the network and transport layers, the CAC2 SPM includes detailed models of the physical and link layers which are used to assess the impact of the physical environment and link-layer countermeasures on the higher layer protocols. This is in contrast to previous network simulations which focused on measuring throughput and grade of service (GOS) as a function of network- and transport-layer protocols. Inclusion of error bursts increase the fidelity of the models and allow a more complete evaluation of communications system performance.

Planned enhancements to the physical-layer models include modifying them to apply errors from an error-burst model (rather than internally generated random errors) to the transmission frame. This will produce a more accurate representation of synchronization and bit errors in the various frame fields.

The performance of the physical-layer model has a direct effect on the behavior of the link layer. If the burst errors in a given scenario occur frequently enough to cause significant changes in message-loss characteristics, values selected for analysis in the model, such as retransmission parameters, may need optimization.

Time-out delays and maximum number of retransmissions in a network should ideally be functions of both the volume of traffic and the frequency of transmission errors. For example, increased traffic volume will result in increased delays in both message transmissions and in the time taken for acknowledgments to be received. Time-out values for retransmission may therefore have to be increased to prevent unnecessary retransmissions. If the dominant source of delays in a network is the retransmission of lost messages and acknowledgments on error-prone links, however, the implication is different. In that case, it may be advisable to decrease the time-out values and increase the maximum number of retransmissions, thereby increasing the number of opportunities to successfully transmit the information in a

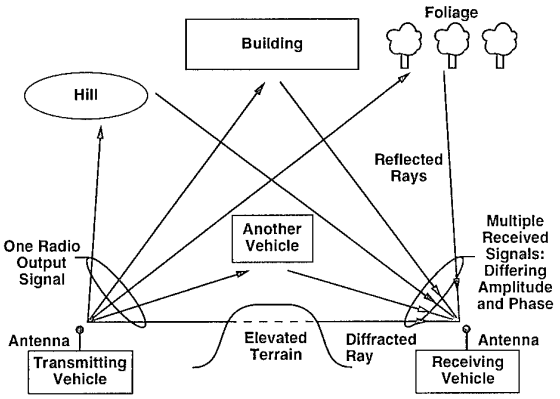


Figure 5. Multiple Received Components for One Transmitted Signal

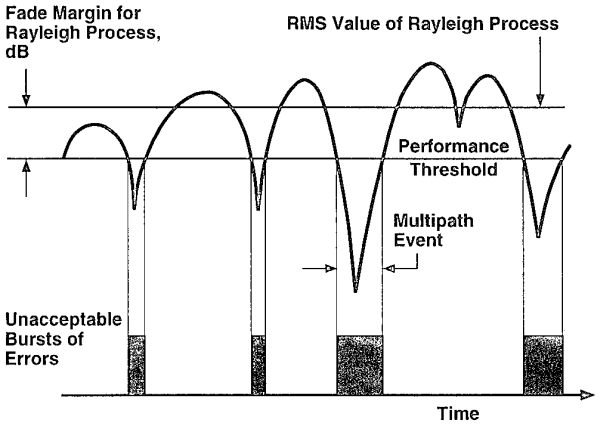


Figure 7. Error-Burst Variability

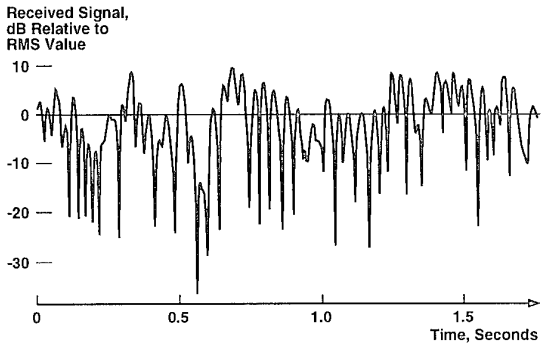


Figure 6. On-the-Move Signal Variation

given time period.

Analysis of a variety of scenarios should be performed to optimize values, such as retransmission parameters, impacted by the incorporation of burst-error effects. For example, values can be obtained for these parameters with respect to low, medium, and high values of transmission errors (or traffic volume). These values could then be dynamically selected for individual networks in a simulation as conditions change, resulting in improved system performance.

Above the data-link layer, end-to-end message accountability is the primary concern (other than the increased delays associated with message retransmissions) regarding messages lost on a link. Communications-channel modeling including all of the needed OTM communications dynamics is essential for successful end-to-end network performance.

Integrated Terrain Environment Multipath Model

Representation of the tactical environment requires deterministic and stochastic models. Simulation of moving platforms dictates determinism regarding the terrain between moving platforms. If the platforms move in such a way that a platform comes between them, the hill must be described accurately and deterministically. Multipath effects from a stochastic model are added to the deterministic effects. Sufficient computing power is not available to deterministically describe the effects of every tree, vehicle, building, cliff, or hill in the neighborhood of the two platforms and in the area of the path between them. One of the advances of the modeling described in this paper is in the integration of the deterministic and stochastic effects. ITEMM uses a mix of deterministic and stochastic models to produce realistic communications effects to determine data transport reliability.

ITEMM is a generic single-link radio model, adapted for use in the CAC2 SPM, that provides detailed communications realism by simulating communications environments and problems that exist in the field. These include signal fades from destructive multipath effects, error bursts, signal propagation blockage due to terrain shadowing, moving platforms, and other factors affecting received signal strength. The model is capable of determining real-time OTM communications effects that degrade system performance. As shown in Figure 9, ITEMM provides dynamic terrain loss calculations for the CAC2 SPM. During the simulation, as the platforms move across the terrain, ITEMM calculates terrain loss on a message-by-message basis and returns a value for received power to the radio submodels. ITEMM is designed for plug-and play simulation, and is capable of providing propagation effects for a variety of waveforms. ITEMM allows the user to define physical-and link-layer parameters (including radio attributes and countermeasures) of many different radios and select trajectories and speed of communications platforms over digitized maps. Outputs on link performance, transmission delays, and message reception/retransmission are provided to other CAC2 SPM submodels while the simulation is running. ITEMM uses the validated Terrain-Integrated Rough-Earth Model (TIREM) model for terrain loss and its multipath models are based on experimentally validated Rayleigh fading performance models for mobile radio systems [1,2].

ITEMM was originally developed as a stand-along proof of concept application. The CRS was developed re-using much of the ITEMM code. Currently the CRS is being used exclusively within the CAC2 SPM to provide dynamic terrain loss. As efforts to validate the model using live test data continue, it is anticipated that ITEMM will also be used to supply multipath and other communications effects as part of the simulation.

Conclusions

Development of digital communications systems for Force XXI requires the use of M&S as a cost-effective expedient for performing network architecture analysis and design. The CAC2 SPM is an effective systems engineering tool to develop and evaluate the systems and technical architectures that support Force XXI. The capability of the CAC2 SPM to faithfully represent the physical and link layers of the protocol stack and their impact on the higher layers provides accurate determination of system performance. The success of the highly mobile forces of Force XXI will depend on the capability to communicate OTM. The CAC2 SPM will support the development of these capabilities by accurately representing the physical propagation environment and identifying appropriate countermeasures.

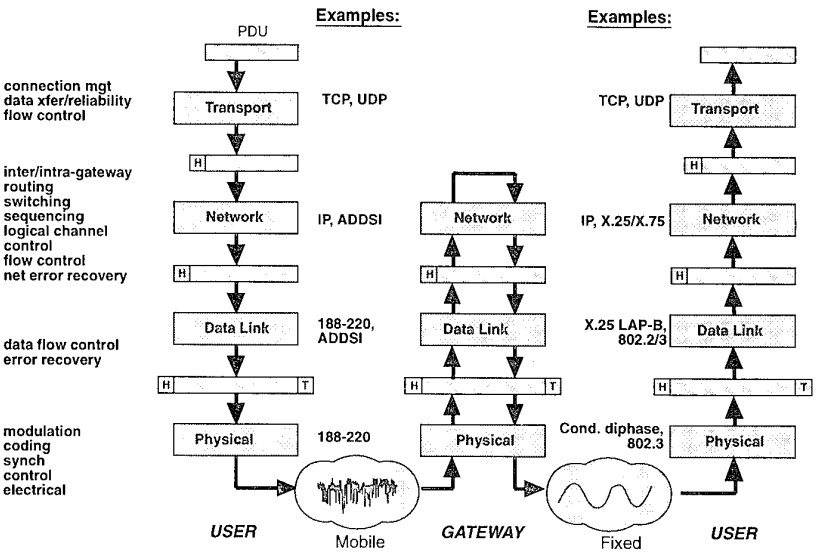


Figure 8. OSI Reference Model

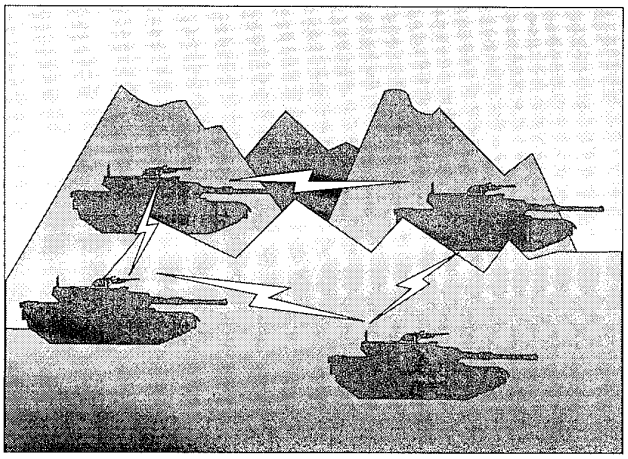


Figure 9. ITEMM Provides Dynamic Terrain Loss Calculations

References

1. W.C.Y. Lee, *Mobile Radio Systems Design*, John Wiley and Sons, 1982.
2. W.C. Jakes, Jr., *Microwave Mobile Communications*, John Wiley & Sons, 1974.

DISCUSSION

Discussor's name: K. Craig

Comment/Question:

Is your propagation channel simulation using the TIREM model?

Author/Presenter's reply:

Yes. The path loss is calculated by TIREM instantaneously on the move, as the vehicles move in a terrain.

Discussor's name: J. Harvey

Comment/Question:

How did you determine what surfaces provided the reflection for the multipath? Did you use some form of ray tracing?

Author/Presenter's reply:

The model does not account for any surface coefficient i.e. reflection coefficient. Multipath calculations are performed using empirical relations and the path loss due to multipath is added to the path loss due to terrain.

We do not use any form of ray tracing in this model.

SYSTEME DE TRANSMISSION NUMERIQUE MULTI-PORTEUSES PAR VOIE IONOSPHERIQUE

S. RIGAUDEAU*, Y. M. LE ROUX
France Télécom / CNET
LAB / PTI / GER
Technopole Anticipa - 2 avenue Pierre Marzin
22307 LANNION CEDEX

1. SOMMAIRE

La propagation dans le canal ionosphérique est de type multi-trajets et évolutive dans le temps. Il en résulte une dégradation des transmissions via ce milieu, accentuée par les différents bruits et brouilleurs présents dans la gamme décamétrique.

L'étalement temporel de ces trajets réduit la rapidité de modulation utilisable. Pour atteindre des débits de 2400 bits/s tout en conservant une vitesse de modulation faible, on répartit l'information sur un grand nombre de sous-canaux. Cette technique du multiplexage fréquentiel orthogonal autorise le chevauchement des spectres d'un grand nombre de sous-canaux. En y associant un codage convolutif correcteur d'erreur et une démodulation cohérente basée sur le critère du maximum de vraisemblance a posteriori, nous obtenons le système de transmission décrit dans cet article.

Différents systèmes de protections ont été mis en oeuvre afin de compenser les sources de perturbations dues au canal ionosphérique. Notre attention s'est particulièrement porté sur l'estimateur du canal pour la démodulation cohérente et sur la correction du décalage de fréquence Doppler.

La mise au point et les premières phases de validation de ce système ont été menées à l'aide du simulateur de canal du laboratoire LAB/PTI/GER du CNET Lannion. Ce simulateur met en oeuvre des canaux soit synthétisés à partir de modèles mathématiques, soit mesurés expérimentalement sur diverses liaisons HF.

2. INTRODUCTION

L'ionosphère, possédant la propriété de réfracter les ondes électromagnétiques de la gamme décamétrique, permet de réaliser des liaisons radioélectriques longues distances (plusieurs milliers de km). Cependant, cette région est constituée d'un plasma dont la densité électronique est inhomogène et soumise à des variations temporelles importantes (surtout pendant le lever et le coucher du soleil).

La propagation, au travers d'un tel canal, s'effectue par trajets multiples (ou modes). Chaque trajet est affecté d'un décalage de fréquence (effet Doppler), d'une dispersion en fréquence et d'atténuations diverses. De même le champ magnétique terrestre, introduit une dépolarisation de l'onde émise, et crée deux modes magnétoïoniques appelés mode ordinaire et extraordinaire. Toutes ces caractéristiques sont à l'origine de fluctuations rapides affectant l'amplitude et la phase des signaux reçus après propagation par voie ionosphérique. Les évanouissements en temps et en fréquence qui en résultent, alliés à la présence de bruits et de brouillages (naturels et artificiels) de forts niveaux dégradent de façon importante les transmissions via ce canal.

Ces différents phénomènes ont le plus souvent limité les transmissions numériques ionosphériques à des débits de quelques centaines de bits par seconde. Cependant, malgré l'existence de systèmes de communication plus performants et offrant des débits supérieurs, le canal ionosphérique dispose d'atoûts non négligeables:

- L'omniprésence du canal autour du globe terrestre et sa quasi-indestructibilité.
- La simplicité et le coût modeste des équipements à mettre en place.

Ainsi, l'objectif de cette étude est la réalisation d'un système de transmission numérique par voie ionosphérique, au débit de 2400 bits par seconde, tout en respectant une bande utile normalisée de 2,3 kHz.

Dans cet article, on présente tout d'abord le principe d'un modem numérique optimisé pour le canal ionosphérique au débit souhaité. Différentes techniques utilisées pour l'estimation de la fonction de transfert du canal sont ensuite décrites, et notamment celles utilisées pour le calcul et le recalage des écarts de fréquence entre émetteur et récepteur (effet conjoint des dérives entre les deux systèmes et de l'effet Doppler ionosphérique). Enfin nous présenterons le procédé de simulation utilisé pour qualifier les caractéristiques du modem. Ceci sera illustré par des résultats de simulations obtenus à partir de réponses impulsionnelles ayant été soit synthétisées par un générateur de canal, soit mesurées par un analyseur de liaisons.

3. LE SYSTEME DE TRANSMISSION

3.1 L'O.F.D.M. (ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING)

L'idée de base consiste à répartir l'information sur un grand nombre de sous-porteuses, tout en conservant une vitesse de modulation faible afin de respecter la bande de cohérence du canal. De tels systèmes, déjà étudiés dans les années 60, présentaient une importante complexité technique pour l'époque. Aujourd'hui, le développement de processeurs dédiés au traitement du signal et l'utilisation d'opérateurs comme la FFT permettent d'en simplifier grandement la réalisation. Cette technique a déjà été présentée en 1989 pour le DAB pour le canal radio-mobile.

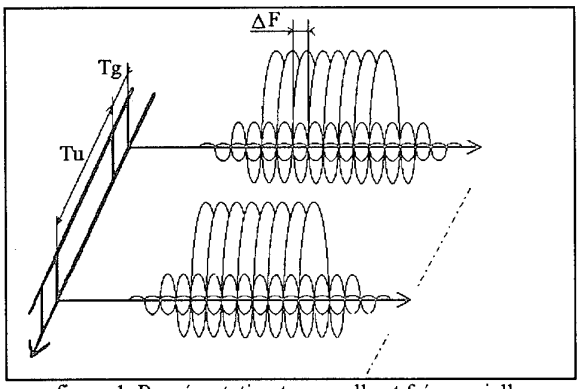
* Les travaux présentés se poursuivent actuellement avec le concours du laboratoire Signal et Communication de Télécom Bretagne.

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

Cette technique du multiplexage fréquentiel orthogonal [2] permet le chevauchement des spectres des différents sous-canaux, sous réserve de respecter la relation d'orthogonalité (voir figure 1) liant la durée du symbole T_u à l'espacement entre les canaux ΔF:

$$T_u = 1/\Delta F \qquad (1)$$

Le choix de T_u est fonction de la bande de cohérence en fréquence du milieu de transmission et doit être très supérieur à l'étalement de la réponse impulsionnelle du canal.



-figure 1: Représentation temporelle et fréquencielle-

Afin de combattre les interférences entre symboles, on insère entre chaque symbole un intervalle de garde T_g de durée supérieure à l'étalement temporel de la réponse impulsionnelle. Chaque trame émise est donc composée d'un intervalle de garde (T_g) suivi du symbole (T_u). La durée totale du symbole est ainsi: $T_s = T_g + T_u$.

Le signal émis est donc:

$$x(t) = Re\left\{ e^{2j\pi f_c t} \sum_{k=-\infty}^{+\infty} \sum_{l=-\frac{N}{2}}^{\frac{N}{2}} X_{kl}\, \Phi_l\left(t - kT_s\right) \right\} \qquad (2)$$

avec:

$$\Phi_l = \begin{cases} e^{\frac{2j\pi lt}{T_u}} & si \quad t \in \left[-T_G, T_U\right] \\ 0 & ailleurs. \end{cases} \qquad (3)$$

et:

f_c: la fréquence centrale du spectre
N+1: le nombre de sous-canaux
X_{kl}: le symbole de la constellation auquel est associé l'amplitude et la phase du "l ième" sous-canal durant le "k ième" intervalle de temps.

La constellation pour chaque porteuse correspond à une modulation à quatre états de phase. En effet, il est préférable de travailler avec une constellation de signaux d'amplitude constante, car d'importants évanouissements peuvent apparaître sur le canal ionosphérique.

Dans ce cas d'un sytème de transmission parallèle, l'utilisation de la bande est optimale (voir figure 1). En effet, plus on utilise de porteuses, plus le spectre du signal tend vers le spectre idéal plat.

3.2 Modulateur et demodulateur O.F.D.M.

Au signal x(t) correspond la suite d'échantillons $x_{n,k}$. Cet échantillon est le "n ème" point composant le "k ième" symbole.

Si on étudie le "k ième" symbole, on obtient:

$$x_n = \left(\frac{1}{M}\right) \sum_{l=-\frac{N}{2}}^{+\frac{N}{2}} X_l\, e^{\frac{2\pi j n l}{M}} \qquad (4)$$

$$n = 0, 1, 2, 3... M-1; \qquad M \geq N+1$$

Il comprend N+1 sinusoïdes modulées par N+1 valeurs complexes, X_l appartenant à la constellation de la modulation de phase choisie. La suite de M valeurs x_k représente un symbole OFDM dans le domaine temporel.

On peut créer le signal x_n à partir d'une transformée de Fourier discrète inverse. En effet on peut écrire:

$$x_n = TFD^{-1}{}_M\,(X) \qquad (5)$$

avec: $X=\{X_0, X_1,..., X_l, 0,..., 0, X_{-l}, ..., X_{-2}, X_{-1}\}$.

Cette opération assure l'orthogonalité du système. Dans le cas ou M est une puissance de 2, le calcul de la FFT inverse peut être efficacement implanté sur un DSP.

Soit H_l la valeur de la fonction de transfert du canal de transmission sur le $l^{ème}$ sous-canal et soit w_n l'enveloppe complexe d'un bruit blanc additif gaussien. Alors l'enveloppe complexe du signal reçu correspondant au "k ième" symbole et en absence de décalage de fréquence et après échantillonnage s'écrit:

$$y_n = \left(\frac{1}{M}\right) \sum_{l=-\frac{N}{2}}^{+\frac{N}{2}} H_l\, X_l\, e^{\frac{2\pi j n l}{M}} + w_n \qquad (6)$$

$$n = 0, 1, 2, 3... M-1;$$

La démodulation de ce signal se fait par l'opération inverse, c'est à dire par la TFD appliquée à la séquence y_n reçue, soit:

$$TFD(y_n) = Y_l = H_l\, X_l + W_l \qquad (7)$$

Dans le cas d'un canal multi-trajets, on insère un temps de garde supérieur à l'étalement de la réponse impulsionnelle. Cette protection évite l'apparition d'interférences entre les différents symboles O.F.D.M. Le temps de garde est créé par simple recopie des derniers éléments au début de la séquence. Le signal final émis est composé de M+Mg éléments:

$$\text{avec: } x=\{x_{M-Mg},...,x_{M-2}, x_{M-1}, X_0, X_l, ..., X_{-M-1}, x_M\}. \qquad (8)$$

A la réception, la séquence x_0 ... x_M est transformée par FFT. On suppose la synchronisation réalisée à ce stade du récepteur. Dans ce cas et en absence de décalage de fréquence entre le signal émis et le signal reçu, la FFT est équivalente à un ensemble de filtres adaptés aux signaux émis sur chaque porteuse.

De plus, la fréquence d'échantillonnage du modulateur et du démodulateur doit respecter la relation (1) afin de conserver l'orthogonalité du système, soit:

$$T_{échantillonnage} = 1/M\,\Delta F = T_u / M = T_s / (M + M_g) \qquad (9)$$

3.3 CODAGE CANAL

Généralement, les données numériques subissent à l'émission un encodage qui, au prix d'une redondance d'information, renforce l'aspect d'unicité du message à transmettre.

Le codage choisi est un codage de type convolutif [3] auquel on associe une cellule d'entrelacement pseudo-aléatoire afin d'améliorer le caractère d'indépendance des erreurs à la réception. Le décodage se fera selon le critère du MAP (Maximum de vraisemblance A Posteriori).

Dans le cas d'une modulation de phase [4] et d'un bruit blanc additif, il consiste à maximiser l'expression suivante:

$$MAX_{\{\hat{X}_k^l\}} \sum_{k=-\infty}^{+\infty} \sum_{l=-N/2}^{N/2} Re\left(\frac{\overline{H}_k^l \hat{X}_k^l Y_k^l}{N_0}\right) \qquad (10)$$

avec:

$\hat{X}_k^l$: k^{ème} symbole émis sur le l^{ème} sous canal.

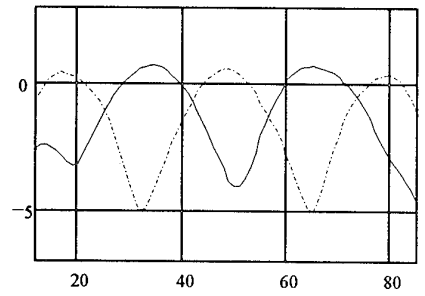
Y_k^l : k^{ème} symbole reçu sur le l^{ème} sous canal.

$\overline{H}_k^l$: expression conjuguée de la valeur de la fonction de transfert du l^{ème} sous canal correspondant au k^{ème} symbole.

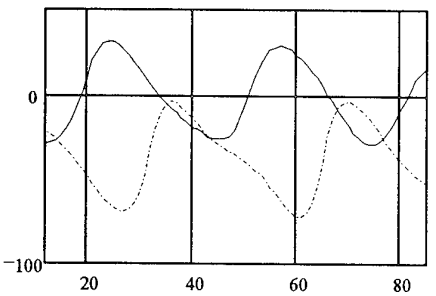
N_0 : densité spectrale bilatérale de puissance du bruit blanc additif.

Le processus de décodage réalise une maximisation de cette expression sur la totalité de la séquence reçue ou sur une partie, sinon le retard de décodage est trop important. Il détermine le meilleur des chemins possibles dans le treillis du code convolutif, en utilisant leur vraisemblance pour critère.

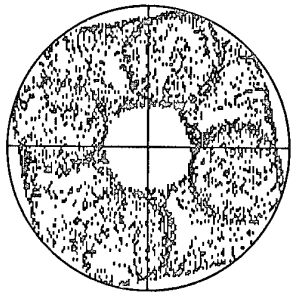
Dans le cas d'un canal sélectif en fréquence, la valeur de H pour certains sous-canaux devient faible, donc sa contribution au critère de décision est faible. En cas de distorsion de phase due au canal de transmission, l'utilisation de l'expression conjuguée dans le critère de décision corrige cette variation.



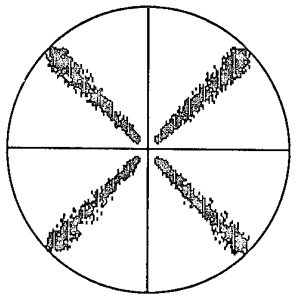
a)- Fonction de transfert du canal (en dB)



b)- Déphasage introduit par le canal (en °).



c) - Constellation des différents sous-canaux du signal reçu.



d) - Constellation fournie au décodage.

- figure2 Illustration du critère de décision -

Ceci est illustré sur la figure 2. La figure 2-a représente le module de deux fonctions de transfert du canal: celle de début et celle de fin de transmission. Ce canal comprend deux trajets d'amplitude relative 1 et 0,5 (évanouissement de 6 dB) avec un décalage de fréquence relatif de 0.5 Hz entre les deux trajets. La figure 2-b présente le déphasage introduit par le canal en début et en fin de transmission. La figure 2-c représente la constellation du signal reçu, et la figure 2-d illustre la constellation obtenue après compensation par la fonction de transfert.

L'utilisation d'un système de décodage à décision douce permet de donner plus d'importance aux points de fortes valeurs, c'est à dire aux porteuses n'ayant pas été atténuées.

Cette technique nous impose de réaliser pour chaque trame une estimation de la fonction de transfert du canal de transmission.

3.4 ESTIMATION DE LA FONCTION DE TRANSFERT

On réserve certains sous-canaux non plus pour transmettre des données mais comme fréquences de références. Ces références, disposées à espacement régulier, permettent de réaliser l'estimation de la fonction de transfert du canal [5]. L'utilisation de références réduit le débit transmissible, mais permet de réaliser une démodulation cohérente.

La disposition des références à intervalles réguliers permet de réaliser un échantillonnage de la fonction de transfert du canal. A l'émission, K porteuses de référence sont positionnées parmi les N porteuses utiles. En réception après la FFT, on isole ces K points et l'on effectue une interpolation afin de disposer d'une estimation de la fonction de transfert du canal. Le nombre de références nécessaires est fonction de

l'étalement temporel maximal de la réponse impulsionnelle du canal (Tc_{max}) et de la durée du temps utile.
Ainsi :

$$K > N.Tc_{max} / Tu \qquad (11)$$

4. SYSTEME DE CORRECTION

4.1 SYNCHRONISATION

Cette opération consiste à déterminer l'emplacement de l'échantillon x_0 dans (8), afin d'éviter l'interférence entre symboles.

On insère, à intervalle régulier, parmi les trames COFDM deux séries complémentaires [6] S1(i) et S2(i) de longueur N. La somme des deux fonctions d'autocorrélation a pour particularité:

$$C_i = \sum_{j=0}^{N-i} S1(j)\,S1(j+i)$$
$$D_i = \sum_{j=0}^{N-i} S2(j)\,S2(j+i) \qquad (12)$$
$$avec \quad \begin{cases} C_i + D_i = 0 & si\ i \neq 0 \\ C_0 + D_0 = 2N \end{cases}$$

Ces deux séquences etant émises si l'on calcule la corrélation en réception, on obtient la réponse impulsionnelle du canal, ce qui permet de déterminer la position du début des trames.

Dans le cas du canal ionosphérique, différents critères sont à prendre en compte:

- Insertion d'un temps de garde entre les deux séquences afin d'éviter l'interférence entre symboles due à l'étalement de la réponse impulsionnelle.
- Au cours de la transmission de la trame de synchronisation, le canal doit pouvoir être considéré comme stationnaire.
- Le canal étant évolutif dans le temps, l'intervalle de temps séparant deux trames de synchronisation doit permettre de visualiser ces variations, il doit être inférieur au temps de cohérence du canal.

4.2 CORRECTION DE FREQUENCE

Dans la suite, on étudie le cas d'un canal de transmission de fonction de transfert H(f), affecté d'un décalage de fréquence relatif ε. Le bruit est additif blanc et gaussien. De plus, on suppose que le décalage de fréquence relatif est inférieur ou égal à la moitié de l'espacement entre les porteuses et que la synchronisation est réalisée, dans ce cas (6) s'écrit:

$$y_n = \left(\frac{1}{N}\right)\left[\sum_{k=-K}^{+K} X_k\, H_k\, e^{\frac{2\pi j n (k+\varepsilon)}{N}}\right] + \omega_n \qquad (13)$$

La démodulation consiste à calculer:

$$Y_k = \sum_{n=0}^{N-1} y_n\, e^{-\frac{2\pi j n k}{N}} \qquad (14)$$

Le k^{eme} élément de la TFD peut s'écrire selon 3 composantes:

$$Y_k = (X_k H_k)\left(\frac{sin(\pi\varepsilon)}{N\,sin\left(\frac{\pi\varepsilon}{N}\right)}\right) e^{j\pi\varepsilon\frac{N-1}{N}} + I_k + W_k \qquad (15)$$

Cette première valeur représente le symbole de modulation X_k affecté par la fonction de transfert du canal de transmission. Il apparaît une réduction d'amplitude et un décalage de phase directement liés au décalage de fréquence.

Le second terme représente l'interférence entre porteuses due à la perte d'orthogonalité du système.

$$I_k = \sum_{\substack{l=-K \\ l\neq k}}^{K} (X_l H_l)\left(\frac{sin(\pi\varepsilon)}{N\,sin\left(\frac{\pi(l-k+\varepsilon)}{N}\right)}\right) e^{j\pi\varepsilon\frac{N-1}{N}} e^{-j\pi\frac{l-k}{N}} \qquad (16)$$

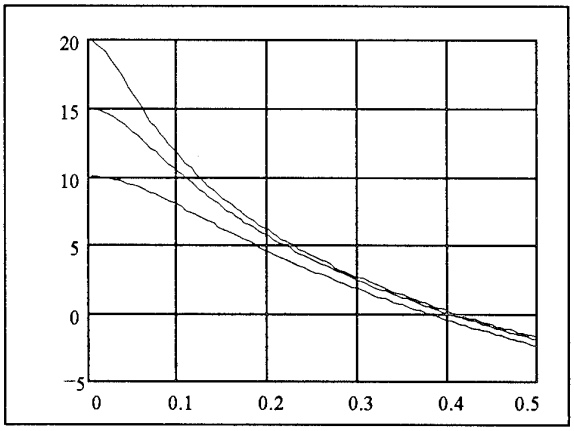
Le troisième terme est l'influence du bruit sur la démodulation.

Moose [7] a calculé le rapport signal à bruit équivalent en sortie de la FFT pour un système OFDM dans le cas d'un canal à bruit blanc additif et avec un décalage de fréquence. Ce calcul prend en compte l'interférence entre porteuses due au décalage de fréquence ainsi que le rapport signal à bruit E_c/N_0 en entrée de la FFT.

Il obtient:

$$SNR_{équivalent} \geq \frac{E_c}{N_0}\left\{\frac{\left[\frac{sin(\pi\varepsilon)}{\pi\varepsilon}\right]^2}{1 + 0.5947\frac{E_c}{N_0}\left[sin(\pi\varepsilon)\right]^2}\right\} \qquad (16)$$

La figure 3 illustre ce résultat en présentant le rapport signal à bruit en sortie de FFT pour des valeurs de E_c/N_0 en entrée de 10, 15 et 20 dB.



-Figure 3- $SNR_{équivalent}$ (en dB) en fonction ε -

On remarque que les dégradations deviennent rapidement importantes. Pour compenser ce phénomène, Moose [7] propose d'émettre pour chaque symbole OFDM deux fois la

séquence des points. Ces deux séquences subissent les mêmes déformations, c'est à dire le même décalage de fréquence, à condition de pouvoir considérer le canal stationnaire.

La contribution de l'interférence entre porteuses (16) étant identique sur ces deux séquences, seule la rotation de phase due au décalage de fréquence et la déformation due au bruit interviennent entre les deux signaux démodulés ($Y1_k$, $Y2_k$) correspondant au symbole k.

A partir de cette constatation, l'estimateur à maximum de vraisemblance de ε est:

$$\hat{\varepsilon} = \frac{1}{2\pi} tan\left\{ \frac{\sum_{k=-K}^{K} Im\left(Y2_k\, YI_k^*\right)}{\sum_{k=-K}^{K} Re\left(Y2_k\, YI_k^*\right)} \right\} \qquad (17)$$

L'estimation du décalage de fréquence est limité à +/- 1/2 Δf, avec Δf espacement entre les porteuses.

Le défaut principal de cette méthode est de néccessiter la répétition des symboles, ce qui réduit le débit utile de moitié. De plus, le domaine de validité de l'estimation, étant limité à +/- 1/2 Δf, cela ne permet pas d'estimer de fortes valeurs de Doppler.

Cette méthode est basée sur la recherche du déphasage moyen obtenu entre les porteuses de deux symboles identiques émis consécutivement. A partir de cette constatation, nous réalisons notre estimation en deux étapes.

Dans un premier temps, on isole les porteuses de référence du reste du signal. Ce sous ensemble n'étant pas utilisé pour transmettre de l'information, reste identique pour chaque trame OFDM. L'algorithme de Moose permet donc d'obtenir ε1 une valeur approchée du décalage de fréquence. En effet, l'interférence des porteuses utiles dues au décalage de fréquence n'est pas identique d'un symbole à l'autre dans le cas présent et vient biaiser l'estimation. De plus, cette valeur étant calculée sur un nombre réduit de porteuse reste sensible au rapport signal à bruit.

On remarque que ce biais sera d'autant plus faible que le décalage de fréquence est lui-même faible. Il est donc préférable de compenser le signal reçu par l'estimation réalisée au cours du symbole précédant avant de commencer celle ci. Ainsi, seul la variation du décalage de fréquence sera estimé. Cette correction étant réalisée pour chaque symbole de façon itérative, l'influence de l'interférence entre porteuses tendra à diminuer, voire à devenir insignifiante.

La compensation entrant dans le processus itératif s'effectue comme suit:

Correction des porteuses utiles par ε1, calcul de $\hat{Y}_k$ et $\hat{Y}_{k-1}$, estimations des symboles émis à l'instant k et k-1.
Calcul de:

$$\varepsilon 2 = \frac{1}{2\pi} tan\left\{ \frac{\sum_{k=-K}^{K} Im\left(Y_{k-1}\, \hat{Y}_{k-1}^*\, Y_k^*\, \hat{Y}_k\right)}{\sum_{k=-K}^{K} Re\left(Y_{k-1}\, \hat{Y}_{k-1}^*\, Y_k^*\, \hat{Y}_k\right)} \right\} \qquad (18)$$

C'est cette seconde valeur ε2 qui permet d'affiner l'estimation. De plus, elle est effectuée sur un nombre plus important de porteuses, ce qui permet de réduire sa sensibilité au bruit.

Le décalage de fréquence est:

$$\varepsilon_k = \varepsilon_{k-1} + \varepsilon 1 + \varepsilon 2 \qquad (19)$$

Pour la méthode de correction presentée ici, la répétition du symbole n'est pas nécessaire. La perte de débit n'est plus de 50% mais directement liée au rapport entre les porteuses de références et les porteuses de données. Cette perte est par ailleurs de toute façon nécessaire pour réaliser l'estimation de la fonction de transfert.

Les limites de cet algorithme étant de +/- 1/2 Δf, il est préférable de ne l'utiliser que pour effectuer une poursuite des fluctuations possibles du décalage de fréquence. Un préambule sera émis en début de transmission afin d'initialiser l'algorithme.

5. LE DISPOSITIF D'EVALUATION

5.1 LE DISPOSITIF D'EVALUATION

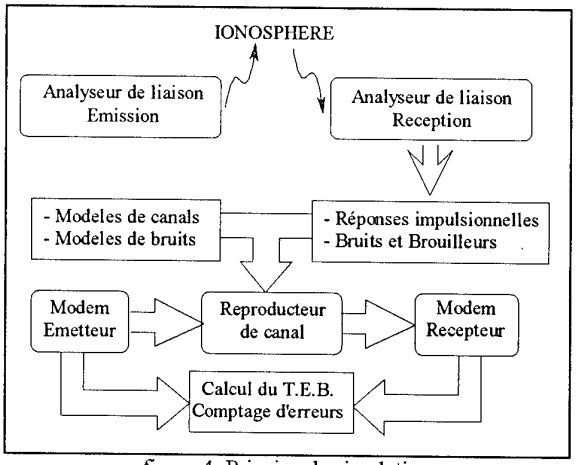
Afin de procéder à une étude approfondie du canal de transmission ionosphérique sur des liaisons expérimentales, un analyseur de liaisons HF a été conçu et réalisé au département GER du CNET Lannion **[8]**.

Ce système permet de mesurer simultanément :

- la réponse impulsionnelle du canal dans une bande utile de 10 kHz avec une période de réactualisation d'une dizaine de millisecondes.
- les différents paramètres de la propagation dans une bande utile de 50 kHz.

Les paramètres mesurés sont principalement les temps de propagation de chaque trajet et pour chacun d'eux, l'atténuation relative, la phase instantanée et le décalage de fréquence Doppler. Ce dernier est obtenu par des analyses spectrales non-linéaires et permet souvent de distinguer les deux modes magnétoïoniques propagés, lorsque ceux-ci sont confondus en temps de groupe.

Le principe de l'analyseur de liaisons et du reproducteur de canal sont représentés en figure 4.



- figure 4:-Principe de simulation -

En effet, l'analyse des performances réelles du système de transmission nécessite de disposer d'un modèle le plus représentatif possible du canal. Cependant aucun modèle ne permet actuellement de décrire de manière complète les différents phénomènes de propagation caractérisant le milieu ionosphérique (limitations du modèle CCIR).

Ainsi donc nous avons retenu, pour réaliser les simulations de transmissions et se rapprocher au mieux des conditions réelles de transmission, une méthode de simulation basée sur le principe de "reproduction de canal de transmission" [9]. Cette approche conduit à utiliser des séquences de réponses impulsionnelles mesurées par l'analyseur sur des liaisons expérimentales.

Ces enregistrements de réponses impulsionnelles sont ensuite exploités en laboratoire pour reproduire fidèlement les variations introduites par l'ionosphère sur les signaux qui y transitent. Avec une telle méthode de simulation, nous nous rapprochons des conditions pouvant être rencontrées lors d'une exploitation réelle sur liaisons ionosphériques.

5.2 CRITERE D'EVALUATION

Le taux d'erreur binaire ou TEB est le critère utilisé ici. Il est évalué par une méthode directe de type Monte-Carlo. Le nombre d'essais N (supérieur à 400.000 informations binaires utiles) de la simulation est choisi de façon à pouvoir estimer des probabilités d'erreur de 10^{-3} avec une confiance de 95% et une précision relative égale à 10^{-1}. Pour un canal unité à bruit blanc additif gaussien, on utilise 600 000 points pour les simulations.

Le TEB est calculé comme étant la moyenne :

$$TEB = 1/N \Sigma e_i \qquad (20)$$

avec: $e_i = 1$ si une erreur apparaît sur le $i^{ième}$ bit.
$e_i = 0$ si le $i^{ième}$ bit est correctement décodé.

6. SIMULATIONS:

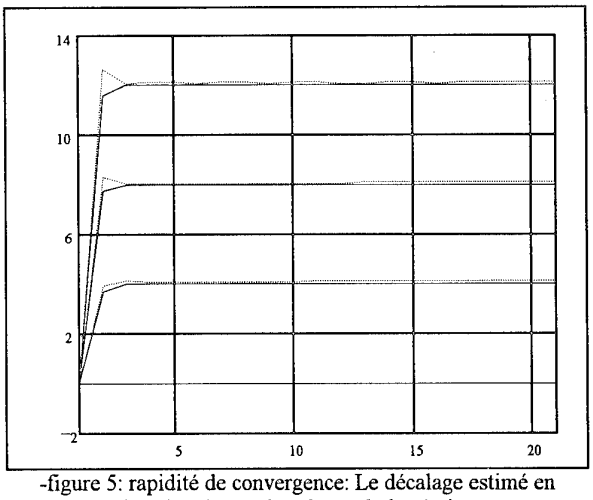
6.1 ETUDE DE LA CORRECTION DE FREQUENCE

La correction de fréquence étant initialisée à zéro en début de transmission, la figure 5 représente la rapidité de convergence de l'algorithme en fonction du décalage de fréquence initial. Ce test a été réalisé pour différents canaux. Les traits pleins correspondent à une réponse impulsionnelle comprennant un trajet unique et les pointillés à un canal comprenant deux trajets d'amplitude relatives 1 et 0.5 espacé de 1 ms. Le signal reçu subit différents décalages de fréquence (0, 4, et 8 Hz).

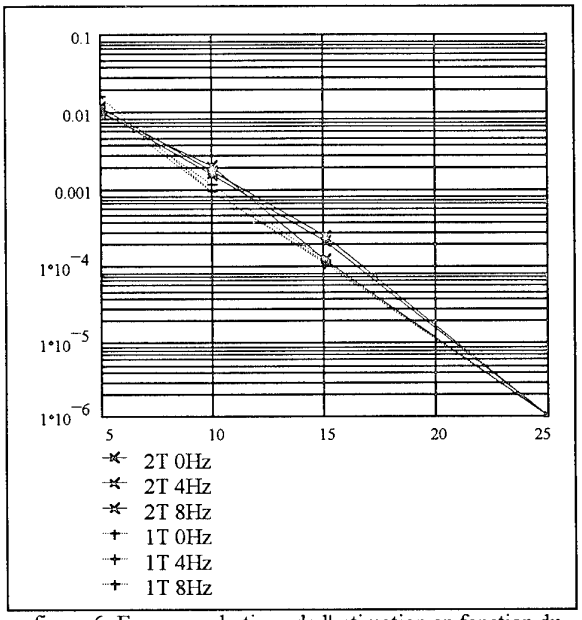
On remarque que l'algorithme converge vers la valeur finale en moins de 5 estimations. Afin d'éviter ce retard de convergence, on peut emettre un préambule permettant d'initialiser l'algorithme au début de la transmission.

le bruit influence sur la précision de l'estimation. La figure 6 permet de visualiser l'erreur quadratique de l'estimation en fonction du rapport signal à bruit en tête du recepteur, ceci représente la résistance au bruit du système d'estimation du décalage de fréquence.

Ce test a été réalisé pour les canaux décris précédement.



-figure 5: rapidité de convergence: Le décalage estimé en fonction du nombre de symboles émis -



- figure 6- Erreur quadratique de l'estimation en fonction du rapport signal à bruit en entrée du recepteur.

On remarque que l'erreur quadratique est directement liée au rapport signal à bruit quel que soit le décalage à compenser. En effet la structure récursive utilisée calcule l'erreur entre la valeur estimée et la valeur réelle. Aprés son initialisation, l'erreur d'estimation devient donc indépendante du décalage à compenser.

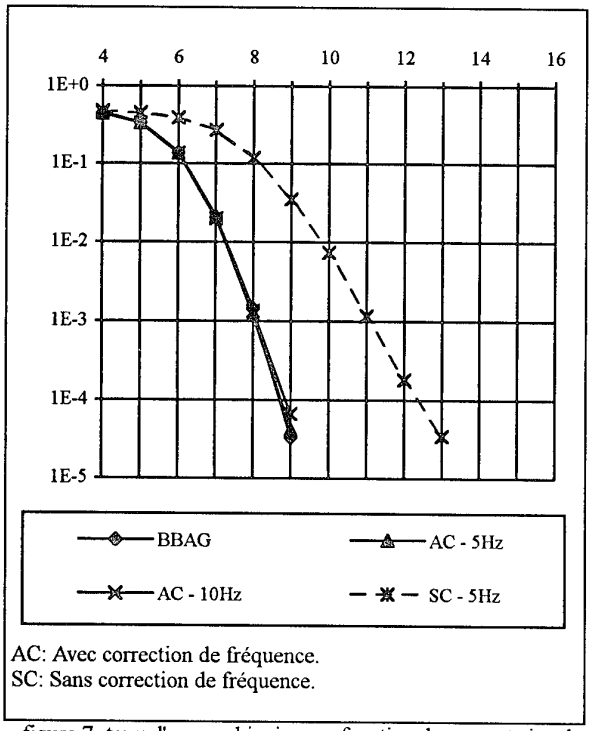
De plus, lorsque l'algorithme a convergé , la précision de l'estimation est indépendante du canal dans les limites de fonctionnement du système (plage d'estimation, temps de garde,..). En effet les traits pleins correspondent à un trajet unique, tandis que les pointillés correspondent à un canal à deux trajets.

6.2 SIMULATIONS OBTENUES A PARTIR DE MODELES

6.2.1 Résistance au décalage de fréquence

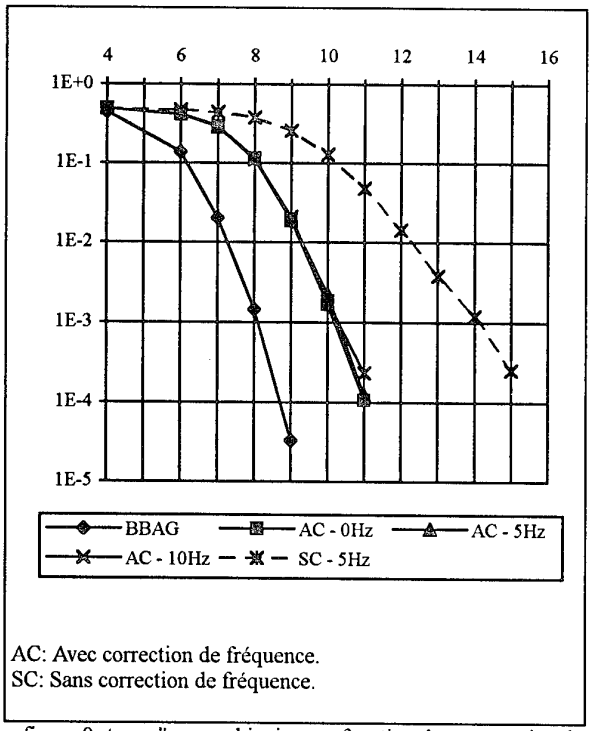
Cette simulation correspond à l'étude du comportement du système avec et sans la correction du décalage de fréquence.

Pour cette étude on reprend les différents canaux définis précédement, soit à un trajet unique (voir figure 7), soit à deux trajets (voir figure 8). La correction du décalage de fréquence est opérationnelle dans le cas des traits pleins et déconnectée pour celle en pointillés.



AC: Avec correction de fréquence.
SC: Sans correction de fréquence.

-figure 7- taux d'erreurs binaires en fonction du rapport signal à bruit en entrée du recepteur pour un canal à 1 trajet.



AC: Avec correction de fréquence.
SC: Sans correction de fréquence.

-figure 8- taux d'erreurs binaires en fonction du rapport signal à bruit en entrée du recepteur pour un canal à 2 trajets.

Dans le cas où la correction de fréquence est opérationnelle, un taux d'erreurs de 10^{-3} est obtenu pour 8,1 dB pour le canal à bruit blanc additif gaussien ainsi que pour les cas de trajets uniques affectés d'un décalage de fréquence.

Sans correction de fréquence, on remarque une rapide dégradation des performances du système.

Pour le cas de deux trajets le TEB de 10^{-3} est obtenue pour un RSB de 10,3 dB. La différence des performances entre les deux exemples est due aux évanouissements sélectifs en fréquences qui affectent ce signal. Pour obtenir de meilleures performances, il suffit d'augmenter le rendement du codage mais cela diminue le débit transmis.

On remarque que le système a des caractéristiques similaires quel que soit le décalage de fréquence (compris dans le domaine de définition déterminé précédement) pour un canal donné. Sans correction de fréquence, des décalages de 5 Hz font apparaitre des dégradations importantes (3,1 dB pour un trajet, 4,1 dB pour le cas à deux trajets).

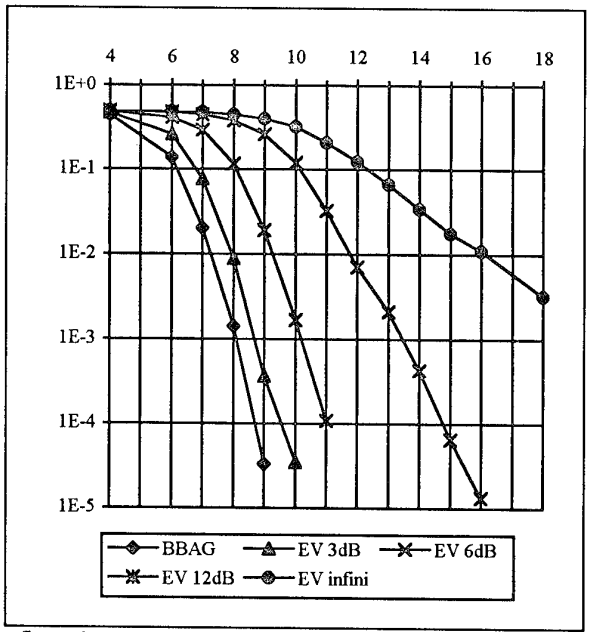
6.2.2 Comportement aux évanouissements sélectifs

On remarque que les performances du système se dégradent avec l'apparition d'un second trajet.

Pour une réponse impulsionnelle comprenant deux trajets d'amplitudes A1 et A2 (avec A1>A2) avec un retard différentiel de propagation τ, la profondeur des évanouissements est fournie par **[10]**:

$$\text{Evanouissement} = -20 \log_{10} [1 - A2 / A1] \qquad (19)$$

Sur la figure 9 on représente les performances du système en fonction de la profondeur des affaiblissements sélectifs en fréquence. Les cas étudiés ont des profondeurs d'evanouissements respectifs de 3dB, 6dB, 12dB et infini.



-figure 9- taux d'erreurs binaires en fonctions du rapport signal à bruit pour différentes profondeurs d'évanouissement.

6.3 SIMULATIONS OBTENUES A PARTIR DE REPONSES IMPULSIONNELLES MESUREES:

Les différents exemples étudiés ont été réalisés, aux cours d'une campagne de mesures en décembre 1992, sur une liaison expérimentale d'environ 1000 kms entre Toulon et Lannion. Les deux exemples simulés correspondent chacun à des données de transmission de 2 minutes, bien que les réponses impulsionnelles soient représentées sur une période de 10 mn.

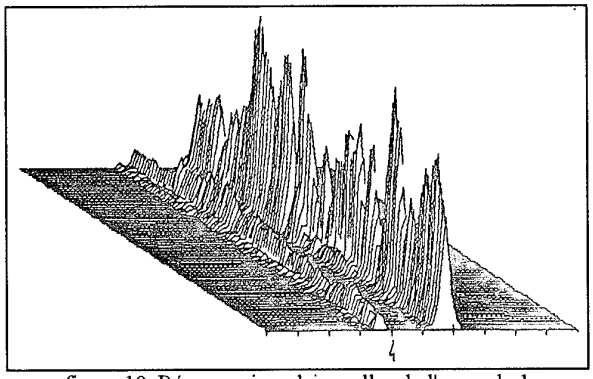
6.3.1 Transmission proche de la MUF

La figure 11 représente les performances d'une transmission réalisée à une fréquence proche de la MUF (Fréquence Maximale Utilisable). En effet le dépouillement des ionogrammes obliques réalisés au cours de cette transmission indique une MUF d'environ 17,5 MHz tandis que la fréquence d'analyse est 15,89 MHz.

Les conditions de transmission sont caractérisées par une structure à deux ou trois trajets de propagation :

- Un mode 1F2 rayon bas ayant un temps de propagation de 3,75 ms avec un décalage en fréquence Doppler d'environ - 0,2 Hz à - 0,38 Hz.
- Un mode 1F2 rayon haut ordinaire possédant un temps de groupe de 4,2 ms et qui tend à diminuer en fin de transmission. Il est affecté par un décalage de fréquence Doppler de - 0,29 Hz à - 0,4 Hz.
- Un mode 1F2 rayon haut extraordinaire que l'on distingue en fin de transmission.

Les évolutions au cours du temps des réponses impulsionnelles utilisées pour la simulation sont représentées sur la figure 10 :

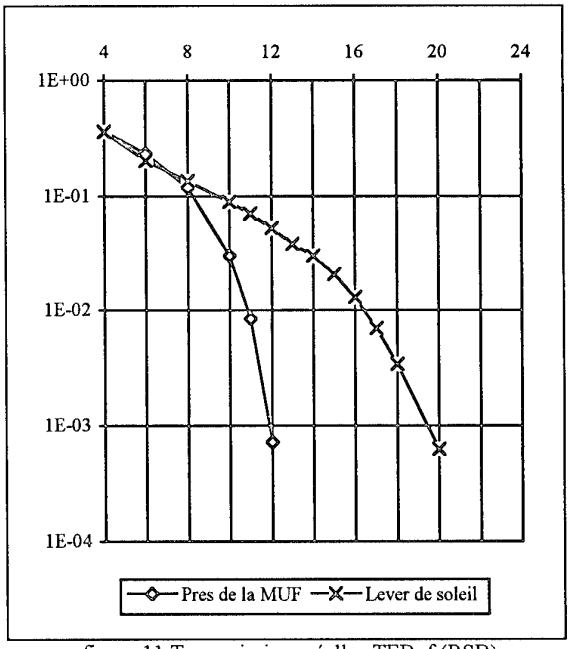


-figure 10: Réponses impulsionnelles de l'exemple 1-

On remarque de fortes variations temporelles de l'amplitude relative du trajet principal et des trajets secondaires. La puissance instantanée du signal subit donc de fortes variations. Des écarts de ± 4 dB entre la puissance moyenne du signal et sa puissance instantanée ont été mesurés.

Les performances sont évaluées en fonction du rapport signal à bruit moyen calculé sur la durée totale de la simulation. Le taux d'erreur sera donc dépendant des variations instantanées du rapport S/B par rapport à cette valeur moyenne.

Nous obtenons un taux d'erreur de 10^{-3} pour un rapport S/B de 11,8 dB. Une observation de la répartition des erreurs non présentée ici nous permet de remarquer de fortes concentrations d'erreurs pour les faibles valeurs de la puissance instantanée.



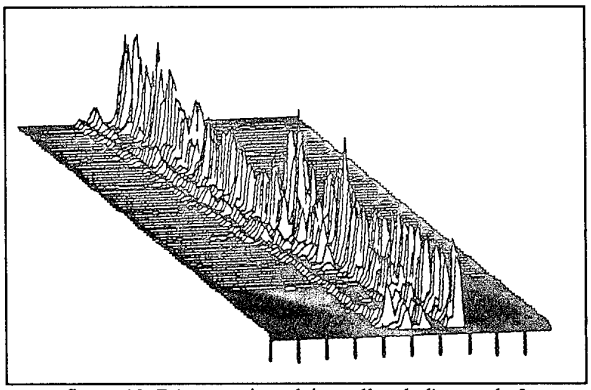
-figure 11 Transmissions réelles TEB=f (RSB)-

6.3.2 Transmission lors d'un lever de soleil

La seconde évaluation a été réalisée sur des mesures effectuées lors d'un lever de soleil. La propagation subit de fortes variations de temps de groupe. En effet les divers modes propagés sont:

- Un mode 1F2, rayon bas ayant un temps de propagation de 3,67 ms. Ce mode subit peu de variations de son temps de groupe.
- Un mode 1F2, rayon haut ordinaire dont le temps de propagation varie de 4,05 ms à 4,27 ms, du début à la fin de la transmission. Cette variation est due à l'augmentation de la MUF pendant le lever de soleil.
- Un mode 1F2, rayon haut extraordinaire dont le temps de groupe varie de 4,17 ms en début de période d'analyse à 4,37 ms en fin d'analyse.

Les réponses impulsionnelles utilisées pour la simulation sont représentées sur la figure 12.



-figure 12: Réponses impulsionnelles de l'exemple 2-

L'analyse des paramètres du milieu de transmission permet d'obtenir plus d'informations sur les conditions de propagation notamment sur le décalage de fréquence Doppler.

En effet, les évanouissements sur l'enveloppe du trajet 1F bas indiquent la présence de deux modes magnétoïoniques propagés. Ils sont confondus en temps de groupe mais ont des Doppler différents. Le Doppler différentiel entre les deux trajets varie de 0,03 Hz à 0,5 Hz. Le décalage de fréquence Doppler moyen prend des valeurs de 0,15 à 0,765 Hz.

D'autre part, les rayons hauts possèdent des décalages de fréquences Doppler plus importants : 0,6 à 0,99 Hz pour le mode ordinaire, et 0,72 à 1,15 Hz pour l'autre mode.

Comme précédemment, la puissance instantanée subit de fortes variations. Les variations entre la puissance moyenne et la puissance instantanée évoluent de + 2,35 dB à - 8,61 dB.

La simulation (voir figure 11), nous donne un taux d'erreur de 10^{-3} pour un rapport S/B de 19,2 dB avec de fortes concentrations d'erreurs pour les faibles valeurs de la puissance instantanée.

7. CONCLUSION

Les performances du système de transmissions numériques multi-porteuses présenté ici ont été obtenues à partir de deux types de représentation du canal ionosphérique. Le premier type s'appuie sur une modèlisation analytique du canal tandis que le second exploite des séquences de réponses impulsionnelles, mesurées sur des liaisons expérimentales. Ceci confère un caractère réaliste aux résultats obtenus, qu'il conviendra de compléter par un test en grandeur réel du système de transmission, celui-ci étant en cours de réalisation.

Les premières conclusions sont que le système de transmission semble bien adapté à la transmission de la parole numérisée via le canal ionosphérique, des taux d'erreur d'environ 10^{-3} étant tolérés dans ce cas. D'autres applications, comme la transmission de données sont également envisagées. En effet pour ce type de transmissions, des performances accrues peuvent être attendues par la mise en oeuvre de techniques plus sophistiquées puisque les contraintes de retard liées au temps de traitement n'existent plus, comme pour la parole.

Ainsi des études se poursuivent pour améliorer l'estimation du canal en présence de bruit. De même les solutions concernant l'initialisation de la synchronisation en temps du système et l'initialisation de la récupération de porteuses sont en cours d'optimisation.

Enfin, des travaux sont menés sur la gestion automatique de fréquence et la recherche de canaux clairs en bande HF pour contribuer à l'amélioration de la qualité et de la fiabilité du type de transmission présenté ici.

8. REFERENCES

[1]: B. Le Floch, R. Halbert-Lassale, D. Castelain; "Digital Sound Broadcasting to Mobile Receivers", IEEE Transactions on Consumer Electronic, vol 35, N°3, August 1989.

[2]: P Scalart, Y.M. Le Roux, " Evaluation of a high speed data communication system over ionospheric channel", 7 th International ionospheric effect symposium, May 1993, Alexandria, Virginia (US).

[3]: G.C. Clark, J. Bibb Cain, "Error-correction coding for digital communication", PLENUS PRESS, New-York, Second Priniting, August 1982.

[4]: P. Scalart, "Elaboration et Evaluation d'un système de transmission numérique à haut débit par voie ionosphérique", Thèse de l'Université de Rennes 1, 18/12/1992.

[5]: B. Le Floch, J.F. Helard, D. Castelain, M. Rivière, "Démodulation cohérente du système de transmission multi-porteuses COFDM dans un canal de radiomobile", Treizième Colloque GRETSI, 1991, pp 413 - 416.

[6]: M. J. E. Golay, "Complementary Series" ; IRE Transactions on Information Theory, pp 82-87, April 1961.

[7]: P. H. Moose, "A technique for othogonal Frequency division multiplexing fraquency offset correction", IEEE Transactions on communications, vol 42, n° 10, October 1994.

[8]: Y. M. Le Roux, P. Lassudrie-Duchesne, R. Fleury, J. P. Jolivet, J. Menard, "Measurement Capabilites of an HF Channel Evaluation System", 5th International conf. on HF Radio Systems and Techniques, IEE 339, pp. 22-25, July 1991.

[9]: Y. M. Le Roux, P. Lassudrie-Duchesne, R. Fleury, J. Menard, J. P. Jolivet, "HF Channel modelling and Simulation" 1990, Proc. IEE, 325, Cambridge W.R., pp 72-76.

[10]: S. Rigaudeau, Y. M. Le Roux, P. Scalart, "Présentation d'un système de transmission numérique multi-porteuses par voie ionosphérique", SEE, Perros - Guirec, mars 1994.

DISCUSSION

Discussor's name: C. Goutelard

Comment/Question:

Pensez-vous utiliser les codes poinçonnés pour réaliser une adaptation aux caractéristiques du calcul?

(Translation:

Have you considered using punched codes to enable adaptation to the calculation characteristics?)

Author/Presenter's reply:

Cette suggestion offrirait en effet des perspectives intéressantes mais nécessiterait une voie de retour.

Cependant la priorité immédiate a été donnée aux tests de turbo-codes en collaboration avec Télécom Bretagne sur les liaisons HF réelles.

(*Translation:*

This is an interesting suggestion but a return path would be needed.

However, top priority has been given to the turbo-code tests being carried out in collaboration with Telecom Bretagne on HF links in use.

Influence des Antennes et de la Propagation sur le Comportement d'un Système de Transmissions Numériques en H.F.

L. BERTEL, O. LEBAILLIF
Laboratoire Radiocommunications
URA CNRS 834
Campus de Beaulieu
Université de Rennes 1
35 042 Rennes Cedex

Y. LE ROUX, R. FLEURY
LAB/PTI/GER
Centre National d'Étude des Télécommunications
22 300 Lannion

RÉSUMÉ

Le but de cette présentation est de décrire l'influence des antennes et de la propagation sur le comportement d'une liaison numérique en HF. Compte tenu du caractère particulier de la propagation en HF, nous illustrons à partir d'un modèle de signal le rôle que jouent les antennes. Une expérimentation a été mise en place avec du matériel opérationnel. Les résultats obtenus sont présentés et analysés. Il est possible de montrer l'intérêt d'un filtrage dit de polarisation équivalent à un traitement vectoriel pour réduire le taux d'erreur binaire.

ABSTRACT

This paper describes the influence of antennas and propagation on the behaviour of H.F. digital communication links. We highlight the role of the antennas with the help of a signal model which takes into account the particularities of propagation in the H.F. range. Results of an experimental link are analysed and we demonstrate the efficiency of a polarization filtering ; this filtering procedure can also be viewed as a vector processing technique allowing a decrease of the BER.

1 INTRODUCTION

Les systèmes de transmission en décamétrique par voie ionosphérique doivent tenir compte des aspects particuliers de la propagation dans cette gamme de fréquence [1][2] : nombreux multitrajets, décalages et étalements doppler, possibilités de deux modes de propagation auxquels sont associés des effets de polarisations différents ayant une conséquence sur le comportement des antennes.

L'analyse des signaux issus de telles liaisons a fait l'objet de nombreuses études. Les effets d'antennes n'y sont en général pas intégrés si ce n'est pour caractériser le bilan de liaison sous forme déterministe ou statistique ; dans ce cas, seul le diagramme de directivité de l'antenne est considéré et non sa réponse complexe.

Des modélisations de signaux incluant la réponse complexe des antennes ont été introduites plus récemment [3][4] en vue d'application à la goniomètrie HF haute résolution ; elles tiennent compte du type des antennes, de leur lieu d'implantation et de la liaison étudiée.

L'objectif de cet article est de mettre en évidence cet effet des antennes et de la propagation sur le comportement d'une liaison numérique réelle. Dans une première partie, on rappelle un modèle de signal mettant en évidence le rôle des antennes ; dans une deuxième partie, on décrit l'expérimentation mise en place et enfin on analyse quelques résultats obtenus ainsi que les conséquences qui en découlent sur la conception de systèmes de transmission futurs.

2 LA MODÉLISATION DES SIGNAUX HF

2.1 Expressions du signal en sortie des antennes

A un instant t_0 pris comme référence, il est possible d'écrire le signal résultant d'une propagation par voie ionosphérique pour un mode k sous la forme suivante :

$$s_k(t) = U_k(t) m(t - \tau g_k) \exp(j[\omega_k(t - \tau p_k)]) \quad (1)$$

Un mode k est défini pour un trajet donné via une couche ionosphérique et pour une polarisation. Ainsi, en général pour un trajet, deux modes de propagation sont possibles ; ils sont dénommés les modes O et X.

Dans l'expression (1), $U_k(t)$ représente l'amplitude complexe du signal fonction de la puissance émise, du type d'antenne utilisé à l'émission et à la réception et de l'atténuation sur le trajet ; $m(t)$ caractérise la modulation, τg_k le retard de groupe, τp_k le retard de phase et ω_k la pulsation du signal reçu affecté d'un décalage doppler. Suivant les conditions de propagation, une dispersion sur τg_k et ω_k peut être observée.

Pour notre étude, nous pouvons expliciter le terme $U_k(t)$; la relation entre onde électromagnétique et signal se fait par l'intermédiaire des antennes : l'onde électromagnétique est une grandeur vectorielle que l'antenne va transformer en grandeur scalaire réelle et vice versa. Pour une

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

liaison donnée, en considérant émission et réception, on peut donc écrire :

$$U_k(t) = F_{ik} G_{lk} A_k(t)$$

où F_{ik} et G_{lk} caractérisent les réponses complexes des antennes à la réception (F_{ik}) et à l'émission (G_{lk}). Les indices i et l se réfèrent au type d'antenne. A l'émission, l'antenne étant figée, le paramètre l est ici inutile.

Les réponses complexes de quelques antennes HF ont été décrites par ailleurs [5] ; leur modélisation est liée :

- à leur type ;
- à leur lieu d'implantation car la réponse dépend de la polarisation de l'onde incidente et donc du vecteur induction magnétique terrestre le long du trajet ;
- à leur environnement et plus particulièrement à la nature du sol.

Pour une antenne i donnée en réception, le signal à la sortie de cette antenne pour un mode k peut s'exprimer sous la forme :

$$s_{ik}(t) = A_k(t) F_{ik} G_k m(t - \tau g_k) \exp(j[\omega_k(t - \tau p_k)])$$

Il faut noter que dans cette expression G_k caractérise le gain de l'antenne d'émission pour le mode considéré (et non en polarisation linéaire). Lorsque l'on a plusieurs modes pour une antenne i donnée, on a donc :

$$s_i(t) = \sum_k s_{ik}(t)$$

$s_i(t)$ étant fonction de F_{ik}, les signaux présents en sortie des antennes sont fonction du type d'antennes utilisées.

2.2 Introduction de la fonction de diffusion du canal

Cette précédente modélisation des signaux ne permet pas toutefois d'étudier simplement l'influence de la réponse des antennes sur le taux d'erreur binaire. D'autres types de repésentations peuvent en effet se révéler plus intéresantes, notamment en exprimant la variabilité en temps et en espace de la réponse impulsionnelle du canal au moyen d'expressions analytiques [6], ou sous forme de séquences de mesures obtenues sur des liaisons expérimentales [7]. Une autre façon pratique de procéder est de supposer le canal stationnaire au sens large et de le représenter dans le plan Retard–Doppler, par sa fonction de diffusion [8]. Toutes ces représentations permettent en effet de tenir compte des effets d'antennes d'émission et de la propagation. Les antennes de réception sont également prises en compte après détermination du filtre de canal équivalent.

Nous représenterons ici le canal ionosphérique par sa fonction de diffusion. Nous pourrons alors obtenir une réalisation du filtre de canal [9] en appliquant :

$$H^{(m)}(f,t) = \iint_{\Re^2} D^{(m)}(\tau,\nu)\, e^{-2i\pi f\tau} e^{2i\pi\nu t}\, d\tau\, d\nu \qquad (2)$$

avec $H^{(m)}(f,t)$ la $m^{ième}$ réalisation du filtre de canal et $D^{(m)}(\tau,\nu)$ la $m^{ième}$ réalisation d'une quantité liée à la densité spectrale de puissance $P^{(m)}(\tau,\nu)$ de la manière suivante :

$$P^{(m)}(\tau,\nu) = D^{(m)}(\tau,\nu) \cdot D^{(m)^*}(\tau,\nu)$$

Cette réalisation de la densité spectrale de puissance s'obtient à partir de la fonction de diffusion du canal $S(\tau,\nu)$ en la multipliant par la $m^{ième}$ réalisation d'un processus aléatoire complexe $r(\tau,\nu)$. On a donc :

$$P^{(m)}(\tau,\nu) = |r^{(m)}(\tau,\nu)|^2 S(\tau,\nu)$$

le processus aléatoire $r(\tau,\nu)$ étant quant à lui défini de la manière suivante :

$$r(\tau,\nu) = \frac{1}{\sqrt{2}}(x(\tau,\nu) + jy(\tau,\nu))$$

$x(\tau,\nu)$ et $y(\tau,\nu)$ sont deux processus aléatoires réels, gaussiens, centrés et de variance unitaire. De tout ceci, nous pouvons déduire que le module du processus $r(\tau,\nu)$ suit une loi de Rayleigh et que sa phase est équirépartie entre 0 et 2π ; ceci nous permet d'exprimer l'espérance mathématique des réalisations de la densité spectrale de puissance par l'expression :

$$\langle P^{(m)}(\tau,\nu)\rangle = S(\tau,\nu)$$

Nous pouvons donc, à partir d'une fonction de diffusion connue, reconstruire une réalisation du filtre de canal en utilisant l'équation (2) dans laquelle nous remplaçons $D^{(m)}(\tau,\nu)$ par la quantité suivante :

$$D^{(m)}(\tau,\nu) = \sqrt{S(\tau,\nu)}\; r^{(m)}(\tau,\nu) \qquad (3)$$

Nous allons maintenant envisager le cas simplifié où la fonction de diffusion $S(\tau,\nu)$ est constituée de deux modes de propagation. En raison de la linéarité du filtre équivalent au canal ionosphérique, on peut écrire :

$$H^{(m)}(f,t) = H_O^{(m)}(f,t) + H_X^{(m)}(f,t)$$

où les indices «O» et «X» se rapportent respectivement aux modes de propagation ordinaire et extraordinaire. Les modifications apportées par les effets des antennes ou par un filtrage de polarisation peuvent alors être introduites sous la forme d'un facteur multiplicatif complexe $\tilde{a}$ dans l'expression précédente, ce qui nous conduit à l'expression finale suivante du filtre de canal :

$$H^{(m)}(f,t) = H_O^{(m)}(f,t) + \tilde{a}\, H_X^{(m)}(f,t) \qquad (4)$$

Nous avons effectué des simulations afin de montrer le lien entre le filtre de canal, la bande de cohérence et donc par conséquence le TEB. Nous avons choisi une réalisation d'une fonction de diffusion d'un cas possible de canal H.F. Dans la figure (1) nous présentons une partie de cette fonction de diffusion où nous pouvons voir les modes O et X considérés ; ils sont localisés dans le plan retard de groupe – doppler par les couples de coordonnées (0.48 ms, 0.6 Hz) pour le mode O et (0.6 ms, 0.85 Hz) pour le mode X. Les dispersions en temps de groupe et en doppler ont été choisies arbitrairement égales pour les deux modes, soit 0.033 Hz pour le doppler et 40 μs pour le temps de groupe. À partir de cette réalisation de la fonction de diffusion décomposée en deux modes distincts comme exposé précédement, nous avons pu calculer les fonctions de transfert, variantes dans le temps, associées à la présence de chacun des modes. Ces fonctions sont présentées aux figures (2) et (3).

Si aucun traitement au niveau des antennes n'est réalisé alors nous avons la présence simultanée des deux modes

et la fonction de transfert du canal est nettement affectée. Nous pouvons voir figure (4) le module de cette fonction de transfert du canal ; on remarquera que l'écart entre les faddings sélectifs dans la bande est directement lié au retard différentiel entre les deux modes et que l'évolution temporelle de ces faddings se fait quasiment à la fréquence du doppler différentiel intermode.

Par contre, si nous effectuons un filtrage de polarisation nous pouvons combiner efficacement les signaux issus des différentes antennes afin d'atténuer très fortement l'un des modes. Un filtre de canal obtenu après traitement est présenté à la figure (5). Nous pouvons constater que ce filtre tend vers celui présenté figure (2) ce qui était attendu car le traitement visait à réduire de 90% l'energie liée au second mode.

Enfin, nous présentons figure (6) l'évolution de la bande de cohérence à 0.9 au cours du temps pour chacun des filtres déjà mentionnés. Nous remarquons les points suivants :

- la bande de cohérence évolue de manière significativement différente pour les deux premiers filtres liés aux modes seuls. Ceci a comme conséquence que les instants d'apparition des erreurs sont bien distints pour chacune des voies de réception ;
- le filtrage de polarisation permet bien d'augmenter la bande de cohérence, donc de réduire le TEB, ceci est mis en évidence par les différences entre les 4^e et 3^e courbes pour lesquelles il y a, ou non filtrage.

2.3 Application à la determination du TEB

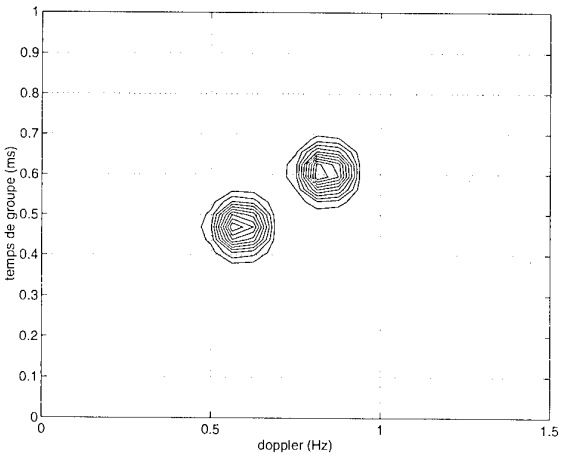
C'est sur la base de cette expression (4) que allons montrer l'influence des effets d'antennes et de polarisation sur les performances des systèmes de communications numériques HF. En effet, nous pouvons désormais calculer la bande de cohérence du filtre de canal. Cette quantité est directement liée au taux d'erreur binaire ; une réduction de la bande de cohérence du canal en deçà de la largeur de bande utile de la modulation entraîne une interférence intersymbole supérieure et donc une augmentation du TEB. Nous montrons que suivant le choix du terme complexe $\tilde{a}$, il est possible de modifier la bande de cohérence du canal et donc de diminuer le TEB.

3 LE SYSTÈME EXPÉRIMENTAL

3.1 La liaison et les constituants du système

Une liaison test Berlin – Lannion (1 242 km) a été utilisée pour cette expérimentation. Elle a été choisie puisque pour un azimut de 64 ° correspondant à cette liaison, la polarisation des ondes incidentes aux antennes de réception est fortement influencée par leur angle d'élévation correspondant [5]. Le dispositif mis en place n'ayant pas comme objectif de tester des récepteurs ou modems, nous avons choisi des constituants simples utilisés sur des liaisons opérationnelles. Les éléments utilisés dans cette campagne de mesure sont décrits dans le tableau (1). Il faut toutefois noter que le système antennaire de réception (antenne tripôle) a été modifié de façon à éliminer tout couplage entre les antennes élémentaires. Ainsi, le mât utilisé est diélectrique et les courants de surface des câbles d'alimentation des antennes ont été éliminés. Le système de réception étant installé à la station ionosphérique de Lannion, nous avons pu disposer durant toute cette expérimentation des résultats de sondages ionosphériques au dessus du site.

FIG. 1 – *Module de la fonction de diffusion du canal H.F. Nous pouvons voir les deux modes placés en 0.6 Hz et 0.48 ms pour le premier avec une dipersion en temps de groupe de 40 µs et une dispersion en doppler de 0.033 Hz. Le second est quant à lui placé en 0.85 Hz et 0.6 ms avec des caractéristiques identiques au premier pour les dispersions en temps de groupe et doppler.*



3.2 Présentation des résultats

Un système de mesure sur deux voies a été entièrement mis au point au CNET [10]. Il permet, sur chacune de ces voies de déterminer :

- le nombre de bits erronnés sur un paquet de 1 024 bits ; au delà de 512 bits erronés, on considère qu'il y a désynchronisation ;
- les instants de désynchronisation des modems ou des cartes de comptage des erreurs ;
- le taux d'erreur binaire moyen sur chaque minute.

Un traitement logiciel différé nous permet, sur de grandes périodes (quelques heures typiquement), de comparer les résultats obtenus sur les deux voies. Des fonctions de répartition des erreurs peuvent être ainsi tracées pour différentes durées de mesure. La figure (7) présente un exemple de résultats obtenus.

L'ensemble du système de mesure se présente donc simplement suivant le schéma de principe présenté figure (8).

3.3 Les tests de validation du système

Afin de s'assurer de la reproductibilité des mesures, nous avons envoyé le même signal provenant d'une des antennes sur les deux voies. Les résultats obtenus restent très voisins. Les fonctions de répartition des erreurs sur les deux voies sont quasiments identiques (voir figure (9)). Les petits écarts peuvent être dus aux caractéristiques techniques légèrement différentes des deux voies de traitement qui étaient, rappelons le, de type opérationnels et n'avaient donc pas subi de réglages particuliers.

3.4 Les possibilités annexes : le filtrage de polarisation

Le système expérimental de réception disposant de quatre antennes actives, nous avons pu débuter des tests visant à

FIG. 2 — *Filtre de canal lié au premier mode seul*

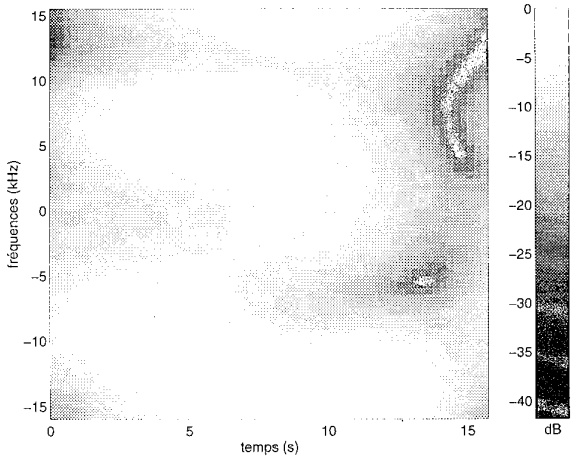


FIG. 3 — *Filtre de canal lié au second mode seul*

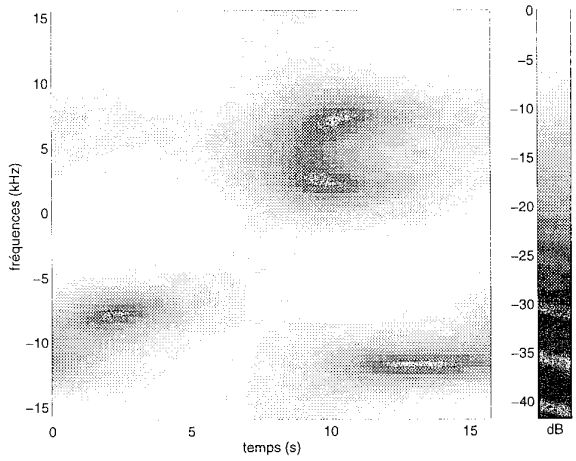


FIG. 4 — *Filtre de canal lié à la somme non pondérée du premier et du second mode*

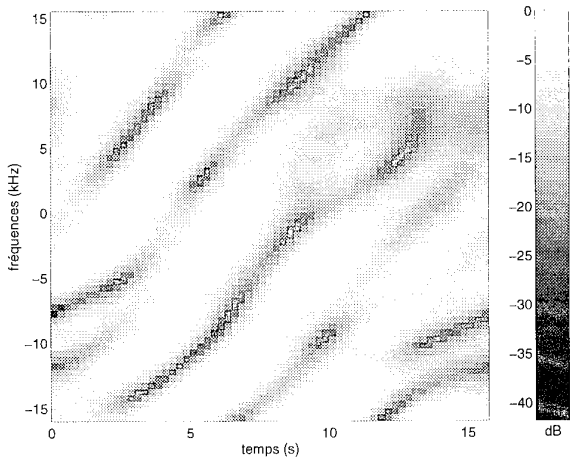
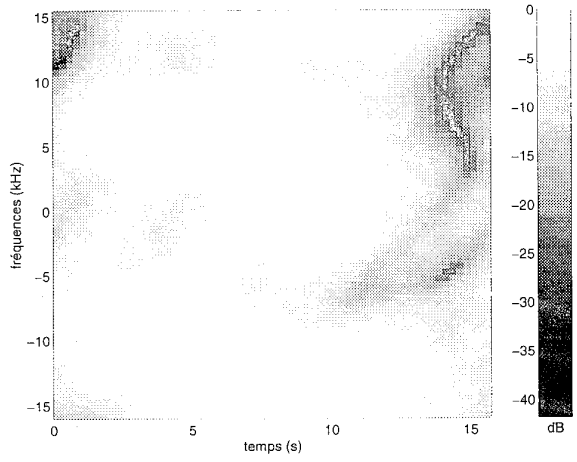


FIG. 5 — *Filtre de canal lié à la somme pondérée du premier et du second mode; la pondération introduite simule un filtrage de polarisation réduisant de 90% la contribution du second mode.*



mettre en évidence le «filtrage de polarisation». Un filtre est constitué en associant les signaux de deux ou trois antennes de façon à amplifier un mode de propagation ou a en atténuer un autre. Les coefficients (complexes) qui interviennent dans le filtre, sont liés aux réponses complexes des antennes et au critère d'optimisation choisi [11].

Dans notre expérimentation, nous disposions de déphaseurs – atténuateurs commandés numériquement. Seule la possibilité déphasage a été utilisée (le système était donc sous-optimal). Deux ou trois antennes ont été associées pendant les essais. Le critère d'optimisation choisi s'appuyait sur les prévisions de propagation pour la liaison : ainsi nous pouvions, en relation avec les prévisions, favoriser la réception d'un mode O ou X par une couche F_2. Lorsque les ondes sont polarisées circulairement et lorsque les antennes sont isolées dans l'espace, les déphasages à introduire sont de $\pm\pi/2$ selon le mode choisi. Dans l'expérimentation, ces déphasages étaient calculés par optimisation pour chaque heure et appliqués automatiquement au système. La figure (10) décrit l'un de ces dispositifs de filtrage.

4 ANALYSE DE QUELQUES SÉQUENCES DE RÉSULTATS

4.1 Le rôle des antennes

L'ensemble des données fait apparaître des résultats très variables d'une antenne à l'autre. Les figures (11 – 12) relatives à des mesures obtenues à partir d'antennes dipôles horizontales disposées en Est–Ouest et Nord–Sud montrent que :

- Les erreurs sur les deux voies n'apparaîssent pas au même moment en général (ce qui est conforme au modèle théorique) ;
- l'importance relative des erreurs peut varier d'un instant à l'autre.

Les fonctions de répartition obtenues à partir de deux séquences successives de deux heures de mesure traduisent cet état de fait comme le montrent les figures (13) et (14).

La comparaison entre les erreurs associées aux antennes horizontales et verticales montre l'avantage à utiliser de

FIG. 6 – *Évolution de la bande de cohérence au cours du temps pour les différents filtres. De haut en bas : filtre lié au 1^{er} mode, filtre lié au 2^{e} mode, filtre lié à la somme des deux modes et enfin filtre après traitement de l'un des modes. L'axe vertical est gradué en kHz.*

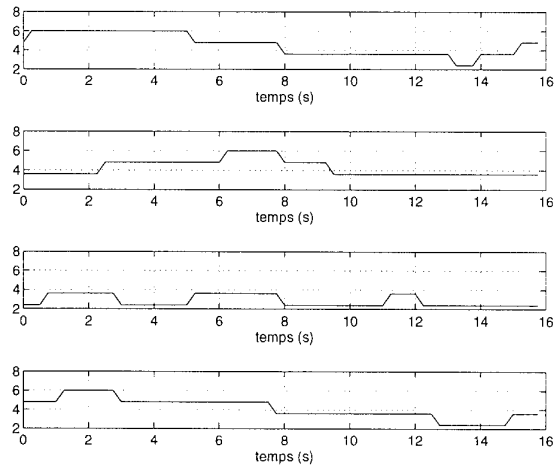
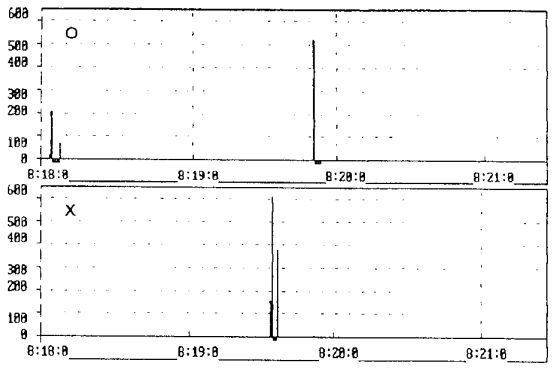


FIG. 7 – *Exemples de détermination du taux d'erreur en fonction du temps sur deux voies. Les instants de désynchronisation sont indiqués sur l'axe des temps en souligné gras. ($f \sim 10.4MHz$)*



jour, sur une telle liaison, l'antenne verticale (exemple figure (15)). Le nombre d'erreurs est en général moindre avec l'antenne verticale quelque soit le moment de la mesure.

TAB. 1 – *caractéristiques techniques des systèmes utilisés*

émission	antennes	losange horizontal
Berlin	puissance émise	1 kW
	fréquences	variables dans la gamme HF de 3 à 15 MHz
	code	pseudo aléatoire sur 9 bits
réception	antennes	antenne tripôle Rodhe Schwarz modifiée + antenne fouet verticale
Lannion	récepteurs	TRT
	modems	série 2 400 bits/s TRT MDM 1224
	système de mesure	réalisation CNET

FIG. 8 – *Schéma de principe de l'ensemble du système de mesure.*

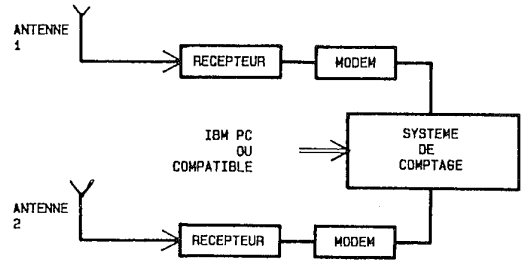


FIG. 9 – *Validation du système : fonctions de répartition des erreurs sur les deux voies, les entrées étant connectées à la même antenne. Durée de la mesure : 4 h. ($f \sim 11.1MHz$)*

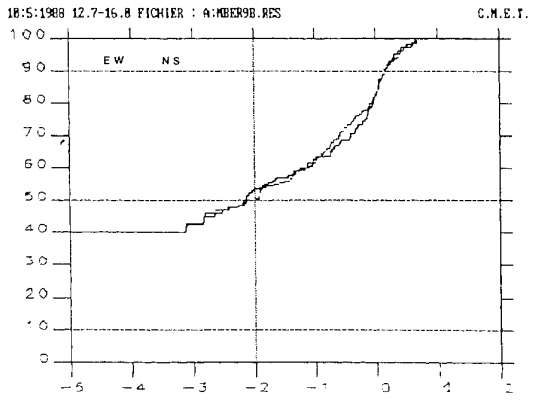


FIG. 10 – *Système de filtrage de polarisation réalisé à partir de deux antennes horizontales*

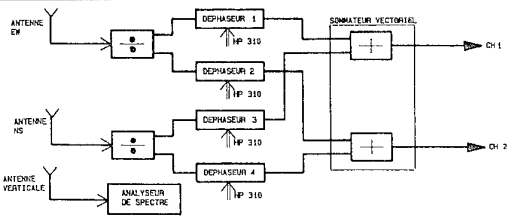


FIG. 11 – *Évolution des erreurs en sortie du système connecté aux antennes actives EW et NS. ($f \sim 11.1MHz$)*

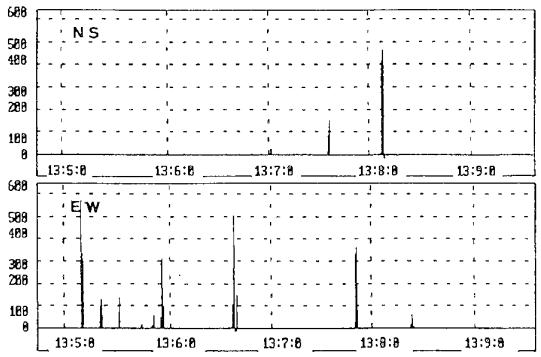


FIG. 12 – *Évolution des erreurs en sortie du système connecté aux antennes actives EW et NS. ($f \sim 11.1MHz$)*

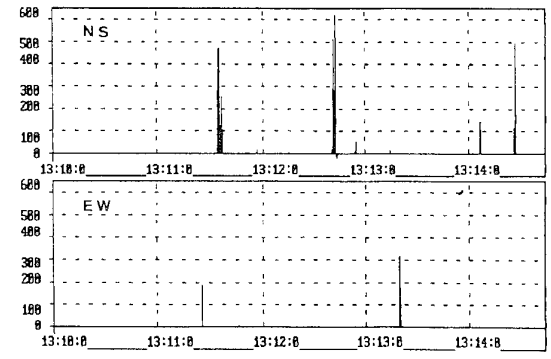


FIG. 13 – *Fonctions de répartitions des erreurs observées en sortie du système connecté aux antennes actives EW et NS. (12 h - 14 h).*

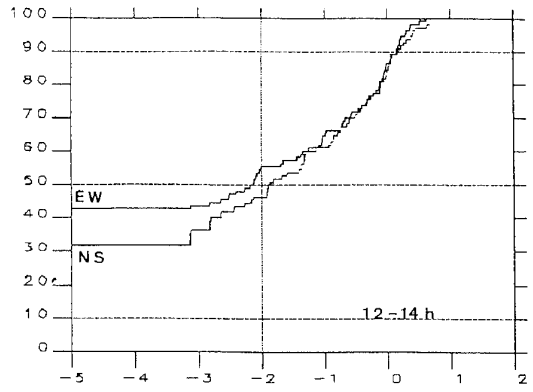


FIG. 14 – *Fonctions de répartitions des erreurs observées en sortie du système connecté aux antennes actives EW et NS. (14 h - 16 h).*

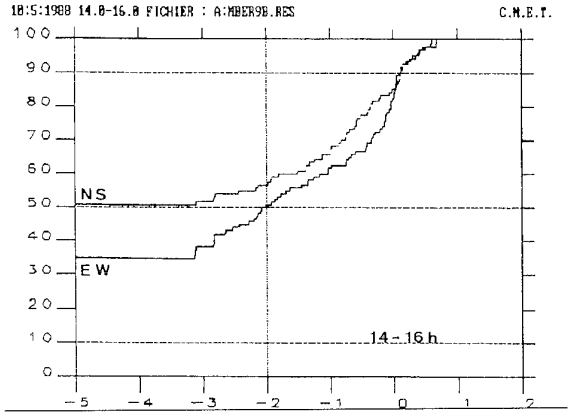
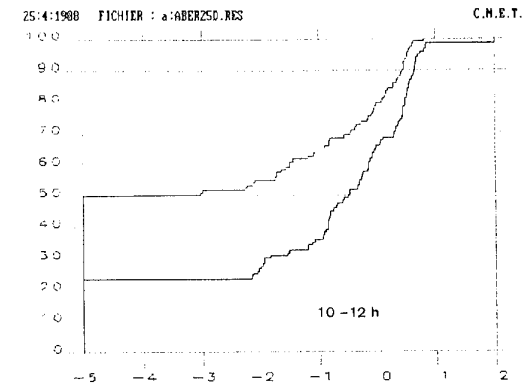


FIG. 15 – *Fonction de répartition des erreurs observées en sortie du système connecté aux antennes actives vertical et NS.($f \sim 13.8MHz$) La courbe supérieure est celle associée à l'antenne verticale, la courbe inférieure est celle associée à l'antenne NS.*



4.2 Introduction au filtrage de polarisation

Compte tenu du système, plusieurs combinaisons fitrage–référence ont pu être testées :

- Filtrage simultané des modes O et X correspondant à un mode de propagation attendu (en général $1F_2$) ;
- filtrage alterné des modes O et X (15 mn pour chaque mode) et comparaison avec une antenne de référence choisie en fonction du rapport S/B observé sur chacune des antennes en début d'expérimentation. En général, de jour une antenne verticale était prise comme référence et de nuit, une antenne horizontale.

De la même façon, plusieurs méthodes de filtrage ont été testées ; elles s'appuient sur la connaissance à priori de la réponse des antennes et des prévisions de propagation. Compte tenu du faible angle d'élévation de réception correspondant au mode $1E$ (environ 8°) et du faible gain de l'antenne d'émission pour cet angle, ce mode de propagation a été ignoré dans le processus de filtrage et seul le mode $1F$ a été pris en compte.

Il ressort des différentes expérimentations les points suivants :

- les filtrages des modes O et X simultanément laissent apparaître des erreurs ou des désynchronisations en des instants différents, sans doute dues à la présence simultanée de modes $1F_1$ et $1F_2$ de même type (O ou X). On retrouve alors ce qui a déjà été observé à partir des antennes élémentaires (voir figure (7)) ;
- suivant les conditions de propagation, nous pouvons obtenir une réduction importante du taux d'erreur. C'est ce que traduisent les figures (16) et (17) où sont résumées une nuit entière de mesures. Durant cette période, nous avons déterminé les erreurs en filtrant alternativement les modes O et X à partir des antennes horizontales Est–Ouest et Nord–Sud, la référence restant une antenne horizontale Nord–Sud.

La réduction du taux d'erreur sur une minute exprimée en puissance de 10 montre, dans ce cas, l'intérêt d'un filtrage de type O et au contraire, le désintérêt d'un filtrage de type X (cf figure (17) présentant quelques séquences de 15 mn de filtrage).

Dans certaines conditions de propagation, cette réduction du taux d'erreur peut conduire à l'absence d'erreur comme représenté figure (18). Dans ce dernier cas, outre une transmission par $1E$ fortement atténuée compte tenu des gains d'antennes, les conditions de propagation ne conduisaient qu'à une propagation par $1F_2$ O et X. L'élimination d'un des deux modes supprime les causes d'erreurs, ce qui est observé.

FIG. **16** – *Évolution de l'effet du filtrage de polarisation sur 10 h de mesures. « 1 » sur l'axe vertical correspond à une réduction du taux d'erreur d'un facteur 10. ($f \sim 4.7MHz$)*

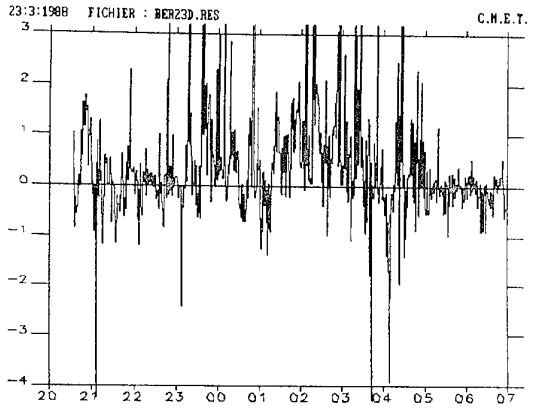


FIG. **17** – *Effet de zoom sur la figure précédente montrant l'intérêt dans ce cas de filtrer le mode O.*

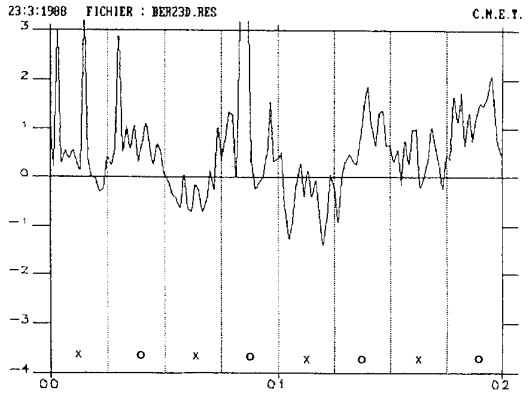
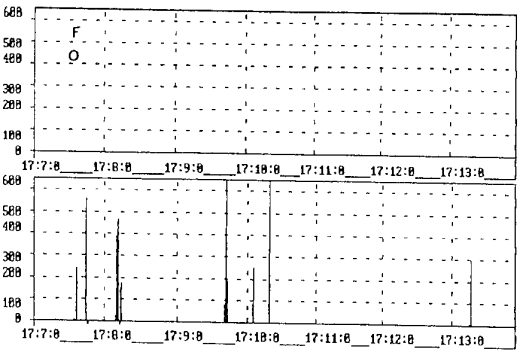


FIG. **18** – *Filtrage de polarisation (mode O) éliminant les erreurs et voie de référence (antenne verticale). ($f \sim 10.4MHz$)*



5 CONCLUSION

Nous avons montré le rôle important de la réponse des antennes dans les liaisons numériques par voie HF. Les essais partiels de filtrage de polarisation ont conduit à des résultats encourageants bien que la méthode utilisée ne soit pas optimale.

Ainsi, dans la conception des systèmes futurs, peut-on envisager des architectures utilisant N entrées parallèles connectées à N antennes colocalisées ou distribuées spatialement ; ces N antennes pourraient être de type différent. Les algorithmes de resynchronisation des voies et de traitement restent à développer.

6 REMERCIEMENTS

L'expérimentation présentée a été réalisée au CNET de Lannion B. Les auteurs remercient la DCN (Direction des Constructions Navales) et la société TRT pour leur soutien dans cette étude. Ils remercient également Mrs POITEVIN, RUIZ, LESTAVEL, DU CHAFFAUD et les personnels de la station d'émission pour leur aide dans la réalisation des systèmes ou campagnes de mesure.

Références

[1] P. SCALART and Y.M. LE ROUX. Evaluation of a high speed digital communication system over ionospheric channels. In 7^{th} *International Ionospheric Effects Symposium*, Washington DC, VA, US, may 1993.

[2] I.K. NISSOPOULOS, C.E. DIMAKIS, S.S. KOURIS, and Y.M. LE ROUX. Preliminary results on performance evaluation of concatenated coding schemes over rayleigh fading ionospheric channels. In 6^{th} *Internationnal Conference on* H.F. *Radio Systems and Techniques, Proc IEE 392*, New York, NY, US, July 1994.

[3] L. BERTEL, A. EDJEOU, V. MASSOT, and Y. ERHEL. Méthode itérative de filtrage de polarisation : application a la goniométrie H.F. In *Quatorzième colloque GRETSI Le Traitement du Signal et des Images*, Juan les pins, 13–16 Septembre 1993.

[4] Y. ERHEL, A. EDJEOU, and L. BERTEL. Contribution of the polarization diversity in H.F. direction finding systems. In *Advanced Signal Processing; Algorithms, Architectures and Implementations*, San Diego CA, 24–29 July 1994. SPIE's 1994 International Symposium.

[5] L. BERTEL, J. ROJAS–VARELA, D. COLE, and P. GOURVEZ. Polarisation and ground effects on H.F. receiving antenna patterns. *Annales des télécommunications*, 44(7–8), Juillet–Août 1989.

[6] Y.M. LE ROUX, J.P. JOLIVET, and J. MENARD. Analyse expérimentale, modélisation et simulations du canal de transmission ionosphérique,. In *Conférence SEE*, Rennes, France, Septembre 1991.

[7] Y.M. LE ROUX, M. NIBERON, R. FLEURY, J. MENARD, and J.P. JOLIVET. H.F. channel modelling and simulation. In 5^{th} *International Conference on Radio Receivers and associated Systems, Proc IEE 325*, Cambridge, July 1990.

[8] P.A. BELLO. Characterization of time-variant linear channels. *IEEE Trans. Com. Sys.*, CS(11), December 1993.

[9] L.E. VOGLER and J.A. HOFFMEYER. A model for wideband H.F. propagation channels. *Radio Science*, 28(6):pp 1131–1142, November–December 1993.

[10] C. LESTAVEL, L. BERTEL, G. DU CHAFFAUT, R. POITEVIN, and J. ROJAS–VARELA. Système pour IBM–PC et compatible permettant l'évaluation des transmissions numériques en H.F. Technical Report 338, NT/LAB/MER, 1988.

[11] L. BERTEL and J. ROJAS–VARELA. Antennes H.F. adaptées à la polarisation des signaux reçus : exemples de résultats. In *J.I.N.A.*, Nice, 1986.

SISP-A Software Tool for Propagation Prediction

Prabhakar M. Koushik, Theodore S. Rappaport, Mansoor Ahmed, and Ning Zhang

Mobile and Portable Radio Research Group
Bradley Department of Electrical Engineering
Virginia Tech, Blacksburg, VA 24061 - 0350, USA

Abstract: Accurate modeling of radio propagation characteristics in and around built-up urban environments is vital to the successful functioning of surveillance and tracking systems utilizing wireless technology. A vast majority of the software tools currently available to military and civilian users are based on very simple models with fairly limited accuracy and are closed architectures. We discuss the design and development of a software tool that provides: a) a framework for modelers to incorporate new models and test their hypotheses, and b) system designers and planners to obtain the propagation characteristics in any environment. The interactive software executes on a workstation, uses topographic maps with building overlays to predict signal coverage and channel characteristics for user-specified antenna locations. Work is in progress to improve the accuracy of the prediction software through better diffraction modeling.

I. INTRODUCTION

While aircraft surveillance has been used extensively for drug interdiction, it is expensive. A surveillance method that's cheap by comparison is tracking suspect cargo and vehicles by placing small hidden transponders on them, capable of relaying position data obtained by a hidden onboard GPS receiver. In such circumstances, the antennas used may well be suboptimal, and it is crucial to know a priori the best locations for beacons or receiving posts within a city to ensure reliable tracking of the suspect. A thorough theoretical understanding and modeling capability of radio propagation characteristics in and around built-up environments is vital to the successful functioning of surveillance and tracking systems utilizing wireless technology. Accurate propagation models, based on real-world phenomena, are needed for rapid installation, hardware development, and spectrum utilization plans for future wireless communication systems. This paper presents SISP, a software tool for propagation prediction and wireless system design.

Radio communication channels are subjected to varying conditions in which the signal travels from the transmitter to the receiver via multiple paths. The received signal strength can be small or large depending upon whether the signals combine destructively or constructively. Thus, it is important to predict the spatial distribution of power, the interference, the propagation power loss, etc., to completely characterize the propagation environment.

Typically, measurements are made to determine the characteristics of the propagation channel. However, these require a partial installation of the system and are time consuming and expensive. Moreover, they cannot be used for determining the performance of the system in the presence of different terrains. Over the years, terrestrial communication systems have been installed using empirical or semi-empirical propagation models. While many experimental or theoretical models have been developed to predict radio propagation in land mobile systems, they are not general enough to characterize the different propagation environments. Site specific propagation models, which utilize information about the environment such as the terrain elevation and morphology of buildings, etc., hold great promise for accurate predictions in different environments.

Site specific propagation models typically utilize large databases of diverse site information. A software tool that facilitates management of these databases can significantly ease the complexity of using them. The visualization of results generated by these models is also significant, since this provides feedback to the user about the performance of the system. Besides, the software tool can be utilized to analyze and compare predicted and measured data which can be used for improving the accuracy of the propagation models.

SISP provides a flexible software platform for propagation prediction, performance analysis, and efficient database resource management. The framework of SISP encompasses propagation prediction modules, the graphical user interface, a suite of display and analysis routines, and data processing routines. The general framework of SISP is shown in Figure 1. The modular structure allows for rapid development, easy extensibility, and enhancements to the software. Section II of this paper discusses the resource organization of SISP and presents the various databases that are used. The principal component of SISP is the propagation prediction module which contains the routines that implement the various propagation models. Section III of the paper discusses the propagation models that SISP currently incorporates. The user

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

interacts with the propagation models through a robust and user friendly graphical interface. Section IV discusses the predictions of the model and compares it to measured results. Section V summarizes the work done and states the objectives of the research effort currently underway.

II. RESOURCE MANAGEMENT

A significant hurdle in the use of site specific models is the diversity of data formats in which the environmental information is available. A considerable amount of work has been done to establish and adopt standards for each type of data. The SISP database comprises terrain elevation data, building contour and elevation data, RF measured and predicted data, and antenna patterns.

A graphical user interface provides the interactive framework of SISP. It furnishes the user with the capability to create and maintain a database of site specific information and establish a simulation environment. The user can specify transmitters and receivers at a particular point, or along a grid, or a track. Parameters such as transmitter power, frequency of operation, height of antenna, etc. can also be entered via the graphical user interface.

In an outdoor microcell environment, terrain and buildings are the dominant surfaces that influence the transportation of energy. Accordingly, the details of the outdoor environment need to be represented accurately in order for the models to provide accurate predictions. AutoCAD [1], a general purpose computer-aided design package is generally used to describe the building data for propagation prediction purposes. The very general way of representing information in AutoCAD represents a significant processing challenge to the prediction routines. Our work has identified a subset of AutoCAD features that is general enough to represent the terrain, morphology, outdoor building databases, and building floor plans while keeping the processing overhead on the propagation prediction routines to a minimum. Details on the Real World Database Format (RWDF) can be found in [2].

The terrain elevation map and the building contour map represent georeferenced data. The geodetic reference system used to reference the locations displayed can be specified by the user. SISP primarily supports the Universal Transverse Mercator (UTM) coordinate system. This is a planimetric coordinate system and has the advantage of using linear decimal units for complete coordinate descriptions. This simplicity of notation offers greater precision and computational convenience. SISP also supports the Geographical Coordinate System (lat-long).

The terrain elevation data for SISP is obtained from the Digital Elevation Models (DEM) supplied by the USGS [3]. The DEM's are available in 7.5' quadrangle coverage and 1^{o} quadrangle coverage. SISP utilizes the 7.5' DEM's, with a scale of 1:24,000. The 7.5' DEM consists of a regular array of elevations referenced horizontally in the UTM coordinate system. The elevation data is stored in profiles in which the spacing along and between the profiles is 30m.

The terrain data format in the SISP database conforms to the GRASS 4.0 raster data format GRASS (Geographic Resources Analysis Support System) is a geographic information system (GIS) designed and developed by the U.S. Army Construction Research Laboratory (USACERL) [4].

The General Data Format (GDF) was developed, for SISP, to provide a standard data format for importing RF measured and predicted data into the database. The GDF is an extensible machine independent ASCII format. It is modular and changes, based on changing requirements, can be easily accommodated. The primary objective in the design of this format is to provide a user editable data format for storing measurements made using various measurement systems.

While GDF is suitable for providing a common format to represent RF measurement and prediction data, it does not lend itself easily to analysis and visualization because of the sequential ordering of data sets that the format enforces and the significant quantity of extraneous descriptive information present in the file. For effective analysis and visualization of data, rapid retrieval is an important requirement. Hence, a indexed file structure that allows random access to the data sets is essential. The data stored in GDF is thus converted to a binary format called the Network Common-Data Format (NetCDF) [5].

III. P OPAGATION MODELS

A few empirical models such as the exponential path loss, and the Hata model have been incorporated into SISP to demonstrate the feasibility of the proposed architecture. The more complex site specific models are now being incorporated into the same framework. Site-specific models that take into account the physical characteristics of the environment are capable of providing more accurate prediction results than statistical models. By using the concept of ray tracing, energy transportation from a discrete point source can be represented by rays, and the interaction of the rays with objects in the environment may be modelled using well known concepts of transmission, reflection, and diffraction. This assumption is reasonably accurate at high frequencies when objects of interest in the propagation environment are far larger than the wavelength.

The signal in a microcellular environment can propagate to the receiver via multiple paths due to reflection, diffraction, and scattering of the radio waves by the objects in the environment. Using simplified geometric optics assumptions,

the prediction software launches and traces rays in three dimensions, and determines the path loss and propagation delay for each ray. The model determines the path through which the direct line of sight, specularly reflected, diffusely scattered, and diffracted rays arrive at a receiver. The received rays are then used to estimate the channel power delay profile and subsequently the path loss at the receiver location. Thus, ray tracing can help estimate not only the deterministic amplitudes of each type of ray, but also the exact time of arrival of the energy at a specific point in the coverage region. This may then be used to compute the statistical frequency and time domain channel characteristics including path loss, and hence received signal strength (RSSI).

The transmitter is modeled as a point source generating rays uniformly in all directions. To model energy transfer correctly from the transmitter, it is necessary to launch rays at all possible angles of departure with a constant angular separation between rays. An elegant approach is to launch rays through the vertices of an icosahedron inscribed in an unit sphere, with each of its triangular faces subdivided into smaller triangles. This method for launching rays provides wavefronts of equal shape and area that can be easily subdivided as may be infered from Figure 2.

A brute force recursive technique is used to trace all the rays after they are launched from the transmitter so as to determine each ray path by which significant levels of energy radiated from the specified transmitter reaches a receiving point. Briefly, the prediction program performs the following:

Check the existence of a line of sight path for every receiver. If such a path exists, the path loss is calculated for the direct ray and the ray parameters are stored. If a direct path does not exist between the transmitter and the receiver, check if a diffracted ray can be received. If a valid diffraction path exists, the received energy due to this path is calculated. To predict energy arriving at a receiver location due to specular reflection or scattering, rays are launched from the transmitter location and each ray is traced till it intersects an object. If no object intersects the ray, a new ray is launched from the transmitter and traced. If a ray intersects an object, a ray path is drawn from the point of intersection to the receiver. If this path is unobstructed, the ray is tested for specular reflection. A reflected ray is received at a point if it lies within a reception sphere around the receiver point. In case the received ray is not specular, it is considered to be a scattered path and the RSSI is computed accordingly. The reflected ray is now launched from the point of intersection and treated as a new ray. This recursion continues until a user specified number of ray object intersections has been exceeded. A new ray is then launched from the transmitter.

The received power at any location is the vectorial sum of the multipath components of the LOS, reflected, diffracted, and scattered rays. The total field at a receiver is determined by the coherent summation of the individual contributions of each ray. The following sections describe the individual effects of the multipath components.

A. Line of Sight (LOS) Rays

The path loss with respect to 1m free space is directly calculated from the distance between the two points (d) using the relation

$$E_{LOS} \propto \frac{1}{d} \tag{3.1}$$

Path loss measurements were made in Rosslyn, VA to test the accuracy of the line of sight model. The predicted path loss for line of sight receiver locations were found to be 8 dB lower (typically) when compared to the measured path loss. Careful analysis of measured and predicted results showed that this difference could be explained by considering ground reflections. It is known that the phase of the plane wave reflection coefficient, for both vertical and horizontal polarization at grazing incidence is 180 degrees. This would mean that the LOS signal and the ground reflected signal arriving at the receiver have a phase difference of 180 degrees. Therefore for such cases the RSSI due to these two paths must be subtracted. Equation (3.2) indicates the new model.

$$E_{LOS} \propto (E_{fs} - E_{tr}) \tag{3.2}$$

where E_{fs} is the free space field as calculated from Equation (3.1) and E_{tr} is the field due to terrain reflection for low lying transmitters.

B. Reflected Rays

Specular reflection or reflection is characterized by the incident and reflected rays making equal angles with the surface normal. Rays r_1 and r_2 in Figure 3 are reflected rays. The relationship of the reflected wave to the incident wave is given by

$$\begin{bmatrix} E^r_H \\ E^r_V \end{bmatrix} = \boldsymbol{R}^T \boldsymbol{D} \boldsymbol{R} \begin{bmatrix} E^i_H \\ E^i_V \end{bmatrix} \tag{3.3}$$

where E^r_H and E^r_V are the horizontally and vertically polarized components of the reflected wave and E^i_H and E^i_V are the horizontally and vertically polarized components of the incident wave. Matrix **R** is the transformation from vertically and horizontally polarized components to components perpendicular and parallel to the plane of incidence and is given by

$$R = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad (3.4)$$

where θ is the angle between the two sets of axes discussed above. The matrix **D** is given by

$$D = \begin{bmatrix} \Gamma_{\perp} & 0 \\ 0 & \Gamma_{II} \end{bmatrix} \quad (3.5)$$

where $\Gamma_{\perp}$ is the reflection coefficient for the wave with E-field in a direction perpendicular to the plane of incidence (perpendicularly polarized) and $\Gamma_{\perp}$ is the reflection coefficient for the horizontally polarized wave. Both the reflection coefficients depend on the angle of incidence and the material of the reflecting surface. The material dependence is limited to the relative permittivity (ε_r) of the material of the reflecting surface for dielectrics, particularly at high frequencies. This method of calculating the reflected energy is based on the assumption that the reflection occurs at the interface between two infinite media.

A reception sphere around the receiver location is used to determine the reception of a specular reflected ray at the point. The radius of the sphere is directly proportional to the total path travelled by the ray and the angular spacing between neighboring rays at the source. Each ray represents the field in the solid angle radiating from the transmitter. If the ray after reflection from one or more surfaces intersects the sphere, it contributes to the total energy received at that location.

C. Scattered Rays (Rough Surface Scattering)

In case a ray intersects a rough surface, the energy is also propagated in directions other than the specular direction. For receiver locations near a wall or a building surface where no specular component is predicted, energy due to the proximity of the surface is predicted by rough surface scattering. The bistatic radar (bistatic refers to the fact that the transmitter and receiver are at different locations) equation may be used to model the scattered energy. The power received due to a scatterer (P_s) at a distance d_r from the receiver and d_t from the transmitter is given by:

$$P_s = \frac{P_t G_t G_r \sigma \lambda^2}{(4\pi)^3 d_t^2 d_r^2} \quad (3.6)$$

where P_t is the transmitted power, G_t is the gain of the transmitter antenna, G_r is the gain of the receiver antenna, λ is the wavelength and σ is the bistatic radar cross section of the scattering object. The bistatic radar cross section is defined as the ratio of power density in the scattered signal in the direction of the receiver to the power density of the signal incident on the scattering object [6].

The models for calculating the bistatic radar cross section of an object are applicable only in the far fields. At 900 MHz and a 10m wall, the far field is 667 m. Hence, the above models cannot be used directly to predict rough surface scattering in microcells where distances of about 1m may be encountered. A heuristic approach [12] that divides a large surface into small facets so that the receiver is in the far field of each facet is used to apply these models. Scattering in non-specular directions is then given by the non-coherent summation of power scattered form each of these facets. The geometry is shown in Figure 3. The size of the facets depends on the angular spacing between the rays at the transmitter and on the distance travelled by the ray to reach the scattering surface. To calculate the bistatic radar cross section of each of the facets the following equation is used:

$$\sigma_i = \frac{4\pi (\Delta x \Delta y)^2 \sigma_r \cos\theta_i}{\lambda^2} \quad (3.7)$$

where θ_i is the angle the scattered ray makes with the specular direction, Δx and Δy are the dimensions of the facet, λ is the wavelength and σ_r is an additional loss factor. This factor is added because most of the surfaces encountered are not perfectly conducting and the material of the surface must be used to calculate this factor.

D. Diffracted Rays

According to geometric optics, reflection, refraction and direct line of sight rays are the three basic mechanisms of energy propagation. The theory breaks down in shadow regions (dominant in urban areas), where neither a direct ray nor a reflected ray is received. When a receiver is heavily shadowed, a significant portion of the received signal is due to energy diffracting over, or around the building edges. Diffraction supplements the geometric optics theory to determine non-zero field in shadow regions. It also eliminates the sharp field transition that is observed between the shadow and lit regions by using GO. The geometrical theory of diffraction [10] extends geometric optics and introduces diffracted rays to account for energy received due to diffraction. Knife-edge and wedge diffraction has commonly been used to account for diffracted energy. The extension of the single knife-edge diffraction theory to multiple edges is involved and a number of approximations have been proposed [7][8][9]. Our software traces diffracterd rays in three dimensions. The field at any point is related to the angle of diffraction with the following model[12]:

$$E_i = E_o \sqrt{\frac{\prod_j A^2(\theta_j)}{\sum_j d_j \prod_j d_j}} \qquad (3.8)$$

Here the attenuation factor $A(\theta_j)$ is calculated using:

$$A(\theta) = \frac{1}{\sqrt{2\pi k}}\left[\frac{1}{2\pi+\theta} - \frac{1}{\theta}\right] \qquad (3.9)$$

where θ is the angle of diffraction and k is the wave number.

IV. RESULTS

Predictions were made using the software described above and compared against actual measurements performed in Rosslyn, Va. Rosslyn is a typical urban setting with a wide diversity of building types. Further, the terrain elevation varied from 15m to 55m over the area considered for the study. Figure 4 shows a projection of the area under study together with street names and the location of the transmitter. Figure 5 shows some of the results obtained.

V. CONCLUSIONS

In this paper, we have presented the structure and salient functional features of SISP. The software tool can be utilized in research for propagation prediction and wireless system design. We are currently working on improving the accuracy and the computational efficiency of the site specific propagation model. Efforts are also being made to provide an antenna library and improved visualization capabilities.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support of the Advanced Research Projects Agency (ARPA) for this project. The authors also wish to acknowledge the work of Kurt Schaubach, Scott Seidel, and Joseph Liberti.

REFERENCES

[1] AutoCAD Release 11, Reference Manual, Autodesk Inc., 1990.

[2] S. Sandhu, "Real World Database Format - MPRG Internal Document, 1995.

[3] Digital Elevation Models -- Data Users Guide, Dept. of the Interior, U.S. Geological Survey, Reston VA, 1987.

[4] James Westervelt et al., "Introduction to GRASS 4.0," *GRASS 4.0 Users' Manual,* U.S. Army Construction Engineering Research Laboratory, Illinois, 1991.

[5] NetCDF User's Guide - An Interface for Data Access, Unidata Program Center, April 1993..

[6] C.A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons, New York, 1989.

[7] K. Bullington, "Radio Propagation at frequencies about 30 Mc", *Proc. IRE*, Vol. 35, No. 10, Oct. 1947, pp. 122-1136.

[8] J. Deygout, "Multiple knife-edge diffraction of microwaves", *IEEE Trans. on Antennas and Propagation*, Vol. AP-14, No. 4, April 1966, pp. 480-489.

[9] J. Epstein and D.W. Peterson, "An experimental study of wave propagation at 850 MC", *Proc. IRE*, Vol. 41, No. 5, May 1953, pp. 595-611.

[10] J.B. Keller, "Geometrical Theory of Diffraction", *Journal of the Optical Society of America*, Vol. 52, No. 2, Feb. 1962.

[11] S. Sandhu, P. Koushik, T.S. Rappaport, "Predicted Path Loss for Rosslyn, VA. - second set of predictions", *MPRG Technical Report for ORD*, MPRG-TR-95-, March 1995.

[12] K.R. Schaubach, "Microcellular Radio channel Prediction Using Ray Tracing", *Masters Thesis*, VPI&SU, MPRG-TR-92-15, Aug. 1992.

[13] M.D. Yacoub., *Foundations of Mobile Radio Engineering*, CRC Press Boca Raton, 1993.

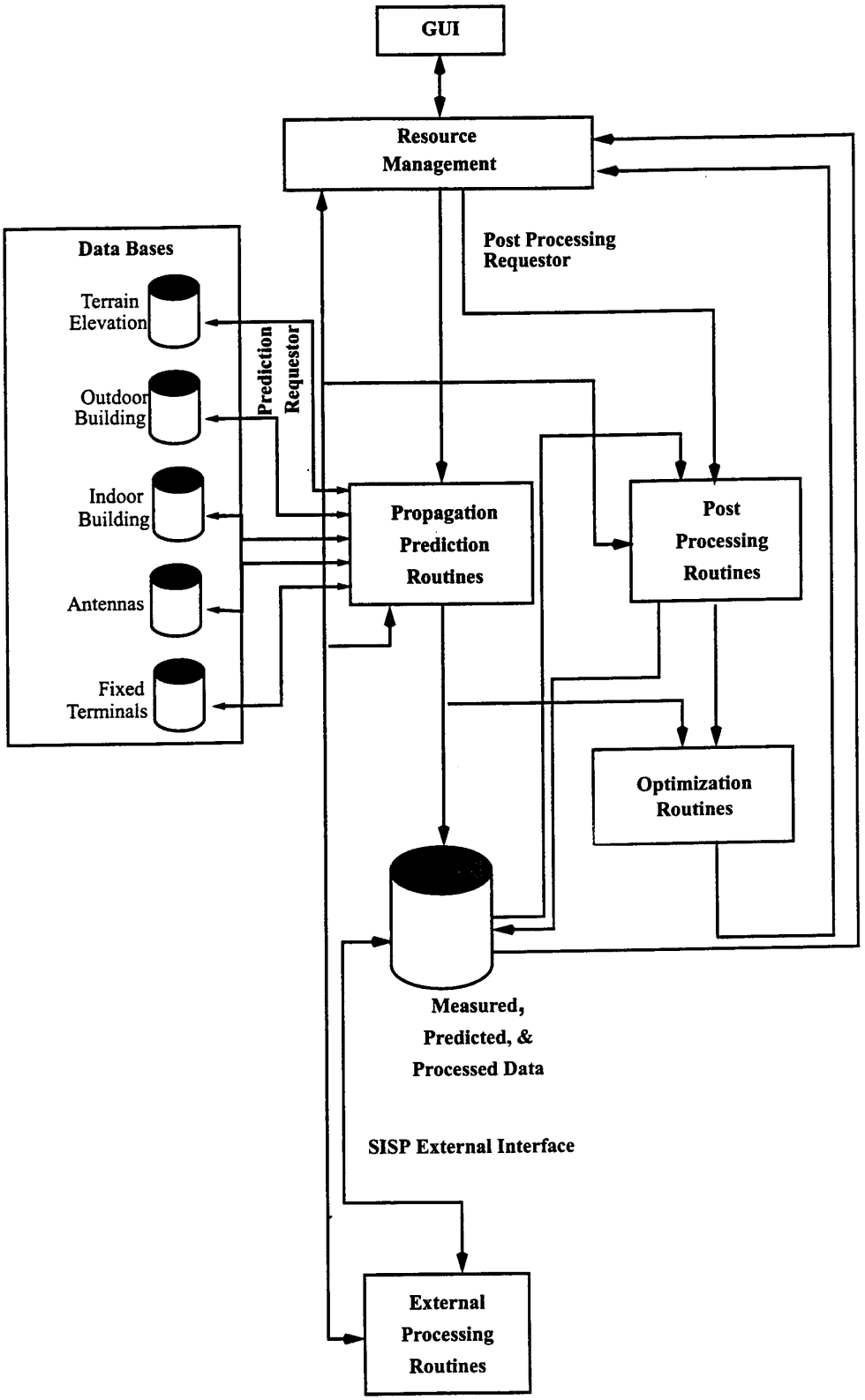


Figure 1. SISP General Framework

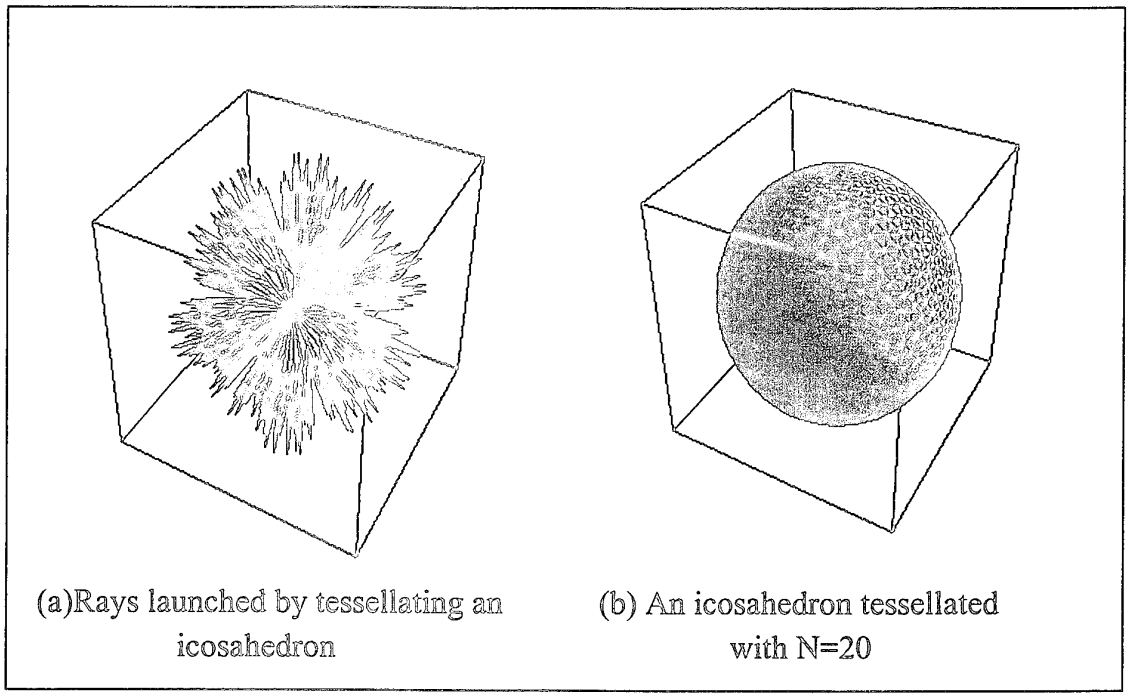


Figure 2. Rays launched through a tesselated icosahedron

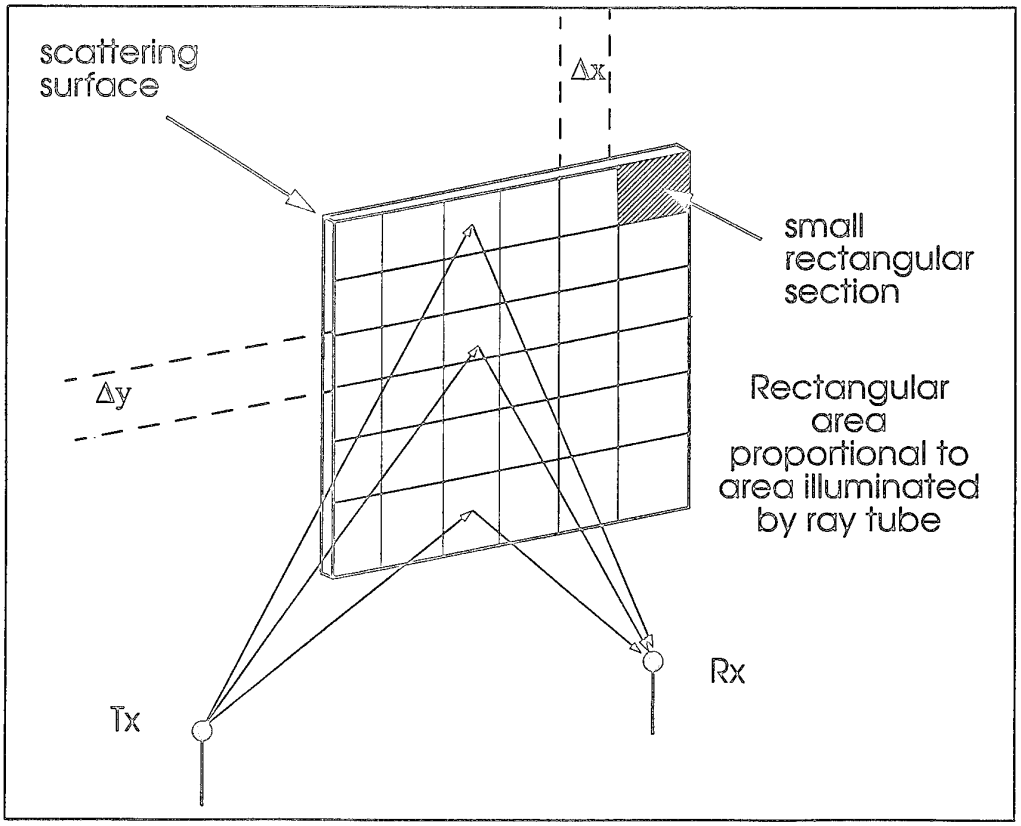


Figure 3. Scattering model for flat surfaces

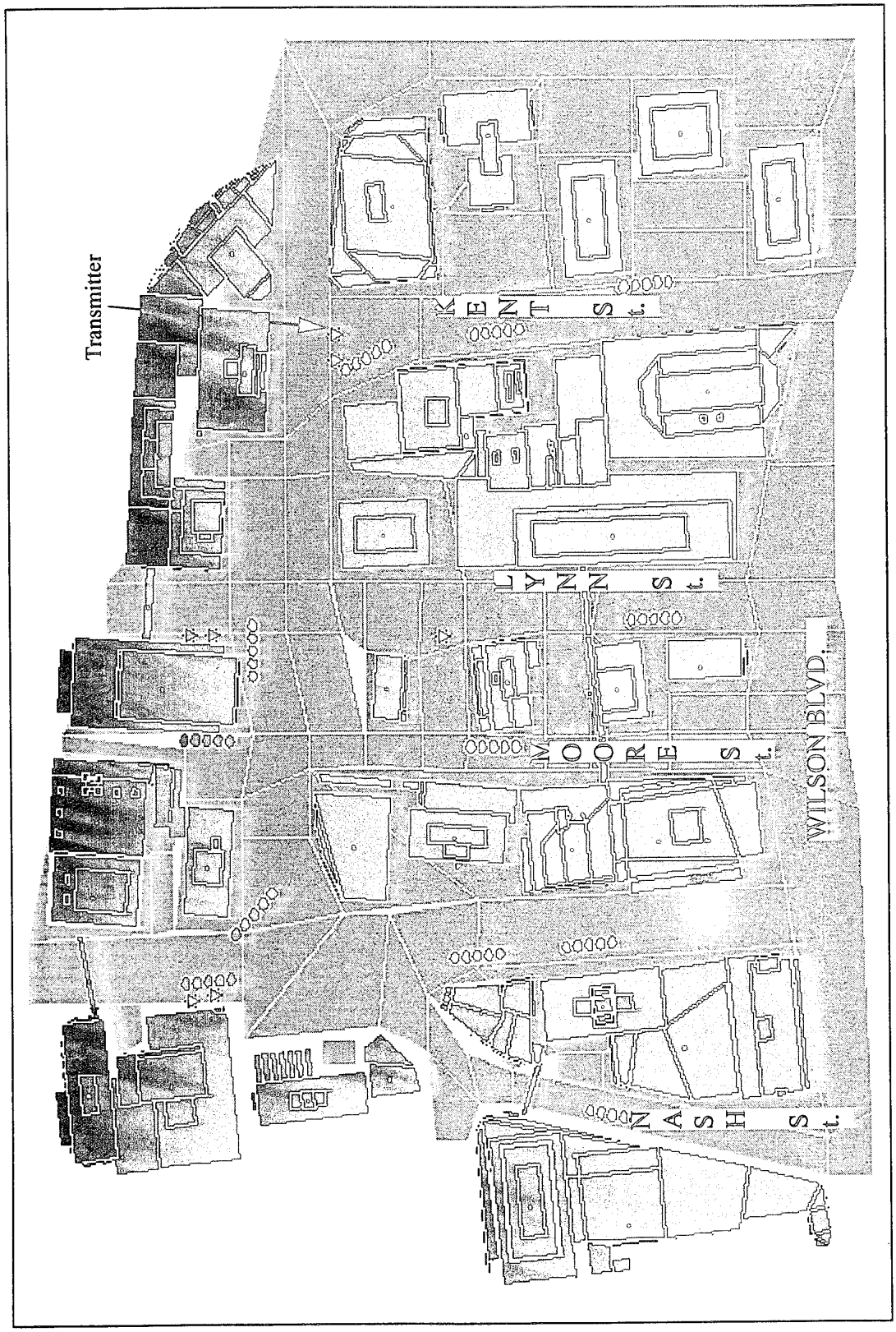


Figure 4 Tx. locations and street names for path loss measurements in Rosslyn, Va.

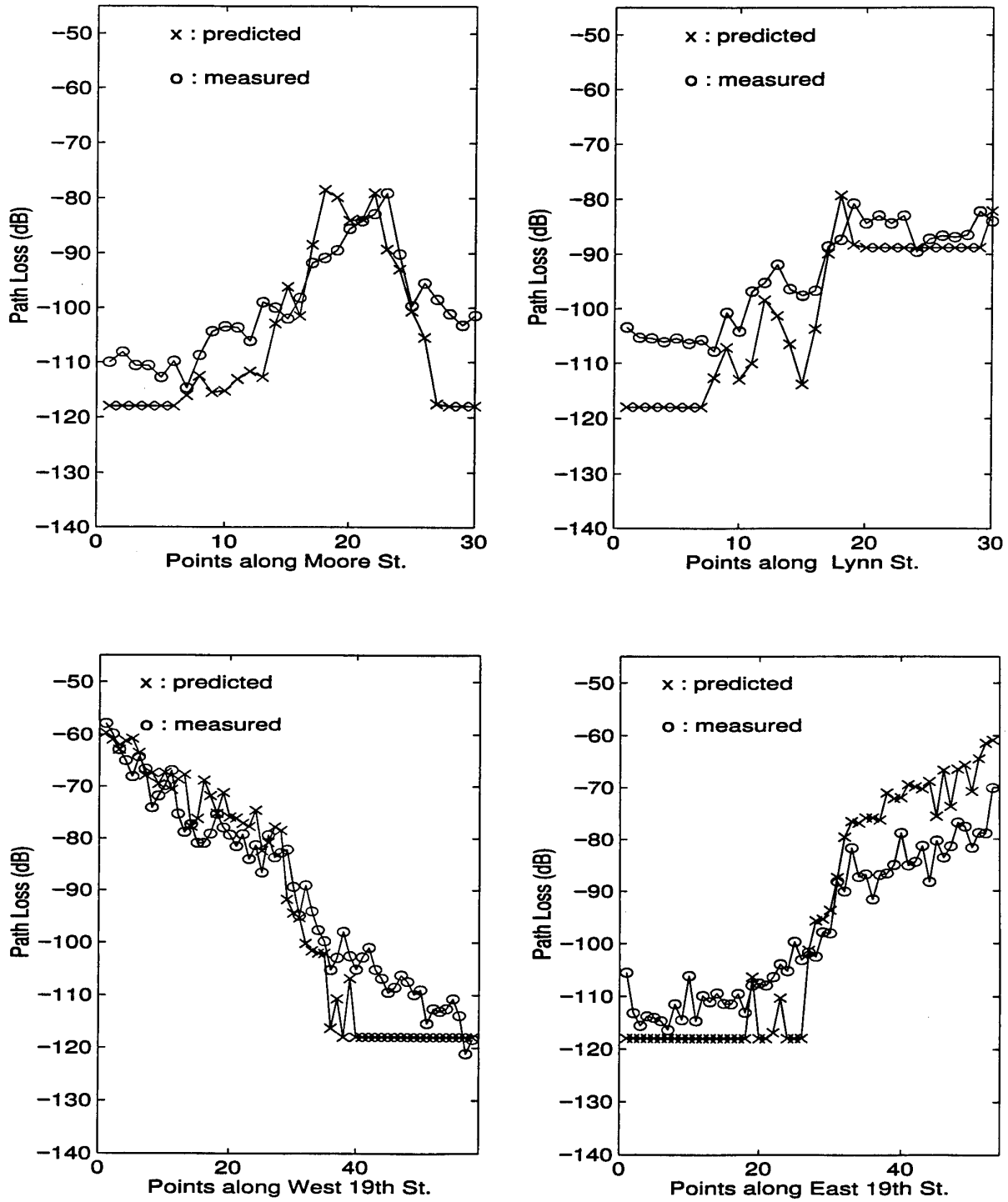


Figure 5. Comparison of measured and predicted path loss along streets in Rosslyn, Va.

DISCUSSION

Discussor's name: K. Craig

Comment/Question:

Commercial site-specific ray trace models are available. Have any of these been included in the validation tests that you (and Professor Luebbers) referred to?

Author/Presenter's reply:

Not to my knowledge, although colleagues in the industry and government indicated to me that those are cruder models than the ones we are testing. I am not certain, however, your suggestion is a good one.

DISCUSSION

Discussor's name: K. Craig

Comment/Question:

Is there a justification for only including non-specular effects at the final reflecting surface? Are not non-specular reflections between any two reflectors equally important?

Author/Presenter's reply:

Our belief is that the last leg of each ray, if it is not offering a specular reflected component, will be the main contributor to the scattering energy. Remember, all rays may offer this type of scattering signal, so the sum of many scattered ray energies will yield the scattered component. Our limited tests show this is very close to measured values, when the RSSI is very weak. However, there is no rigorous justification for this.

DISCUSSION

Discussor's name: J. Harvey

Comment/Question:

1. How many orders of reflection do you typically follow?

2. Could you comment on differences between your model and Professor Luebbers?

Author/Presenter's reply:

1. Twelve are more than sufficient for 1km x 1 km.

2. We use Bonding Volume Hierarchy and consider diffuse scattering on the "last leg" of each ray. We currently use a simpler wedge diffraction model, but have 3-D capability and polarization. We have a variable reception sphere and shoot out rays with a specified angle/spatial separation, as opposed to using Professor Luebbers' technique of specifying a number of rays which nearly guarantees receipt of ray types.

DISCUSSION

Discussor's name: A. Altintas

Comment/Question:

Did you test the sensitivity of your algorithm on the building model? What I mean is that, did you try the algorithm with the slight shifting or rotating of the buildings?

Also, you are developing a commercial code, in that case is it a good idea to use parallel processing techniques?

Author/Presenter's reply:

1. We have not rigorously tested this, although an interesting paper on this appeared in the 1995 IEEE Vehicular Technology Conference Proceedings from a European group. This is an important question since it impacts accuracy and resolution issues when running the models.

2. Parallel processing will be common place in a few years, we believe. Right now, it's best for use in the research laboratory.

DISCUSSION

Discussor's name: E. Van Lil

Comment/Question:

How do you discriminate between the different buildings like concrete, aluminium, etc. Do you include this effect in your program or do you intend to do so?

Author/Presenter's reply:

Right now, we must manually choose to model each building with a specific ε_r. We use a default value of concrete. Ideally, the photogrametry data base would have this information passed to the model via the data base.

HYBRID DIRECT SEQUENCE - FREQUENCY HOPPING CELLULAR SYSTEM

K. Kehagias
Hellenic Air Force Research Center (KETA)
Post Office Terpsitheas
16501 Athens
Greece

SUMMARY

The performance of a hybrid Direct Sequence - Frequency Hopping system for mobile communications is described through computer simulation results. The term hybrid implies the combination of two separate systems operating together within the same physical channel and not a Direct Sequence (DS) signal whose center frequency hops periodically. The DS part supports the high data rate services and the Frequency Hopping (FH) part the low ones.

Initially, a pure DS Code Division Multiple Access (CDMA) system is considered, capable of supporting data communications in a variety of bit rates. The capacity of this system (number of users in the allocated spectrum vs. bit error rate) is calculated. Consequently, low data rate services are removed from the DS system and are transmitted using a FH CDMA technique within the same bandwidth and for the same expected bit error rate performance. Overlapping and non-overlapping hopping patterns are considered. The overall (DS + FH) capacity is calculated and compared with the capacity of the pure DS system.

The simulation model for the DS CDMA system is based on the specifications of CODIT (COde DIvision Testbed), a proposal for next generation's cellular communication system. For the FH CDMA the simulation model is based on a SFH system employing Phase Shift Keying (PSK) modulation.

1. INTRODUCTION

Transfer of information has become more and more important in our society. The general, personal desire to compute or communicate at any time ant from any place, has supported the growth of cellular mobile telephone network, the emergence of wireless PBX, wireless Local Area Networks (LAN's) and many other Personal Communications Services (PCS).

Wireless technology used in these applications faces a number of challenges such as spectrum overcrowding, privacy considerations, fading mobile channels, complexities of the propagation environment in urban areas or inside buildings, etc.

Many of these problems have been tackled for military and aerospace applications, where often complex and costly solutions have been considered acceptable, if the desired performance was achieved. Spread Spectrum (SS) communications have been one of the most intriguing and exiting technologies to emerge from these efforts. However, rapid advances in LSI technology have made it possible to implement the complex functions required for SS within size and cost constraints that make it attractive for consumer products.
Success of commercial SS products such as Global Positioning System and enormous growth potential of mobile communications market shifted the research interest from military to commercial applications.

In mobile communications in particular, the characteristics of SS signaling appear to be well suited to mitigating the problems of multiple access and harsh propagation conditions encountered in a mobile environment.
In terrestrial cellular mobile communication systems, the first SS proposal was for a Slow Frequency Hopping (SFH) system in late 70's. [1]. Another SFH system [2] was a strong contender against GSM for the pan-European digital mobile radio network.
In recent years proper changes in regulations such as worldwide frequency allocation for Future Public Land Mobile Telephone Services (FPLMTS) and the Federal Communications Commission (FCC) frequency allocation for Personal Communication Systems (PCS) within USA, prompted numerous SS proposals for future terrestrial and satellite digital mobile communication systems. Most of these systems employ DS/CDMA as multiple access method with various PSK schemes for data modulation. There are also proposals for hybrid DS/FH [3] and also for FH/TDMA (GSM incorporates a FH option). All new systems are designed to offer best communication quality maximizing bandwidth efficiency. New design features include:

Support of emerging telecommunication services through the exploitation of the concept of 'bandwidth on demand' i.e. the ability to handle information exchange in a variety of data rates according to the provided service (voice, fax, low rate data, high rate data, etc.).

Proper design of the radio interface in order to offer coexistence with other users and systems with little frequency coordination.

COde DIvision Testbed (CODIT), a proposal for a third generation DS/CDMA system for cellular mobile communications is examined in this paper. The project is supported by Ericson. IBM and Italtel. So far a testbed has been implemented and Swedish Authorities allocated the frequency spectrum for initial trials. CODIT is designed to support "bandwidth on demand" by adjusting the spreading factors according to the information rates. However, use of large spreading factors for low capacity users may be considered a waste of bandwidth. Thus, a way to increase

Paper presented at the AGARD SPP Symposium on "Digital Communications Systems: Propagation Effects, Technical Solutions, Systems Design", held in Athens, Greece, 18-21 September 1995 and published in CP-574.

the capacity of CODIT using an embedded FH system for low capacity users is examined.

2. CODIT SPECIFICATIONS

A short description of CODIT's radio interface specifications (layer 1 and some parts of layer 2 of OSI model) follows in this section.

Information is carried through *logical channels* which are divided to traffic and control channels. *Traffic channels* (TCH) are intended to carry either encoded speech (TCH/S) or user data (TCH/D). For speech the maximum allowable net rate is 16Kbit/sec. Choice between variable rates (400, 3200, 7200, 8500, 12000, 12500 and 16000 bits/sec) depends on information about source significance, physical channel state, traffic load and other parameters provided by the resource manager. For data three different net bit rates are supported: 9.6, 64 and 128 Kb/sec. Choice depends on user application. *Control channels* (CCH) carry various information used by the system such as handover commands and measurement data. Transmissions are performed at a net bit rate of 9.6 Kbit/sec.

For each logical channel a different *coding* scheme is adopted. The general structure of channel coding is presented in Figure 1.

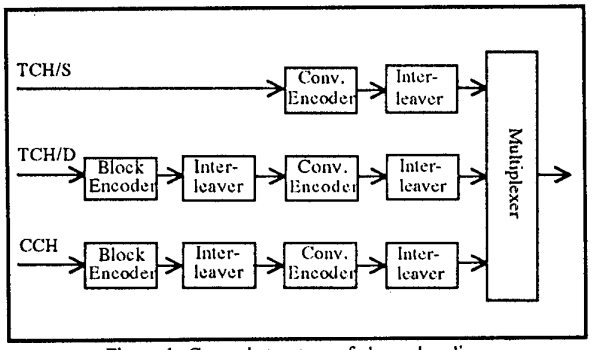


Figure 1. General structure of channel coding

TCH/D and CCH channels are encoded first with a block code followed by a convolutional code. Interleaving is performed after block encoding and after convolutional encoding. TCH/S channels are encoded with a convolutional code and are then interleaved.

Coding parameters and channel rates for THC/D channels are presented in Table 1. For TCH/S the rate of the convolutional code may vary according to the sensitivity of the transmitted bits to the channel errors.

Channel type	Information rate	RS code	Convolutional Code	Channel rate
TCH/D 9.6	9.6 Kbits/s	RS(60,48)	1/2 d_{FREE}=10	24.8 kbit/s
TCH/D 64	64 Kbits/s	RS(100,80)	1/2 d_{FREE}=10	161.6 kbit/s
TCH/D 128	128 Kbits/s	RS(100,80)	1/2 d_{FREE}=10	323.2 kbit/s

Table 1. Coding parameters of TCH/D channels

Mapping of logical to physical channels. Logical channels after coding are formed to a sequence of *data blocks*. Each data block is transmitted within a certain period of time, a *CDMA frame*. Mapping of logical data blocks into the CDMA frame is done using a *physical packet*. Each logical data block is multiplied with an appropriate spreading code for expansion of the channel bit rate to the chip rate of the physical packet.

There are totally six physical packets. Three of them are called mediumband and three narrowband. Mediumband packets have a chip rate of 5.115Mchip/s and can be used for all types of logical channels. Narrowband packets have a chip rate of 1.023 Mchip/s and cannot be used for TCH/D 64 and TCH/D 128 logical channels due to inadequate spreading.

Spreading factor and chip rate parameters for various logical channels of a mediumband physical packet are presented on Table 2:

Data rate	Spreading Factor	Chip rate	Application
4	1,278	51,120	TCH/Speech 0.4
10	511	51,100	TCH/Speech 3.2
18.8	272	51,136	TCH/Speech 7.2
22	232	51,040	TCH/Speech 8.5
24.8	206	51,088	TCH/Data 9.6
30	170	51,000	TCH/Speech 12
31.2	163	50,856	TCH/Speech 12.5
32	159	50,880	TCH/Speech 16
161.6	31	50,096	TCH/Data 64
323.2	15	48,480	TCH/Data 128

Table 2. Parameters of a mediumband physical channel

For transmission each physical packet requires a certain bandwidth which is used in the allocation of a corresponding *physical channel*. The frequency band allocated for CODIT (2221.45 to 2240.88 MHz) is split into six radio frequency channels. Narrowband physical channels have a bandwidth of 1.023MHz and wideband physical channels a bandwidth of 5.115MHz. Down link has a frequency offset of 100.5 MHz form uplink.

Channel modulation for the TX path is QPSK with different spreading codes for I and Q channels and OQPSK with coherent demodulation for the RX path. A square root cosine filter is used for pulse shaping with a roll of factor of 0.35. There are three classes of TX power for MS 0.02, 0.2 and 2 Watts.

CODIT capacity in traffic data channels.

The capacity of CODIT in TCH/D channels presented here has been evaluated via simulation studies. The purpose of the simulation is to provide results that will be used to compare the performance of CODIT with the performance of the hybrid system. Only the parameters which are necessary to estimate the multiple access capacity of the system are incorporated into the simulation model. Results describe the number of traffic data channels that can be simultaneously supported by a mediumband physical channel (5.115 MHz) versus the bit error rate probability for each data channel.

The following assumptions were made in developing the simulation:

- Only TCH/D channels are considered to exist within the physical channel. A single communication link is modeled and by application of Gaussian approximation the multiple access performance of the system is evaluated. Increased data traffic does not block active communication channels but has the effect of a 'graceful degradation' in the overall bit error rate performance.
- Coding and interleaving techniques of CODIT described in the previous section have not been fully implemented in the simulation model in an attempt to reduce simulation time. To compensate for this the spreading factors for each channel have been increased according to the expected coding gain. The modulation scheme for the RX path has been modeled as QPSK since in linear systems OQPSK has the same bit error rate probability and power spectral density as normal QPSK[
- The propagation model includes Additive White Gaussian Noise (AWGN). The results can . be transformed for a log-normal shadowing or Rayleigh fading propagation environment with the addition of a uniform performance degradation for all types of TCH/D's. This can be done since the same digital signal processing is used for every data channel.
- Perfect power control and synchronization between spreading sequences is assumed.

Simulation results are presented in Figures 2 to 5.

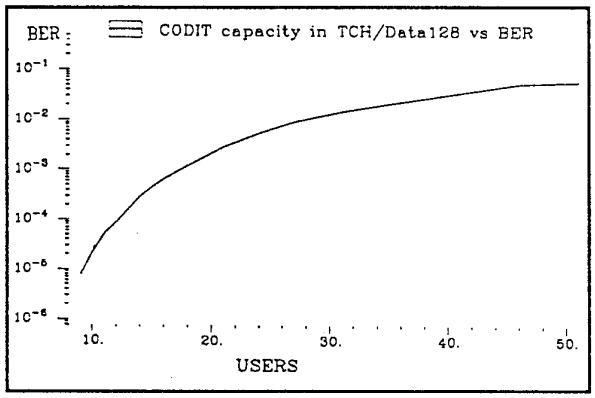


Fig. 2 CODIT mediumband channel capacity in TCH/D128

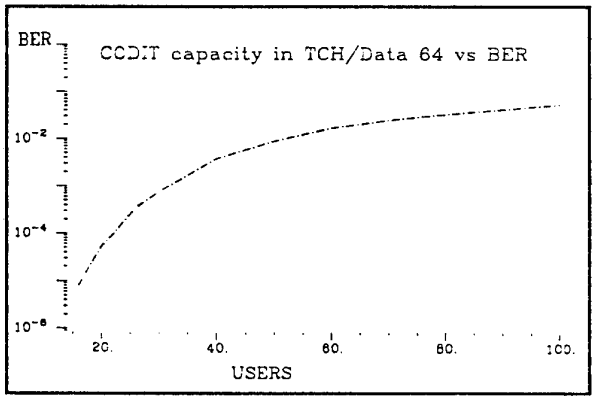


Fig. 3 CODIT Mediumband channel capacity in TCH/D64

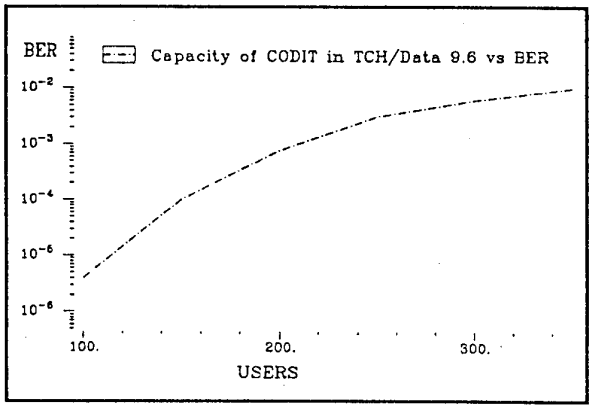


Fig. 4 CODIT Mediumband channel capacity in TCH/D9.6

In figure 5 the presented results refer to a mixed data users scenario. Each number in the x -axis is normalized and is equivalent to one 128kb/sec user, two 64kb/sec or twelve 9.6 Kb/sec users.

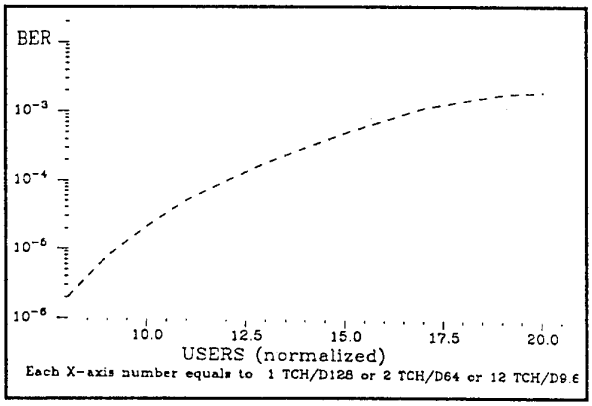


Fig. 5 Normalized CODIT mediumband capacity in TCH/D channels

3. FREQUENCY HOPPING SYSTEM

In a frequency hopping system the allocated spectrum is divided into a number N of equally spaced intervals and at a specific time the transmitted signal occupies only one of the N available channels. A P.N code sequence fed into a frequency synthesizer generates the actual carrier frequency. In the receiver a synchronized P.N code generator and an identical synthesizer provide the same hopping pattern in order to dehop the received signal before demodulation. The parameters which define the performance of a FH system are:

- The number N of available frequency channels
- The hopping rate
- The type of channel modulation
- The type and length of hopping patterns
- The type of Forward Error Control (if used).

Compared to DS-SS, FH-SS provide better performance in fading channels and jamming environments (especially when the hopping rate is fast, i.e. hopping rate > twice the symbol rate) but their multiple access capability under AWGN is worse [9,10] due to interference between multiple hopping users. This is the main reason why FH as standalone systems are better suited for military rather than commercial applications [8].

In this section, the parameters of the frequency hopping system that will be overlaid with CODIT are described. Since the FH system must operate from the same platform as CODIT, the hopping parameters must be defined in coordination with the radio interface specifications of CODIT.

The total *hopping bandwidth* is set to be equal to the bandwidth allocated for the three mediumband channels of CODIT, i.e. 5.115MHz * 3 = 15.345MHz.

The FH system will be used for low data rate communication. Thus the *information rate* of a logical channel is set equal to the TCH/D9.6 of CODIT. The coding and interleaving scheme of CODIT, presented in Figure 1, may be also used in the FH system resulting in a chip rate for each channel at 24.8 Kbit/sec.

The *hopping rate*, i.e. number of hops per chip must satisfy two conditions: (a) to be fast enough to overcome the fading rates encountered in a mobile environment. (Assuming a maximum mobile speed of 100mph, the fading rate at the TX frequency of 2220 MHz is 330 fades/sec). (b) to be slow enough to allow the adoption of a coherent modulation scheme.

Thus a hopping rate of 1hop / 2symbols is used.

Channel modulation. All types of modulation schemes, FSK, DPSK and PSK have been proposed for use in FH systems [5]. However, in fast hopping systems coherent modulation schemes are not feasible without the use of channel equalization techniques. The hopping rate of 0.5 hops/sec chosen for the specific system is slow enough to allow the adoption of a coherent modulation scheme. For better spectrum efficiency, and compliance with CODIT the same modulation scheme with CODIT is adopted.

With a chip rate of 24.8 Kb/sec and QPSK modulation, each logical channel will have a -3db bandwidth of 13.1KHz. Thus the total *number of discrete channels* within the available hopping bandwidth is N=1170.

The *hopping rates* are random. In order to achieve a hopping distance greater than the coherence bandwidth of the channel, the hopping patterns are designed to change to a different mediumband channel in successive hops.

Performance analysis of FH QPSK

The bit error rate for a phase coherent FH spread spectrum signal employing QPSK modulation in an AWGN channel may be derived as follows:

The signal at the transmitter output is described by: (1)

$$s_d(t) = \sqrt{2P}\sum_{n=-\infty}^{\infty} p_T(t-T)\cos[2\pi f_n t + \phi_n + \theta_d(t)]$$

where P is the average power of the signal,
p_T is a unit pulse of duration T
f_n is the frequency during the nth hop interval
ϕ_n is a random phase during the nth hop interval

and

$\theta_d(t)$ is an arbitrary coherent data modulation

The channel will add noise and assuming that the dehopper is perfectly synchronized with the hopper both in phase and in frequency, the received signal input to the bandpass filter after dehopping may be written: (2)

$$r_d(t) = \left\{\sqrt{2P}\sum_{n=-\infty}^{\infty} p_T(t-T)\cos[2\pi f_n + \phi_n + \theta_d(t)] + n_n(t)\right\} * \left\{2\sum_{n=-\infty}^{\infty} p_T(t-T)\cos[2\pi f_n t]\right\}$$

where $n_n(t)$ is a frequency hopped noise signal with instantaneous one sided bandwidth equal to the bandwidth W of the instantaneous frequency hopped transmission.

Since the received noise spectrum covers the entire FH band, the despreader output noise power spectral density is constant for all frequencies, so the output of the bandpass filter may be written:

$$u(t) = \sqrt{2P}\cos[\theta_d(t)] + n_n(t) \quad (3)$$

where $n_n(t)$ is band limited AWGN noise with two sided spectral density of 1/2 N.

From (3) it is shown that after the despreading we get a normal QPSK signal. Thus u(t) which is the input to the QPSK demodulator will have the error performance of a narrowband QPSK signal. That means that the bit error rate of the coherent demodulated Slow Frequency Hopped QPSK is the same as that of a narrowband QPSK. Therefore:

$$P_{FH_{AWGN}} = Q\left(\sqrt{\frac{2E_b}{N_o}}\right) \quad (4)$$

where Q(x) is the complementary error function

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty}\exp\left(-\frac{u^2}{2}\right)du \quad (5)$$

In a fading channel which apart from AWGN noise has also amplitude variations described by Rayleigh statistics, the probability of error for a SFH QPSK signal may be derived as follows:

If K out of N (0<K<N) descrete frequency channels experience a deep fade, the probability that a transmission is faded is K/N whilst 1-K/N is the probability of not being faded. The average probability of error may be written:

$$P_{SFH_{RAY}} = \frac{K}{N}(P_{RAY} + P_{FH_{AWGN}}) + \frac{N-K}{N}P_{FH_{AWGN}} \quad (6)$$

where the probability of error in a slow fading channel for coherent PSK is [11]

$$P_{RAY} = \frac{1}{4\bar{\gamma}} \quad \text{where} \quad \bar{\gamma} = \frac{E_b}{N_0}E(r^2)$$

and $E(r^2)$ is the average value of r^2 and r is the fading amplitude which has a pdf:

$$p(r) = \frac{r}{\frac{r_m^2}{2}} \exp\left(-\frac{r^2}{r_m^2}\right) \qquad (7)$$

Simulated performance of the FH QPSK in AWGN and Rayleigh fading channels is presented in Figures 6 to 8.

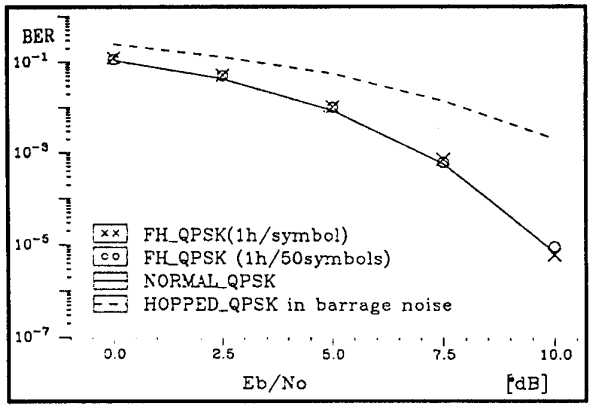


Fig. 6 Performance of FH QPSK in AWGN

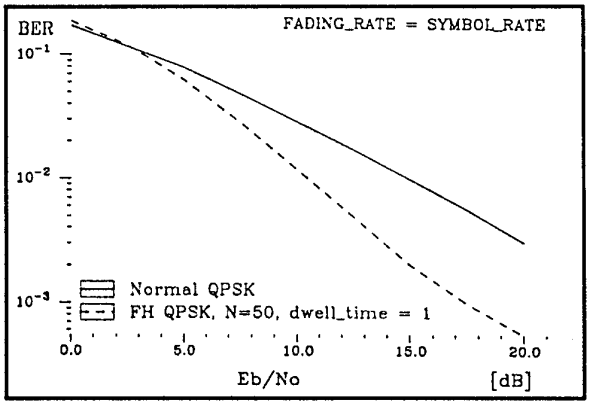


Fig 7. Performance of SFH QPSK in a fading channel

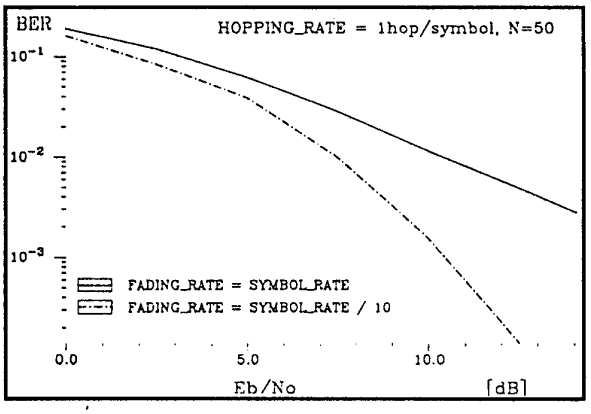


Fig. 8 SFH_QPSK in Rayleigh channels of different fading rate

4. HYBRID DS/FH SYSTEM

A description of the hybrid system and simulation results for it's performance are presented in this section.
A conventional hybrid DS/FH system is presented in Figure 9.

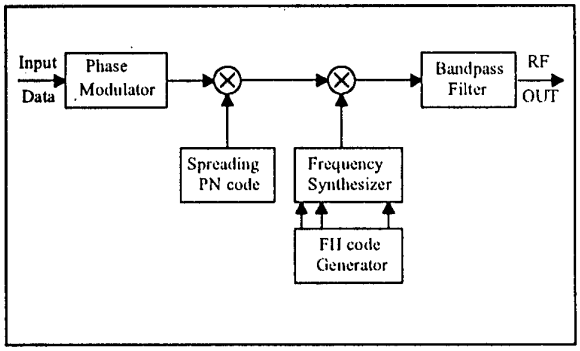


Fig. 9 Transmitter of a Hybrid DS/FH system

Information bits are first spread by the modulo 2 addition with a spreading sequence and the TX frequency is hopped according to a hopping sequence which feeds a frequency synthesizer. The performance of those systems in terms of multiple access has been thoroughly analyzed in [10]. In general, under normal propagation conditions, their multiple access capability is worse than pure DS and worse than pure FH.

The hybrid system which will be examined here is presented in Figure 10.

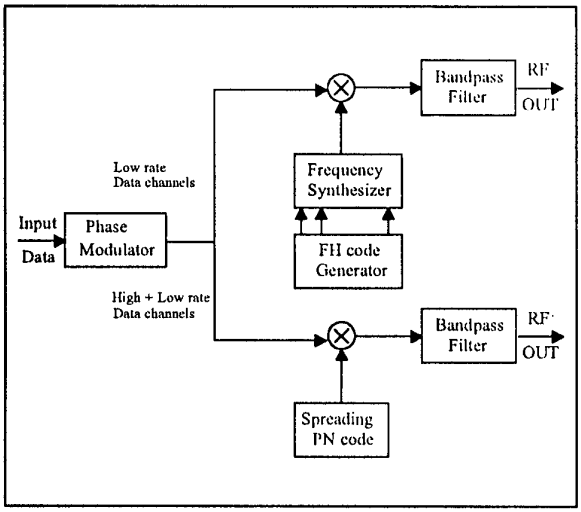


Fig 10 Transmitter part of the proposed hybrid

This Hybrid is effectively an overlay between a FH system and a DS one. The type of TCH/D channels supported by the hybrid is the same as for CODIT but the overall traffic load is divided between the two systems. TCH/D64 and TCH/D128 channels are transmitted with the DS part of the hybrid. For TCH/D9.6 channels the resource manager will

decide if they will be transmitted with the DS or the FH part. In this way the traffic load of the DS system is reduced resulting in a better bit error rate performance for the remaining channels. On the other hand overlaid FH transmissions will interfere to the DS communications causing a performance degradation to the DS system . The question is if the overall spectrum efficiency with this configuration is better compared to CODIT.

A simulation model has been created to study the performance of the hybrid system of Figure 10. The simulations were run with the following settings:

The DS model is the one used in section 2 to simulate the capacity of CODIT. The FH model has been designed according to the description of the FH system in the section 3.

Transmitting power for each FH user is set to a level that will enable hopping users to operate with a bit error rate in the region of 10E-5 to 10E-4.

Multiple hopping users are simulated as individual hopping transmitters. This approach requires excessive simulation time but is necessary in order to obtain a reliable outcome since noise due to overlaid hopping transmissions can not be considered uniformly distributed.

The propagation channel is an AWGN channel.

Simulation results are presented in Figures 11 to 13.

The performance of the system as the number of the FH transmissions is increased is shown in Figure 11. The hopping patterns of the FH users are not allowed to overlap. If CODIT was to operate with a bit error rate in the are of 10E-5 to 10E-4 it is observed that for a small number of overlaid FH links (up to 12) the bit error rate of the remaining CODIT connections is improved. The best improvement that can be obtained is about 10%, when the DS system is operating with a bit error rate in the region of 10^{-5}. As the number of hopped users is increased the interference caused by the FH transmissions overtakes the improvement due to reduced traffic and the result is a performance degradation.

In Figure 12 the hopping patterns of the FH users are allowed to overlap. The maximum overlapping allowed in this case means that the occasional number of hopping users cover a bandwidth half that the bandwidth covered in the non-overlapping case. The simulation result shows that we can increase the maximum number of hopping users to an amount directly proportional to the overlapping degree. However the maximum performance gain remains the same, while FH communication with overlapping hopping patterns may not be accepted for commercial applications (FCC regulations for commercial FH systems in ISM bands).

In Figure 13 the effect of an increased TX power of FH transmissions is presented. In this case the DS system performance is decreased in favor of a better FH communication.

For he performance of he FH connections, since the total number of FH transmissions is less than 10% of the available discrete channels we can assume that problems of mutual interferenece they will not be affected the bit error rate.

6. CONCLUSIONS

The performance of a Hybrid DS/FH system in the form of overlaid FH and DS transmissions has been examined. It has been shown that such a hybrid may offer a small improvement in spectrum efficiency when the number of FH transmissions is sufficiently small. As the number of FH transmissions is increased the overall spectrum efficiency is getting worse compared to pure DS. The higher hardware complexity for the implementation of such a system may not be justified if spectrum efficiency is the objective. But a hybrid system of the proposed configuration has the flexibility to offer improved services to users when privacy and resistance to interference are of importance.

Acknowledgments.

The author would like to thank Dr. S. Atungsiri and Mr. R. Tafazoli members of CSER, University of Surrey, UK for their help and advice during the preparation of this work.

All simulation models have been created in a the digital communications simulation engine COSSAP and simulations were run in computer facilities at University of Surrey, UK.

REFERENCES
AND BIBLIOGRAPHY

[1] G. Cooper and R. Nettleton "A Spread spectrum Technique for High Capacity Mobile Communications", IEEE transactions on vehicular technology, Nov. 1978, pp. 264 - 275.

[2] T. Magill, F. Natali, G. Edwards, "Spread Spectrum Technology for Commercial Applications", Proceedings of the IEEE, 1994, pp. 572-584

[3] J. Wang and M. Moeneclay, "Hybrid DS/SFH Spread-Spectrum Multiple Access with Predetection Diversity and Coding for Indoor Radio" IEEE Journal on selected areas of communication , May 1992, pp. 705 - 713.

[4] P. Rasky, G. Chiasson, D. Borth, R. Peterson, "Slow Frequency - Hop TDMA/CDMA for Macrocellular Personal Communications", IEEE Personal Communications Second Quarter 1994, pp. 26 - 35.

[5] S. Tachikawa, "Modulation in Spread Spectrum Communication Systems", IEEE Transactions on Communications, June 1992, pp. 448-449.

[6] P. Teder, "Radio Interface specifications for the Testbed" R2020/ERA/RI/DS/R/024/b1 Jan. 1994, pp. 17-85

[7] R. Dixon, "Spread Spectrum Systems with Commercial Applications" John Wiley & Sons inc., 1994.

[8] A. Aghvami, "Digital modulation techniques for mobile and personal communication systems" ,Electronics 7 Communication Engineering Journal, June 1993, pp. 128-129.

[9] Geraniotis E. Coherent Hybrid DS-SFH Spread Spectrum Multiple Access Communications , IEEE journal on sel. areas of communications, Sept. 1985, p 871

[10] T. Vlachos, E. Geraniotis "Performance Study of Hybrid Spread - Spectrum Random - Access

Communications" IEEE transactions on communications, June 1991, pp. 990-995
[11] J. G. Proakis, "Digital communications" New York: McGraw-Hill, 1983
[12] L. C. Palmer, P. Y. Chang , "Simulating spread Spectrum Systems "Comsat Technical Review, Volume 19 Number 2, Fall 1989, pp.163-193
[13] A. Correia, A. Albuquerque, "Frequency Hopped Spread spectrum multiple access over dispersive fading channels", IEE Proc.-Communications, April 1994, pp. 79-88

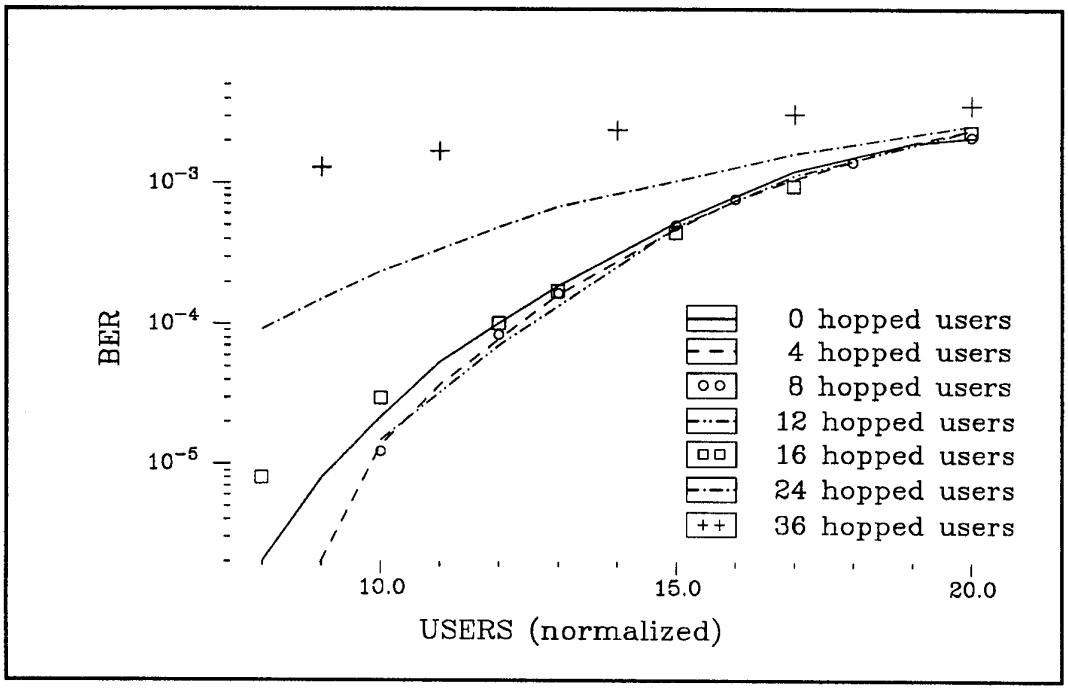


Fig 11. Bit error rate performance of the DS part of the hybrid as # of FH users is increased

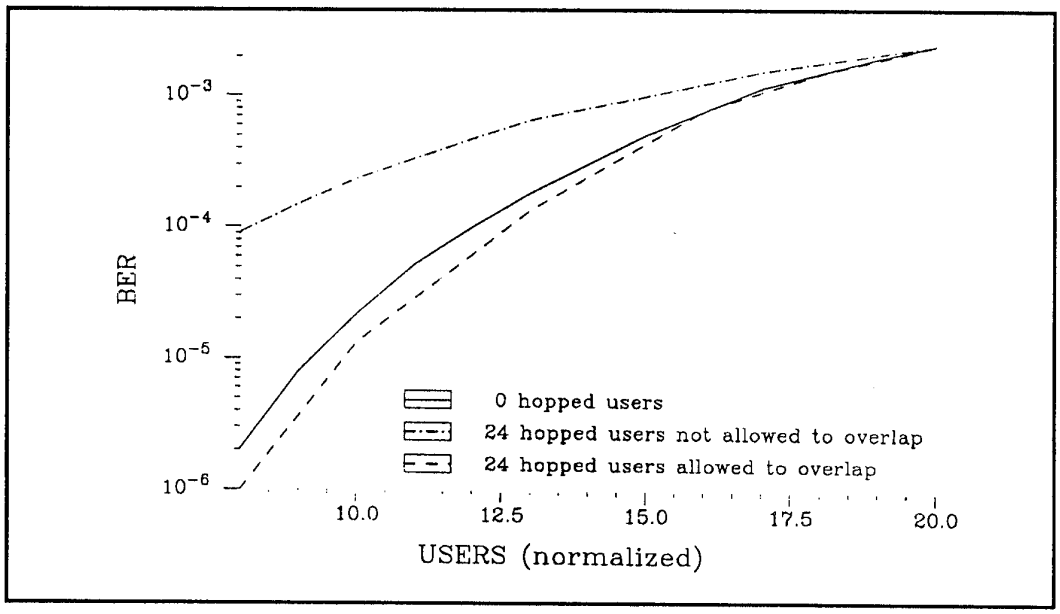


Fig 12. Comparison between overlapping and non overlapping hopping patterns

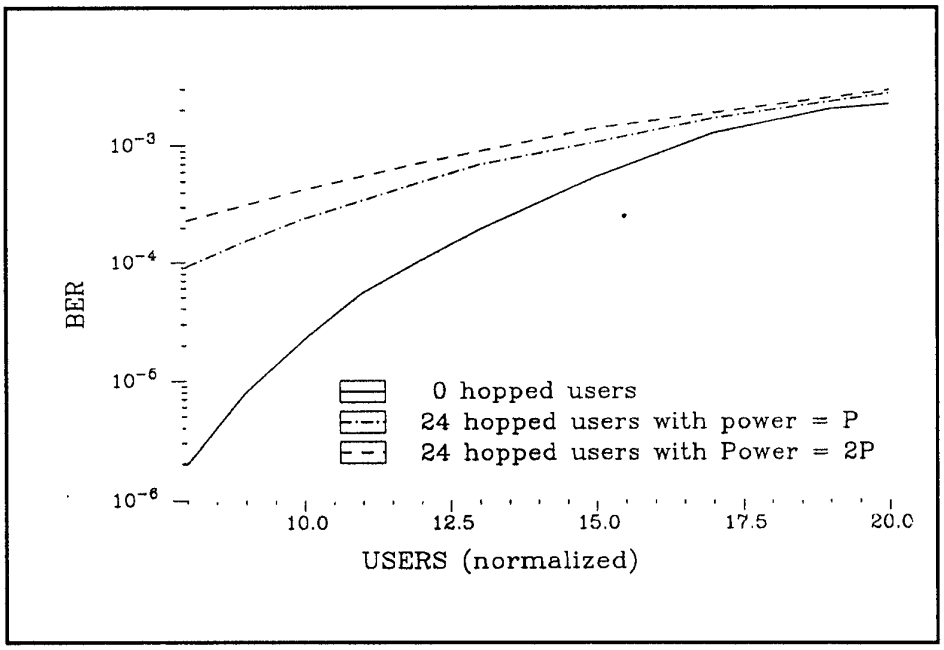


Fig. 13 Effect of Frequency Hopping TX power on performance

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-574	ISBN 92-836-0023-1	UNCLASSIFIED/ UNLIMITED

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

6. Title

Digital Communications Systems:
Propagation Effects, Technical Solutions,
Systems Design

7. Presented at/sponsored by

The SPP Symposium held in Athens, Greece, 18-21 September 1995.

8. Author(s)/Editor(s)	9. Date
Multiple	April 1996
10. Author's/Editor's Address	**11. Pages**
Multiple	484

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other AGARD unclassified publications is given on the back cover.

13. Keywords/Descriptors

Digital communication	Data transmission systems
Propagation	Satellites
Systems engineering	Urban areas
Telecommunication channels	Communications networks
Mathematical models	Adaptive systems
Computerized simulation	Concepts
Signal processing	Trends

14. Abstract

This publication contains the unclassified papers presented to a specialists' meeting sponsored by the Sensor and Propagation Panel of AGARD.

The topics covered included:

- Limitations imposed by propagation on digital communications systems. All frequency bands;
- Recent advances in propagation evaluation and in propagation models;
- Simulation of communication channels;
- Signal processing methods for digital communication systems;
- Data digital communication systems: urban zones, satellites, networks, adaptive systems;
- New concepts for systems design.

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Télécopie (1)47.38.57.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS

AGARD NON CLASSIFIEES

Aucun stock de publications n'a existé à AGARD. A partir de 1993, AGARD détiendra un stock limité des publications associées aux cycles de conférences et cours spéciaux ainsi que les AGARDographies et les rapports des groupes de travail, organisés et publiés à partir de 1993 inclus. Les demandes de renseignements doivent être adressées à AGARD par lettre ou par fax à l'adresse indiquée ci-dessus. *Veuillez ne pas téléphoner.* La diffusion initiale de toutes les publications de l'AGARD est effectuée auprès des pays membres de l'OTAN par l'intermédiaire des centres de distribution nationaux indiqués ci-dessous. Des exemplaires supplémentaires peuvent parfois être obtenus auprès de ces centres (à l'exception des Etats-Unis). Si vous souhaitez recevoir toutes les publications de l'AGARD, ou simplement celles qui concernent certains Panels, vous pouvez demander à être inclu sur la liste d'envoi de l'un de ces centres. Les publications de l'AGARD sont en vente auprès des agences indiquées ci-dessous, sous forme de photocopie ou de microfiche.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordonnateur AGARD-VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Directeur, Services d'information scientifique
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Headquarters
Code JOB-1
Washington, D.C. 20546

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
Netherlands Delegation to AGARD
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
2700 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

Le centre de distribution national des Etats-Unis ne détient PAS de stocks des publications de l'AGARD.
D'éventuelles demandes de photocopies doivent être formulées directement auprès du NASA Center for AeroSpace Information (CASI) à l'adresse ci-dessous. Toute notification de changement d'adresse doit être fait également auprès de CASI.

AGENCES DE VENTE

NASA Center for
AeroSpace Information (CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
Etats-Unis

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris
France

The British Library
Document Supply Division
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Les demandes de microfiches ou de photocopies de documents AGARD (y compris les demandes faites auprès du CASI) doivent comporter la dénomination AGARD, ainsi que le numéro de série d'AGARD (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Veuiller noter qu'il y a lieu de spécifier AGARD-R-nnn et AGARD-AR-nnn lors de la commande des rapports AGARD et des rapports consultatifs AGARD respectivement. Des références bibliographiques complètes ainsi que des résumés des publications AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
publié par la NASA Scientific and Technical
Information Division
NASA Headquarters (JTT)
Washington D.C. 20546
Etats-Unis

Government Reports Announcements and Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 22161
Etats-Unis
(accessible également en mode interactif dans la base de
données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Telefax (1)47.38.57.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED

AGARD PUBLICATIONS

AGARD holds limited quantities of the publications that accompanied Lecture Series and Special Courses held in 1993 or later, and of AGARDographs and Working Group reports published from 1993 onward. For details, write or send a telefax to the address given above. *Please do not telephone.*

AGARD does not hold stocks of publications that accompanied earlier Lecture Series or Courses or of any other publications. Initial distribution of all AGARD publications is made to NATO nations through the National Distribution Centres listed below. Further copies are sometimes available from these centres (except in the United States). If you have a need to receive all AGARD publications, or just those relating to one or more specific AGARD Panels, they may be willing to include you (or your organisation) on their distribution list. AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
2700 Amadora

SPAIN
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Headquarters
Code JOB-1
Washington, D.C. 20546

The United States National Distribution Centre does NOT hold stocks of AGARD publications.
Applications for copies should be made direct to the NASA Center for AeroSpace Information (CASI) at the address below.
Change of address requests should also go to CASI.

SALES AGENCIES

NASA Center for
AeroSpace Information (CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
United States

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris
France

The British Library
Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Requests for microfiches or photocopies of AGARD documents (including requests to CASI) should include the word 'AGARD' and the AGARD serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Note that AGARD Reports and Advisory Reports should be specified as AGARD-R-nnn and AGARD-AR-nnn, respectively. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Division
NASA Headquarters (JTT)
Washington D.C. 20546
United States

Government Reports Announcements and Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic
Database or on CD-ROM)

Printed by Canada Communication Group
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-836-0023-1