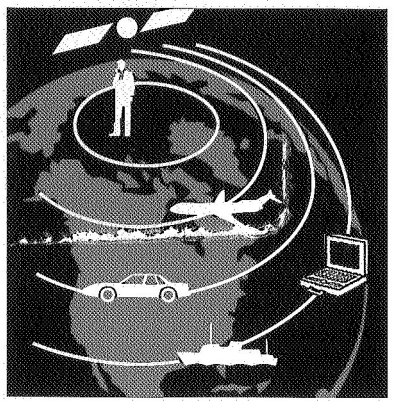
THE NEXT GENERATION

IMSC '97



Proceedings of the fifth

International Mobile Satellite Conference 1997

Pasadena, CA June 16-18, 1997

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology

Industry Canada Industrie Canada

Communications Research Centre

JPL Publication 97-11

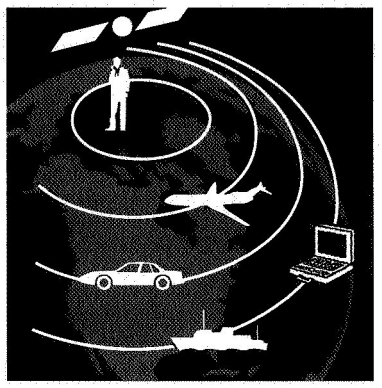
National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology

Industry Canada Industrie Canada

Communications Research Centre

IMSC '97



Proceedings of the fifth

International Mobile Satellite Conference 1997

Pasadena, CA
June 16-18, 1997

Co-sponsored by the
Communications Research Centre, Canada
and the Jet Propulsion Laboratory, USA

Compiled by:
T. Jedrey,
Technical Committee Co-Chairman, JPL

J. Rigley,
Technical Committee Co-Chairman, CRC

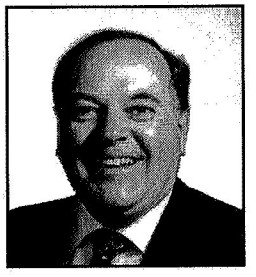
Edited by:
Louise Anderson, JPL

IMSC '97 Organizing Committee

Thomas Jedrey, JPL
Technical Co-Chairperson

Jack Rigley, CRC
Technical Co-Chairperson

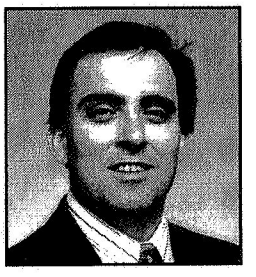
Stephen Townes, JPL
Conference Co-Chairperson



Bob Huck, CRC
Conference Co-Chairperson

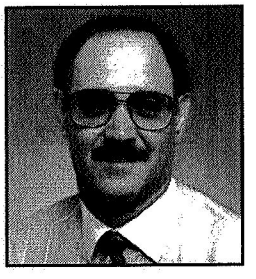
Juliann Gibson, JPL
Conference Organizer

Patricia McLane, JPL
Conference Logistics



Christopher Pasqualino, JPL
Web Page Administrator

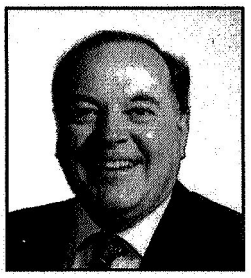
Louise Anderson, JPL
Conference Proceedings Co-Editor



Barry Levitt, JPL
Conference Proceedings Co-Editor

Message From the IMSC '97 Conference Co-Chairs

Stephen Townes
Jet Propulsion Laboratory



Bob Huck
Commnication Research Centre

On behalf of NASA's Jet Propulsion Laboratory and Industry Canada's Communications Research Centre, we welcome you to Pasadena for the Fifth International Mobile Satellite Conference. The Organizing Committee has done an excellent job of putting together a conference that will stimulate thought and discussion about our theme of "The Next Generation" of satellite-based mobile communications. We hope that you will use this opportunity to exchange ideas with colleagues and enjoy the lovely Pasadena area.

This being our *fifth* IMSC (at two-year intervals), it is somewhat ironic that only now do we have the theme of "The Next Generation." It has been a long road since the late seventies and early eighties, when many of us first began working in the area of satellite mobile communications. The reality is that first-generation systems like AMSC, TMI, INMARSAT, Optus, and Orbcomm are now on line, Iridium and Globalstar are beginning to launch this year, and ICO, Odyssey, and others around the world are coming soon.

There are two points to make about the first-generation systems. First, having operational systems available will give us a much clearer picture of issues related to how well the current technology performs and of the market for the current and future mobile satellite services. Second, since many of these first-generation systems have been in planning for many years, it is clear that we should already be planning the second generation so that the concepts and technologies will be in place when needed.

How has technology changed since the first generation? Can we make the currently offered services better (less expensive, more ubiquitous, etc.) by appropriate insertion of technology and market research? What are the new services that might be offered, and what new technologies will we need to enable them? How will mobile satellite services become an integral part of the information infrastructure, offering multimedia services? What are the regulatory challenges to be faced by mobile satellite services as the information infrastructure goes global? These questions are the cornerstones of IMSC '97, and we invite you all to participate in the definition of "The Next Generation."

Contents

Session 1
Plenary

How MSAT Came About

Roy E. Anderson
P.O. Box 2531, Glenville, NY 12325
Phone: 518-384-1212 FAX: 518-384-1211
email: regainc@aol.com

Orest S. Roscoe
Vistar, Suite 1410, 427 Laurier Ave. W., Ottawa, Ontario K1G 3J4
Phone: 613-230-4848 FAX: 613-230-4940
email: roscoe@vistar.ca

ABSTRACT

In the 1960s engineers in Canada and the United States concluded that a satellite system for land, maritime, and aeronautical services was technically feasible and would be economically viable. They identified land mobile needs that would not be served by terrestrial means.

In the following decades there were technical, regulatory, and business hurdles to overcome. Technical feasibility was proven by the use of NASA and other satellites. Regulatory issues were addressed in the U.S., Canada, and at worldwide conferences. A UHF band was allocated internationally to the mobile satellite service, but in the U.S. pressures by commercial and public service entities caused it to be diverted for terrestrial use. An L-band allocation for aeronautical mobile satellites was modified to include maritime and land use in the U.S. and Canada, but only after years of acrimonious international debate.

When the debate was over, the U.S. and Canadian governments authorized generic satellite systems in the L-band. More than two years of vigorous competition for the license in the U.S. was finally resolved by requiring applicants to form a consortium that became the American Mobile Satellite Corporation. In Canada, the process was more straightforward. Telesat Mobile Inc. was established by Telesat Canada and several other investors to implement MSAT on a commercial basis, following an undertaking by the Canadian government to purchase a significant amount of its capacity. AMSC and TMI built identical systems to provide MSAT services for North America.

INTRODUCTION

It has been the tradition in mobile communications to have completely separate systems for land, maritime and aeronautical communications, and even for different users within each of those categories. Competition for radio frequencies caused the tradition of separate systems to be firmly ingrained in the minds of providers and users of mobile radio services. They carried this prejudice over to satellites: there should be separate satellites for each service.

Early in the space age a few engineers who were interested in mobile communications realized that a satellite could be used like an improved ionosphere. Just as you could have all kinds of users bouncing signals off the ionosphere you could have all kinds of users relaying their traffic through a satellite, and with greater capacity and better control than the ionosphere.

By 1965 it was evident that satellites could improve communications for ships and aircraft far from land. Few people thought that satellites were needed for land mobile communications. There were separate developments of satellite mobile services for air and maritime users. They eventually came together in the generic MSAT service as the result of influence from the engineers who believed that generic satellites would serve all users better than separate systems. MSAT is the mobile satellite service for North America that is offered by the American Mobile Satellite Corporation and TMI Communications. It took a long time to bring it about.

AERONAUTICAL MOBILE SATCOM

In 1937 when commercial service started, and as it is today, oceanic air traffic communications was in the high frequency, short wave band. For the North Atlantic there are traffic control base stations at Gander, Newfoundland, Reykjavik, Iceland, and Shannon, Ireland. Air traffic controllers and airplane pilots talk to each other using the ionosphere to reflect their signals beyond the horizon.

Communications are often bad, and occasionally wiped out by solar storms that are especially troublesome at the high northern latitudes of the busy transoceanic routes. In 1964 the first geostationary satellite, Syncom, relayed messages from an aircraft far out over the Pacific. The success of the experiment led to the inclusion of VHF transponders on the NASA ATS-1 and ATS-3 satellites for more aeronautical satellite experiments.

The experiments were highly successful. Boeing installed VHF antennas on the first production 747's, and the Airlines Electronic Engineering Committee wrote standards for aircraft equipment. Airlines and governments teamed up to launch an AEROSAT satellite, and a contractor was selected to build it. Uplink and downlink frequencies in the L-band were allocated internationally for the exclusive use of safety related aeronautical communications through satellites.

The AEROSAT program was hampered by disagreements among its participants. Some wanted it to be at VHF in the bands used for direct air-ground communications over land, and some wanted it to be in the L-band. A final blow came with the mid-70's oil crisis. Because of a big jump in fuel costs, the airlines could no longer fund their part of the program and it was terminated.

After AEROSAT the interest in aeronautical satellites waned to a level near zero. It was revived with a vengeance in 1983 when the first application for mobile satellite service requested the generic use of the aeronautical L-band. Aviation entities throughout the world teamed up to retain the band for en route air traffic control even though there was no realistic prospect of its use for that purpose. The ensuing regulatory battles contributed years of delay to the implementation of MSAT.

There is a limited use of satellites by aviation, but not for air traffic control. Administrations that are responsible for air safety are evaluating the use of the Global Positioning System combined with digital communications for improved traffic control. When all of their stringent requirements can be met and the system can be built, satellites will play their part in better air traffic control. A target date is 2005.

MARITIME MOBILE SATCOM

Just as with aviation, the first plans for maritime satellites were based on the concept of exclusive radio frequencies and satellites for their industry. L-band frequencies adjacent to the aeronautical L-band were assigned for maritime use only. That band eventually became generic, not through an international regulatory battle, but rather in response to demand without much objection by maritime users.

Prior to the introduction of satellites, ships at sea communicated by Morse code in the HF band. The vagaries of ionospheric propagation complicated and sometimes disrupted communications. It took hours, and sometimes days for a message to reach its destination.

In 1976 COMSAT Corporation introduced worldwide commercial maritime communications with three Marisat satellites. In 1979 the International Maritime Satellite system, Inmarsat, became the worldwide provider of maritime services. As time went on, the demand for land mobile and aeronautical communications became so great that Inmarsat changed its charter to include those services, and thus in fact became a worldwide generic satellite service even though it operated in a band that was assigned exclusively to maritime satellite communications.

In the 1960's most land mobile communications served public safety and business. There was a very limited mobile telephone service. The ideas that led to cellular and PCS were in their earliest incubation. So were the ideas that led to land mobile satellites. Prevailing thought was that land mobile satellite communications would never be needed, that all land mobile needs would be met with terrestrial systems.

MOBILE SATCOM IN CANADA

In Canada, the interest in mobile satellite communications was largely motivated by a requirement for reliable, thin route and mobile communications to serve the vast, sparsely populated regions of the country. Approximately 80% of the Canadian land mass, north of the 55th parallel in the west and the 50th parallel in the east, is home to less than 500,000 residents. Only 100,000 people live north of 60° latitude, in an area about half as large as the 48 contiguous United States. Until the mid '70s, the communications needs of the area were served with HF radio, which was even more unreliable than elsewhere at these high latitudes because so much of the auroral zone lies over Canadian territory. Eventually the fixed, community communications needs were met with the Anik satellites. The mobile needs, however, remained.

The potential of satellite communications technology to meet the Canadian needs was first experienced in May, 1967 when Canada joined the U.S. sponsored cooperative research and development program to explore the potential of satellites for tactical military communications. Over the next five years, experimental programs were conducted to test Canadian developed aircraft, ship, and manpack stations with the U.S. Lincoln Laboratory experimental UHF satellites, LES-5 and LES-6, and the operational TACSATCOM-1. System performance R&D included tests of operation at low elevation angles, ionospheric scintillation fading, airborne multipath, and modems. A number of firsts were achieved in the Canadian tests, including the first air-to-air mobile satellite communications using phased array antennas on a C47 and a C130 Hercules aircraft, in July 1969. The program demonstrated the feasibility of mobile satellite communications at UHF and resulted in the development in 1972 of a proposal for a Canadian military satellite similar to the U.S. FLTSATCOM system. However, because one satellite would not be able to meet the global needs of Canadian commitments to NATO and to peacekeeping roles, the proposal was not implemented.

On the civilian side, solutions were sought by the Canadian Department of Communications (DOC) for the problem of providing reliable communications in the north, and in 1970 DOC proposed that a UHF transponder be included on the joint DOC-NASA communications technology satellite program (CTS-Hermes), in addition to the primary mission 12 GHz transponders for direct broadcast TV and other small terminal services. However, the UHF transponder, which was to have been used for demonstration of thin route and mobile communications to small terminals, had to be abandoned because of satellite launch weight constraints.

Interest in a Canadian UHF satellite persisted, however. Requirements were broadened to include civil government, military, and commercial maritime user needs for mobile, data acquisition, and other specialized communications services in rural and remote land areas and Arctic and coastal waters. Traffic studies indicated that the aggregated needs of the government departments and agencies could be met by a geostationary multipurpose UHF satellite system, dubbed MUSAT. Economic feasibility resulted from system cost sharing by the many user agencies.

Services planned to be provided by MUSAT included two-way voice and low-rate data to ships, aircraft, and land terminals for mobile and transportable communications in the 240-400 MHz allocations, data collection in the 401-403 MHz band, emergency beacon monitoring in the 406-406.1 MHz band, maritime mobile service in the L-band to complement INMARSAT coverage which was missing in Canadian Arctic waters, and military fixed satellite service in the 7250-7750 and 7900-8400 MHz bands. To meet Department of National Defence needs, interoperability with the U.S. FLTSATCOM and LEASAT systems was included as well as some anti-jamming capabilities. Thus it can be seen that MUSAT was truly multipurpose, although it served government departments and agencies primarily. Its planning spanned the mid 1970s, in parallel with the AEROSAT program, but implementation was stalled by the onset of the early stages of government spending restraints.

The MUSAT proposal also suffered from its inability to provide public services in the 240-400 MHz band because the band was allocated in Canada to meet military and civil government needs and in the U.S. strictly to serve military needs. This was remedied by the 1979 World Administrative Radio Conference (WARC) which, based on proposals by Canada and the U.S., allocated 806-890 MHz to the mobile satellite service on a primary basis in the Americas (ITU Region 2), as well as in some Region 1 and 3 countries. The allocation was shared with the mobile terrestrial service, limited to national boundaries, subject to agreement between the administrations concerned, and did not include the aeronautical mobile satellite service. Nevertheless, the allocation permitted an expansion of the planned MUSAT system to meet a much broader range of mobile communications requirements. It also allowed the prospect of economies of scale from sharing at least some technology with the terrestrial mobile service.

DOC-NASA COOPERATION

In the same time period as DOC was developing the MUSAT concept, through the '70s, NASA conducted experiments with its Applications Technology Satellites which were used for numerous land mobile communications voice, data, and position surveillance tests. The experiments demonstrated technical feasibility of land mobile satellite communications and revealed numerous applications which would benefit from the technology.

Given the possibility of public services in the new UHF allocation and an interest by NASA, DOC initiated discussions with NASA in February, 1980 which resulted in a concept for a joint DOC-NASA demonstration satellite program. Canada would provide the satellite, modified to meet NASA requirements, and NASA would provide the launch and launch support services. Following the model set by Hermes, use of the satellite would be shared. Launch would be in the late 1980s. The objectives of the joint program were to demonstrate the technical and economic feasibility of mobile satellite services, formulate the institutional arrangements for a follow-on commercial system, and provide a period of interim operational services prior to the implementation of a commercial system.

Phase A studies commenced in September, 1980, carried out cooperatively between DOC and NASA. These established technical feasibility and indicated a strong potential commercial demand. One of the studies was NASA sponsored and carried out under contract by General Electric. Its objective was to determine the best way to provide ubiquitous mobile communications in the United States. The study concluded that one half of the contiguous states land area would not be served by cellular telephone because of too low a population density. About one tenth of the population lived in this unserved area. Service to that population as well as to transportation and remote fixed locations would comprise a market large enough to justify a mobile satellite service.

In early 1982, a government exchange of letters took place, agreeing that at the request of Canada, the United States and Canada would enter into negotiations to permit the introduction of mobile satellite services in the 806-890 MHz band. Phase B (project definition) was authorized in July, 1982, with the principal objectives of refining the studies of market demand and commercial viability carried out in Phase A, developing performance requirements and specifications for the spacecraft and ground segment, developing the required technologies, and developing a technical, cost, and management proposal for program implementation. Launch of the satellite was planned for 1988.

The basic Canadian service requirements were for coverage of the Canadian land mass including the far north within range of a geostationary satellite and coastal waters to the 200 mile offshore limits. Principal services were mobile radio, mobile telephone and mobile data, requiring an estimated allocation of 12 MHz in the 806-890 MHz band. Data collection was to be provided in the 401-403 MHz band, and emergency beacon monitoring in the 406-406.1 MHz band. The Phase B studies confirmed that there was a growing, strong user demand for mobile satellite services.

Notification of the International Frequency Registration Board concerning the Canadian UHF mobile satellite was first made in January, 1983, of a six beam system using an allocation of 8 MHz. and capable of service in both Canada and the U.S., in anticipation of the joint DOC-NASA project. This was subsequently amended in November, 1983, reducing the six beams to four with domestic coverage only. In 1984, in a Canada Gazette Notice, DOC proposed that a further 12 MHz in the 806-890 MHz band and 30 MHz at L-band be reserved for second generation expansion.

Corresponding regulatory events in the U.S. commenced with a NASA petition for rulemaking filed with the FCC in September, 1982 to provide a U.S. frequency allocation for mobile satellite service. The internationally allocated UHF radio frequency band was recommended for the service. Public comments requested by the FCC and received in March, 1983 indicated considerable support. However, objections were also raised with NASA about the planned joint program with DOC, on the grounds that the program was in conflict with commercial interests and would compete for limited spectrum. This resulted in postponement of an April, 1983 planned signing of a negotiated agreement for cooperation between DOC and NASA.

COMMERCIALIZATION OF MSAT

The rapidly building commercial pressures in the U.S., coincident with the results of the Canadian Phase B demand study results confirming a potential for commercial viability, resulted in the MSAT program being reoriented from a government-led DOC/NASA demonstration program to a commercially-led program, with DOC and NASA falling back to support and coordination roles. Canadian studies in progress were redirected to support development of a commercial system in which Telesat Canada and a U.S. company would procure jointly the space segment of a two satellite system providing service in both Canada and the U.S. The commercial model proposed was eventually implemented. Each company would own and operate

one of the satellites, which would be a back-up to the other in case of failure as well as supply spare capacity in the event that one experienced a traffic overload.

Commercial activity was inaugurated in the U.S. with the formation of the Mobile Satellite Corporation in July, 1982, by Roy Anderson who was the principal participant in the satellite experiments and in the NASA study by GE, and John Keisling, a major participant in the study. Mobile Satellite Corporation filed an application for Mobile Satellite Service at the Federal Communications Commission on March 9, 1983. The application proposed land, aeronautical and maritime mobile satellite communications for the United States including Alaska, Hawaii, Puerto Rico, the U.S. Virgin Islands and territorial waters out to 200 miles. It requested that two radio frequency bands be allocated for transmissions of all kinds between mobiles and satellites. One, the UHF band, the other, the L-band, (1545-1559 and 1646-1660 MHz) that was reserved exclusively for aeronautical safety communications. Neither band was in use. Technical and operational assurance of preemptive access for safety services would be achieved by continuous network control of every mobile unit.

Four months after Mobilesat filed its application for a license, Skylink Corporation filed an application that did not include a request for the L-band. Two geostationary satellites would serve North America.

The initial system configuration selected by Telesat was for two hybrid satellites to serve North America, each with two UHF beams and four L-band beams. Feeder links to base and gateway stations were to be at Ku-band. The UHF band was to be used primarily for moving vehicle applications and the L-band for transportable and mobile applications where higher gain antennas could be provided and line-of-sight to the satellite was assured. To the mobile radio, telephony, and data services to land and ship terminals initially targeted for satellite service, aeronautical mobile, fixed rural and remote telephony, and wide area paging were added.

REGULATORY BATTLES

With two applications on file, the Federal Communications Commission responded to the NASA request. It announced that it had adopted a Notice of Proposed Rulemaking on November 21, 1984. The Commission stated that it planned to allocate "....eight megahertz (821-825 and 866-870 MHz) [UHF-band] for a land mobile satellite service. Spectrum in the 1.5 GHz [L-band] range is expected to increase the capacity and offerings of such service either in the initial system or future generations."

The UHF-band requested by Mobilesat and Skylink was between the band allocated to cellular mobile telephone and a band used by business and public safety mobile radio. It was the band that the FCC planned to allocate to the mobile satellite service in its November, 1984, Notice of Proposed Rulemaking. The requested band was in reserve - no licenses had been granted for its use. It was allocated internationally for mobile satellites. Canada and the United States had an agreement that it would be used for that purpose.

What seemed like a reasonable request by the two mobile satellite applicants met with a storm of opposition. From one side cellular radio telephone service providers and equipment manufacturers argued that their band was not wide enough to meet their future needs and they should get the reserve band that Mobilesat and Skylink wanted. On the other side Motorola, a leading manufacturer of cellular and other mobile radio equipment, had expected to persuade the FCC to extend the business and public service band to include the reserve band.

Motorola was especially aggressive in its opposition. They argued that all mobile communications needs everywhere in the country would be met with non-satellite mobile radio systems. Motorola, with the Associated Public-Safety Communications Officers, Inc. (APCO) stated that the reserve band was desperately needed for public safety communications. APCO orchestrated a campaign to have large numbers of public safety officers file comments to the FCC opposing allocation of the band to mobile satellites. Motorola, APCO, and Mobilesat lobbied Congress and the FCC. Motorola and APCO won. On July 24, 1986, the FCC issued a Report and Order allocating the band to public safety.

Mobilesat had estimated that the UHF-band alone could not accommodate enough mobile communications to justify investment in the satellite system. Its application for the aeronautical L-band, argued that the only way the aeronautical community could get their safety service was to license the L-band for air, land and maritime service so that investment in a satellite system would be justified. Aeronautical safety service could be assured by

designing the system to give aeronautical safety communications priority and preemptive access.

There was a national and international outcry. Satellite communications had been on aviation's back burner since the demise of AEROSAT, but suddenly it became of prime importance. Committees were formed to conduct studies to determine the width of the band needed for aeronautical safety communications. They all concluded that it was just a little more than the entire L-band. It was absolutely essential to keep the "pizza trucks" out of "their" band.

Within the Federal Aviation Administration the persons responsible for radio frequency coordination viewed the Mobilesat application unfavorably. Other FAA managers and each Administrator that served during the years-long proceedings did not oppose allocation of the band to generic use. They agreed with Mobilesat's argument that aviation would not have the advantage of satellite communications unless it was willing to share the use of the band so that an economically viable system could be built.

On January 28, 1985, the Federal Communications Commission released the Notice of Proposed Rulemaking it had announced on November 21, 1984, proposing to establish a Mobile Satellite Service for the United States. It invited interested parties to submit applications for the license to provide the service. Although the L-band allocation to MSS had not yet been made, applicants were encouraged to include that band. The Commission agreed with NASA and Mobilesat that the UHF-band alone was not sufficient for development of multiple services and efficient use of the spectrum.

On April 30, 1985, twelve applications for the Mobile Satellite Service license were filed at the FCC. The applicants were a mix of large companies with unquestioned qualifications, start-up companies with a varied range of qualifications but a serious commitment to the project, and some speculators who knew there was a chance to make money from of the FCC licensing procedures.

One application was mostly a photocopy of the 1983 Mobilesat application, even retaining the Mobilesat name, with a letter stating that the applicant would do what was described therein. The FCC did not dismiss the application. At a NASA sponsored meeting held at the Jet Propulsion Laboratories in Pasadena, CA in November, 1985, the applicant announced the terms of the financial settlement he would make with the company that got the license. He warned that if the company did not agree, the legal fees to get rid of him would be more than the financial settlement.

On July 24, 1986, the Federal Communication Commission allocated the L-band to the Mobile Satellite Service including land, maritime and all aeronautical uses. It specified rules on the use of the band to insure priority and preemptive access for aeronautical safety.

The FAA did not object to the allocation. Public announcement of FAA policy was finally made in the August 15, 1988, issue of Aviation Week. It agreed to the sharing. It spelled out the requirement for priority and preemptive access and the procedure for reserving adequate satellite capacity for aviation safety when the demand for non-satellite communications increased.

The FCC would issue only one license for the mobile satellite service. It could select the licensee by lottery, as it had done with cellular radio telephone licenses, or it could hold comparative hearings to determine which applicant had the best proposal. Neither one seemed like a good choice. The lottery might select a poorly qualified applicant. Comparative hearings would take years of time and costly legal proceedings with appeals of every decision.

The Commission attempted to narrow the field of applicants by requiring each to place five million dollars in a escrow account, the money to be used toward the construction of the system. Eight of the applicants complied.

The Commission then decided that instead of a lottery or hearings it would require the applicants to form a consortium and file a single application on behalf of the consortium. The FCC dismissed the individual applications and directed the Consortium to file its application by July 27, 1987.

There were many points of contention among the applicants and they split into two factions, each submitting an application. The FCC dismissed the applications of each faction and set September 4 as the date for submitting a unified application. Again two applications were filed, one by a group of five who called themselves the United States Mobile Satellite Consortium. The five divided ownership,

marketing, administrative and procurement businesses among themselves, including the construction and operation of a dedicated $145 million satellite. The other group of three stated that the plan violated the antitrust laws and asked the FCC to invalidate the contracts among the five.

On February 9, 1988, the eight applicants jointly filed a technical amendment describing the system proposed by the group. The geostationary orbit was selected with one large satellite at 101°W, near the longitude of the central part of the U.S. Service continuity would be assured by an agreement with Telesat Mobile Inc. of Canada, which would place a similar satellite nearby at 106.5°W.

IMPLEMENTATION

Public announcement and formal signing of the Joint Ownership agreement between the eight applicants was consummated at a meeting in Pasadena, California, on May 3, 1988. With that action, the American Mobile Satellite Consortium was formed, subsequently incorporated in 1989 as American Mobile Satellite Corporation (AMSC). The meeting was held in conjunction with an annual Mobile Satellite Conference sponsored by NASA at its Jet Propulsion Laboratories where many studies and experiments in satellite mobile communications were conducted in cooperation with some of the applicants and with Canadian organizations. On August 4, 1989, the FCC issued a license to the Consortium. The allocation of the L-band to all mobile services was contingent upon the assurance that aeronautical safety would have priority and preemptive access to the use of the satellite.

In Canada, the commercialization process proceeded with far less drama than in the U.S. Telesat Canada reached agreement with DOC that Telesat would lead commercial implementation of MSAT, subject to a Canadian government undertaking to lease capacity on the satellite for services to government departments and agencies. The government agreed to purchase airtime up to a value of C$ 126.5 million, to be consumed over the operating life of MSAT, on a prepayment basis to assist in financing implementation. In addition, the government committed C$ 50 million for technology and applications development to assist commercial success and to garner industrial benefits. Telesat Canada subsequently incorporated Telesat Mobile Inc. (TMI) in 1988, in which it held 50% ownership, with Canadian Pacific and Itochu Corporation holding 30% and 20%, respectively. Canadian Pacific subsequently withdrew, and TMI was later reorganized as TMI Communications Limited Partnership (TMIC).

AMSC launched the first MSAT satellite, which was named AMSC-1, on April 7, 1995. TMIC launched its MSAT-1 on April 20, 1996.

REFERENCES

[1] O.S. Roscoe, *A Small Terminal Satellite Communication System for Remote Areas,* Canadian Electronics Engineering, December 1970.

[2] C.A. Franklin and P.M. Boudreau, *Multipurpose UHF Satellite System (MUSAT),* presented at the Communications Satellite Workshop - Canadian Experience and Australian Planning, Canberra, Australia, August 22-24, 1979.

[3] A.F. Brisken, R.E. Anderson, R.L. Frey, and J.R. Lewis, *Land Mobile Communications and Position Fixing Using Satellites,* IEEE Transactions on Vehicular Technology, Vol. VT-28, No. 3, pp. 153-170, August 1979.

[4] R.E. Anderson, R.L. Frey, J.R. Lewis, and R.T. Milton, *Satellite-Aided Mobile Communications: Experiments, Applications, and Prospects,* IEEE Transactions on Vehicular Technology, Vol. VT-30, No. 2, May 1981.

[5] R.E. Anderson, R.L. Frey, and J.R. Lewis, *Satellite-Aided Mobile Communications Limited Operational Test in the Trucking Industry,* prepared for National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, MD. Contract No. NAS5-24365, Final Report for Period Nov. 1977 - July, 1980, SRD-81-005, July 1980.

[6] C.A. Franklin, *MUSAT to MSAT: Evolution of Mobile Satellite Communications in Canada,* presented at Fibresat 86, International Conference on Satellite and Fibre Optic Communications, Vancouver, Sept. 9-12, 1986.

[7] A.E. Winter, M.J. Zuliani, and D.J. Sward, *Land Mobile Satellite Communications via MSAT,* Earth Oriented Applications of Space Technology, Vol. 5, No. 4, pp. 345-349, 1985, Pergamon Press.

[8] R.E. Anderson, *A Long Journey to AMSC,* March 1995. (Available from the Author)

Session 2
Networking and Protocols I: Mobile and Hybrid Networks

Session Chairperson—*Keith Smith*, ICO Global Communications, UK
Session Organizer—*Mike Moher*, Communications Research Centre, Canada

Integration Towards a Future Terrestrial and Mobile Satellite Based Communication System - An Analysis

Alexander Guntsch
Communication Networks, Aachen University of Technology, 52064 Aachen, Germany;
Phone: +49.241.80.7916, Fax: +49.241.8888.242,
email: aeg@comnets.rwth-aachen.de

ABSTRACT

The aim of this paper is to provide a comprehensive performance analysis for candidate integrated GSM based and mobile satellite communication systems. After a brief introduction of terrestrial and mobile satellite communication networks a basic set of common services envisioned for integrated system operation will be presented. Two possible integration scenarios able to provide this set of services will be identified and discussed. Appropriate evaluation models as well as performance parameters will be presented. In the analysis section of this paper a comprehensive performance evaluation of call handling procedures like inter network handover and location management in these integrated systems will be given. Based on the identified performance parameters the restrictions and limitations will be discussed. The paper concludes with a summary of the results and an outlook into the field of future integrated terrestrial and mobile satellite system solutions.

1 INTRODUCTION

The GSM (Global System for Mobile Communication) network is nowadays readily available in most west-European and Asian countries and is currently extending its reach even into the United States of America, Canada, Africa and Russia. The coverage though within these countries is many times limited to specific areas where the population and therefore also the customer density is very high. Extending the coverage to less populated regions provides less return on investment to a potential network operator and/or service provider of such a system and is therefore proceeding at a slower pace or is not proceeding at all.

During the past years, an increased demand for personal mobile communication paired with a steady increase in communication technology and aided by corresponding regulatory decisions have led to the introduction of new cellular communication networks including soon to come mobile satellite systems. All of these new systems, including the old ones, are envisioned to be part of a new global communication system architecture - UMTS - interconnected as illustrated in the proposal illustrated in figure 1. Several access systems paired with different service specific backbone networks will provide environment specific telecommunication services from indoor urban to open nomadic environments.

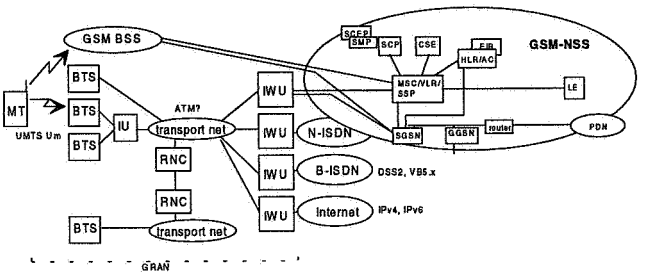


Fig. 1 Envisioned UMTS communication infrastructure for the 21st century according to ETSI/SMG3 (still under discussion)

In this approach systems like IRIDIUM[tm], GLOBALSTAR[tm], ICO[tm], and ODYSSEY[tm] will be part of a hierarchical cellular concept providing umbrella like coverage on a global or near global basis [1, 3].

The integration of already existing cellular networks like GSM and proposed mobile satellite systems following this concept will therefore significantly enhance the functionality and flexibility, as well as the availability of global telecommunication services. The satellite systems involved will be able to provide global or near global coverage and wide-area roaming to users outside their individual terrestrial cellular coverage areas. They can also complement existing cellular radio systems like GSM or DCS 1800/1900 in specific areas or regions. Whenever a terrestrial mobile cellular communication network is only operated in an „insular[1]“ mode for example, the mobile satellite system can insure the possibility to be reached or to perform a series of telecommunication services [7].

In order to be able to propose an efficient structure for such an integration of terrestrial cellular and mobile satellite communication systems a thorough analysis and discussion of the protocols and procedures involved, stating the possibilities and restrictions of each considered integration scenario is required. The scope of this paper is to perform such an analysis for two proposed inter-segment handover protocols in an integrated LEO (Low Earth Orbit) satellite system and GSM.

[1] Insular in this context means the coverage of small regions or towns only, avoiding the coverage of less populated surrounding areas.

2 SERVICES AND USER REQUIREMENTS

Mobile satellite systems that are currently being developed will provide services which are a mix of satellite services (e.g. radiodetermination) and services similar to those provided by GSM. In the following table an overview of the services envisioned for mobile satellite systems and GSM are presented.

Table 1: Mobile Satellite System and GSM services [3]

Service/System	GSM	Mob. Satellite
Voice	13 kbps	2.4 - 4.8 kbps
Data	9.6 kbps[2]	2.4 kbps
Fax	Group 3	Group 3
Paging	SMS	Available
position/location	-	Intrinsic/GPS

Integrated GSM and mobile satellite PCN will aim at harmonising the columns presented in table 1. In general terms this will lead to an increase in the number or quality of possible services with respect to each single network (e.g. positioning services for former „GSM only" users or higher available bit rates for former „satellite only" users). User requirements in this field indicate that in addition to the new/better services made available, additional functionality in the network should be provided to manage the interaction between both original networks. A call drop or poor speech quality are not acceptable at any point in time. In addition a high degree of billing transparency and security within the network is desired.

3 INTEGRATION SCENARIOS

When considering the integration of GSM and a possible LEO satellite PCN the question of level of integration for this envisaged system arises. User requirements such as global availability, single number roaming, ease of use and value for money as well as network operator requirements such as signalling and power efficiency, location register interworking and a minimum of required GSM modifications need to be taken into account [1].

As a result of requirements both an integration at *network* and *system* level infer [1,4]. Corresponding to these two levels of integration two integration scenarios can be defined.

The first scenario, integration at *network* level (figure 2), provides an integration of two separate networks. They are connected via a fixed terrestrial network and are each operable on its own without having to make use of components from each other. All exchange of necessary data-information may be achieved via especially designed interworking functions. Having a large degree of independence, this scenario is very efficient in terms of hard and software for each independent segment since each segment may be individually designed to provide an optimum solution for the special requirements of the segment.

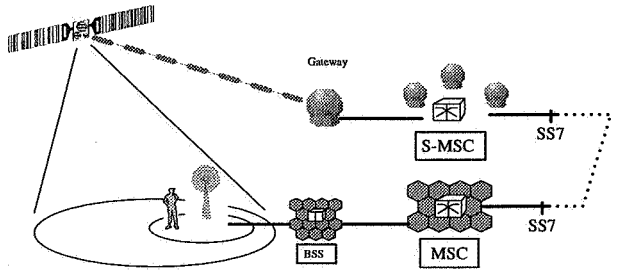


Fig. 2 GSM and mobile satellite system network level integration scenario

A second scenario, integration at *system* level, is characterised by both sub-systems (terrestrial and satellite) using the same data base (HLR- Home Location Register and VLR - Visitor Location Register) and using the same main switching facility (MSC - Mobile Switching Center). In contrast to the network level integration scenario though, the system level integration scenario does strongly depend upon the satellite system network architecture and requires the implementation of an interworking unit at MSC level. This is mainly because the part of the satellite network involved in the integration process may not provide all of the required functionality (i.e. full BSS - Base Station Subsystem functionality).

It is the aim of this paper to make use of the first integration scenario providing a basis for evaluating the integration performance of the system in terms of call handling including handover and location management issues.

4 INTEGRATED NETWORK FUNCTIONALITY

The integrated network functionality can be defined based on the desired set of user requirements to be fulfilled (see section 2). Within the following subsections this functionality shall be identified and discussed. Performance parameters shall be found in order to evaluate the new functional behaviour.

4.1 FUNCTIONAL REQUIREMENTS

Based on the user requirements identified two new functional requirements of the Integrated GSM and mobile satellite PCN emerge:

1. To avoid ongoing call drops with the combined network due to lack of coverage or traffic congestion, procedures should be implemented that allow to switch between both available networks: inter-network handover.
2. To be able to be reached at any desired time and location and to enable user selection of the communication network to be used initially,

[2] 9.6 kbps is the normal circuit switched data rate in GSM. Future packed transmission (e.g. GPRS) will allow much higher data rates.

procedures should be implemented allowing for the appropriate user data flow: inter-network roaming.

It will be seen in the following sections if and how these two functional requirements can be fulfilled.

4.2 PERFORMANCE PARAMETERS

To be able to evaluate the performance of procedures implemented to fulfil the desired functional requirements a set of performance parameters has to be identified.
For the inter-network handover for example it is important to evaluate the handover interruption time as well as the handover failure rate. For a potential network operator a signalling analysis would be of additional interest.
For the inter-network roaming procedures it is important to know how often a location update has to be performed for a specific scheme and how much signalling is involved in the update as well as potential paging procedures.

4.3 INTER-NETWORK HANDOVER

In order to perform an inter-network handover within an integrated GSM and mobile satellite PCN (**P**ersonal **C**ommunication **N**etwork) an especially designed communication protocol between the two networks has to be used. It has to account for transmission errors over the GSM and mobile satellite radio channels by allowing several repetitions of the PDUs (**P**rotocol **D**ata **U**nit) sent between the mobile terminal and the respective transmission counterparts in both systems.
One candidate protocol is a modified GSM inter-MSC handover protocol given in figure 3 below.

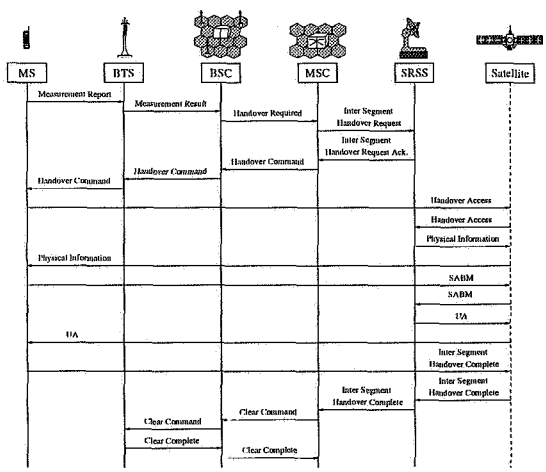


Fig. 3 First inter-network handover protocol based on the GSM inter-MSC handover.

Modification have to be made to several of the protocol timers taking the special propagation environment and the different frame structure and bit-rates of the satellite network into account. The $LAPD_m$-timer T200 and the *physical information* timer T3124 were modified to 220 ms and 660 ms respectively. In addition the max. number of I-frame repetitions within the FACCH/F $LAPD_m$-protocol were set to 17 accounting for a different IRIDIUM like frame length of the satellite network (see GSM recommendation 04.06).

4.4 LOCATION MANAGEMENT

A major area of concern to the network layer protocols of an integrated GSM and satellite PCN is the definition of location areas and the management of the mobility of a mobile user within the joint network [5].
Wherever the GSM and the satellite network are overlapping, it is preferable to make use of already present definitions of location areas (in GSM a location area may be any subset of coverage areas of BTSs-Base Transceiver Station connected to a single BSC-Base Station Controller) and use existing procedures to update them accordingly. This may require conversion between GSM location areas and (virtual) satellite ones, but this functionality is quite easy to handle with respect to a specialised satellite segment mobility management.
In areas where the GSM segment is not present though these specialised procedures need to be used in order to provide the same quality of service (QoS) to users there as compared to the GSM system.

4.4.1 Single Satellite Location Area

One possibility of defining a location area for use if in an integrated network solely satellite coverage is present is the satellite coverage zone approach shown in figure 4 below.

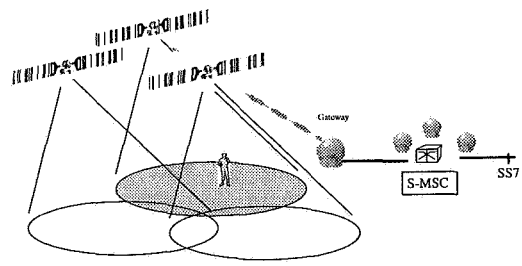


Fig. 4 „Satellite" Scenario: The coverage zone of the satellite makes up a particular location area

In this approach the location areas are satellite based and each satellite carries his location area with him as he passes over the globe at very high speeds due to his dynamic behaviour (see Introduction). This causes users to move in and out of location areas quite rapidly. dependent upon the satellite selection strategy and the speed (height) of the constellation used.
In order to obtain a first feel for the amount of location updates required though this location update scenario was chosen for reference purposes.

4.4.2 Gateway Location Area

Another possibility of defining a location area (without making use of user localisation methods) is to choose as one location area all satellites that are being serviced (in terms of connection management) by a specific gateway

(see figure 5). Since all satellites are connected to a gateway (no ISL's) at a given point in time this approach is valid.

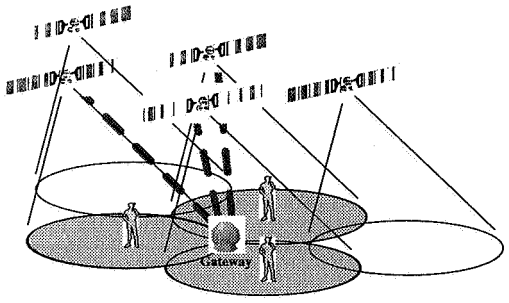


Fig. 5 „Gateway" Scenario: The coverage zone of all satellites connected to a specific gateway make up a particular location area

4.4.3 Geometric Location Area

A third possibility to define a location area is to use localisation information provided by the satellite system (elementary localisation of every user is available in the network for billing purposes) to generate a for every user specific location area. This location area will be much smaller than in any of the previous definitions (see figure 6).

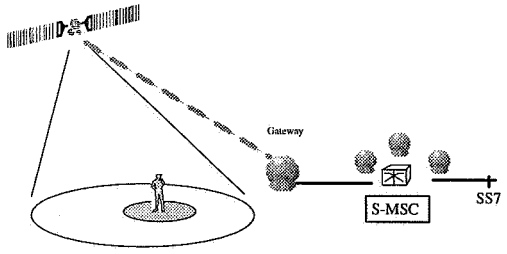


Fig. 6 „Geometric" Scenario: A specific uncertainty radius around the users periodic reported position makes up a location area

The procedure works in such a way, that every user defines a location area when his/her terminal is turned on for the first time by taking the approximate present position and drawing a circle with a certain uncertainty radius around it. Whenever the user senses that his/her present position is outside this original defined location area a location update is performed and a new location area is generated.

Every one of these definitions has its advantages and disadvantages. With the help of stochastic simulation insight may be gained as to the optimal way of defining a location area according to the methods specified.

5 SIMULATION MODELS

For the simulations performed in the evaluation process several assumptions for the user, the satellites system and the transmission environment were made. For the user it was assumed that his movement be of random nature with respect to the regular periodic movement of the mobile satellite system. An average talk time of 120 seconds was assumed (negative exponentially distributed). The activity of each user was assumed to be 10 mErlang.

In order to model the terrestrial propagation loss and the corresponding BER within the GSM system the propagation prediction model developed by Okumura and Hata combined with an estimation for the co-channel interference within the GSM system was used. The transmission probabilities of a GSM layer 2 frame for a given *CIR ratio* can be calculated according to the ETSI GSM pattern files which are available for a CIR ratios of 3 dB, 5 dB, 7 dB, 9dB and 11 dB. A modified version of these pattern files were used within the context of this paper.

The GSM propagation loss for an urban propagation environment according to Okumura and Hata can be estimated according to the following formula:

$$L_{urban}(dB) = 69.55 + 26.16 \cdot \log_{10} f_c - 13.82 \cdot \log_{10} h_b - a(h_m) + (44.9 - 6.55 \cdot \log_{10} h_b) \cdot \log_{10} R \quad (1)$$

with:

frequency	150 MHz ≤ f_c ≤ 1500MHz
base station height	30 m ≤ h_b ≤ 200 m
mobile terminal height	1 m ≤ h_m ≤ 10 m
distance to base station	1 km ≤ R ≤ 20 km

The correction factor

$$a(h_m) = (1.1 \cdot \log_{10} f_c - 0.7) \cdot h_m - (1.56 \cdot \log_{10} f_c - 0.8) \quad (2)$$

for an antenna height *hm* = 1.5 m equals to zero and is not considered for the rest of the calculations.

From formula 1 the propagation loss for suburban propagation environments:

$$L_{sburban}(dB) = L_{urban} - 2 \cdot [\log_{10}(f_c / 28)]^2 - 5.4 \text{ dB} \quad (3)$$

and the propagation loss for open propagation environments:

$$L_{open}(dB) = L_{urban} - 4.78 \cdot (\log_{10} f_c)^2 + 18.33 \cdot \log_{10} f_c - 40.94 \text{ dB} \quad (4)$$

can be derived. Interference within the GSM scenario is estimated by a normal distribution around -148 dBm.

The satellite environment used for the performance evaluation of the protocols for call setup and termination can be characterised by essentially a two state model: the channel showing Ricean behaviour for non shadowed and Rayleigh/Log-Normal behaviour for shadowed conditions [3]. The combination of Rayleigh/Log-Normal behaviour for the shadowed or bad" channel state is sometimes also called Suzuki-process.

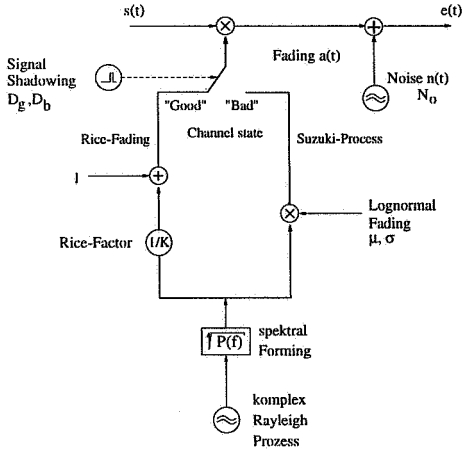


Fig. 7 Gilbert Elliot (two state) channel model for the L/S band mobile satellite radio propagation environment

The combined probability density function for the combine model equals to [5]:

$$pdf(S) = (1 - A) \cdot pdf_{Rice}(S) + A \cdot \int_0^\infty pdf_{Rayleigh}(\frac{S}{S_0}) pdf_{LN}(S_0) dS_0 \quad (5)$$

with A being the shadowing factor representing the probability of being either in the „good" or „bad" state.

Assuming a matched filter reception (Viterbi decoder) the E_b/N_0 *ratio* can directly calculated and the *BER* can be estimated using a receiver curve like the one shown in fig. 8.

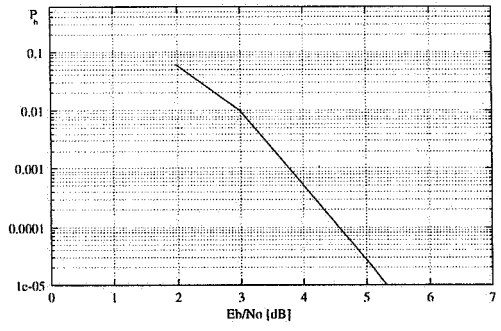


Fig. 8 Viterbi decoder receiver BER curve [2]

For the mobile satellite radio access system a bit rate of 150 kbit/s with slot duration of 800 bit using a TDD/TDMA scheme were assumed. When using a ¾ coder this amounts to two GSM layer 2 frames in every mobile satellite system layer 2 frame. For first time access to the mobile satellite system a pure Aloha channel allowing a maximum of 4 access repetitions was used.

6 SIMULATION RESULTS

In this section an exert of the simulation results to be expected for the full version of this paper will be presented.

6.1 INTER-NETWORK HANDOVER

The simulation results obtained for the modified GSM inter-MSC handover protocol are given in the following figures. For an open environment (fig. 9 and 10) the mean handover interruption time was found to be 264.6 ms (ranging from 200 to 4200 ms) with a handover failure rate of 1.38 %.

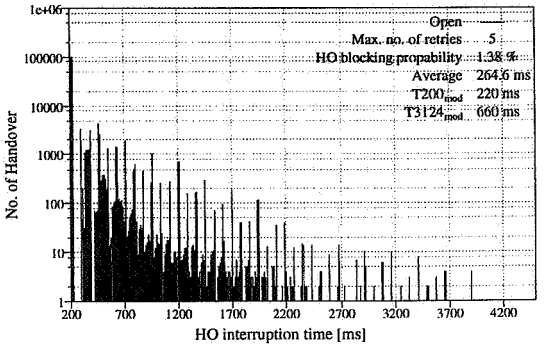


Fig. 9 Inter-network handover simulation results for the modified GSM inter-MSC handover protocol in an open environment

The cumulative probability density function (fig. 10) shows a large step around 200 ms almost down to 0.1 and afterwards a steady exponential decay up to 4200 ms.

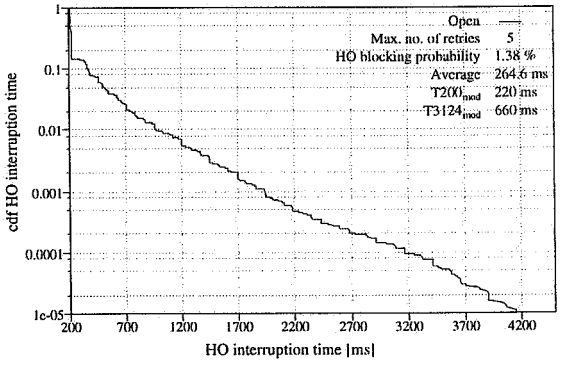


Fig. 10 Cumulative probability density function for the inter-network handover simulation results of the modified GSM inter-MSC handover protocol in an open environment

Summarising it can be concluded that in an open environment the inter-network handover is possible when making use of the modified GSM inter-MSC handover. For any other environment [6] though the inter-network handover blocking rate increases drastically (12.93% for a suburban environment) so that a handover seams only reasonable for very few cases where a connection has to be maintained at all costs and with all means. Another possibility would be to introduce a new seamless handover scheme but this would require major modifications to the GSM standard implemented.

6.2 LOCATION MANAGEMENT

Figure 11 shows the simulation results of the simulation for the satellite coverage location area approach.

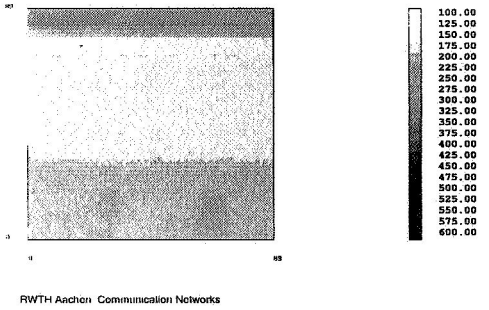


Fig. 7 Geometrical Location Update Distribution (0 - 90^0 Long. and 0 - 90^0 Lat.)

Since for the satellite coverage location area approach no terrestrial geometries need to be taken into account, the amount of location updates do only depend upon the constellation (here IRIDIUM) of the satellite system.
Figure 12 shows a first set of simulation results for the gateway based location area approach using 3 gateways.

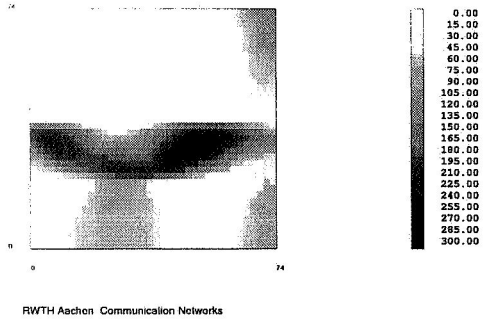


Fig. 12 Geometrical Location Update Distribution for 0 to 75^0 Longitude and 0 to 75^0 Latitude using 3 gateways

As a conclusion out of both simulation runs it can be stated that for a mobile satellite system on the average without user localisation many more location updates have to be performed. It may also be concluded that a terrestrial, gateway based location area definition is quite superior to a pure satellite system based one. The paging efficiency for an Iridium type system would be 1/48 (the Iridium system has 48 cells per coverage area) for the satellite coverage area approach and about 3 (much more paging) for the gateway (22 gateways) based solution (several satellites are connected to one gateway). The geometric location area solution follows directly from a refinement of the gateway based solution and seems to be a good compromise because the amount of location updates does only depend of the ratio of the size of the geometric segment to the average speed of the end-user. First simulation results for users within a confined area (Europe) show a linear relationship between the average-users speed (0-100m/s) and the amount of location updates necessary (0-38 updates/day for a 200km location area) Further analysis is required in this area.

7 SUMMARY AND CONCLUSION

As a conclusion of the analysis and performance evaluation of the candidate inter-network handover protocol inter-network handover between a mobile satellite and the GSM network seams feasible in open propagation environments. A handover in suburban or even urban environments appears to be unreasonable due to the high percentage of handover failure.
For the location management a terrestrial geometry based solution seems to perform best. Further investigation though is necessary, especially in the area of paging vs. location update frequency.
For future integrated terrestrial cellular and mobile satellite systems an improved decentrally organised seamless handover solution seams reasonable since the large handover interruption times can be avoided and an increased number of handover tries becomes possible because the connection can be maintained via the original communication system until the handover is successful..

8 ACKNOWLEDGEMENT

This paper is based in part on work performed within the ACTS/INSURED project of the European Commission in which the author is participating. The views expressed in this paper are those of the author and do not necessarily represent those of the project as a whole. The author wants to thank his workpackage colleagues for their comments and discussion on the subjects of this paper.

9 REFERENCES

[1] A. Guntsch et al, *Architectures and Functionalities of an Integrated GSM and Satellite System Environment*, Mobile Communications International 31,.Mai 1996, pp 62-65.

[2] Wu, W. W., *Elements of Digital Satellite Communication*, Vol. 2, Rockville: Computer Science Press Inc.,1985.

[3] G. Maral and M. Bousquet, *Satellite Communication Systems*. John Wiley & Sons, 2nd edition, 1993.

[4] A. Guntsch, *Signalling Analysis for an integrated GSM and Satellite PCN*, Proceedings IEEE WCSS 95, Smithtown, USA Nov. 27-28, 1995.

[5] A. Guntsch, *Mobility Management in an Integrated GSM and Satellite PCN*, Proceedings 46th IEEE VTC 96, Atlanta, USA, pp 1225-1229.

[6] A. Guntsch, *Inter-Segment Handover Analysis in an Integrated GSM and Mobile Satellite PCN,* Proceedings 47th IEEE VTC 97, Phoenix, USA.

[7] Enrico Del Re, F. Delli Prescoli, P. Iannucci, *Architectures and Protocols for an Integrated Satellite-Terrestrial Mobile System*, IMSC 1993, Passadena, June 16-18, 1993.

Guaranteed Handover (GH) Service in a Non-Geo Constellation with "Satellite-Fixed Cell" (SFC) Systems

Joaquin Restrepo
Universidad Pontificia Bolivariana (U.P.B.)
and
COLCIENCIAS. Colombia
email : restrepo@tlse.enst.fr

Gerard Maral
Ecole Nationale Superieure des
Télécommunications (E.N.S.T.)
Site de Toulouse. France
email : maral@tlse.enst.fr

ABSTRACT

A procedure that allows implementing a guaranteed handover (GH) service in a satellite-fixed cell non geostatioary satellite constellation is presented. Conditions to be satisfied at the call set-up demand time from a GH service subscriber are listed. Simplified implementations of the procedure are discussed, and applications of the technique are indicated. Offering GH service leads to an increase in satellite capacity depending on the proportion of GH users in the network.

INTRODUCTION

Several non-geostationary satellite constellations will offer to mobile users switched circuit services (paging, voice, fax, data) [1]. These are :

- IRIDIUM (start of service : mid 1998)
- GLOBALSTAR (start of service : early 1999)
- ICO (start of service : early 2000)
- ODYSSEY.

In all the above constellations, satellites are equipped with multibeam antennas, where each beam covers an area on the earth surface, called *satellite cell*. Should each beam remain pointed in a given direction with respect to the satellite body, the cell is called a *satellite-fixed cell.*. Satellite cells move along with the satellite and on-going calls must be handed over from one beam to the next (*beam handover*), and eventually to the next satellite (*satellite handover*).

In the event the next beam or satellite has no idle circuit to take over the handed-over call, the call is terminated (forced call termination), and this event is referred to as a *handover fail.*.

Current mobile users in Public Land Mobile Networks (PLMN) experience handover fails, but only as a result of their own mobility. PLMN fixed users do not, as they always remain in a given cell. Now, future non-geostationary satellite constellations will eventually complement PLMN for the purpose of increasing the service area. The system designer should therefore take into account the presence among candidate subscribers of fixed users, and design the system so that it offers a Quality of Service (QoS) acceptable not only to any mobile user [2], but also to any fixed one. For the latter, the reference QoS would be that of Public Switched Telephone Networks (PSTN) or current Geostationary Satellite Networks (GSN). This implies avoiding any forced call termination due a to handover fail.

A simple method to guarantee the success of any fixed user handover consists in reserving at the call set-up a channel in all the cells the user will visit. However, the duration of the call is not necessarily known at call set-up, and the number of concerned cells is consequently uncertain. Moreover the method leads to overdimensioning the satellite capacity, as the reserved circuits would remain unused most of the time.

In this paper, a technique allowing to guarantee the success of a handover, is presented. The technique is based on circuit reservation and makes use of the minimum satellite capacity during the minimum time. The corresponding service will be hereafter called *Guaranteed Handover (GH)* service, and can be offered to any category of users (fixed and even mobile). A user subscribing to the service is called a GH *user*. The duration of the GH service subscription should be known by the network management system at the instant the GH user initiates a demand for call set-up.

The remaining of the paper is organised as follows. The GH technique is first presented with general assumptions, and a procedure is derived. Then simplified implementations of the procedure are discussed for relevant cases. Examples of applications of the GH technique are indicated. Finally, results obtained by simulation implementing the GH technique in a simple network geometry highlight the increase in the required satellite capacity as a result of having to service a given proportion of GH users.

GUARANTEED HANDOVER TECHNIQUE

The proposed technique guarantees to any GH user the success of all handovers during the time interval he has

subscribed to the GH service. So doing, the technique only reserves the capacity strictly necessary, and during the minimum time, in the cells visited by that user during the specified subscription time interval. At the end of his subscription, the GH user turns into a regular user, and is not protected any more against handover fails. The GH service can be offered to any fixed or mobile user, under the condition that the user's position be known at call set-up, and for a mobile user that his speed be significantly lower than that of the sub-satellite point, i.e. negligible when compared to 7 km/s for LEO constellations and 2 km/s for ICO ones. This is achieved in practice for most mobile users (pedestrians, land and maritime mobiles, subsonic airplanes).

It is assumed that a position determination technique is available which allows the network management system to determine the position of the GH user at call set-up demand time. From the known satellites trajectories, the network management system identifies the cells that will be visited by the considered GH user during the subcription time. The originality of the proposed GH technique resides in that a circuit is not reserved in all of these cells, but only in those cells for which the GH user is at a distance less than a critical distance D_{max} which will be defined here below. If no circuit is available in one of these cells, the call is blocked. Otherwise, a circuit is reserved in each one, the call is set up, and at the time the GH user enters any of these cells, the reserved circuit is used to convey the handed over call in progress. This guarantees the success of all handovers within these cells. For any cell the GH user is farther away than D_{max} at call set-up demand time, the GH user waits until he becomes positioned at a distance D_{max} from the considered cell entry point and then requests a circuit in that cell. If the cell has idle circuits, one circuit will be reserved for his call to be handed over successfully. Otherwise, the GH user request is queued with others of his kind, and any circuit that is liberated later on, before handover occurs, is used to fulfil in priority the requests in that queue, according to a FIFO discipline. The requests in the queue have priority over both new calls and handed over calls of regular users. The number of GH calls must be kept within limits, according to the procedure discussed further below, in order to guarantee that all queued requests from GH users are satisfied in time with liberated circuits.

Definitions

Figure 1 shows the trajectory of a GH user. For convenience the figure assumes fixed satellites and a user moving to the right with velocity equal to the resultant velocity of subsatellite point and earth rotation. As this resultant velocity is identical for all users, user trajectories are parallel lines (neglecting the mobile user velocity). The arrow X(i) represents the trajectory of user (i) through cells (j), (k), (l) and (m) from his position at call set-up demand time to the end point of his GH service subscription. Other users would have different trajectories and possibly would visit a different set of cells. Therefore circuit reservations are dependent of the position of the user at call set-up demand time, and the specified time interval the GH service is subscribed for. Cell contours are represented by ellipses, although any other realistic shape could be considered. Such a contour represents a constant radiofrequency parameter (antenna gain, satellite EIRP or G/T, power flux density). Adjacent cells overlap, and therefore a transition zone exists between cells within which the handover is performed, thus ensuring the required minimum link quality at all times. To reduce the number of handovers, it is considered in the figure that the handover is initiated as late as possible. The entry point of a cell on the trajectory of the user is therefore taken to be the point of the contour that is at the extreme right of the transition zone. For a given cell (j) one defines the *longest path*, $D_{max}(j)$, as the longest path through the cell for all possible trajectories.

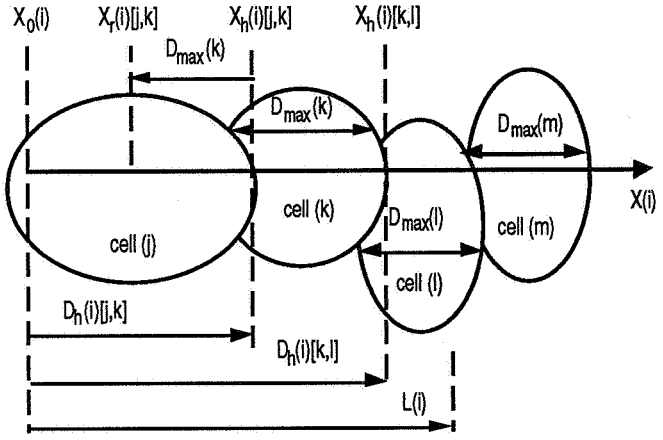


Figure 1 : trajectory of GH user (i)

The notations in Figure 1 are as follows :

- $X_0(i)$: position of user (i) at call set-up demand time.
- L(i) : length of the trajectory where the GH service is subscribed by user (i), starting from $X_0(i)$.
- $X_h(i)[j,k]$: handover point of user (i) from cell (j) to cell (k)
- $X_r(i)[j,k]$: position of user (i) at the time he should request a circuit in cell (k).
- $D_h(i)[j,k]$: distance between position of user (i) at call set-up demand time to the handover point from cell (j) to cell (k). This distance is calculated by :

$$D_h(i)[j,k] = X_h(i)[j,k] - X_0(i) \qquad (1)$$

Figure 2 shows the position of the GH user (i) and three cells (k), (l) and (m), at call set-up demand time. The regions and zones illustrated in the figure result from considering in the figure that users are fixed and cells move to the left.

- Region R_b (*'before' region*) : this region is obtained by translating cell (k) to the left so as to position the GH user (i) on the left side of the contour of the translated cell (k). Region R_b extends to the right of the right part of the contour of the translated cell (k). Its extension in the direction perpendicular to the users trajectories (vertical direction in the figure) is equal to the height of cell (k). Any user, be it regular or GH, positioned in region R_b at the call set-up demand time from the GH user (i) will have left cell (k) *before* the GH user (i) enters it. Therefore, it is certain that a circuit will be liberated before the GH user (i) enters cell (k).

- Region R_a (*'after' region*): this region is obtained by translating cell (k) to the left so as to position the GH user (i) on the right side of the contour of the translated cell (k). Region R_a extends to the left of the left part of the translated cell (k) contour. Its height is equal to that of region R_b. Any user, be it regular or GH, positioned in region R_a at call set-up demand time from the GH user (i) will enter cell (k) *after* the GH user (i) has left it. The circuit liberated by the GH user will be available to one of these incoming users.

- Region R_d ('*during*' region): this region is the region comprised between R_a and R_b. All users in this region will be located within cell (k) *during* some time interval when the GH user (i) is himself within cell (k). This means that the users in region R_d and the GH user (i) will be competing for a circuit in cell (k).

Note that the above definitions of the three regions have been illustrated in the case of cell (k), and that there would be three other different regions associated with each one of the other cells (l) and (m), as well as with any other cell visited by the GH user (i).

Figure 2 also displays the following zones :

- $Z_k\{X_0(i)\}$: zone served by cell (k) at call set-up demand time from the GH user (i)
- $Z_k\{X_h(i)[j,k]\}$: zone served by cell (k) at the instant the call from the GH user (i) is handed over from cell (j) to cell (k).
- $Z_k\{X_h(i)[k,l]\}$: zone served by cell (k) at the instant the GH user (i) is handed over from cell (k) to cell (l).
- $Z_k\{X_k\}$: zone served by cell (k) at the instant the GH user (i) is positioned at $X(i) = X_k$. This zone lays somewhere in between the two above zones, i.e. :

$$X_h(i)[j,k] \leq X_k \leq X_h(i)[k,l] \qquad (2)$$

Hence, $Z_k\{X_k\}$ can be considered as a scanning cell (k) over region R_d.

Other definitions involve the following ones :

- $N_{GH}(i)[Z_k\{X_k\}]$: number of active GH users in zone $Z_k\{X_k\}$ at call set-up demand time from the GH user (i). These GH users keep their status of GH users while their calls are being handed over from cell (j) to cell (k).
- C(k) : capacity of cell (k)
- W(i)[k] : length of the queue of GH users requesting a circuit in cell (k) at call set-up demand time from the GH user (i).

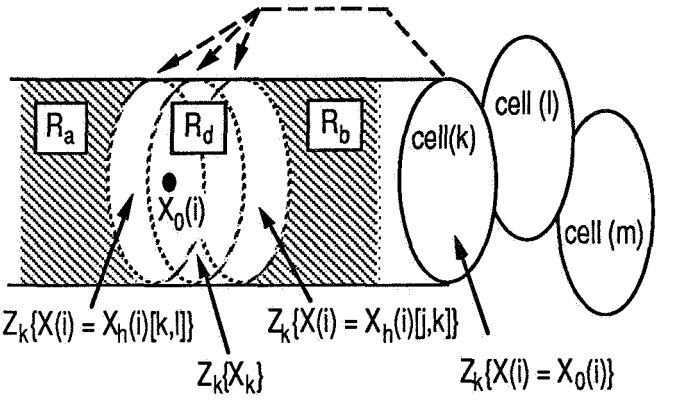


Figure 2 : Definition of regions and zones

Conditions for handover success for the GH user (i) :

In order to experience a successful handover, the GH user (i) must request a circuit in cell (k) soon enough for at least one circuit to be liberated before he enters cell (k). In the worst case, cell (k) is fully booked when the request is issued by the GH user (i). Now all users in cell (k) will travel at most the distance $D_{max}(k)$ within cell (k) before exiting cell (k) and releasing their circuits. Consequently, the GH user (i) should issue his request when reaching the position $X(i) = X_r(i)[j,k]$ at a distance D_{max} from its entry point in cell (k). Then $X_r(i)[j,k]$ is given by :

$$X_r(i)[j,k] = X_h(i)[j,k] - D_{max}(k) \qquad (3)$$

This is illustrated in Figure 1. Two situations can be experienced :

1 - Non-overlapping zones :

This happens when zones $Z_k\{X_h(i)[j,k]\}$ and $Z_k\{X_0(i)\}$ are distinct. Therefore, at call set-up demand time the GH user (i) is positioned at a distance larger than $D_{max}(k)$ from its entry point in cell (k). Hence :

$$D_h(i)[j,k] > D_{max}(k) \text{ then } X_0(i) < X_r(i)[j,k] \qquad (4)$$

Once the call has actually been set-up, the GH user (i) waits until he becomes positioned at a distance $D_{max}(k)$ from cell (k) entry point and then requests a circuit in cell (k). If a circuit is available, it is immediately reserved for him. Otherwise, his request is placed in a queue, with a FIFO discipline. The requests in the queue are prioritised with respect to any other call, be it that of a regular handed over user, or a new call from both regular or GH users. In this manner any new call attempt or regular handed over call is not accepted unless the queue is empty.

Therefore, channels liberated in cell (k) are firstly allocated to GH users whose requests are standing in the queue. Nevertheless, it must also be guaranteed that the request of the GH user (i) when entering the queue will be satisfied before his call is handed over from cell (j) to cell (k). This imposes that the number of active GH users in zone $Z_k\{X_h(i)[j,k]\}$ be less than the capacity of the cell, i.e. :

$$N_{GH}(i)\,[Z_k\{X_h(i)[j,k]\}] < C(k) \qquad (5)$$

When this condition is satisfied, the call set-up demand from the GH user (i) is accepted at the time it is initiated and the user will be allocated a circuit in cell (k) at the latest when handover occurs. If (5) is not satisfied at the time of call set-up demand, there exist a possibility for the GH user (i) to face congestion when entering cell (k) and as the GH service cannot be ensured, the call set-up demand of the GH user (i) is rejected. This is perceived by the user as call blocking.

2 - Overlapping zones :

Here, zones $Z_k\{X_h(i)[j,k]\}$ and $Z_k\{X_0(i)\}$ intersect or are tangent. It means that at call set-up demand time the GH user (i) is positioned at a distance from the entry point of cell (k) less than or equal to $D_{max}(k)$. Hence :

$$D_h(i)[j,k] \le Dmax(k) \text{ then } X_0(i) \ge X_r(i)[j,k] \qquad (6)$$

In this case, the GH user (i) immediately requests a circuit in cell (k). If there is an idle circuit in cell (k), this circuit is reserved at once for the GH user (i) in order to guarantee the success of his handover from cell (j) to cell (k). Otherwise, the call must be blocked : indeed, it is not certain that at least one circuit will be liberated, and even so, it could happen that the liberated circuit be allocated to another GH user whose request is standing in the queue and has priority over new calls.

Connditions for handover success for GH users others than (i)

It is also necessary to make sure that the acceptance of the demand for call set-up from the GH user (i) does not prevent GH users with calls in progress from experiencing a successful handover from cell (j) to cell (k). This refers to those active GH users whose request for a circuit has not yet been initiated and introduced in the queue because they have not yet reached the critical distance $D_{max}(k)$ to their own entry point in cell (k). Those users are likely to be further away than the GH user (i) from the handover point.

In the worst case, it may happen that, in the absence of the call from the GH user (i), the request of some other GH user, say (n) with $n \neq i$, would have filled the queue up to its maximum acceptable size where the capacity C(k) of cell (k) would be just large enough to accommodate all GH users, excluding the GH user (i). Now, let us consider that the call set-up demand of the GH user (i) has been accepted, and he arrives at the point where he should request a circuit in cell (k) before the GH user (n) does. Then his request loads the queue to its maximum acceptable size. Next the GH user (n) comes to the point where he should request a circuit in cell (k), and now the size of the queue exceeds C(k).

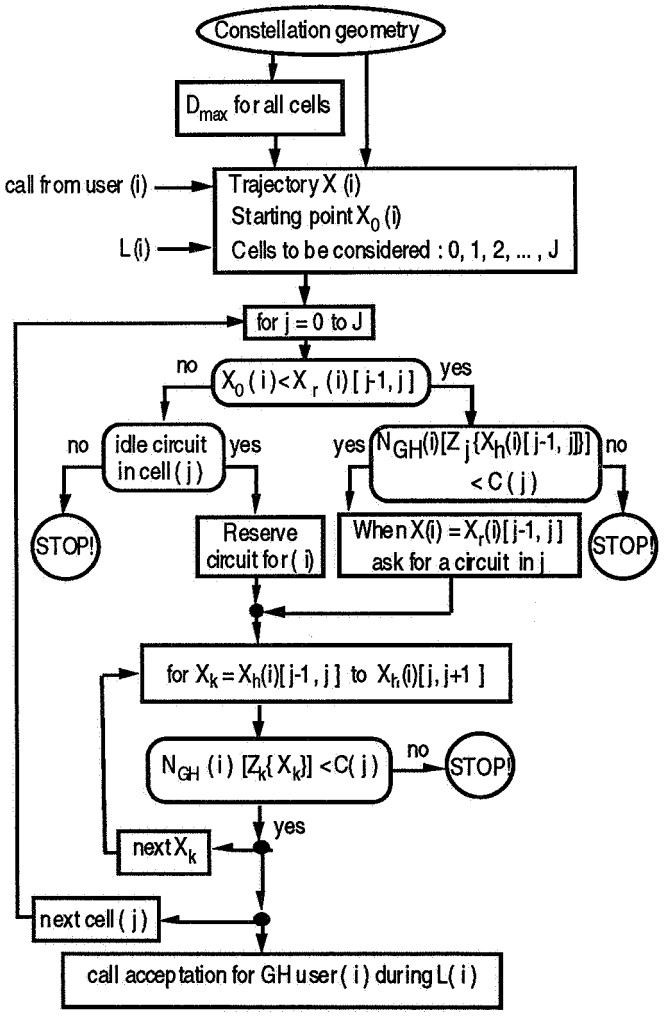


Figure 3 : Procedure for ensuring a GH service

In this way, the GH user (i) has 'stolen' the circuit of the GH user (n) who will not get a circuit and will experience a handover fail. As the GH user (n) has been accepted as a GH user prior to the set-up of the call from the GH user (i), this represents a default in the procedure. To prevent this, the number of accepted GH calls must be kept within limits, and this is reflected in the following restriction :

$$N_{GH}(i)[Z_k\{X_k\}] < C(k)$$
$$\text{for } X_h(i)[j,k] < X_k \leq X_h(i)[k,l] \qquad (7)$$

If condition (7) is not satisfied for all values of X_k in the specified interval, then the call set-up demand of the GH user (i) shall be refused.

Figure 3 summarises the complete procedure which must be followed in order to ensure a GH service. This procedure must be applied for all cells pertaining to the distance L(i). Should the procedure indicate that at a given step the call set-up demand from user (i) must be refused, then all reservations made at previous steps should be cancelled.

SIMPLIFIED IMPLEMENTATIONS

Several cases allow important simplified implementations of the procedure described above.

Identical cells

With cells identical in shape and capacity, if all cells are crossed in a similar manner by the trajectory of the GH user (i), condition (7) needs to be satisfied just for one cell, as it would then be verified for all cells .Similar considerations apply to condition (5). This reduces the number of cells to be evaluated.

Scanning effort reduction:

When the cells are rectangles aligned in parallel way with respect to the satellite trajectory, the scanning operation over all values of X_k defined in condition (7) can be reduced to a check limited to the two extreme values of the specified interval, i.e. :

a.- $X_k = X_h(i)[j,k]$ user (i) enters cell (k) (8)
b.- $X_k = X_h(i)[k,l]$ user (i) leaves cell (k) (9)

Thus doing, one will consider that once a given cell (n) is overlapped by cell (k) after cell (k) has been translated to any of these extreme positions, all GH active users in cell (n) are dwelling in the overlapping area, and are included in the calculation of the total number of active GH users, i.e. :

$$N_{GH}(i)[Z_k\{X_k\}] = \Sigma_n N_{GH}(n)[Z_k\{X_h(i)[j,k]\}] + \Sigma_n N_{GH}(n)[Z_k\{X_h(i)[k,l]\}] \qquad (10)$$

where :
- Σ_n : sum operator over n cells
- n : numbering of cell partially or totally overlapped by the translation of cell (k) to the two extreme positions defined by (8) and (9). The user (i) is included only in one of the above terms.
- $N_{GH}(n)[Z_k\{X_k\}]$: number of GH active users in cell (n). considering only those GH users who maintain their GH status in cell (k).

The above practice is pessimistic, and definitely generates a higher call blocking probability, but it leads to a drastic simplification in the network management system computational load, both by reducing the cell scanning to only the two extreme positions, and by eliminating the need for introducing the exact position of GH users others than (i).

Rectangular cells, aligned, equal size, equal capacity

Now all cells have the same rectangular shape and length, they are aligned in parallel way with respect to the satellite trajectory, then $D_{max}(k)$ is a constant, D_{max}. Additionally, they have the same channel capacity, then C(k) is a constant, C. Users separated by more than D_{max} never will be jointly served at a given time by a unique cell. This simplifies the procedure in the following way :

1- At call set-up demand time, the GH user (i) requests a circuit in the cell he is presently in and in the adjacent one. If a circuit is not available simultaneously in both cells, call set-up is refused. Otherwise the available circuit in the user's present cell is used to set up the call and the available circuit in the adjacent cell is reserved and used at call handover.

2- At every handover, the GH user (i) requests a circuit to be reserved in the subsequent cell for the next handover.

MAIN APPLICATIONS

Leased lines :

An operator may wish to use non-geostationary satellite networks as a substitute to long distance terrestrial lines. This means considering a number of GH users with infinite trajectory length L(i). The procedure for a GH service must be applied to all cells in the constellation, and for all possible trajectories.

Using a resonant constellation reduces the computational effort as the trajectories of the satellites repeat every sidereal day, and it is enough to define L(i) for a sidereal day (23

hours and 56 minutes). The set-up instant is not critical because the operator can wait for necessary conditions to be present.

Circuit allocation

The GH procedure applies both for cells with a fixed number of circuits (Fixed Circuit Allocation, FCA), or with circuits being assigned dynamically (Dynamic Circuit Allocation, DCA).

- *FCA systems :* The procedure is applied cell after cell. Satellite handover can be considered as the handover from the last cell of the current satellite to the first cell of the subsequent one.
-
- *DCA systems :* Cell capacity now corresponds to the maximum capacity that can be assigned to a single cell, C_{max}. Although C_{max} could be equal to the total satellite capacity, C_{sat}, in practice it must be lower, to avoid that other cells be impeded to offer any service. Beam handover is not a problem, as the user keeps his allocated circuit while changing cell into the same satellite.

Satellite handover

- *Handover between co-orbiting satellites* : The geometry of satellite footprints is fixed and the concept of "street of coverage" applies [4]. By representing satellite footprints as rectangles of equal size, and considering that the satellite speed is greater than the earth rotation, one can assume trajectories parallel to satellite tracks, and apply the simplifications proposed in the previous section.

- *Handover between co-rotating satellites* : This applies to adjacent polar orbits (except at the seam). As the orbits converge to the pole, the cell size is a function of the latitude.

- *Handover between non--co-rotating satellites :* This applies to polar constellation for seam orbits and non polar constellations. The user trajectory in Figure 1 is no longer a straight line, but a succession of segments with 'breaks' in between at every handover. Distances must be measured along the trajectory, regardless of the breaks.

SATELLITE CAPACITY DIMENSIONING

Simulation runs implementing the procedure for GH service in the simplified case of rectangular and equal aligned cells have delivered results reported previously [3] (in that article, GH users are called *"fixed"*, and regular ones are called *"mobile"*) . It has been shown that :

- the call set-up blocking probability for GH users is twice that of regular ones.
- the probability of call forced termination due to handover fail is zero for GH users (as expected) and nearly equal to the call set-up blocking probability for regular users.

if the service provided to regular users is taken as a reference, maintaining a constant QoS requires that the satellite capacity be increased according to the following rule of thumb :

$$\Delta C(\%) = A\, k_{GH}(\%) \qquad (11)$$

where :
– ΔC : channel capacity increase
- k_{GH} : proportion of GH users in the network,
- A : Design parameter, depending of the QoS

For a call set-up blocking probability, P_b, comprised between the interval : $1\% < P_b < 5\%$,
and a call dropping probability, P_{drop}, comprised between the interval : $0.1\% < P_{drop} < 1\%$,
the value of A is included into the interval : $0.8 < A < 0.9$.

CONCLUSIONS

This paper has presented a procedure that allows implementing a GH service in a satellite-fixed cell non geostatioary satellite constellation. Conditions to be satisfied at the call set-up demand time from a GH user have been identified. Simplified implementations of the procedure have been discussed, and applications of the technique have been indicated. It has been shown that offering GH service leads to an increase in satellite capacity depending on the proportion of GH users in the network.

REFERENCES

[1] Kevin P. Corbley, 'Accessing satellites and cellular systems', Via Satellite, Feb 96, pp 76-90.
[2] ITU-T Recommendation E 771, "Network grade of service parameters and target values for circuit-switched land mobile services" ITU-T Blue Book, 1995.
[3] J. Restrepo and G. Maral, "Providing appropiate grades of service to fixed and mobile users in a non-geo satellite-fixed cell system", Proceedings of the 2nd European workshop on Mobile/Personal Satcoms, Rome, 9-11 Oct, 1996.
[4] L. Rider, "Analytic design of satellite constellations for zonal earth coverage using inclined circular orbits", The Journal of the Astronautical Sciences, Vol 34, No 1, Jan-Mar 1986, pp 31-64.

Channel Adaptive Satellite Diversity for Non-Geostationary Mobile Satellite Systems

Hermann Bischl, Markus Werner
German Aerospace Research Establishment (DLR)
Institute for Communications Technology
P.O. Box 11 16, D-82230 Wessling, Germany
Phone: +49 8153 28-{2884,2826} Fax: +49 8153 28-1442
Email: {Hermann.Bischl,Markus.Werner}@dlr.de

ABSTRACT

The fundamental trade-off between handover and satellite diversity for the operation of the mobile user link in non-geostationary satellite PCNs (NGSO-SPCNs) is discussed with respect to the three highly important criteria bandwidth consumption, service availability, and signalling complexity. A combined strategy called channel adaptive satellite diversity (CASD) is proposed, which promises a better reduction of service impairment than handover, and which is not as bandwidth consuming as permanent dual satellite diversity. The basic idea of CASD is to use dual satellite diversity only during periods, where the channel conditions are critical. The crucial benefits and drawbacks of this method will be discussed and quantitatively analyzed.
The performance evaluation is based on a specially developed software tool, which allows the simulation of the non-geostationary satellite constellations and which includes an elevation-dependent Rice-Rayleigh/lognormal model for two correlated mobile satellite channels. All parameters of the model have been derived from channel measurement campaigns conducted for different elevation angles and different environments.
Representative simulation results are presented for the Globalstar and ICO constellation. For the sophisticated CASD scheme, some alternatives are discussed with respect to parameter settings and an optimal solution is assessed.

I. INTRODUCTION

The availability and quality of service in geostationary and non-geostationary mobile satellite systems (NGSO-MSS) is crucially influenced by the particular characteristics of signal propagation in the link between the mobile or personal user and the satellite. Specifically in the low/medium earth orbit (LEO/MEO) satellite scenario, the behaviour of the channel – and hence the parameters of any shadowing and fading processes for a channel model – are expected to be closely coupled with the varying elevation angle of the user link.
In this critical shadowing and fading environment, both *handover* and *satellite diversity* are good means for reducing possible service impairments. However, both methods have their advantages and drawbacks. Typically, satellite diversity reduces the service impairment better than handover, if realistic implementations of both are considered. On the other hand, handover is less bandwidth consuming than satellite diversity, because with satellite diversity one has to allocate bandwidth for every involved satellite at the same time, whereas with handover bandwidth as to be allocated only for one satellite instantaneously. Another important parameter to be considered is the complexity of the required signalling for both schemes.
Especially in the NGSO-MSS environment low bandwidth consumption, low terminal complexity and maximum QoS are all highly important. Consequently, we investigate the trade-off between handover and satellite diversity for different LEO and MEO systems with respect to these parameters. A new scheme called channel adaptive satellite diversity (CASD) is proposed, which combines the advantages of the two "pure" strategies.
Two of the proposed first generation S-PCN systems, namely Globalstar [1] and ICO [2], include the use of satellite diversity as fundamental system characteristic. In the course of this paper, representative simulation results are presented for these constellations. Table I summarizes the relevant parameters.

TABLE I
CONSTELLATION PARAMETERS AS CONSIDERED IN THIS PAPER.

	Globalstar	ICO
Orbit altitude	1414 km	10354 km
Orbit period	114 min	360 min
Number of satellites	48	10
Number of orbits	8	2
Inclination	52°	45°
Min. user elevation angle	20°	20°
Min. gateway elevation angle	5°	5°

II. NARROWBAND CHANNEL MODEL

The narrowband channel model considered in this paper is essentially based on earlier channel modelling work, cf. [3], [4], [5]. This section shortly recalls the most important characteristics, since it is helpful for the understanding of the CASD scheme introduced later on. Moreover, also the evaluation of simulation results is clearly facilitated with a clear picture of the underlying channel model. For a comprehensive and detailed discussion of the channel model the reader is referred to [5].

The model takes into account the following three main characteristics of the non-geostationary MSS channel(s):

1. The channel depends on the user environment and the user velocity (like the geostationary LMS channel).
2. The channel depends on the time-varying elevation angle (which is in fact quasi-static in the GEO case).
3. Channels from one user to different satellites are correlated, mainly with respect to their azimuth angle difference.

A. Elevation-dependent Single MSS Channel Model

The general Rice-Rayleigh/lognormal channel model used for a single LMS channel has been introduced and discussed in [3]. Fig. 1 illustrates the basic idea: The fading process is switched between Rician fading, representing unshadowed areas with high received signal power (good channel state) and Rayleigh/lognormal fading, representing areas with low received signal power (bad channel state). The two global channel states are driven by a two-state Markov chain.

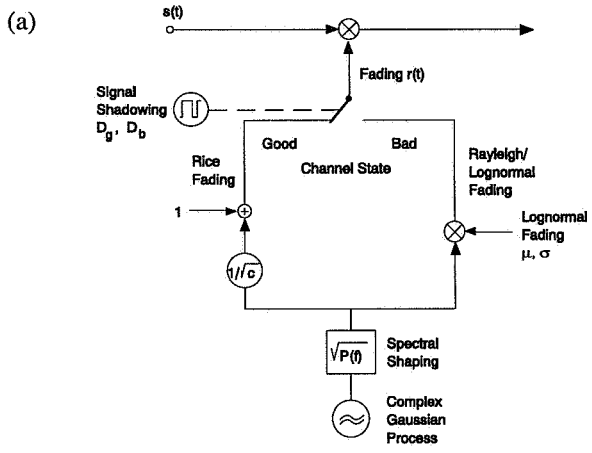


Fig. 1. Dynamic MSS channel model based on Rice-Rayleigh/lognormal fading processes.

The original model parameters have been derived for a European geostationary satellite, evaluating measurement data for different environments. However, the general model approach can easily be applied to non-geostationary LMS systems as well by assessing elevation-dependent parameter sets for the typical LEO/MEO scenario. Aiming at this goal, several dedicated measurement campaigns have been conducted for different elevation angles and different environments [6],[7],[8]. The channel measurements indicate a strong dependence on the elevation angle, especially in urban and suburban environments. As an example, Fig.2 provides a simplified set of measured and fitted elevation-dependent data for all five parameters of the model, which are valid for urban environment.

B. Correlation of Two MSS Channels

Exploiting multiple satellite visibility on Earth, the service availability (the percentage of time when the service is available) may substantially be improved. Of course, gain in service availability can only be achieved if the considered satellite channels behave *different*. Therefore, any dependency between the channels influences the benefit of satellite diversity.

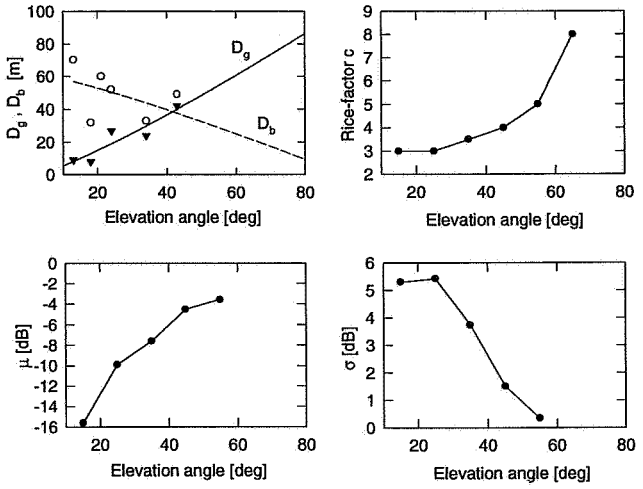


Fig. 2. Elevation-dependent channel parameters for urban environment.

In [4] a concept for modelling the shadowing behaviour of two statistically dependent satellite channels was developed, which is based on a four-state Markov chain, Fig.3. In this context, the mathematical notation and analytical treatment of dual satellite diversity has been discussed introducing a *correlation coefficient*, which depends on the elevation angles and on the azimuth separation of the two satellites.

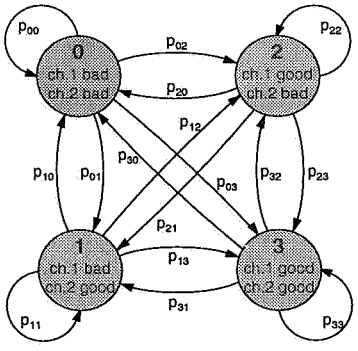


Fig. 3. Four-state Markov model for the shadowing of two correlated land mobile satellite channels.

Again, the numerical implementation of the correlation model and its associated parameters is validated by results from real measurement campaigns. Fig. 4 illustrates the dependency of the correlation on both azimuth and elevation separation. The correlation decreases with increasing azimuth separation and is smaller if the satellites have different elevation angles.

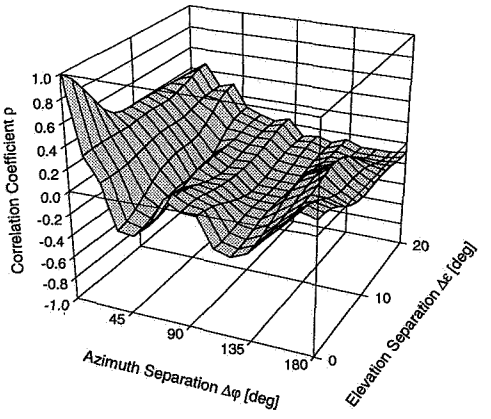


Fig. 4. Correlation of shadowing vs. azimuth and elevation separation in urban environment; $\Delta\varepsilon = \varepsilon_1 - \varepsilon_2, \varepsilon_2 = 35°$

III. Channel Adaptive Satellite Diversity

A. Earlier Work and Reference Schemes

Satellite systems offering single coverage of the service area may rely on extensive link margins to overcome not too heavy shadowing (tree shadowing, e.g.). Other systems such as Globalstar or ICO essentially provide double coverage of the earth. This feature enables the application of satellite diversity, i.e., the simultaneous communication with a user via two or more satellites. If one of these satellites is shadowed, there is some chance for another satellite being still in view to the user and maintaining the service. In this way, satellite diversity can substantially improve service availability (the percentage of time when the service is available).

A "simple" implementation of this concept, namely a non-adaptive *full diversity*, has been extensively discussed and evaluated in former work [5]. "Full" refers to the fact that with this scheme both diversity channels are *always* used when the geometrical visibility is given, i.e. they both consume link resources permanently. Considering for instance the case of both channels being in good channel state and a pure selection diversity being implemented, this scheme does obviously not use the expensive user link bandwidth resources very sparingly. On the other hand, one gains significant improvement in terms of service availability. The essentials of the full diversity scheme are given by the solid line components of the flowchart in Fig. 5.

A less bandwidth consuming alternative to the full diversity operation is *pure satellite handover*. The major trade-off is then between bandwidth consumption and signalling complexity. Considering service availability, ideal (hard) handover shows the same performance as full diversity. For realistic implementations of both, however, service impairments are slightly better reduced through diversity. This is mainly due to non-ideal (hard) handover because of the delay for signalling message exchange.

A more comprehensive discussion and a numerical performance comparison of these two alternative schemes can be found in [9].

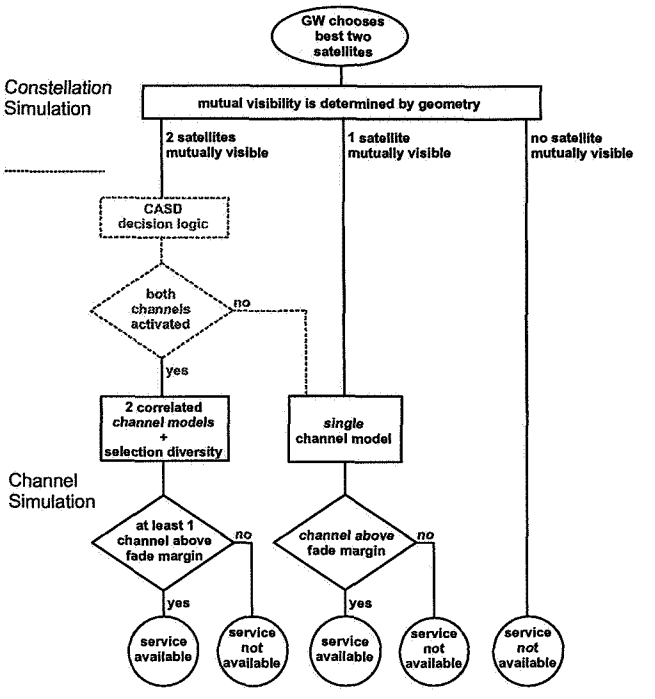


Fig. 5. Approach for satellite diversity simulation and performance evaluation in terms of service availability and required link capacity. Replacing the direct flow line by the dashed line components makes up the adaptation of full diversity toward CASD.

B. CASD Scheme

The driving force behind the development of a new and more sophisticated scheme is the prospect of combining the advantages of full diversity and pure handover without including their respective drawbacks, i.e., essentially: minimize the bandwidth consumption while keeping the service availability to a maximum (as achieved with full diversity). Considering for instance a situation where both diversity satellites are typically unshadowed during long time periods (as in highway environment) one can easily conclude that a straightforward approach to reach this goal is to release one of the channels since it does not affect service availability. However it is also obvious that the bandwidth saving potential of this scheme is closely related to the user environment, with some inherent reduction for critical environments such as urban. It is therefore highly important to elaborate a sophisticated implementation of this principle in order to meet the characteristics of the respective environments. Already before looking on any concrete solution, one can generally formulate the demand for an adaptive solution. From here, it is only a last consistent step to use the channel conditions themselves for deriving the decision to activate or release a channel, because the channel behaviour also reflects the user environment.

We propose such a *channel adaptive satellite diversity (CASD)* scheme as a consistent adaptation of the full diversity solution. The dashed line elements in Fig. 5 illustrate the

required adaptation of the full diversity scheme (solid lines) on a general level: whenever two satellites are geometrically visible, an additional logic is introduced that decides if only one or both of the corresponding channels are activated. The decision is derived from the channel conditions and results in activation and release procedures of the channels. The behaviour of this logic is therefore very similar to a handover scheme, where an old channel is forced to be released or dropped, and a new channel has to be set up. From this viewpoint, the CASD may also be seen as a hybrid solution between full diversity and pure handover.

Fig. 6 provides insight into the details of a possible realization of CASD. The process is driven by two timers T_{activate} and T_{drop} which are implemented in the receiving branches of the user terminals. The starting point of the depicted time interval is given with the situation that due to the CASD scheme channel 2 is inactive (i.e. no link capacity is used). As soon as the received power level in channel 1 drops below a given threshold (bad channel state), the timer T_{activate} controls the set-up of channel 2 via the partner diversity satellite. If channel 1 keeps below the threshold during the complete timer period, channel 2 is activated. In an operational system, of course, the complete channel set-up signalling has to be performed within this period. Thus the signalling protocol together with the round trip delay effectively set a lower bound to an applicable T_{activate}. Section IV will provide some respective numerical considerations. Once both channels are in use, a T_{drop} timer is initialized for both fading processes whenever the power threshold is exceeded. If one of the channels experiences a timeout while staying above the threshold for the whole timer period, the *partner* channel (if there is a second active diversity channel) is released. This can be modeled to happen immediately after $t = T_{\text{drop}}$ by assuming that the required release message(s) are sent just before the timeout.

The dashed parts of the fading processes in both channels illustrate the link capacity saving compared to the full diversity scheme.

IV. Performance Evaluation

We have performed extensive simulations for one gateway (GW) at -100°W/40°N, i.e. somewhere in the centre of the United States, and a number of user positions (UPs) uniformly distributed in a circle around the GW, forming its service area. The constellation parameters for the two investigated systems, Globalstar and ICO, have been set to the values given in Table I. Furthermore, we assume that the GW always selects the two satellites with the highest elevation angle for potential dual satellite diversity operation toward mobile users. The channel model is according to Section II, with a link margin of 7 dB for both systems. More details on the general simulation program flow and basic service area considerations can be found in [10], whereas a dedicated presentation and discussion of numerical results for full diversity is given in [5].

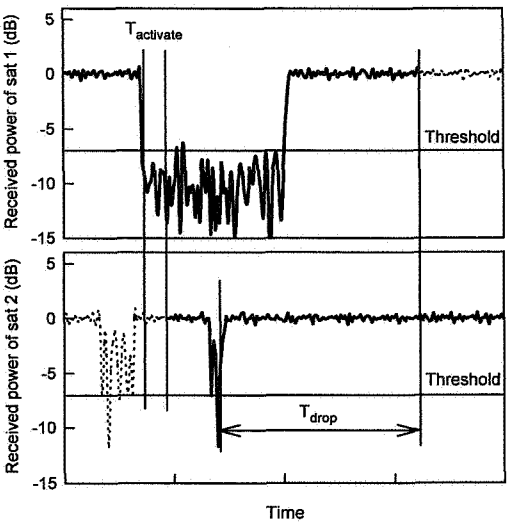


Fig. 6. CASD operation: Timers T_{activate} and T_{drop} used in the fading processes of the two satellite channels. Solid lines indicate an active channel, whereas in dashed line periods link capacity is saved.

In this paper we focus on the performance evaluation of the CASD scheme compared to its "parents" (1) full diversity scheme and (2) pure (non-ideal) handover operation. The three performance quantities are

(a) service availability,
(b) link capacity (resp. bandwidth) requirements, which correspond to the average number of used satellites, and
(c) channel set-up signalling requirements.

In a first step it is worthwile investigating the influence of the timers on these quantities in order to fix reasonable values. Fig. 7 provides corresponding simulation results for three representative environments and user velocities, showing the influence of both timers at the same time. Given an environment/velocity pair, the task of finding optimum timer values could of course consist of solving a multidimensional optimization problem subject to a certain target function in analytical form. But then the problem is basically to formulate *one* such target function since there is obviously no unique "optimum" taking into account the three given quantities; rather there is some inherent fuzziness.

A more pragmatic approach is to consider in a first step natural lower bounds for the timer values and then to reduce possible choices step by step by taking into account the target performance quantities in prioritized manner.

First of all, a lower bound for T_{activate} and T_{drop} is given by the duration of signalling message exchange for channel set-up and channel release, respectively. Here a simplified version of the formerly proposed forward satellite handover signalling protocol [11] can be applied, then basically requiring one message being sent from the mobile terminal to the GW and vice versa. Neglecting processing delay, the accumulated propagation delay alone yields a maximum (for the

worst case, i.e. minimum elevation angle) of roughly 40 ms for LEO and 180 ms for MEO constellations.

Given this lower bounds for the timer values, one may next allow a maximum of 1% decrease in service availability to be introduced by CASD compared to full diversity. (Note that the ideal full diversity case is represented in the figures by $T_{\text{activate}} = 0$ and $T_{\text{drop}} = \infty$.) Applying this to the availability surface regions, Fig. 7(a), the applicable timer pairs are immediately reduced. In the next step, it seems reasonable to give complete priority to the link capacity requirements (b) over the signalling requirements (c), because bandwidth saving is what the whole CASD is primarily for. Minimizing the link capacity requirements within the allowed timer pair region finally yields an "optimum pair". Despite some differences with respect to the environments and user velocities one finds that $T_{\text{activate}} = 100$ ms and $T_{\text{drop}} = 2$ s is a generally attractive choice for LEO systems. Specifically, these values also yield a low channel set-up rate although this criterion has not really been used for the selection.

With these timer settings extensive simulations have been performed to numerically evaluate the performance of CASD in comparison with full diversity (FD) and pure handover (PH). For the remainder of this paper we restrict ourselves to the presentation and discussion of simulation results for urban environment and a user velocity of 5 km/h.

In Figs. 8 and 9 (a) the service availability, (b) the average number of used satellites (reflecting the bandwidth consumption), and (c) the channel set-up rate (as indicator for signalling requirements) are all recorded versus the distance between user and gateway in order to catch the general performance decrease when approaching the edge of the service area. Fig. 8 provides the respective curves for Globalstar, Fig. 9 shows the ICO results for comparison. Note that "handover due to channel condition" (i.e. pure non-ideal handover) is given by $T_{\text{activate}} = 100$ ms (200 ms) and $T_{\text{drop}} = 0$.

Qualitatively, the performance of CASD – compared with FD and PH – is the same with respect to all three criteria for both systems. The figures prove that – as intended – CASD is capable of significantly reducing FD capacity and PH signalling requirements at the same time. With suitable timer settings, the respective drawbacks of both parent schemes can be largely avoided. At the same time, the decrease of service availability is virtually negligible. Other timer settings would shift the respective CASD curves either closer to the FD or to the PH ones. This characteristic can be utilized in a real implementation by introducing another level of adaptivity: the timer values need not be fixed but may adapt to different environments and/or traffic load situations dynamically during the operational phase of the system.

V. Conclusions and Outlook

A new channel adaptive satellite diversity (CASD) scheme has been proposed for non-geostationary mobile satellite systems. The basic idea is to use dual satellite diversity only during periods with critical channel conditions. The new scheme shows a better reduction of service impairment and requires less signalling than pure handover, and at the same it is not as bandwidth consuming as full dual satellite diversity. A numerical performance evaluation has been based on extensive computer simulations for two candidate systems. Besides an exact constellation calculation, the simulator includes as core module a sophisticated environment- and elevation-dependent channel model taking into account the correlation of two MSS channels.

A promising extension of the CASD scheme is an approach that does not only adapt to the propagation channel conditions but also to the traffic load variations encountered in the serving satellites or spotbeams. In case of high traffic load the extended scheme could reduce the average number of used satellites at the expense of service availability. A higher service quality could be provided by changing toward full diversity operation during periods of low or moderate load.

References

[1] J. Schindall. Concept and implementation of the Globalstar mobile satellite system. In *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, pages A11–A16, Ottawa, Canada, June 1995.

[2] P. Poskett. Satellite system architectures. In F. Vatalaro and F. Ananasso, editors, *Mobile and Personal Satellite Communications 2, Proceedings 2nd European Workshop on Mobile/Personal Satcoms (EMPS '96)*, pages 485–500, Rome, Italy, Oct. 1996. Springer.

[3] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke. The land mobile satellite communication channel – recording, statistics, and channel model. *IEEE Transactions on Vehicular Technology*, 40(2):375–386, May 1991.

[4] E. Lutz. A markov model for correlated land mobile satellite channels. *International Journal of Satellite Communications*, 14(4):333–339, July/Aug. 1996.

[5] H. Bischl, M. Werner, and E. Lutz. Elevation-dependent channel model and satellite diversity for NGSO S-PCNs. In *Proceedings 46th IEEE Vehicular Technology Conference (VTC '96)*, pages 1038–1042, Atlanta, Georgia, USA, Apr./May 1996.

[6] A. Jahn and E. Lutz. DLR channel measurement programme for low earth orbit satellite systems. In *Proceedings 3rd International Conference on Universal Personal Communications (ICUPC '94)*, pages 423–429, San Diego, CA, USA, Sept./Oct. 1994.

[7] A. Jahn. Propagation data and channel model for LMS systems. Final Report, ESA Purchase Order 141742, ESA/ESTEC, Noordwijk, The Netherlands, Jan. 1995.

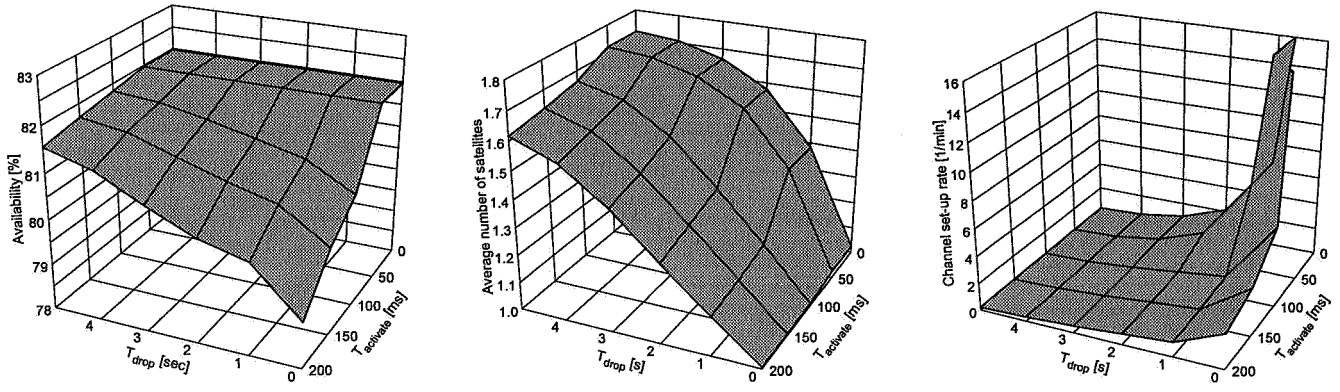
[8] A. Jahn, M. Sforza, S. Buonomo, and E. Lutz. Narrow and wideband channel characterization for land mobile satellite systems: Experimental results at L-band. In *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, pages 115–121, Ottawa, Canada, June 1995.

[9] M. Werner, H. Bischl, and E. Lutz. Handover and satellite diversity in personal satellite communications systems. In *Proceedings European Personal and Mobile Communications Conference (EPMCC '95)*, pages 140–145, Bologna, Italy, Nov. 1995.

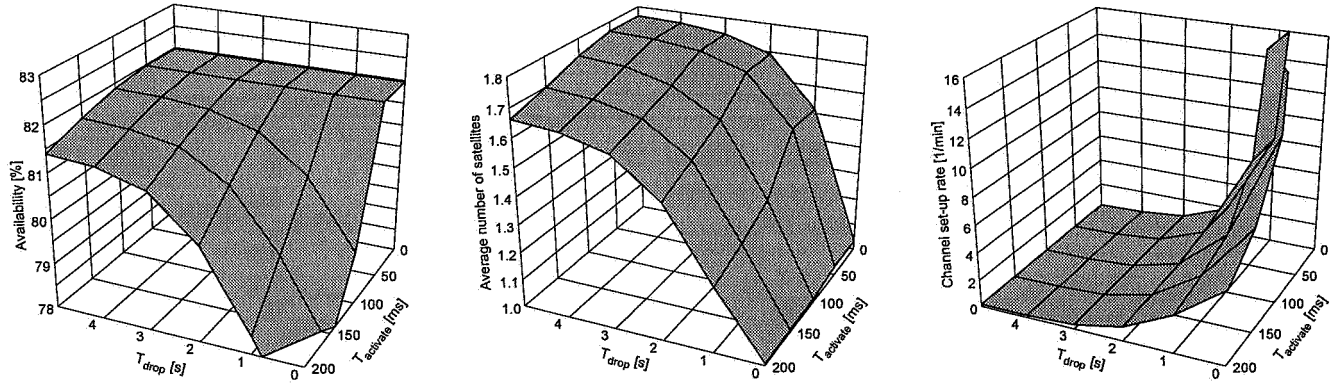
[10] H. Bischl, M. Werner, and E. Lutz. On satellite diversity and mobile user environment for NGSO S-PCNs. In *Proceedings 16th AIAA International Communications Satellite Systems Conference (ICSSC '96)*, pages 762–767, Washington, DC, USA, Feb. 1996.

[11] M. Eisenschmid. Handover in LEO/MEO-Satellitennetzen. Master's thesis, Technical University Munich, Institute of Communication Networks, Munich, Germany, Mar. 1996.

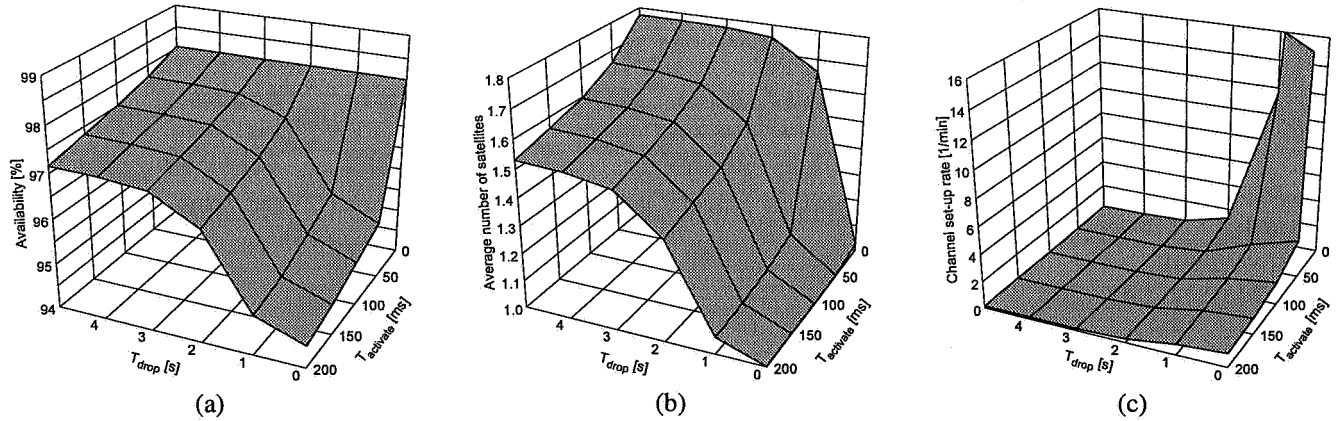
City, $v_u = 5$ km/h



City, $v_u = 50$ km/h



Highway, $v_u = 90$ km/h



(a) (b) (c)

Fig. 7. Influence of the timers T_{drop} and T_{activate} on (a) service availability, (b) satellite link capacity and (c) channel set-up signalling. Three typical environments and user velocities; Globalstar system, UP distance from GW = 400 km.

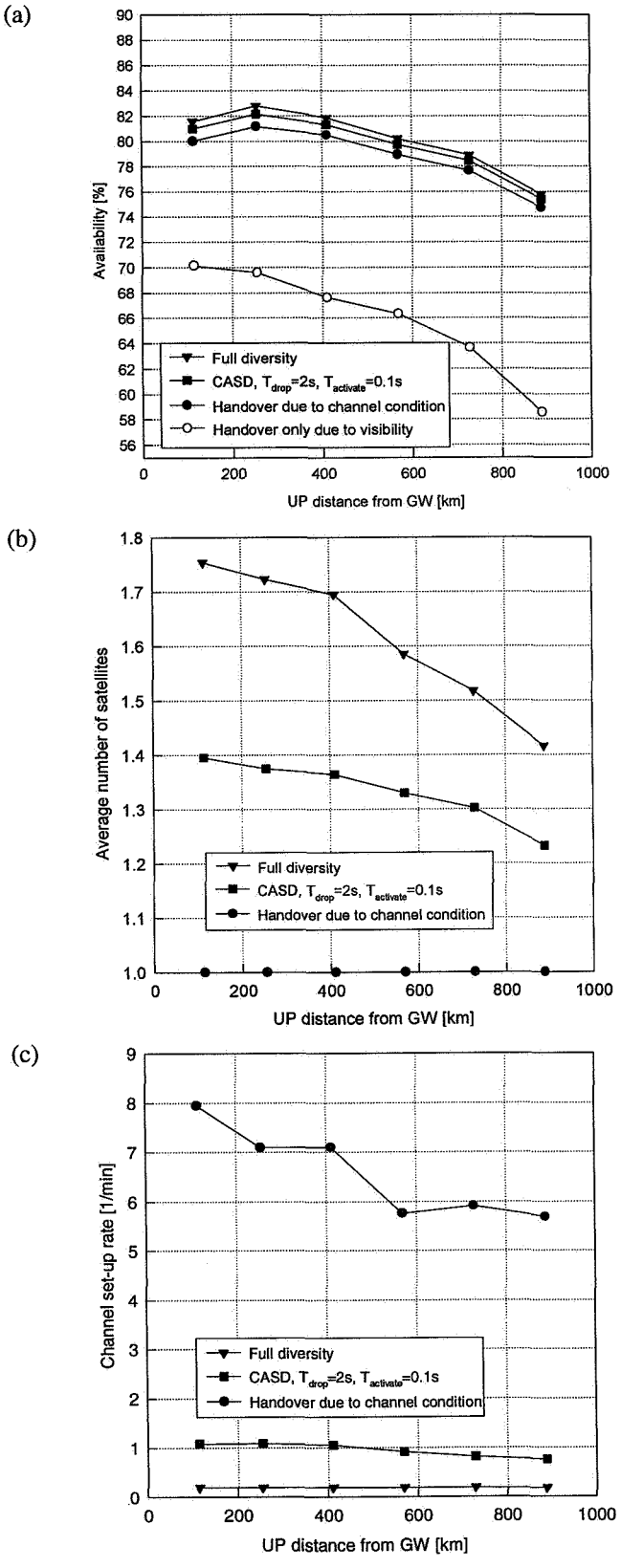


Fig. 8. Performance comparison of different diversity / handover schemes for Globalstar, city environment, $v_u = 5$ km/h: (a) service availability, (b) satellite link capacity, (c) channel set-up signalling.

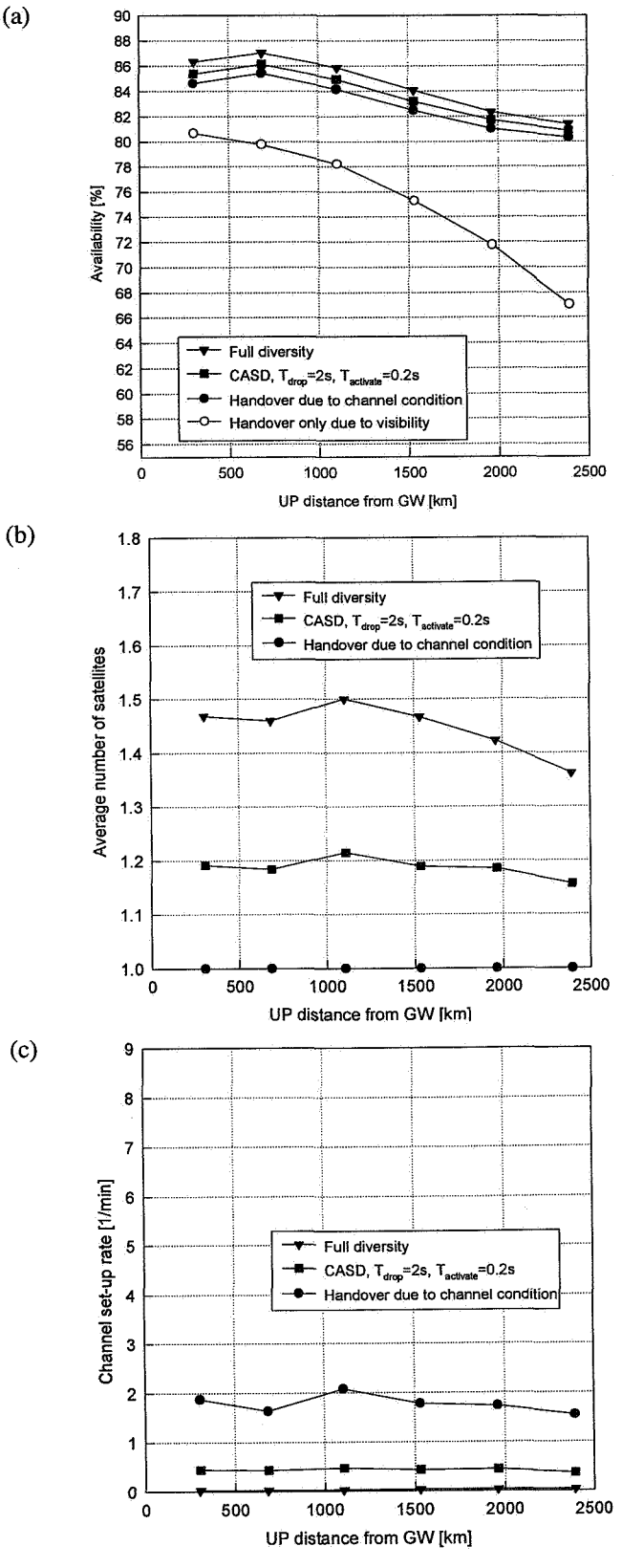


Fig. 9. Performance comparison of different diversity / handover schemes for ICO, city environment, $v_u = 5$ km/h: (a) service availability, (b) satellite link capacity, (c) channel set-up signalling.

Mobility Management and its Impact On Call Routing in Dynamic Satellite Personal Communication Networks

C.Meenan, R.Tafazolli, B.Evans
Mobile Communications Research Group
Centre for Communications Systems Research
University Of Surrey
Guildford, England, GU2 5XH
Phone: +44(0)1483 259808 Fax: +44(0)1483 259504
Email: c.meenan@ee.surrey.ac.uk

Abstract

This paper investigates the problem of call routing in a dynamic satellite personal communications network. A previously proposed scheme is presented, and its limitations highlighted. We present a new definition of an FES service area based on highest satellite connectivity. In-call FES handover rates are evaluated for a range distances from FES at various latitudes. A mobility management scheme suitable for dynamic S-PCN is described. A novel method of determining the probability of a future in-call FES handover at location updating time is presented. Finally the variation in size of the area around an FES where location updating results in a probability of no future in-call FES handover being above a range of thresholds is presented. Results are presented for two different types of dynamic satellite constellations. They are a Globalstar-like LEO system and a ICO-like MEO system.

Introduction

First generation mobile satellite communication networks will be the first systems to offer truly global roaming. These new satellite networks will provide communications to both developed and developing areas of the world where there is little or no telecommunications infrastructure or where it is not economically viable to offer terrestrial cellular coverage due to low population density. The services offered by these new systems will be similar to those currently available in second generation terrestrial systems like GSM[1] such as low bit rate telephony, paging, fax and data.

When a mobile communication network receives a mobile user terminated call, the call is routed to a Mobile Switching Center that is associated with the users last registered location area. As a location area is associated with one MSC only, this routing strategy does not involve any further inter MSC re-routing during call set-up. Any further inter MSC handovers during the call are due only to the mobility of the user.

In a dynamic satellite personal communication network it is highly likely that each Fixed Earth Station (FES) in the ground network will have MSC functionality. On receiving a user terminated call it is preferable to route the call to a FES that has a high probability of maintaining the call throughout its duration. Inter FES handover during a call is highly undesirable as the handover is likely to involve international re-routing of the call to the new FES. The re-routing of the call will have an associated delay and possible higher call charge rates as the new FES may be owned by a different network operator, both of these properties compromise the quality of service to the user. Unlike terrestrial networks an inter MSC handover can be initiated not only by a mobile terminal but also a stationary terminal. This is caused by the dynamic connectivity property of a non-GEO satellite constellation. As satellites come in and out of view, a mobile terminal will attempt to use the satellite that is providing the most optimal channel even if this satellite does not have direct connectivity with the FES that is currently handling the call. To restrict the user terminal's choice of satellites to those only offering connectivity with the current FES can compromise the quality of service.

Network Architecture

We assume that the ground segment of the satellite network consists of two types of FES. The first type of FES is the Primary Earth Station (PES). PES are responsible for the co-ordination and control of satellite resources. We assume that there are enough PES distributed around the globe to ensure that each satellite is always in view of at least one PES. The second type of FES is the Traffic Earth Station (TES). The TES is

responsible for carrying S-PCN speech and data to terrestrial networks such as GSM,PSTN and ISDN. The results in this paper refer to the TES type earth station.

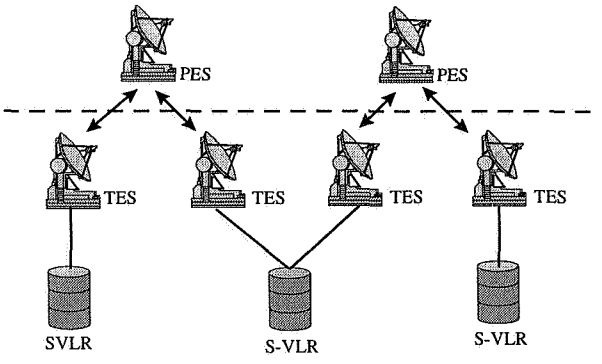


Figure 1 S-PCN Network Architecture.

Described below in Table 1 are the constellation parameters of the two systems we have considered in this paper.

	ICO-like	Globalstar-like
Number of Sats.	10	48
Inclination Angle	45°	52°
Altitude	10349km	1414km
No Of Planes	2	8
FES Min Elev.	5°	10°
Terminal Min Elev.	10°	10°

Table 1 Constellation Parameters.

Guaranteed Coverage Area

As described above due to the dynamic motion of the satellites in a non-GEO constellation, the area around the FES where a terminal can connect with the FES via a satellite is continuously changing in time. The dynamic connectivity property introduces complexities into the routing of mobile terminated calls, that are not found in terrestrial systems such as GSM. In order to reuse the terrestrial approach to call routing in a S-PCN system the concept of the Guaranteed Coverage Area (GCA) was introduced in [2]. The GCA is defined as the area around the earth station where a mobile terminal can connect to the earth station via at least one satellite that is above the minimum elevation angle of the terminal and FES. Although the instantaneous coverage area of the FES will be larger than the GCA, outside the GCA region connection with the FES cannot be guaranteed. Therefore while a terminal is roaming in the GCA of an FES, calls to the mobile are to routed to the FES. A network of FES's is in presented [2] for the Globalstar and ICO-like systems respectively. The FES's are arranged to have minimal overlap between GCA's, while ensuring that the earth's land mass is completely covered by the GCA's.

Shown below Figure 2 and Figure 3 are the CGA's for an FES at 40° latitude in the ICO and Globalstar-like systems.

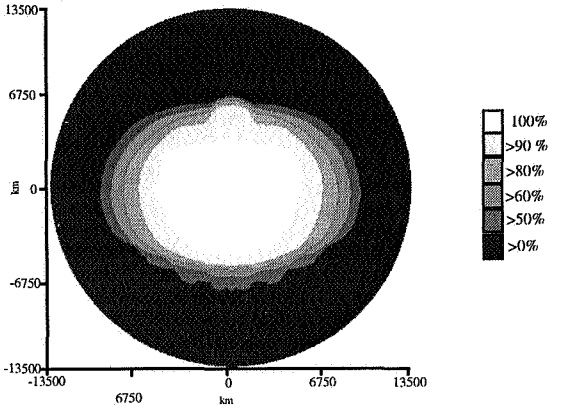


Figure 2 ICO-like CGA.

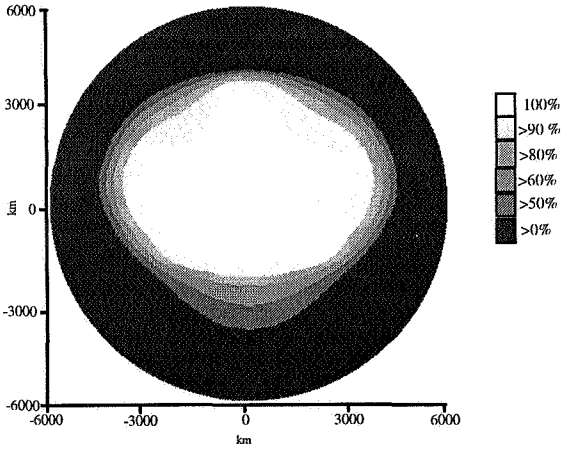


Figure 3 Globalstar-like CGA.

The size and shape of the GCA's varies with latitude due to elevation angle distributions changing with latitude. This method of defining an FES service area, requires that the mobile terminal use a satellite that is currently providing connectivity with the FES. Connection with the FES through any satellite above the minimum elevation angle of the terminal requires that the minimum elevation angle of the FES is lower than that of the terminal. This is the case with the ICO-like system where the difference in minimum elevation angle between the FES and terminal results in an circular area of radius 550km around the FES, where a terminal can connect to the FES through any visible satellite. Therefore while roaming in the GCA the mobile terminal maybe required to use a satellite that is not optimal (i.e. highest elevation angle) or to use a

satellite which is currently not available due to shadowing which becomes more prevalent at lower elevation angles[3] &[4]. These two cases could potentially lead to an FES handover, while inside the GCA of the FES. Or an alternative result is the call being dropped.

Optimum Satellite Connection Area

As described above the GCA requires the terminal use a satellite that is currently providing a connection to the FES, not the satellite that the terminal is most likely to use. As shadowing becomes more prevalent at lower elevation angles the satellite with the highest elevation angle to the terminal is the most likely to be providing a good channel. Shown below in Figure 4 and Figure 5 are the areas around and FES at 40° latitude in the ICO and Globalstar -like systems, where a terminal can connect to the FES through the satellite with the highest elevation w.r.t. the terminal.

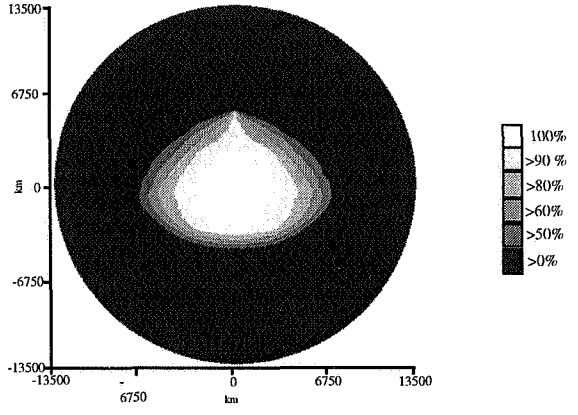


Figure 4 ICO-like highest satellite connectivity.

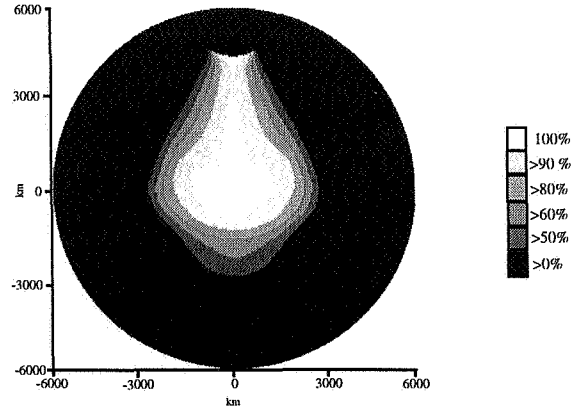


Figure 5 Globalstar like highest satellite connectivity.

When Figures 4 and 5 are compared Figures 2 and 3 respectively the extent to which the GCA restricts the terminal from using the highest satellite becomes clear. When Figure 3 and Figure 5 are compared it is clear that optimum satellite coverage area extends to a higher latitude than the GCA, this is due to the fact that a satellite is not always available above 70° latitude in the Globalstar-like system, therefore the GCA does not include this region in its definition, as connection with the FES cannot be guaranteed. However the optimum satellite coverage area is concerned only with the highest satellite loosing connectivity the FES. Although above 70° latitude a terminal may not always have a satellite visible, when one or more satellites are visible connectivity with the earth station is guaranteed through the highest satellite, in the region shown in Figure 5.

In-Call FES handover

In this section we evaluate the probability that the highest satellite w.r.t. the terminal will lose connectivity with the FES during a call and therefore induce an FES handover. Calls within the 100% connectivity area will always complete without loss of FES connectivity. To evaluate loss of connectivity with an FES during a call in the region where connectivity is less than 100%, we must consider the duration of the connection periods in relation to the call length. Consider a connection period *i*, where the user terminal is connected to an FES via the highest satellite for a period *Ct* seconds. If we assume that the call distribution length is negative exponential with a mean of *m* seconds, and that the starting point of each call is uniformly distributed within the connection period, then the probability of a call completing within the connection period is given by ;

$$P_{Complete}\{i\} = \frac{1}{mCt\{i\}} \int_{\tau=0}^{Ct\{i\}} \int_{t=0}^{\tau} e^{\frac{-t}{m}} dt\, d\tau \qquad (1)$$

The overall probability of a call successfully completing without loss of connectivity with the FES, requires that the call begins in a connection period and is completed within the connection period. Therefore the probability of a call completing successfully within a series of N different connection periods is given by;

$$P_{NoFESHandover} = \sum_{i=1}^{N} P_{Complete}\{i\} P_{Connect}\{i\} \qquad (2)$$

Where $P_{connect}\{i\}$ is the probability of the connection period *i* occurring.

Shown below are the successful call (m = 90 seconds) completion rates for FES at 40 degrees latitude in the Globalstar and ICO-like systems.

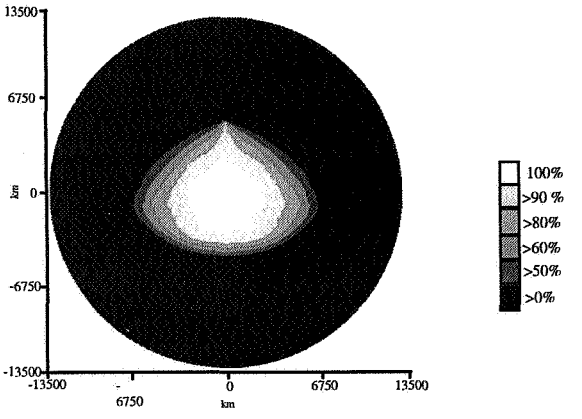


Figure 6 ICO-like Call completion rate without loss of connectivity with FES.

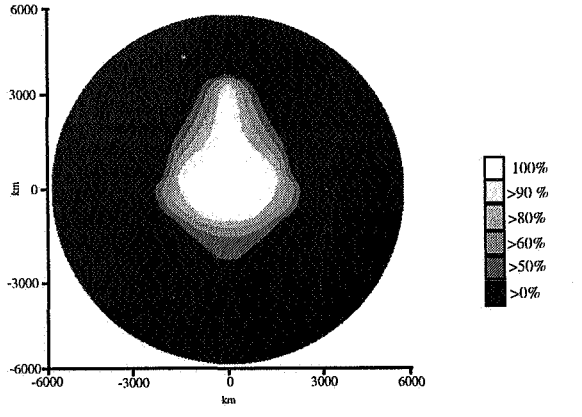


Figure 7 Globalstar-like Call completion rate without loss of connectivity with FES.

When Figure 4 is compared with Figure 6 is it clear that there is little difference between the connection probability and successful call completion probability. This indicates that the connection periods are significantly longer than the call duration. However this is not the case when Figure 5 and Figure 7 are compared. The successful call completion rate area is noticeably smaller than the connection probability area, indicating that the connection periods are nearer to being of the same order as the call duration periods. This is due to the fact that the average duration of an ICO-like satellite pass is much longer than that of a Globalstar-like satellite. Also comparing Figure 5 and Figure 7 the successful call completion rate above 70° latitude is not 100% although the connection probability is. This is due to the fact the call cannot be completed because the terminal loses connectivity with the satellite, as the Globalstar like system does not provide continuous satellite coverage above 70° latitude.

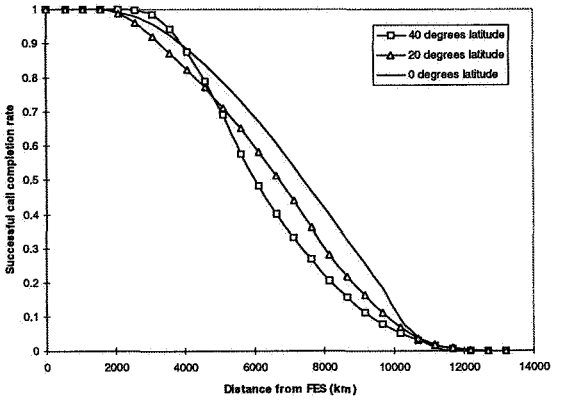


Figure 8 ICO-like FES handover probability vs distance.

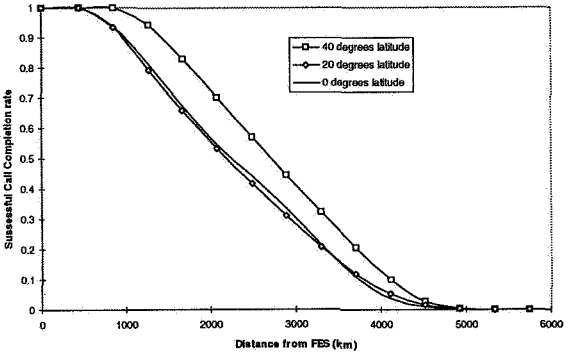


Figure 9 Globalstar-like FES handover vs. Distance.

Shown above Figures 8 and 9 are the call completion rates without FES handover against distance (averaged over azimuth) from FES's at various latitudes.

Mobility Management

Due to the dynamic motion of satellites in a non-GEO constellation re-use of the GSM approach to mobility management results in a large increase in mobility management signalling. To overcome this problem a mobility management scheme was proposed in [1] & [5], In this scheme mobile terminals make a location update after moving a predetermined distance from their last point of contact with the network. Upon receipt of a user terminated call the network must page the terminal through all spotbeams that provide coverage over the location area. This is necessary as the terminal could be anywhere inside the location area and therefore could be

monitoring the paging channel on any spotbeam. The size of each users location area can be varied to reach an optimum trade-off between location updating and paging signalling. This results in a significant reduction in mobility management signalling when compared with GSM type method, it does however assume that the mobile terminal will have the ability to measure its position while in idle mode. It should also be noted that once the mobile terminal makes a location update the new location area is a circle of a given radius LAR centred at the point of the location update. This results in the location area being positioned relative to the user location and not to any given network entity as is the case with GSM. As a result the location area cannot be guaranteed to be contained wholly within the coverage area of one FES. Therefore the S-PCN system must evaluate which FES's are suitable to have the terminal registered at. As shown earlier we can evaluate the probability of a user terminated call being successfully completed without FES handover at a given distance and azimuth away from the FES. Given that the network has no knowledge of when the next call will arrive for the terminal, and therefore where the terminal will be in the location area when the call arrives, the probability of successful call completion without FES handover must be evaluated over the entire location area, which is given by;

$$P_{FES} = \int_{\theta=0}^{2\pi} \int_{r=0}^{LAR} P_{No\,FES\ Handover}(r,\theta)\, P_{User}\,(r,\theta) r dr d\theta \qquad (3)$$

The probability P_{FES} represents the probability that a terminal making a location update, will successfully complete a user terminated call without loosing connectivity with the given FES, at some point in time after the location update. The function $P_{user}(r,\theta)$ relates to the users position probability distribution within the location area.

The above function describing P_{FES} can be evaluated by using the *Virtual Paging Cell* method introduced in [6] and the data shown in Figures 5 & 6. In this method at location updating time the location area is divided into VPC's, the probability of the user being located in one of the cells some time later (e.g. when a mobile terminated call arrives) is evaluated. Once the probability of the user being in a cell when a user terminated call arrives has been calculated, the probability of completing a call without FES handover can be found for the centre of the cell from the data in Figures 6 & 7. In this paper we assume that each user will travel in a constant direction from the centre of the location area.

The P_{FES} function was evaluated in simulation and compared with FES handover rates for various mean call duration's in the Globalstar and ICO-like systems. Shown below are the parameters used to described each mobile terminals mobility and call arrival rate.

Number of Users	20000
Direction of travel	Constant between 0 and 2π
Speed	100-150 km/h
Location Area Radius	200km
Call Arrival Rate / user	1 call/hr
Simulation Time	24 hours
FES Position	40° lat. 0° long.

Table 1 User mobility parameters.

Shown below in Figures 10 & 11 are the results for mean call duration's of 90, 400 and 800 seconds.

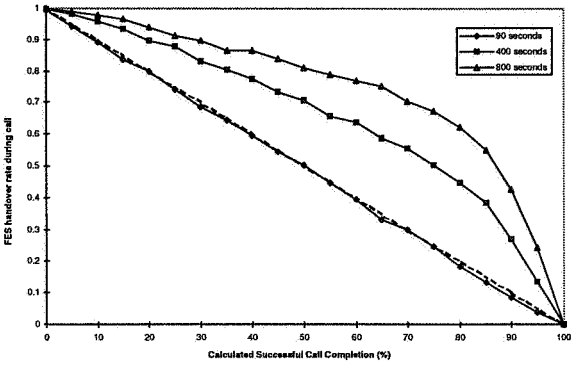


Figure 10 Globalstar-like FES handover rate vs P_{FES}.

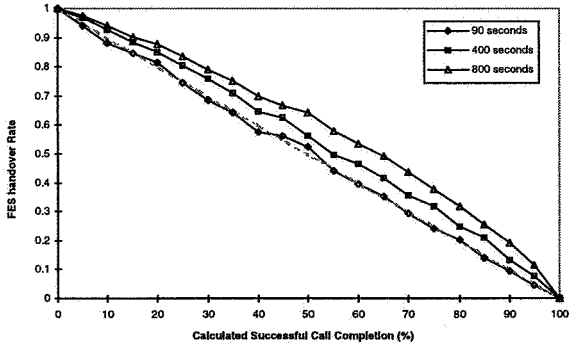


Figure 11 ICO-like FES handover rate vs. P_{FES}.

As expected for a call duration of 90 seconds both sets of results demonstrate a close match to the expected result, as the data used to calculate P_{FES} was based on a mean call duration of 90 seconds. The effect of increasing the call duration results in a higher than expected FES handover rate. The effect is much more dramatic in the Globalstar-like system, this is again due to the relative shorter duration of satellite passes in this system.

Given that we have demonstrated that the probability of a call being successfully completed without FES handover at location updating time, we can evaluated the size of the areas around the FES where a mobile terminal can make a location update with a given location area radii with a probability P_{FES} greater than a specific threshold value.

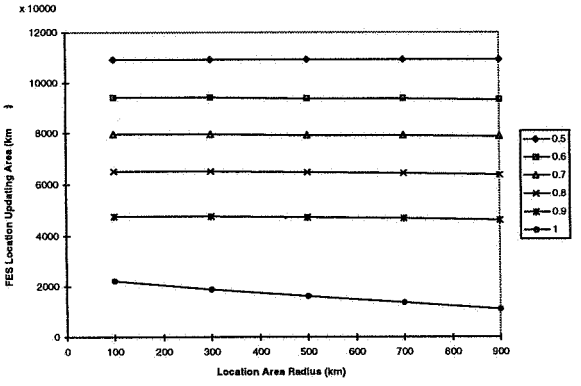


Figure 12 ICO-like FES location updating area vs. location area radius.

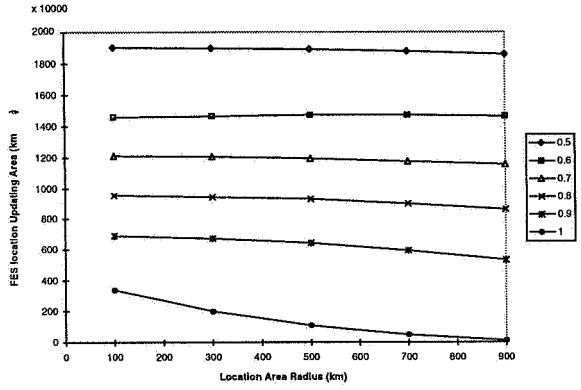


Figure 13 Globalstar-like FES location updating area vs. location area radius.

The results shown above relate to an FES at 40° latitude. They demonstrate the impact the size of the location area has on the area around the FES where a location update can be made with a value of P_{FES} above a given threshold. This effectively means that in determining the users optimum location area radius, the users position relative to the FES must be considered as well as the call arrival rate and mobility of the user.

Conclusions

In this paper we outlined the GCA approach to call routing a dynamic satellite personal communications network. The GCA method was shown to prevent the mobile terminal from using the highest satellite a significant proportion of the time, at certain distances from the FES. Consequently we presented the Optimum Satellite Connectivity area as the area around the earth station, where the satellite with the highest elevation to the mobile terminal will provide connectivity to the FES. We then presented FES handover probabilities for FES at various latitudes for a typical mobile call length distribution.

A mobility management scheme suitable for S-PCN based user positioning was revisited. A novel method of calculating at location updating time the probability of a future user terminated call requiring FES handover, for a given location area radii was illustrated. Results demonstrated the effect of longer mean call duration's on the accuracy of the algorithm. Finally we showed how size of the area around the FES where a terminal can make a location update with a probability of no future in-call FES handover being required above a certain threshold value varies with location area radii. This would be of interest to operators of FES who would wish to have users remain registered at their FES, only while the probability of any future call being completed without handover to another FES is high.

For systems that wish to exploit satellite diversity the second highest satellite can be included in the calculations.

References

[1] M.Mouly and M.B.Pautet, "The GSM System for mobile communications",ISBN:2-9507190-0-7.

[2] Saint , Race Project 2117, Deliverable No. 15

[3] E.Lutz et. al. , 1191. "The Land Mobile Satellite Communication Channel-Recording , Statistics, and Channel Model", IEEE Transactions on Vehicular Technology, Vol 40,No.2

[4] M.Parks et. al., 1993,"High Elevation Angle propagation results, applied to a statistical model and an enhanced empirical model", Electronic letters, Vol.29,No.19

[5] A.Sammut, et. al., 1994, "Mobility Management Related Signalling for a MAGSS-14 Based Satellite Personal Communications Network (S-CN)",COST227D,NationalTechnology University Of Athens Greece.

[6] C.Meenan, et. al., 1997, "Intelligent Paging Schemes For Non-GEO Satellite Personal Communication Networks", VTC 97, Phoenix, Arizona.

A Scheme to Improve Throughput for ARQ-Protected Satellite Communication

Daniel Friedman and Anthony Ephremides
Center for Satellite and Hybrid Communication Networks
Institute for Systems Research
University of Maryland
College Park, MD 20742 USA
Telephone: (301) 405-7900 Fax: (301) 314-8586

ABSTRACT

Automatic-repeat-request (ARQ) error control is often employed to assure high fidelity information transmission. However, ARQ error control can provide poor throughput for satellite multicasting. The throughput in such communication may be improved by the combination of a terrestrial network parallel to the satellite network and a judiciously modified ARQ protocol. In particular, retransmitted ARQ frames can be sent terrestrially in such a hybrid network, allowing higher throughput than in a pure-satellite network. This work presents analytic results to establish the potential for improving the throughput of satellite multicast communication employing ARQ error control by the adoption of such a hybrid network architecture.

INTRODUCTION

A satellite is excellently suited for multicast communication. As in all communication systems, an error control scheme is required for multicasting. Such schemes may be broadly classified as forward error correction (FEC) and automatic repeat request (ARQ) protocols. While numerous error control schemes for point-to-point communication (unicasting) appear in the literature, relatively few for point-to-multipoint communication (multicasting) have been presented (see [1, 2, 3, 4, 5, 6, 7, 8] for a representative selection). Error control for multicasting is hence fertile research territory.

*This work was supported by the Center for Satellite and Hybrid Communication Networks under NASA Grant NAGW-2777 and by the Institute for Systems Research under National Science Foundation Grants NSFD CDR 8803012 and EEC 940234.

We have chosen to begin our venture into this territory by examining ARQ protocols for multicast delivery of data. The typical problem in a multicast ARQ system is that since retransmissions are sent over the multicast channel, those required by only one receiving station do not benefit the other receivers. Accordingly the throughput for the system falls drastically as the number of receivers increases. Furthermore, if one receiving station is a "poorer listener" than other stations, i.e. suffers a relatively high frame error rate, then the throughput to all stations is essentially limited to the throughput achievable to that poorer listener [9].

If the retransmissions could somehow be sent only to the receivers which require them, the throughput might be improved considerably. It is natural, then, to suggest supplementing a satellite multicast system with a set of point-to-point terrestrial links between the transmitter and each receiver, as depicted in Figure 1. In such a system, retransmissions may be sent terrestrially instead of via the satellite multicast link, and an improvement in throughput might be possible. Further, if the ARQ acknowledgements are sent terrestrially as well, then the receiving stations do not require satellite transmission capability, and the cost of such stations may be correspondingly reduced.

In this article, we examine the throughput offered by such a *hybrid* (satellite and terrestrial) network configuration for unicast and multicast selective-repeat ARQ operation. In the next section we examine the throughput for unicasting and multicasting in pure-satellite and hybrid networks. Numerical examples are presented in the following section. Finally, we conclude with some thoughts for future work.

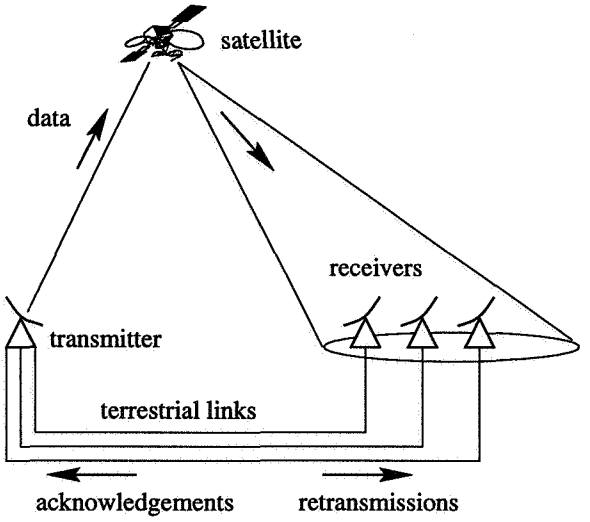


Figure 1: Multicasting in a hybrid network.

ANALYSIS

Point-to-Point Communication

We first examine unicasting in a pure satellite network. We make the following assumptions and notational definitions:

1. Infinite buffering, infinite window size; ideal selective-repeat ARQ protocol.
2. All acknowledgements are delivered without errors.
3. The satellite frame error rate (the probability a frame sent via satellite arrives in error at the receiver) is p_s, while the terrestrial frame error rate is p_t.
4. There are ℓ information bits and h non-information (overhead) bits per information frame sent either via satellite or via a terrestrial link.
5. In the hybrid network, all retransmissions are sent terrestrially.
6. We define the *throughput*, η, as the expected value of the ratio of the number of information bits delivered to a receiver per bit sent to that receiver. We will attach subscripts to η to denote the number of receivers (1 or M) and the type of network (satellite or hybrid).

We remark this last assumption can be demonstrated by straightforward analysis to be valid for plausible, implementable combinations of satellite link and terrestrial link bandwidths and error rates.

Let β denote the expected number of frames sent to a receiver per frame delivered to that receiver. (In point-to-point pure-satellite and hybrid networks, and in a pure-satellite multicast network, β is equivalent to the number of frames transmitted per frame delivered.) For the pure-satellite network we have [10, 11]:

$$\beta_{1,satellite} = \sum_{i=1}^{\infty} i\,(1-p_s)\,p_s^{i-1} = \frac{1}{1-p_s}.$$

The throughput is then

$$\eta_{1,satellite} = \frac{\ell}{\ell+h}\frac{1}{\beta_{1,satellite}} = \frac{\ell}{\ell+h}(1-p_s).$$

In the hybrid network, each frame is sent initially via satellite, and all retransmissions are sent terrestrially, so we have:

$$\begin{aligned}\beta_{1,hybrid} &= 1\,(1-p_s) + p_s\,(1-p_t)\sum_{i=2}^{\infty} i p_t^{i-2} \\ &= (1-p_s) + \frac{p_s}{p_t}\left[\frac{1}{1-p_t} - (1-p_t)\right] \\ &= \frac{1-p_t+p_s}{1-p_t}.\end{aligned}$$

This yields for the throughput:

$$\eta_{1,hybrid} = \frac{\ell}{\ell+h}\left[\frac{1-p_t}{1-p_t+p_s}\right].$$

Point-to-Multipoint Communication

For analyzing multicast networks, we preserve the assumptions of the point-to-point analysis and add the following:

1. There are $M > 1$ receivers.
2. The noise processes experienced by all receivers are independent and identical.
3. There is no competition among receivers for access to the acknowledgment channel.
4. The propagation delays for acknowledgements traveling from the receivers to the transmitter are the same for all acknowledgements.

5. The transmitter maintains a history of which stations have acknowledged which frames. Accordingly, if receiver m ($m = 1, 2, \ldots, M$) has positively acknowledged receipt of frame $\mathcal{F}$, an acknowledgement is not required from m for any retransmissions of $\mathcal{F}$ which may be required for other receivers in the network.

In the multicast pure-satellite network, the transmitter continuously sends frames via the satellite multicast channel to the M receivers, which generate respective acknowledgments to send to the transmitter. Upon receiving acknowledgments from the receivers, the transmitter retransmits the frame if one or more receivers so requested through their acknowledgements. Otherwise a new frame is sent.

Let m_j denote the number of receivers which successfully receive some frame $\mathcal{F}$ after exactly j multicast transmission attempts to deliver $\mathcal{F}$. Also let $\gamma(j)$ denote the probability with which the frame $\mathcal{F}$ is successfully delivered to all M receivers with j or fewer transmissions. Then, by counting all possible combinations of the number of transmissions required to deliver $\mathcal{F}$ to each of the M receivers, given $\mathcal{F}$ was transmitted j times, we obtain

$$\gamma(j) = \sum_{m_1=0}^{M} \cdots \sum_{m_j=0}^{M} \left[\binom{M}{m_1, m_2, \cdots, m_j} \times \prod_{k=1}^{j} \left[p_s^{k-1} (1 - p_s) \right]^{m_k} \right]$$

$$\sum_{h=1}^{j} m_h = M$$

where the multinomial coefficient is given by

$$\binom{M}{m_1, m_2, \cdots, m_j} = \frac{M!}{m_1! m_2! \cdots m_j!}.$$

Suppose a random variable A assumes the value j if the transmitter must send frame $\mathcal{F}$ exactly j times to elicit positive acknowledgements for $\mathcal{F}$ from all M receivers. If we define $\gamma(0) = 0$, then $\gamma(j)$ is the cumulative distribution function for the random variable A. Then we may calculate β, the expected number of frames sent per frame delivered to all receivers, as:

$$\beta_{M,satellite} = E[A] = \sum_{j=1}^{\infty} j[\gamma(j) - \gamma(j-1)]$$

Hence the throughput for multicasting in a pure-satellite network is

$$\eta_{M,hybrid} = \frac{\ell}{\ell + h} \frac{1}{\beta_{M,satellite}}$$

with $\beta_{M,satellite}$ calculated as above.

In the hybrid network, each frame is initially sent via satellite and all retransmissions are sent terrestrially. Hence, for multicasting in a hybrid network, the expected number of frames sent to a receiver per frame delivered to that receiver is the same as for unicasting in the hybrid network:

$$\eta_{M,hybrid} = \eta_{1,hybrid} = \frac{\ell}{\ell + h} \left[\frac{1 - p_t}{1 - p_t + p_s} \right].$$

Numerical Examples

We now turn to some numerical examples to better understand the throughput expressions derived above. For these examples, we will make the following further assumptions:

1. Binary symmetric channel (BSC) models characterize the terrestrial channels and the logical satellite channels between the transmitter and each receiver. The crossover probabilities (bit-error rates, BERs) are q_s for all logical satellite channels and q_t for all terrestrial channels.

2. The terrestrial channel BER is $q_t = 10^{-5}$.

3. There are $\ell = 2200$ information bits and $h = 48$ overhead bits in all ARQ information frames, whether sent via satellite or via a terrestrial link. (The value of h was chosen supposing the ARQ frame has a 16-bit sequence number and a 32-bit CRC for error detection. The value of ℓ was chosen to maximize the throughput in a point-to-point satellite network, which is the reference network for comparison purposes, as calculated by a straightforward differentiation method presented in [12].)

4. In finding $\beta_{M,satellite}$, we approximated the infinite summation by truncating the summation at the minimum j such that $\gamma(j) > 1 - 10^{-3}$. (We justify this truncation not only as a fair approximation, but also because, in an actual network, a station which requests retransmissions too frequently would likely be recognized by the transmitter as suffering from excessive noise, and would accordingly be disconnected from the communication.)

Calculated throughput values for point-to-point communication are presented in Figure 2. As shown in the figure, the throughput in the hybrid

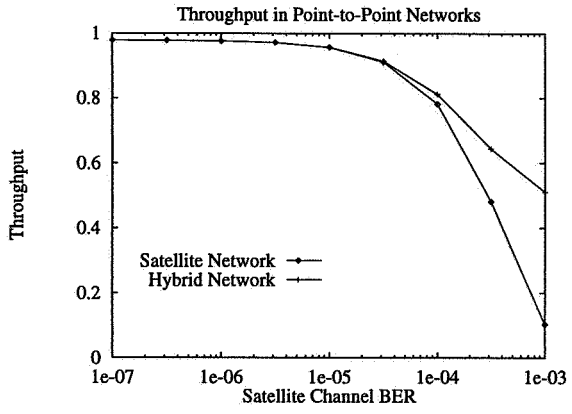


Figure 2: Throughput in point-to-point networks ($\ell = 2200$, $h = 48$, $q_t = 10^{-5}$).

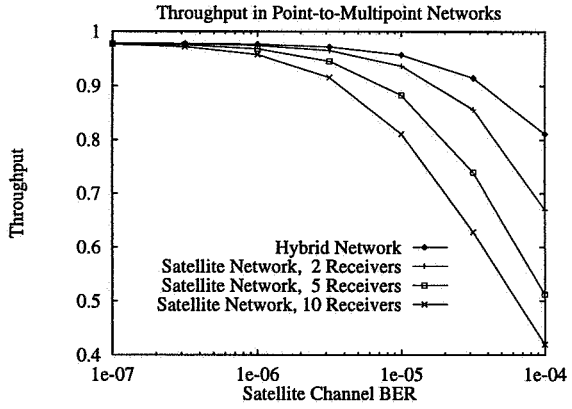


Figure 3: Throughput in point-to-multipoint networks ($\ell = 2200$, $h = 48$, $q_t = 10^{-5}$).

network comes to exceed that in the satellite network as the satellite channel BER increases. This is easily explained by the terrestrial link having lower BER than the satellite link and by the shifting of retransmissions onto the terrestrial link with the adoption of a hybrid network. Note, however, that as the satellite channel BER increases, the terrestrial bandwidth required to support retransmissions approaches the satellite channel transmission rate.

Figure 3 presents throughputs for multicast networks of two, five, and ten receivers. As is characteristic in satellite multicasting, the throughput is seen to fall rapidly as the satellite channel BER increases, especially as the number of receivers increases. However, the hybrid network provides throughput significantly superior to that available in the satellite network. While remarks concerning required terrestrial bandwidth as in the unicast case apply in the multicast case as well, the throughput improvement achievable with a hybrid network in the multicast case can be appreciable.

Additional Considerations

The inherent problem in ARQ multicasting, as stated in the introduction, is that retransmissions sent over the multicast channel do not benefit stations which do not require them. Consequently the throughput falls drastically as the number of receivers increases. In this work we have suggested a solution to this problem, namely retransmissions be sent over a system of point-to-point terrestrial links between the transmitter and each receiver. However, many considerations remain to be studied.

We have not, for example, yet examined the effect of packet lengths on throughput. While the frame length which maximizes throughput in a point-to-point satellite network is easily calculated ([12]), the optimal frame length for unicasting in a hybrid network, and for multicasting in satellite and hybrid networks, remains to be found. Adaptively changing the frame length may offer a throughput advantage, particularly at high bit error rates in the satellite channel.

We have also not yet studied terrestrial network topologies other than a star topology. Our proposed solution does not necessarily preclude other configurations. On the contrary, other topologies are not only acceptable, but perhaps even desireable. In particular, suppose the terrestrial network is a tree of terrestrial links, with the transmitter at the root node and a receiver at each non-root node. Such a tree could not only support multicasting in a hybrid network as we have described above, but would also allow a retransmission request sent by one receiver node to be serviced by the nearest ancestor node having the requested frame. The transmitter's load in servicing retransmission requests would then be reduced.

Similar possibilities arise if the terrestrial network is a wireless network, as in, for example, the case of mobile receiving nodes. For example, mobile receivers, with omnidirectional antennas, can broadcast retransmission requests to other receivers possibly nearby and receive frames over the terrestrial wireless channel. A terrestrial tree for retransmissions, albeit a continuously changing tree, is perhaps applicable for mobile receivers as well.

Hybrid ARQ schemes for multicasting, which employ FEC techniques for improving throughput

have appeared in the literature recently, and these suggest possibilities in the context of hybrid networks [7, 8, 13]. (The reader is cautioned that the term "hybrid ARQ," which is standard in the literature for ARQ schemes incorporating FEC, is not related to our term of "hybrid network" for a parallel arrangement of satellite and terrestrial networks.) In [7], for example, an adaptive type-II multicast hybrid ARQ scheme is proposed. Rate-compatible BCH codes are used for error correction in this scheme. Each time another retransmission is requested for a particular frame, the transmitter sends an increasing number of parity digits, which, when combined with the original data frame, from a series of BCH codewords of decreasing rate. Such an FEC technique would not only improve throughput, it would reduce the bandwidth required for retransmissions in a hybrid network. This is particularly important since, as we remarked in discussing our numerical examples which use a pure ARQ protocol, the bandwidth required for terrestrial retransmissions in a hybrid network approaches the satellite channel bandwidth as the satellite channel deteriorates.

There are clearly many aspects of multicast ARQ to explore. In addition to exploring such aspects, we intend to also consider how to tolerate and/or recover from errors in systems where the multicasted information has delay constraints, such as voice and video multicast systems. Because of the delay constraints, ARQ is not suited well for error control in such settings, and other schemes for mitigating error effects must be devised.

*

References

[1] S. B. Calo and M. C. Easton, "A broadcast protocol for file transfer to multiple sites," *IEEE Transactions on Communications*, vol. 29, pp. 1701–1707, November 1981.

[2] I. S. Gopal and J. M. Jaffe, "Point-to-multipoint communication over broadcast links," *IEEE Transactions on Communications*, vol. 32, pp. 1034–1044, Sept. 1984.

[3] K. Sabnani and M. Schwartz, "Multidestination protocols for satellite broadcast channels," *IEEE Transactions on Communications*, vol. 33, pp. 232–240, Mar. 1985.

[4] R. H. Deng, "Hybrid ARQ schemes for point-to-multipoint communication over nonstationary broadcast channels," *IEEE Transactions on Communications*, vol. 41, pp. 1379–1387, Sept. 1993.

[5] J. L. Wang and J. A. Silvester, "Optimal adaptive multireceiver ARQ protocols," *IEEE Transactions on Communications*, vol. 41, pp. 1816–1829, Dec. 1993.

[6] M. A. Jolfaei, S. C. Martin, and J. Mattfeldt, "A new efficient selective repeat protocol for point-to-multipoint communication," in *IEEE International Conference on Communications (ICC '93)*, vol. 2, pp. 1113–1117, 1993.

[7] A. Shiozaki, "Adaptive type-II hybrid broadcast ARQ system," *IEEE Transactions on Communications*, vol. 44, pp. 420–422, April 1996.

[8] H. Liu, Q. Zhang, M. E. Zarki, and S. Kassam, "Wireless video transmission with adaptive error control," in *1996 International Symposium on Information Theory and its Applications (ISITA '96), Victoria, British Columbia*, pp. 371–374, 1996.

[9] Y. Yamauchi, "On the packet radio multicast scheme for the personal communications era," in *International Conference on Communication Systems (ICCS '94), Singapore*, pp. 576–580, IEEE, 1994.

[10] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications.* Prentice-Hall, 1983.

[11] S. B. Wicker, *Error Control Systems for Digital Communication and Storage.* Prentice-Hall, 1995.

[12] M. Schwartz, *Telecommunication Networks: Protocols, Modeling, and Analysis.* Addison-Wesley, 1987.

[13] H. Zhao, T. Sato, and I. Kimura, "A hybrid-ARQ protocol with optimal adaptive error control for multidestination satellite communications," in *International Conference on Communication Systems (ICCS '94), Singapore*, pp. 420–424, 1994.

Satellite Based Mobile Transport Networks for Use in Feeding Wireless Local Loop Systems - An Analysis

Alexander Guntsch
Communication Networks, Aachen University of Technology, 52064 Aachen, Germany;
Phone: +49.241.80.7916, Fax: +49.241.8888.242,
email: aeg@comnets.rwth-aachen.de

ABSTRACT

The aim of this paper is to provide a comparison between terrestrial and mobile satellite based broadband network planning for use in feeding terrestrial wireless local loop (WLL) systems. After presenting a typical WLL urban scenario the terrestrial design for connecting the WLL nodes to a terrestrial backbone network via two pops (point of presence) is given. This design is followed by introducing the envisioned TELEDESIC[tm] broadband mobile satellite network including its technical features and limitations. Based on this mobile satellite broadband network a design for connecting the same WLL nodes as in the first case is provided. Special limitations due to the characteristic requirements of the 20/30 GHz transmission environment as well as dynamic channel allocation issues are taken into account.
The paper concludes with a summary of the results and an outlook into the field of satellite based mobile broadband transport networks and their possible use in supporting already existing network solutions.

1 INTRODUCTION

Within Europe and many other parts of the world heavily populated areas are serviced by one or several mobile cellular network providers in addition to the regular wirebound communication technology. In recent years with the coming liberalisation of the telecommunication market a strong competition has grown and alternative communication network providers have challenged the market. One planned approach for „getting to the customer" without having to rely on wirebound techniques is to make use of wireless local loops (WLL) or radio in the local loop (RLL)
Traditional WLL/RLL systems are based on a mix between wirebound and wireless communication techniques. The last mile to the customer is realised e.g. with a DECT relais or similar wireless communication system whereas the backbone is realised using fiberoptical and copper wire technology mixed with directive transmissions on a point to point basis. Using wireless local loop systems instead of copper wire dug into the grounds provides a fast and cost effective method for challenging the high costs of connecting the core network to the point of service, the end-user.
Instead of using terrestrial directive transmissions to connect the DECT relais to the telecommunication backbone, envisioned broadband mobile satellite communication (TELEDESIC, M-Star) could be used providing a challenging solution with the advantage of very short setup times to the potential customer. The connection to the terrestrial segment could be performed via modified terrestrial access nodes providing satellite instead of directive terrestrial transmission links.
It is the aim of this paper to compare for a given real life scenario for the town of Gelsenkirchen, Germany a terrestrial based and a mobile satellite based design. A mixed quantitative and qualitative analysis will provide the possibility to compare both solutions on a cost as well as system performance level.

2 SERVICES IN RLL SYSTEMS

Services being offered in telecommunication systems which make use of radio in the local loop technology have to at least match present wirebound services in terms of availability, QoS, security and pricing [9, 10].
Within the context of this paper, a WLL/RLL system based on DECT (Digital European Cordless Telephone) technology is used. Table 1 shows a comparison of the services currently being offered in normal analog wirebound and within the DECT communication network.

Table 1: Present analog and DECT type services

Services	Analog	DECT
voice	4 kHz (64 kbps PCM)	32 kbps ADPCM
data	300 - 28.8 kbps (33.4[1] kbps)	32 kbps
fax	group 3	group 3
supplem.	yes	yes

It can be seen that the requirement of equal or better services is achieved and that for the case of data transmission on the average a higher data rate can be offered using DECT. Through careful design and by providing appropriate link margins the comparability in terms of BER and transmission delay can be achieved as well.

[1] 33.4 kbps can only be achieved using a V34bis type modem.

3 THE „GELSENKIRCHEN" SCENARIO

In order to have a basis for cost estimation in the comparison process of the analysis a real life scenario, the „Gelsenkirchen" scenario was chosen as a basis for further analysis. In Gelsenkirchen, Germany a test net based on WLL/RLL technology is currently being investigated by one of the large German telecommunication companies.
In figure 1 the town map of Gelsenkirchen is shown. For the context of this paper only the traffic generated within the city limits of Gelsenkirchen was taken into account.

Fig. 1 Map of the city of Gelsenkirchen, Germany to be serviced via wireless local loop technology (topologic height differences < 25m).

Investigating the population distribution of Gelsenkirchen and the mix between residential and business areas within the city a potential worst case traffic profile may be generated. Since a potential WLL/RLL access network is not seen as a replacement, but as competition to normal wirebound telecommunication the aimed for market segment of the overall telecommunication traffic generated was set at 10%.
In addition it was assumed that every user considered produces up to 75 mErlang of telecommunication traffic outside his/her local premises. In real life large local company-wide telephone systems as well as a restructuring of the telecommunication tariff has lead to change in telecommunication behaviour for residential areas leading to a slight decrease of this rather high traffic value.
In figure 2 the calculated offered traffic distribution is portrayed.

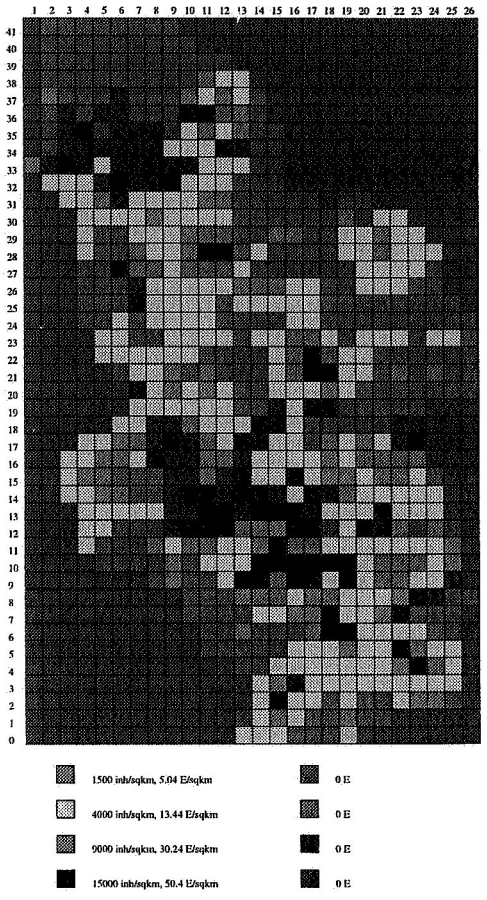


Fig. 2 Population/desired telecommunication traffic (10%) distribution for Gelsenkirchen (1 square = 383 m X 383 m).

For the purposes of the evaluations within this paper it was assumed that 10% of all city areas regardless of specific location should be covered. A pre-selection of specific areas to be covered would result in a different, easier and cheaper approach for carrying the desired telecommunication traffic.

4 TELECOMMUNICATION BACKBONE NETWORKS

In the following two sub-sections the specifics that need to be taken into account when performing a terrestrial or mobile satellite based connection planning of the DANs/SANs (Satellite Access Nodes) to the core telecommunication network (pops) shall be discussed.

4.1 TERRESTRIAL NETWORKS

The terrestrial backbone connection to the DANs (DECT Access Nodes) is most easily performed by using directive transmissions on a point to point basis from the DANs to the other network elements required for collecting and routing the traffic to the pops desired.
Of course in real life fibre optic or other connections are used wherever available but for the purposes of the comparison a pure directive transmission based terrestrial

network shall be considered. Figure 3 shows the principal hierarchy of the terrestrial backbone network.

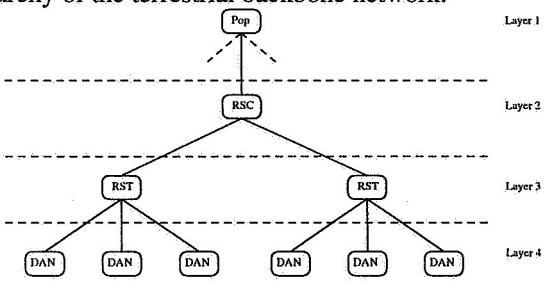


Fig. 3 Four stage terrestrial hierarchy

The structure resembles a tree like architecture including the DANs as leaves and two stages of routers/multiplexers collecting the traffic to be fed to the terrestrial core network pop.
Typical DANs will allow a maximum of 60 Erlang of traffic to be collected with the respective loads of 300 and 1500 Erlangs of traffic for the two stages of routers/multiplexers.

4.2 BROADBAND MOBILE SATELLITE NETWORKS

Future broadband mobile satellite systems like TELEDESIC are envisioned to be complementary to terrestrial wireless networks. They are a further development of first generation mobile satellite systems like IRIDIUM, GLOBALSTAR or ICO and will allow to carry an increased amount of traffic providing high rate speech and data services. Their main advantage though over any terrestrial solution is the ability to reach many parts of the population world-wide which would due to very high investment costs of terrestrial networks never get access to advanced digital applications through terrestrial means.
Figure 4 shows the TELEDESIC LEO (Low Earth Orbit) constellation of 840 satellites at 700 km orbital height operating at 20/30 GHz (down-/uplink)[8].

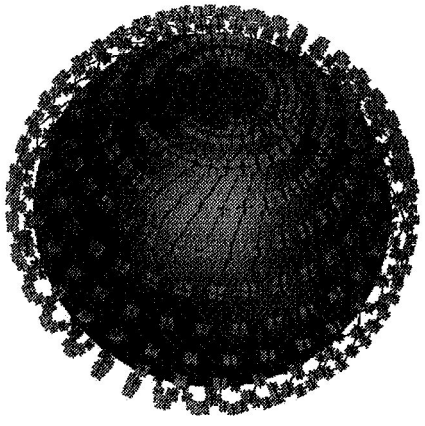


Fig. 4 TELEDESIC Constellation (840 satellites !) [7]

The 840 satellites are arranged in 21 planes à 40 satellites plus spares. Orbital synchronisation will be random which allows to save fuel otherwise necessary to maintain the exact position over time and space. Every satellite is equipped with a set of intelligent phased array antennas which will allow him to provide coverage to fixed earth cells compensating the orbital satellite motion over time. Whenever another satellite „sees" one of these fixed earth cells under a higher elevation angle than the present one a handover is performed. At hat equator every satellite services a total of 64 super-cells (160 km x 160 km per super-cell) which themselves are subdivided into 9 smaller cells. These 9 cells are scanned using a SDMA approach with a hop time of 2.276 ms (512 bit) per hop including a 0.292 ms guard time [7,8]. These 512 bits will allow to transport one 53 byte ATM (asynchronous transfer mode) packet per burst including some extra bits for FEC and signalling.

A possible scenario for making use of the TELEDESIC system for servicing a terrestrial RLL/WLL system is shown in figure 5 below.

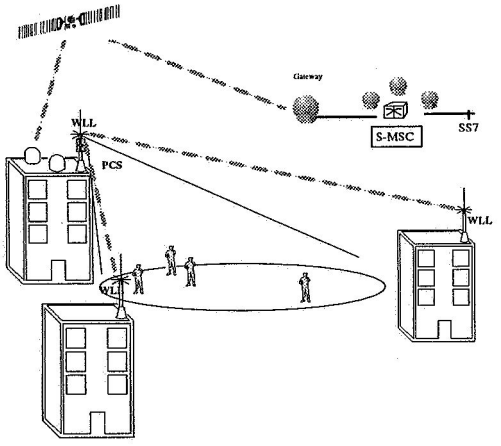


Fig. 5 WLL/RLL scenario with a broadband mobile satellite based connection to the pop

Making use of 500 MHz of spectrum in the 20/30 GHz band the TELEDESIC system is able to support 900 32 kbps channels per cell. For these channels up- and downlink are handled in a different way. Where as the down link make use of a frequency division multiplex (FDM) scheme, the uplink uses an advanced time division multiple access (ATDMA) scheme which is packet based.

5 SIMULATION MODEL AND ASSUMPTIONS

For the terrestrial calculations a combination of a center of gravity algorithm for the placement of the DANs [2] and a genetic network design algorithm [1] for the calculation of the optimal routing and router placement position was used. The traffic values for the DANs and the router/multiplexer network elements were assumed as stated in the previous section.
For the broadband mobile satellite network the most challenging area encountered was the development of an appropriate rainfall model for the area considered. Rain accounts for the highest fading losses within the LOS (line of sight) propagation environment at 20/30 GHz.

Figure 6 gives an overview of the different rainfall rates and probabilities responsible for the propagation fades over Europe according to CCIR Rep. 563-2.

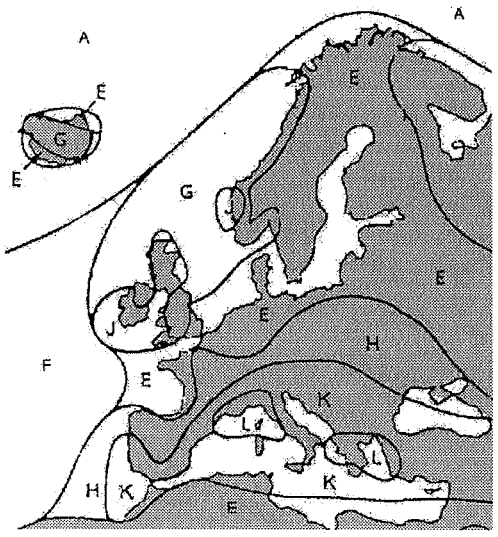


Fig. 6 Rainfall areas over Europe according to CCIR Rep. 563-2

The corresponding table indicating the % of time that the given rainfall rate for every zone is exceeded is given in table 2 below.

Table 2: Rainfall parameters for the rainfall zones in figure 6 according to CCIR Rep. 563-2.

% of time	A	E	F	G	H	J	K
1.0	0	1	2	0	0	0	2
0.3	1	3	4	7	4	13	6
0.1	2	6	8	12	10	20	12
0.03	5	12	15	20	18	28	23
0.01	8	22	28	30	32	35	42

The predicted attenuation with the given rain rate R_0 can be calculated following the CCIR approach [3] to:

$$A(R_0) = a \cdot R_0^b \cdot L_{Eff}. \quad (1)$$

$$\text{with } L_{Eff} = \begin{cases} \dfrac{L_{Actual}}{1-\exp(-cL_{Actual})} & \text{for } R_0 \le 10 \text{ mm/h} \\ c & R_0 > 10 \text{ mm/h} \end{cases} \quad (2)$$

$$c = \gamma \cdot b \ln(R_0 / 10) \cos\varepsilon \quad (3)$$

ε := elevation angle to the satellite

$\gamma = 1/22$

$$a(f) = 4.21 \times 10^{-5} \cdot f^{2.42} \quad , 2.9 \le f \le 54 \text{ GHz} \quad (4)$$

$$b(f) = 1.41 \cdot f^{-0.0779} \quad , 8.5 \le f \le 25 \text{ GHz} \quad (5)$$

Taking the right noise temperature at the receiver and the technical specifications of a candidate mobile satellite system (LEO) into account the following graph portraying the received E_b/N_0 over time can be calculated:

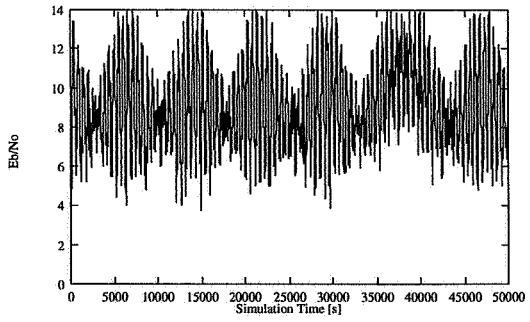


Fig. 7 Calculated E_b/N_0 over time for transmission at 20 GHz without power control.

From this received E_b/N_0 distribution over time the appropriate BER and packet loss rates can be derived. Using a 7/8 BCH code for example to protect a possible ATM cell, 1 bit error per packet could be corrected so the packet error rate would equal the probability of having 2 or more bit errors within the 512 bits/burst.

In order to compare the resulting probabilities for dropped and blocked calls to the terrestrial planning scenario it has proven to be useful to define a measure of quality for the mobile communication scenario. Based on a similar approach used for DECT type communication scenarios [6], a quality of service (QoS) measure was defined to:

$$QoS = \frac{n_{connection\_block}}{n_{connections\_attempted}} + 10 \times \frac{n_{connection\_drop}}{n_{ongoing\_connections}} \quad (6)$$

It can be seen in (6) that within this quality measure a dropped call is estimated to be 10 times more undesirable than a dropped call, i.e. an unsuccessful call attempt. It is further assumed that a QoS figure of 0.01 is the highest still acceptable QoS figure to a potential end-user. With this figure in mind the mobile satellite based solution can be compared to the purely terrestrial one.

Since not only poor weather condition do account for a high QoS value but also the blockage due to insufficient channels per super-cell under special user demand conditions, the channel allocation scheme has to be taken account as well. Within the context of this paper a first available dynamic channel allocation (FADCA) strategy was taken as reference. With FADCA a potential channel m_i is only allocated if the following inequality for the carrier to interference ratio (CIR) of the channel m_i holds true:

$$CIR_i \le CIR_{threshold} \quad (7)$$

CIR_i can be calculated by summing up all of the potential co-channel interference for a potential chosen channel m_i where as $CIR_{threshold}$ is a given value that indicates how much CIR can be tolerated during reception over any particular channel within the system. The antenna which is mostly responsible for the level of received CIR within the TELEDESIC system was estimated to be sufficient

narrow to allow the full allocation of the 900 channels within every cell of a super-cell at a given point in time. This way the full spectrum can be used in every cell during every 2.276 ms time slot (every 9th) that the spectrum is available to it.

6 ANALYSIS AND SIMULATION RESULTS

In a first step the appropriate location of the DANs/SANs has to be calculated. As mentioned before this was done using a center of gravity (COG) algorithm [2] and the traffic distribution for the given scenario depicted in figure 2.
Figures 8 show the results of DAN/SAN geometric distribution for the given scenario.

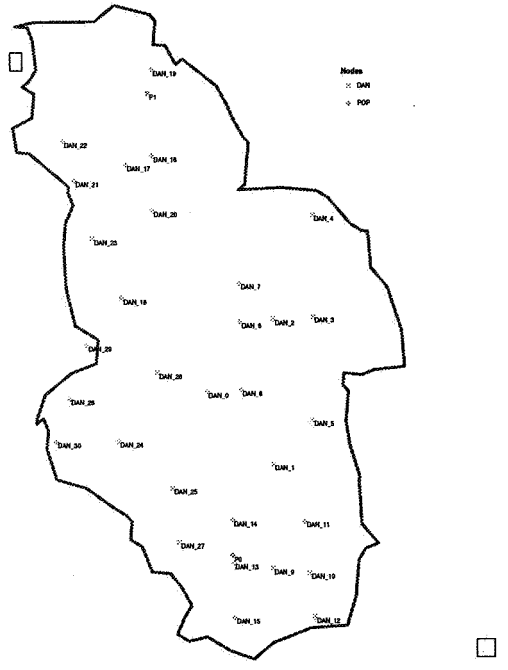


Fig. 8 Calculated DAN locations using the COG algorithm
A total of 30 DANs/SANs where needed to achieve the desired coverage of 10%.

This distribution of the DANs/SANs was taken as a basis for both the terrestrial as well as the mobile satellite based evaluations. It was estimated that the costs per DAN are equivalent to the cost per SAN. It was further assumed that a directive terrestrial connection would be as expensive as a connection to a mobile satellite. Therefore for the quantitative cost comparison the DAN/SAN network will not have to be taken into account and only the additional terrestrial hardware and the extra links necessary to connect this hardware to the DANs will have to be compared to the system costs of the potential broadband mobile satellite solution.
In figure 9 the results of the placement of the multiplexers/routers is given. In addition to the traffic distribution given in figure 2 the topology of the city was taken into account. The net was optimized using a genetic algorithm [1]. A total of 1000 generations were calculated for the terrestrial net until this final (cheapest with single link type) solution was achieved. It is made up of 10 first level, 3 second level and 92 directive connections between the network elements.

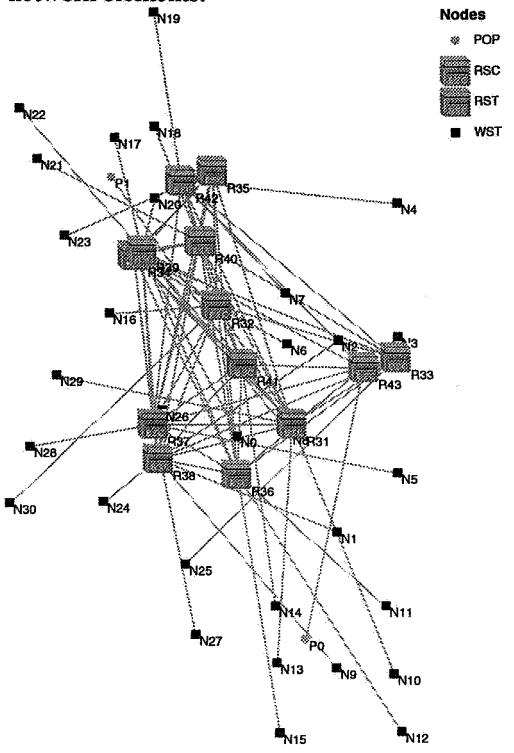


Fig. 9 Optimized terrestrial network using a genetic algorithm developed at the Institute for Communication Networks [1]

The additional cost generated by the supplementary routers/multiplexers and the additional directive connections amounts to about 1400 $ per Erlang carried (typical market values for directive connections, multiplexers and routers were assumed) which can be further interpreted to about 100 $ per user (70 mErlang peak traffic per user). This value constitutes the costs per user that a potential broadband mobile satellite solution could consume assuming the same pricing for calls carried as in the terrestrial network.
For the Gelsenkirchen scenario an average of 9.2 Erlangs traffic per km^2 is carried via the WLL/RLL system solution. The population density amounts to 2760 inhabitants per km^2. For Germany as a whole the average population is less than 1/10 of this value: 230 inhabitants per km^2. This illustrates that careful planning of the fixed earth cells is necessary in order to avoid very high loads of traffic within one super-cell and only low amounts of offered traffic in an adjacent one.
Using elementary mathematics the TELEDESIC system would be able to carry the equivalent load of 14 super-cells over the area of Germany which amounts to a total of 113400 simultaneous calls that could be served. The average number of calls per super-cell that can be served taking the potential weather conditions into account shall be looked into next. In figure 10 the simulation results for

the GoS parameter assuming the full use of the spectrum (500 MHz) and two different availabilities dependent on weather conditions (99.9% and 99.99%) are portrayed over the offered traffic per square kilometer. The increased availability from 99.9% to 99.99% is obtained by adding extra link margin which enables to withstand heavier rain but drawing heavily on the mobile satellites power supply.

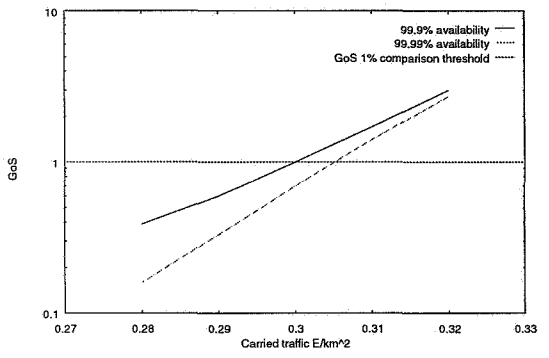


Fig. 10 QoS simulation results using 900 channels per super-cell (500 MHz) and two different availabilities due to weather conditions (99.9% and 99.99%)

In figure 10 it can clearly be seen that neither solutions offers the required amount of telecommunication resources on average to support a telecommunication scenario as depicted in figure 1. Even though this result could have been expected taking the amount of free channels per super-cell and the traffic to be carried into account, the simulation results also show that the average traffic per km^2 in Germany taking a 10% market share into account can also not be serviced sufficiently (0.92 E/km^2 rush hour traffic).

As a result it can be concluded that either due to traffic channel limitations only regional traffic or specific customers (corporate networks) could be serviced or that the amount of spectrum currently planned would have to be extended. Another alternative would be to make use of even smaller antenna spots, but due to limitations in antenna technology (and due to subsequent pricing) this alternative seams rather futuristic. With regional coverage for example, the scenario as depicted in figure 1 can be offered sufficient traffic channels because the placement of the scenario on the border of four cells within one super-cell would four fold the local capacity amounting to 3600 traffic channels to support the 2055 Erlangs of offered rush hour traffic.

7 SUMMARY AND CONCLUSION

In this paper a comparison between a terrestrial and broadband mobile satellite based planning for connected a given WLL/RLL scenario to the telecommunication backbone has been provided.

For both approaches an individual solution was calculated first using the same boundary conditions for both cases. As main conclusion of the comparison process it can be stated that the broadband mobile satellite solution will only be able to service a fraction of the expected offered traffic (10% market share) with its current allocated spectrum as compared to the terrestrial solution. For a country like Germany an average market share of 3.3% could be carried which is much less than the desired 10% commonly assumed for new telecommunication network providers. In urban areas even less percentage can be carried due to a strong concentration of the offered traffic per unit area.

Due to this limiting behaviour it can be concluded that only regional coverage or specific customer/corporate network solutions can be considered where a cost of about 100$ per customer serviced could be used to build an equivalent broadband mobile satellite solution. If different than terrestrial telecommunication tariffs are offered to the end-user connected over such a solution, this number may vary of course.

8 REFERENCES

[1] S. Can, *Topologieentwurf von lokalen Zugangsnetzen*, Internal ComNets report, November 1996.

[2] Roshan L. Sharma, "Network Topology Optimization", The Art and Science of Network Design, Van Nostrand Reinhold, 1990, pp. 104-105.

[3] L. Nestler, *Dimensionierung, Analyse und Optimierung eines Telekommunikationssystems mit funkgesteuertem Zugangsnetz und Satelliten im Weitverkehrsbereich*, Internal ComNets report, August 1996.

[4] R. C. Cooper, *Introduction to Queuing Theory*, New York: The Macmillan Company, 1972, ch. 4.

[5] B. Walke, *Kommunikationsnetze und Verkehrstheorie II*, Lecture Notes, 1997.

[6] C. Plenge, I. Lenzen, *Entwicklung und Bewertung eines Relais-Konzeptes als Erweiterung des DECT-Standards*, in Proceedings ITG-Fachtagung Mobile Kommunikation 135, Neu-Ulm, 26th-28th September, 1995.

[7] TELEDESIC™ Internet WWW homepage at http://www.teledesic.com.

[8] M. A. Sturza, *Architecture of the TELEDESIC satellite system*, in Proceedings International Mobile Satellite Conference - IMSC'95, pp. 212-218, 1995.

[9] A. Beccera, A. Guntsch, et. al., *Service Definition in the Integrated UMTS Environment*, in Proceedings Tenth International Conference On Digital Satellite Communications, Brighton, United Kingdom, IEE, May 1995.

[10] B. Walke, *Radio in the Local Loop - Technology Trends*, Internal ComNets report, 1996.

Session 3
CDMA

Session Chairperson—*Eric Lutz*, German Aerospace Research Establishment, Germany
Session Organizer—*Norman Lay*, Jet Propulsion Laboratory, USA

On Channel Coding for Satellite Multimedia Systems with a Synchronous CDMA Access Scheme

C. Valadon, R. Tafazolli, B.G. Evans,
Mobile Communications Research Group
Centre for Communications Systems Research (CCSR)
Guildford, Surrey GU2 5XH, UK
Tel:+44 1483-259808, Fax:+44 1483 259504
E-mail: C.Valadon@ee.surrey.ac.uk

ABSTRACT

This paper investigates the performance of channel coding schemes targeting the provision of multimedia services by satellite using a Synchronous CDMA (S-CDMA) access scheme. An upper bound on the performance of convolutionaly coded S-CDMA transmission is first derived. It is shown that spreading and channel codes have different influences on the performance of S-CDMA transmissions. Hence, the trade-off between channel coding and spreading is investigated. Finally, the performance of concatenated *Reed-Solomon (RS)*-convolutional codes is presented.

I. INTRODUCTION

Third generation mobile communication systems aim at providing a wide range of multimedia services (telephony, video conferencing, audio and video broadcasting ...) to a large number of users. The provision of these services will require the transmission of very high data rates while achieving extremely low Bit Error Rate (BER) values (between 10^{-6} and 10^{-10} for high quality video broadcasting for example). In order to keep the mobile terminal complexity as low as possible, it is essential to reach the required Quality of Service (QoS) with very low values of the Signal to Noise Ratio (SNR). Hence, the choice of the optimum channel coding technique with the best possible access technique is one important technique in realising such performance limits.

In recent years, there has been a growing interest in the capabilities of Code Division Multiple Access (CDMA) for both terrestrial and Satellite Personal Communication Networks (S-PCN). CDMA presents several interesting features for mobile communications such as flexible frequency reuse, the capability of performing soft-handover and a lower sensitivity to interference. Moreover, a novel and more efficient access technique for satellite networks based on CDMA has recently been proposed [1]. Improved spectral efficiency is achieved by synchronisation of the spreading code epochs at the satellite input, hence giving rise to the term Quasi-Synchronous CDMA (S-CDMA).

In Asynchronous CDMA (A-CDMA), the users start their transmission in a completely uncoordinated manner. Hence, the Multiple Access Interference (MAI) arising from the non-orthogonality of the different spreading waveforms is randomised by the variation in the propagation delays and channel conditions. The spreading codes of the interfering users are seen as pseudo random codes at the receiver side, and the level of MAI is highly independent from the choice of the codes used to spread the signal. Therefore, for A-CDMA transmissions the use of very low rate channel codes provides good coding gains without any reduction in the spectral or power efficiency [2,3].

On the other hand, the efficiency of S-CDMA lies in the full exploitation of the cross-correlation properties of the spreading codes. The use of channel coding will alter the cross-correlation properties of the spreading codes and will not prove as beneficial as in the case of A-CDMA. The trade-off between channel coding and spreading for S-CDMA is investigated in this paper. An upper bound on the BER of convolutionaly coded S-CDMA is derived in section two. In section three, the respective influence of spreading and channel codes on the performance of convolutionaly coded S-CDMA are identified. Then, the performance of concatenated *RS*-convolutional codes is presented in section four. Finally, conclusions are drawn in section 5.

II. CONVOLUTIONAL CODING FOR S-CDMA

In this paper, an Additive White Gaussian Noise (AWGN) channel shared by K users is considered. It is now important to point out that S-CDMA can only be efficiently used in frequency non-selective channel. The multipath characteristics of frequency selective channels would alter the correlation properties of the spreading codes. Hence, S-CDMA is to be used for satellite communications where coherence bandwidths are large enough (of the order of 10 MHz) to keep the delay spread negligible compared to the chip duration, hence justifying the channel model used in this paper.

In this section, a convolutional code with transfer function $T(D,N)$ and free distance d_{free} is considered. If R denotes the channel code rate, the energy per coded symbol of the kth user E_k^s is equal to R times the energy per bit E_k^b. It is important to stress that, in order to preserve the low cross-correlation properties of the spreading sequences, the product between the chip duration T_c and the spreading code length N must be equal to the coded symbol duration T_s. Moreover, the number of spreading sequences with good cross correlation properties is limited by the code length [1]. Hence, assuming a constant bandwidth, the use of a Forward Error Correction (FEC) code of rate R will reduce the number of available sequences by a factor of $1/R$. Nevertheless, studies carried out in the frame of the European project *ACTS-SECOMS* showed that the number of available codes was higher than the expected number of users in each spotbeam [4]. These findings depend on the *SECOMS* system characteristics and may not hold for other multi-spotbeam satellite CDMA system. In such case, the use of trellis coded modulations should be preferred to convolutional codes [5].

The spread data of the kth user can be written as:

$$s_k(t) = \sum_{i=-\infty}^{\infty} d_i^k \cdot c_k\left(t - iRT_s\right), \tag{1}$$

where d_i^k is the coded symbol sent during the ith signalling interval, and $c_k(t)$ is the spreading waveform of the user k.

The received signal can now be expressed as:

$$r(t) = \sum_{i=1}^{K} A_i \cdot s_i\left(t-\tau_i\right) \cdot \cos\left(\omega_o t + \theta_i\right) + n(t). \tag{2}$$

K is the number of active users in the system. A_i is the amplitude of the kth user signal level. θ_i is the initial phase of the ith modulator and is assumed to be uniformly distributed in the interval $[0,2\pi]$. ω_o is the carrier angular frequency. $\{\tau_i\}_{i\in\{1,\ldots,K\}}$ is the set of propagation delays affecting the different users. It is assumed, without any loss of generality, that $\tau_k = 0 \bmod [T_s]$. $n(t)$ represents the AWGN with two-sided power spectral density equal to $N_0/2$.

The output of the kth user correlator during the signalling interval 0 is equal to:

$$y_0^k = d_0^k \frac{E_i^s}{A_i} + \phi_0 + \tilde{n}_0^k. \tag{3}$$

The first term in Equation (3) corresponds to the information bit to be recovered by the receiver and the second term ϕ_0 represents MAI due to the non orthogonality of the different user spreading waveforms. Since the MAI results from the combination of the signals send by all the users (minus one) present in the system, it can, according to the weak law of large numbers, be modelled as a Gaussian random variable with a zero mean and variance equal to:

$$\mathrm{var}\left(\phi_0\right) = \sum_{\substack{i=1 \\ i\neq k}}^{K} \frac{A_i^2}{8} E\left(\rho_{i,k}^2\left(\tau_i\right)\right), \tag{4}$$

where the expectation is taken over all the data and spreading symbols and the relative delays, and:

$$\rho_{i,k}\left(\tau_i\right) = \int_0^{T_s} s_i\left(t-\tau_i\right) \cdot s_k(t). \tag{5}$$

Finally, $\tilde{n}_0^k$ is a zero mean Gaussian random variable with variance equal to $N_o E_c/4$, where E_c denotes the energy of the spreading waveforms and is equal to $\int_0^{T_s} s_k(t)^2\,dt$.

Convolutional codes with low constraint length can efficiently be decoded using the Viterbi algorithm which uses soft information inputs. Since convolutional codes are linear, it can be assumed that the zero sequence has been transmitted for the derivation of the bit error probability. The first-event error probability $P_2(d)$ will first be derived. The first event-error probability is the probability that a path that differs from the all-zero path in d bits has a higher metric than the all-zero path [6]. It can be written as:

$$P_2(d) = \Pr\left(\sum_{l=0}^{d-1} y_l^k > 0\right). \tag{6}$$

Equation (6) has been written assuming, without any loss of generality, that the bits differing from the all zero path have indexes ranging from 0 to d-1.

Equations (3) and (4) are used to derive the first-event error probability:

$$P_2(d) = Q\left(\sqrt{\frac{d\cdot\left(E_k^s/A_k\right)^2}{\frac{N_0 E_c}{4} + \sum_{\substack{i=1 \\ i\neq k}}^{K} \frac{A_i^2}{8} E\left(\rho_{i,k}^2\left(\tau_i\right)\right)}}\right). \tag{7}$$

Equation (7) can be written as:

$$P_2(d) = Q\left(\sqrt{\frac{2\cdot d\cdot R\cdot E_k^b/N_0}{1+\alpha\cdot\gamma\cdot(K-1)\cdot R\cdot E_k^b/N_0}}\right), \tag{8}$$

where $\alpha = \frac{1}{K-1}\sum_{\substack{i=1 \\ i\neq k}}^{K}\frac{E_i^b}{E_k^b}$ is the Average Interfering Energy Ratio (AIER), and $\gamma = E\left(\rho_{i,k}^2(\tau_i)\right)/E_c^2$ depends on the cross-correlation properties of the spreading waveforms.

In general, the transfer function of convolutional codes can be expressed as:

$$T(D,N) = \sum_{d=d_{free}}^{\infty} a_d D^d N^{f(D)}. \quad (9)$$

Hence, an upper bound on the bit error probability can be written as [6]:

$$P_b \le T(D,N) = \sum_{d=d_{free}}^{\infty} \beta_d P_2(d), \quad (10)$$

with $\beta_d = a_d f(d)$.

Using the following approximation of the Q function [7]:

$$Q\left(\sqrt{x+y}\right) \le Q\left(\sqrt{x}\right)e^{-\frac{y}{2}}, \quad (11)$$

an upper bound on the bit error probability of convolutionaly coded S-CDMA transmissions can be found:

$$P_b \le Q\left(\sqrt{\frac{2\cdot d_{free}\cdot R\cdot E_k^b/N_0}{1+\alpha\cdot\gamma\cdot(K-1)\cdot R\cdot\frac{E_k^b}{N_0}}}\right)\times e^{\frac{d_{free}\cdot R\cdot E_k^b/N_0}{1+\alpha\cdot\gamma\cdot(K-1)\cdot R\cdot\frac{E_k^b}{N_0}}} \times \frac{dT(D,N)}{dN}\bigg|_{\substack{N=1 \\ D=e^{\frac{-d\cdot R\cdot E_k^b/N_0}{1+\alpha\cdot\gamma\cdot(K-1)\cdot R\cdot\frac{E_k^b}{N_0}}}}}. \quad (12)$$

As seen in the analysis presented above, the performance of S-CDMA depends on the choice of spreading waveforms. Moreover, in order to achieve good spectral efficiency, satellite communication systems aiming at providing multimedia services will have to use large antennas and cover the service area with a large number spotbeams. Hence, the choice of the best spreading waveforms should be made considering the necessity to have a set of code large enough to enable the communication with all the users in the different spotbeams. One good solution is to use preferentially phased *Gold* codes that are particularly well suited for multi-beam applications [1]. In fact, it is possible to generate a certain number of preferentially phased Gold families with quasi-orthogonal cross-correlation properties inside the family, and pseudo random correlation properties among elements of different families [8]. Users in a given spotbeam will be allocated spreading codes belonging to the same family, and users in other spotbeams will use spreading codes from the other families. The in-beam multiple access interference is then almost eliminated. And the multiple access interference generated by adjacent spotbeams will be attenuated by the beam isolation. Note that throughout the calculations, no parameter for adjacent spotbeam MAI is used. However, this interference can be incorporated into the total thermal noise density as it neither depends on the spreading waveforms nor on the channel coding scheme [3]. The use of rectangular pulse shaping and preferentially phased *Gold* codes is assumed in this paper.

The performance of convolutionaly coded S-CDMA obtained through Equation (12) will now be compared to simulation results. Figure 1 compares the simulated performance of the best half rate convolutional code with constraint length 7 to the upper bound (12). The spreading code length has been assumed equal to 31, and perfect time synchronisation between the different users is considered.

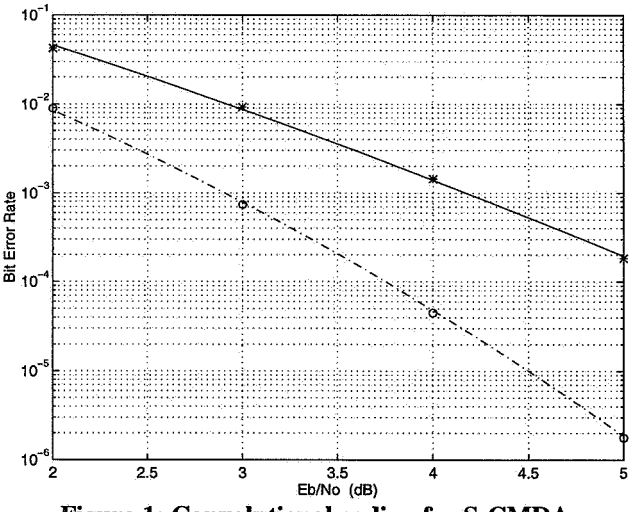


Figure 1: Convolutional coding for S-CMDA Comparison between simulation and analysis

- AIER=10 dB, 29 users: - analysis; * simulation
- AIER= 6 dB, 10 users: -. analysis; o simulation

The performance upper bound provided by Equation (12) matches very closely the results obtained by simulation. Hence, Equation (12) will now be used to investigate the trade-off between channel coding and spreading for S-CDMA transmissions.

III. Trade-off between Channel Coding and Spreading for S-CDMA

The following two figures present two different spreading/convolutional coding schemes that have the same processing gain. Moreover, in order to keep the complexity of the two decoders comparable, both codes have the same constraint length. The performance of the best half rate convolutional code of constraint length 5

used with a *Gold* spreading code of length 63 is compared to that of the best convolutional code of length ¼ and constraint length 5 used with a *Gold* code of length 31. Both the case of 10 users transmitting under perfect power control conditions and the case of 30 users with an AIER of 15 dB are presented. In Figure 2, perfect time synchronisation between the different users has been assumed. A maximum timing error of one third of a chip has been assumed for Figure 3.

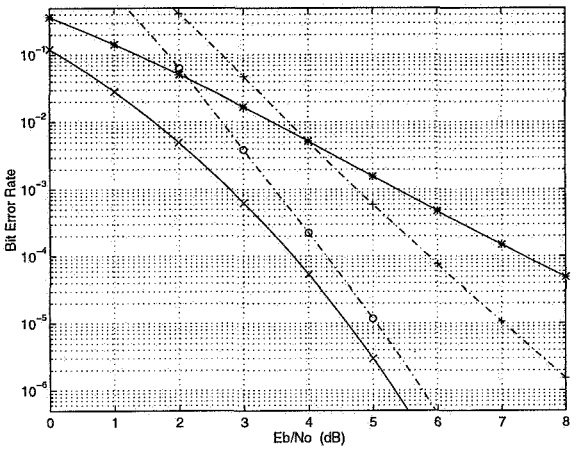


Figure 2: Trade-off between spreading and coding, perfect time synchronisation

- ¼ cc, *Gold* 31: **x** α =0 dB, *K*=10; * α =15 dB, *K*=30
- ½ cc, *Gold* 63: **o** α =0 dB, *K*=10; **+** α =15 dB, *K*=30

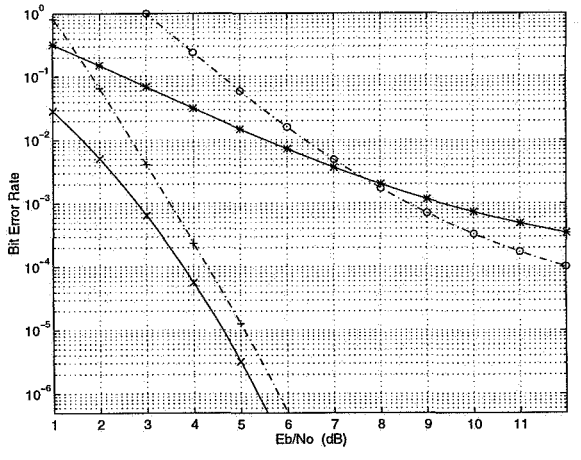


Figure 3: Trade-off between spreading and coding, maximum timing error equal to 1/3 of a chip

- ¼ cc, *Gold* 31: **x** α =0 dB, *K*=10; * α =15 dB, *K*=30
- ½ cc, *Gold* 63: **o** α =0 dB, *K*=10; **+** α =15 dB, *K*=30

The above figures indicate that under perfect power control conditions the S-CDMA transmission using the lowest convolutional code rate has the best performance. Even under imperfect power control conditions, the coding scheme with the lowest FEC code rate presents the best performance when the E_b/N_o value is low. On the other hand, the S-CDMA scheme with the highest spreading code length is the least sensitive to power control errors. When the AIER is equal to 15 dB and the system is highly loaded, the coding scheme with the highest spreading code length outperforms the coding scheme with the lowest FEC code rate when low values of the BER are to be reached. Hence, the trade-off between spreading and FEC coding depends on the convolutional codes used, the spreading code rates, the power control error, the number of active users and the accuracy of the synchronisation scheme.

These observations will now be proven mathematically.

Under the assumption that $\alpha \cdot (K-1) \cdot R \cdot E_b/N_o << 1$, the upper bound (12) can be approximated by:

$$P_b \le Q\left(\sqrt{2 \cdot d_{free} \cdot \frac{E_s}{N_o}}\right) \times e^{d_{free} \cdot \frac{E_s}{N_o}} \left.\frac{dT}{dN}\right|_{\substack{N=1 \\ D=e^{-E_s/N_o}}}, \quad (13)$$

which is the upper bound of a convolutionaly coded single user transmission. Hence, for low values of the AIER and low E_b/N_o, the coding scheme with the best convolutional code will have the best performance.

On the other hand, for high values of the E_b/N_o, Equation (11) can be used to closely approximate the BER by:

$$Q\left(\sqrt{\frac{2 \cdot d_{free} \cdot E_s/N_o}{1+E_s/N_o \cdot \alpha \cdot (K-1)} - \ln\left(\beta_{dfree}^2\right)}\right). \quad (14)$$

Let us consider a code C_1 defined by its rate R_1 and its free distance d_1 used with a spreading code of length N_1, and a second code C_2 defined by R_2 and d_2 used with a *Gold* code of length N_2. The processing gain P_g is constant and equal to $N_1/R_1 = N_2/R_2$. It is also assumed that $R_1 > R_2$. The convolutional code C_1 outperforms the code C_2 when:

$$\Delta\left(E_b/N_o\right) \ge \ln\left(\beta_{dfree}^{C_1}/\beta_{dfree}^{C_2}\right), \quad (15)$$

with

$$\Delta(x) = \left(\frac{d_1 \cdot R_1 \cdot x}{1+R_1.x \cdot \alpha.\gamma_1 \cdot (K-1)} - \frac{d_2 \cdot R_2 \cdot x}{1+R_2.x \cdot \alpha.\gamma_2 \cdot (K-1)}\right) \quad (16)$$

In order to evaluate when Equation (15) is valid, the behaviour of the functions $f(x) = x \cdot \left(\frac{1}{1+x} - \frac{\alpha}{1+\beta \cdot x}\right)$ will be examined. Moreover, in the case studied in this paper, the following relations hold:

$$\alpha = \frac{d_2 \cdot R_2}{d_1 \cdot R_1}, \quad (17)$$

and $\beta = \frac{N_2 \cdot \gamma_2}{N_1 \cdot \gamma_1} = \frac{\lambda_2}{\lambda_1}$. (18)

When $\beta^2 - \alpha \leq 0$, the function f is decreasing with x. Moreover, by considering the function limit as x tends to infinity, it can easily be shown that Equation (15) is never valid. Hence, the code C_2 will outperform the C_1 even for high values of the E_b/N_o ratio.

On the other hand, assuming that $\beta^2 - \alpha > 0$, the function f is increasing with x. Again, in order to know whether Equation (15) holds, the limit of f has to be examined. The sign of the limit depends on the sign of $\beta - \alpha$, and two different cases have to be examined. If $\beta - \alpha > 0$, $\lim_{x \to +\infty} f(x) > \ln\left(\beta_{dfree}^{C_1} / \beta_{dfree}^{C_2}\right)$, unless the AIER takes unrealistic values. Hence, there is an E_b/N_o value above which the code C_1 will outperform the code C_2. But when $\beta - \alpha < 0$, Equation (15) can not be valid, and the transmission with the lowest FEC code rate will achieve the best performance. Finally by noticing that $\beta^2 - \alpha < 0$ implies that $\beta - \alpha < 0$, it is possible to conclude that the parameter driving the choice of the best FEC code/spreading combination is the sign of $\beta - \alpha$.

Figure 4 presents the value of $\beta - \alpha$ for the two codes considered in the two previous figures against the maximum synchronisation error between the different users. The ratio between the BER of the two different FEC code/spreading combinations is also plotted in dB. The BER values have been calculated for an E_b/N_o of 10 dB, assuming the presence of 30 active users and an AIER equal to 10 dB.

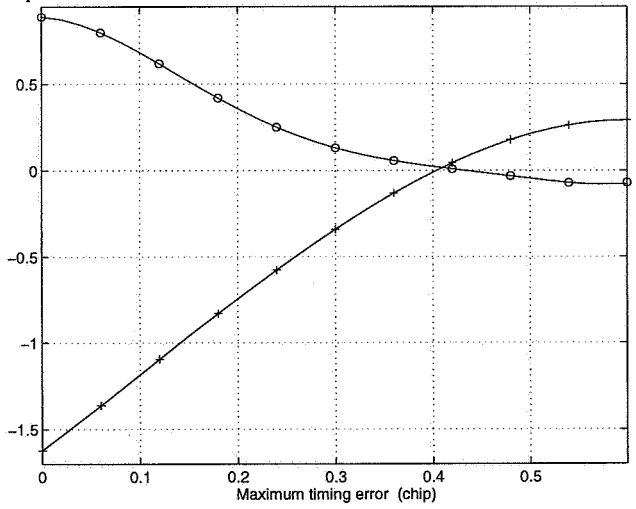


Figure 4: Trade-off between FEC coding and spreading vs. synchronisation error

o $\beta - \alpha$; + $10 \cdot \log\left(P_b^{C_1} / P_b^{C_2}\right)$

When the synchronisation error between the different users is low, $\beta - \alpha$ is positive and the transmission with the longest spreading code has the best performance. On the other hand, when the maximum timing error exceeds 0.4 of a chip, the sign of $\beta - \alpha$ changes and the FEC code with the lowest rate outperforms the transmission with the longest spreading code.

The results presented in this section can be summarised as follow. Assuming an accurate synchronisation between the different users, the use of a low rate FEC code is beneficial for low values of the MAI, and when the MAI level is high it is important to spread the data with long codes in order to increase the SNR at the input of the *Viterbi* decoder. But, if the time synchronisation error can not be kept within 0.4 of a chip, S-CDMA will behave similarly to A-CDMA with respect to channel coding, i.e. the use of the lowest rate FEC code will allow the achievement of the best performance whatever the level of MAI.

IV. CONCATENATED CODING SCHEMES FOR S-CDMA

The power that can be radiated by the satellite is usually the limiting factor in the downlink. It is therefore essential to use very powerful FEC coding techniques in order to achieve the very low BER required by multimedia applications with E_b/N_o values as low as possible. One possibility is to serially concatenate a convolutional code with an outer block code separated by an interleaver in order to randomise the position of the block coded symbols in error at the input of the block decoder.

RS codes are very interesting candidates to perform the outer block coding. First, these codes are maximum distance codes [7]. Hence, an *RS (n,k,b)*, which coded symbols belong to $GF(2^b)$, can correct up to $t = \lfloor (n-k)/2 \rfloor$ erroneous symbols. It should also be stressed that shorten *RS* codes are also maximum distance codes. Moreover, efficient and relatively low complexity hard-decoding algorithms exist to decode long *RS* codes.

An upper bound on the performance of concatenated convolutional-*RS* codes with a S-CDMA scheme will now be derived. The errors at the output of the Viterbi do not arrive in an independent manner. It is therefore important to interleave the data between the output of the Viterbi decoder and the input of the *RS* decoder. Perfect interleaving will first be assumed. With this assumption, the probability of incorrect decoding or decoder failure for hard decision decoding can be upper bounded by:

$$P_T \leq \sum_{i=t+1}^{n} \binom{n}{i} p_s^i (1-p_s)^{n-i} = 1 - \sum_{i=0}^{t} \binom{n}{i} p_s^i (1-p_s)^{n-i}, \quad (19)$$

where p_s is the symbol error probability and can be upper bounded by:

$$p_s \leq b \cdot \overline{p}_b . \quad (20)$$

$\overline{p}_b$ is the bit error probability at the input of the *RS* decoder and can be approximated by Equation (12).

In order to estimate the BER, the following assumptions are made [9]:

- On average, each *RS* symbol in error corresponds to $b/2$ bits in error.
- When the error capability of the code is exceeded, the decoder may add up to t erroneous symbols.

Using the above assumptions, an upper bound on the bit error probability at the output of the *RS* decoder is derived:

$$P_{eb} \leq \sum_{i=t+1}^{n} \frac{i+t}{2 \cdot n} \binom{n}{i} p_s^i \left(1-p_s\right)^{n-i} . \quad (21)$$

Figure 5 plots the performance of the *RS(204,188,8)* code concatenated with the convolutional code of rate ½ and constraint length 7. It is interesting to note that the *RS(204,188,8)* code has been recommended by the Digital Video Broadcasting Commercial Module [10]. The *Gold* spreading code length has been taken equal to 63. The presence of 60 active users has been assumed.

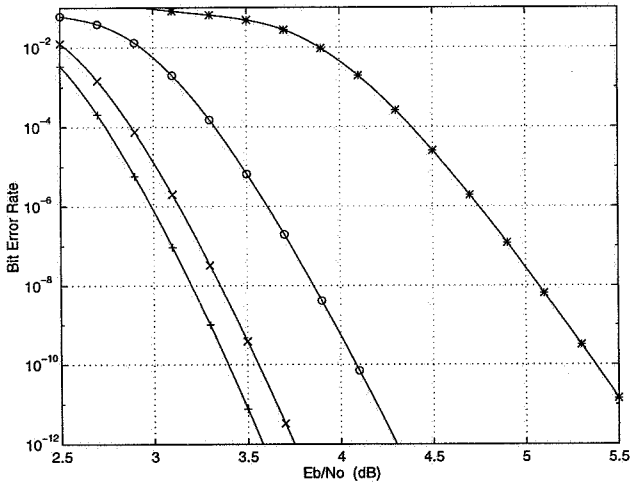


Figure 5: *RS(204,188,8)* + CC(1/2, L=7)
+ single user bound; **x** perfect power control
o AIER=6 dB; * AIER=10 dB

With this concatenated coding scheme, it is possible to achieve a BER of 10^{-10} for relatively low values of the E_b/N_o. Nevertheless, it is important to stress the following two points. First, the slope of the BER curve is very high. Hence a small degradation in the E_b/N_o will result in an important performance degradation. It is therefore important to incorporate adequate shadowing and atmospheric loss margins in order to keep the QoS at an acceptable level even when the propagation attenuation is high. Moreover, even though S-CDMA is less sensitive to power control errors than A-CDMA, it is still important to have a proper power control scheme to avoid important performance degradation under severe near-far conditions.

V. CONCLUSIONS

The performance of convolutionaly coded S-CDMA has been presented. The trade-off between FEC coding and spreading has been investigated. It has been shown that, unlike for A-CDMA transmissions, the choice of the best possible FEC coding/spreading combination for S-CDMA depends on a number of factor such as the MAI level, the FEC code performance and the synchronisation accuracy. Finally, the concatenation of *RS*-convolutional codes has been studied. It has been shown that they allow the achievement of low BER for reasonable E_b/N_o values.

REFERENCES

1. **R. De Gaudenzi, C. Elia,** and **R. Viola,** *"Bandlimited Quasi-Synchronous CDMA: A Novel Satellite Access Technique for Mobile and Personal Communication Systems"*, IEEE Jnl. On Selected Areas in Communications, vol. SAC-10, No. 2, pp 328-343, Feb. 1992.
2. **J.Y.N. Hui,** *"Throughput Analysis for Code Division Multiple-Accessing of the Spread Spectrum Channel"*, IEEE Jnl. On Selected Areas in Communications., vol. SAC-2, No. 4, pp 482-486, July 1984.
3. **A.J. Viterbi,** *"When Not to Spread Spectrum-a Sequel"*, IEEE Communications. Magazine, vol. 23, No. 4, pp 12-17, April 1985.
4. European project ACTS-AC004, *"WP3200: Link Design and Dimensioning: Preliminary Link Design Report"*, document number AC004/TOR-DEL/DRR/001/a1, 29/08/96.
5. **R. De Gaudenzi, F. Giannetti,** *"Analysis and Performance Evaluation of Synchronous Trellis-Coded CDMA for Satellite Applications"*, IEEE Trans. On Communications, vol. COM-43, No. 2/3/4, pp 1400-1408, Feb/March/April 1995.
6. **J.G. Proakis,** *Digital Communications-second edition*, McGraw-Hill International Editions, 1983, ch.5, pp 460.
7. **S.B. Wicker,** *Error Control Systems for Digital Communications and Storage*, Prentice Hall International, 1995.
8. **W.W. Peterson** and **E.W. Weldon,** *Error Correcting Codes*, Cambridge, MA: MIT Press, 1961.
9. **S. Bellini,** *"On the Interleaver Depth in Concatenated Coding Schemes"*, presented at the EMPS'96 conference, Oct. 1996.
10. **G.M. Drury,** *" DVB Channel Coding Standards for Broadcasting Compressed Video Services"*, IEE Electronics & Communication Journal, vol. 9, No. 1, pp 11-20, Feb. 1997.

Multimedia CDMA Studies for Ka-band and Millimeter-wave Satellite Communications

Tetsushi Ikegami, Eihisa Morikawa, Takashi Takahashi,
Huan-Bang Li, Eiji Okamoto, and Shin'ichi Yamamoto

Communications Research Laboratory,
Ministry of Posts and Telecommunications
893-1 Hirai, Kashima, Ibaraki 314 JAPAN
Telephone: +81-299-84-7124 **Fax**: +81-299-84-7158
e-mail: ikegami@crl.go.jp

ABSTRACT

A variable data rate CDMA system is studied for mobile and personal multimedia random access satellite communications in Ka-band and millimeter-wave region. This system utilizes a constant occupied bandwidth and has variable data rates and processing gains to mitigate communication link impairments such as fading, rain attenuation and interference as well as to handle variable data rate on demand.

A proof of concept hardware system for a 6 MHz bandwidth transponder is developed, which uses MSK-type constant envelope direct sequence spread spectrum modulation and handles data rates of 4k to 64 kbps. The receiver is designed with coherent matched filter technique to achieve fast code acquisition, AFC and coherent detection in a single matched filter circuit. This receiver structure facilitates variable data rate on demand during a call without the aid of system control earth stations.

The proposed system is described, along with the CDMA performance of the prototype equipment with the non-geostationary ETS-VI satellite at the S-band (2.3/2.1 GHz) with Ka-band (30/20 GHz) feeder link. Both the IF loop back and satellite link test performances of the prototype equipment prove it has the potential for personal multimedia satellite applications. Continued experimental program of next test satellite COMETS is also briefly described.

1. Introduction

Spread spectrum (SS) code division multiple access (CDMA) systems with random access capability and without coordination stations are attractive for personal and mobile satellite communication systems. Distributed systems which handle one-hop communication links between small earth stations are also attractive for personal and multimedia communications in private or non-profit networks [1]. A variable data rate CDMA system is proposed for mobile and personal satellite communications in a random access mode, and a prototype hardware system is developed[2].

This paper deals with experimental results of the prototype hardware using the non-geostationary ETS-VI satellite [3] at the S-band (2.3/2.1 GHz) with Ka-band (30/20 GHz) feeder link. The use of FEC reduces the effect of multiple access interference, and enhances the bit error performance under CDMA conditions. Experimental CDMA link is established under real Doppler shift environment of non-geostationary satellite. Finally, experimental program of next test satellite, called COMETS, is briefly described.

2. Variable Data Rate CDMA System

The SS system has the advantage of random access capability without the aid of coordination stations. This advantage is attractive for distributed personal and mobile multimedia satellite communication systems with user-to-user one-hop links. Multimedia communications in which various amounts of data should be transmitted for video, text, and audio transmission, need variable data rate transmission systems. The SS technique can be used for variable data rate systems because various channels which handle different data rates can co-exist in the same SS bandwidth, and receiver implementation for variable data rate systems is relatively easy.

Based on the study mentioned above, a variable data rate CDMA system is proposed for one-hop multimedia mobile and personal random access satellite communications. This system utilizes a constant spreading bandwidth and a fixed PN code length. To obtain necessary link margins at appropriate data rates, fading and rain attenuation can be compensated for by changing the data rate and transmitting power. The appropriate data rate and quality can be achieved on demand whenever a communication link is established.

Automatic Frequency Control (AFC) at the receiving earth station is mandatory for a one-hop system

without the aid of network coordination stations or pilot signals. The necessary frequency range for AFC is around 10 kHz in geostationary satellite systems at the L-band (1.6/1.5 GHz) and tends to 100 kHz in low earth orbit satellite systems at S-band (2 GHz) frequencies. Synchronization of code and carrier should be accurate under fading and low C/N conditions which are typical in mobile satellite communication channels.

A coherent matched filter (CMF) receiver [4][5] is suitable for this system because it facilitates receiver AFC and coherent detection without the aid of a pilot SS signal. The CMF technique makes it possible to recover a carrier and to demodulate data coherently from low S/N SS signals in fading channels. In a CMF receiver, two correlators for spreading code are inserted into both the I and Q arm filters of the Costas loop demodulator.

3. Hardware

A prototype system is developed based on the proposals mentioned above. Table 1 summarizes the general specifications of the prototype system. MSK-type constant envelope and ordinary offset-QPSK direct sequence spread spectrum systems are examined for the non-linear effect of HPAs of earth stations and transponders. Spreading sequences are chosen from augmented Gold code or arbitrary code of length 1024. The spreading code clock is fixed at 4.3 Mbps. The user data rate is variable on demand and ranges from 4 kbps to 64 kbps with FEC (8-128 kbps without FEC). Rate 1/2 convolutional FEC code and a Viterbi decoder are utilized for error protection of the data. The user data rate is changed during a call in accordance with the performance monitor output. It can also be changed on request from both the transmitter and receiver sides whenever a link is established.

An RS-422 data port is available for multimedia communication via workstations or PCs. Also, an H.261 video codec is installed for TV conference along with low rate voice codecs. A ranging function is installed to measure the range between two earth stations via a satellite.

The signal format is designed as a fixed frame structure of 120 ms length as in Fig. 1. It consists of a unique word (UW) for synchronization of data, data rate indicator bits, variable data rate status bits, and the information bit stream. Except for the information bit stream, the operating bit rate is 8.5 kbps. Within an information bit stream of 112.94 ms, 960 to 15360 bits of data (including rate 1/2 FEC and overhead) run at 8.5 kbps to 136 kbps respectively.

The receiver is designed using the coherent matched filter technique with multibit quantized digital correlators. It performs code synchronization, AFC, and coherent detection of data in a single circuit. This receiver structure facilitates the provision of variable data rates on demand during a call. The range for AFC is +/-60 kHz. The correlation process for MSK-type spreading is similar to OQPSK. It uses a quadrature detection scheme with correlators inserted in both the I and Q arms as in Fig. 2. The tap coefficient of correlators for the reference spreading code is one bit, that is, the reference PN code pulse shape is rectangular, whereas the exact matched pulse for MSK-SS is a raised-cosine shape. This greatly reduces the hardware size with minimal mismatch loss.

Bit error rate performance is measured in the IF loop back mode. Figures 3 and 4 show the bit error rates versus C/N_0 in single-user and equal-power-eight-user environments. It should be noted that the hardware loss is within 1 dB, and the use of FEC enhances the performances especially in eight-user CDMA environment. Initial acquisition time including AFC over +/-60 kHz is found to be about two seconds at average when C/N_0 is greater than 43 dBHz.

4. ETS-VI Satellite Tests

The Engineering Test Satellite-six (ETS-VI) was launched in August 1994. It failed to reach a geostationary orbit because of apogee-engine failure. It's orbit was elliptic (8500 km at the perigee and 38700 km at the apogee) and the satellite was available for only 3 to 4 hours every 3 days by July 1996 because of the satellite attitude limitation (Figs. 5, 6). The main mission of ETS-VI at a geostationary orbit was inter satellite communication experiments in the S-band (2.3/2.1 GHz) with Ka-band (30/20 GHz) feeder links and mm-wave (43/38 GHz) communication experiments [3].

Thanks to the non-geostationary nature of the orbit, land mobile satellite experiments, for example, propagation measurements at high elevation angles and Doppler compensated communications using ETS-VI at the S-band were performed intensively[3]. CDMA experiments using S-band and Ka-band earth stations were performed as one of the above experiments. Experimental configuration is shown in Fig. 7. In these experiments, a Ka-band feeder link earth station plays the role of a user station in the one-hop communication link of the proposed system. The Doppler shift due to satellite motion is a significant factor in this system. As the feeder link between a feeder link station and the satellite uses higher frequency

of Ka-band (30/20 GHz), the maximum Doppler shift is -150 kHz and the Doppler rate at the receiving earth station is more than 10 Hz/sec during the experiments. As the Doppler shift exceeds the designed AFC range, a frequency offset of -80 kHz at receiver local oscillator is applied to meet the designed AFC range. Eight-user CDMA environment is configured by four SS terminals and a user simulator which generates SS signals of up to four users.

Figure 8 to 10 show the bit error rate performances of single-user and equal-power-eight-user CDMA environments using the ETS-VI return link (2.3 GHz up/20 GHz down). Even in a real Doppler shift environment, the performance is similar to the static performance in IF loop-back tests, and the use of FEC greatly reduces the error floor caused by multiple access interference. The synchronization is robust: as the synchronization is controlled at the lowest bit rate of 8.5 kbps, the link is kept in lower C/N_0 condition without the aid of pilot signals even when the bit error rate of the higher data rate users tend to 10^{-1}. This advantage can be applied for fading and rain attenuation compensation. The lowest C/N_0 for initial acquisition is less than 45 dBHz in eight-user CDMA environments.

Dual link experiment is performed using forward (30/2.1 GHz) and return (2.3/20 GHz) links of the ETS-VI. Designed functions are tested satisfactorily: the dual link for H.261 video or voice communications is established on demand, and user data rates of both links are controlled from both forward and return link terminals. Due to time limitation of ETS-VI experiments, detailed evaluation of adaptive data rate control function is left for the next experimental items.

5. COMETS satellite program

COMETS is an engineering test satellite for advanced mobile satellite communications, advanced satellite broadcasting and inter-orbit communications in much higher frequency bands. The satellite will be launched into geostationary orbit at 121 degrees east longitude in August 1997[6][7]. Experiments on mobile communications will be carried out mainly by the Communications Research Laboratory (CRL). The objectives of the COMETS advanced mobile satellite communications experiments are the development of higher frequency bands for future mobile and personal satellite communications at Ka-band (30/21 GHz) and mm-wave Q-band (47/43 GHz) in single hop configuration. Specifications of COMETS advanced mobile satellite communications mission are summarized in Table 2 and conceptual sketch of the satellite is shown in Fig. 11. The system for the experiments consists of a space segment, some mobile/personal stations and two hub stations for reference and/or pseudo personal stations.

The experiments on variable data rate CDMA will be continued using COMETS satellite at Ka-band and Q-band. Modification and evaluation of the hardware are scheduled; application of GMSK for spreading, variable rate FEC, interference cancellation, elastic buffer for multimedia source, etc.

6. Conclusions

A variable data rate CDMA system is tested for personal multimedia satellite communications. Experimental results of the prototype hardware are shown. The use of FEC reduces the effect of multiple access interference, and enhances the bit error performance under CDMA conditions. Experimental CDMA links at the S-band (2.3/2.1 GHz) with Ka-band (30/20 GHz) feeder link are established under real Doppler shift environment of non-geostationary satellite. Both the IF loop back and satellite link test performances of the prototype equipment prove it has the potential for personal and mobile satellite applications. This study will be continued using COMETS satellite at Ka-band and millimeter-wave, which will start in late 1997.

The author would like to thank the members of NASDA and CRL for their help in conducting ETS-VI experiments.

References

[1] H. Wakana, T. Ikegami, S. Yamamoto, and N. Hamamoto, "An Overview of ETS-V/PARTNERS Project", Int'l Conf. on Telecommunications, ICT'95, pp. 399-402, Bali, Apr. 1995.

[2] T. Ikegami, T. Takahashi, Y. Arakaki, and H. Wakana, "Adaptive Data Rate SSMA System for Personal and Mobile Satellite Communications", Int'l Mobile Satellite Conf., IMSC'95, pp. 20-25, Ottawa, June 1995.

[3] T. Takahashi, H. Li, S. Yamamoto, Y. Arakaki, E. Okamoto, T. Ikegami, and H. Wakana, "S-band / O-band Satellite Communications Experiments using ETS-VI", 20th Int'l Sympo. on Space Technology and Science, ISTS, 96-h-10, Gifu, May 1996.

[4] N. Hamamoto, R. Suzuki, I. Nishiyama, R. Miura, and T. Nishigaki, "PN-SS Equipment for Satellite Communications with Data Demodulation by Matched Filter", IEICE Trans. Vol. J69-B, 11, pp. 1540-1547, Nov. 1986.

[5] T. Ikegami, R. Suzuki, N. Hamamoto, and N. Sato, "Experiments on a Coherent Matched Filter Receiver for Spread Spectrum Mobile Satellite Communications", IEICE Trans., Vol. E74, No. 5, pp. 1130-1136, May 1991.
[6] S. Ohmori, S. Isobe, M. Takeuchi, and H. Naito, "Advanced Mobile Satellite Communications Using COMETS Satellite in MM-wave and Ka-band", IMSC'93, pp. 549-553, Pasadena, June 1993.
[7] M. Takeuchi, Y. Hase, C. Ohuchi, H. Mineno, M. Nishida, H. Saito, N. Obara and T. Nishigaki, "Mobile / Personal Terminals for COMETS Advanced Mobile Satellite Communication Experiment", Fourth ICUPC, pp. 568-570, Tokyo, Nov. 1995.

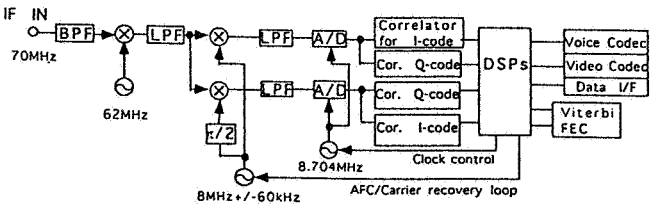


Fig. 2 Block diagram of receiver

Table 1. Hardware specifications

IF Frequency	70 MHz
Modulation	MSK-type constant envelope DS
Spreading code	Augmented Gold code or arbitrary, 1024 chips
Code clock	4.352 Mbps
Data Rate	Variable, adaptive or on demand 4, 8, 16, 32, 64kbps
FEC	Rate 1/2 Viterbi
Demodulation	Coherent matched filter, coherent detection
Matched Filter	Digital correlater, 1024 stages
Data Terminal	RS-422 I/F, H.261 Video codec (64 kbps), Voice codecs (4.8k, 9.6kbps)
AFC range	+/-60 kHz

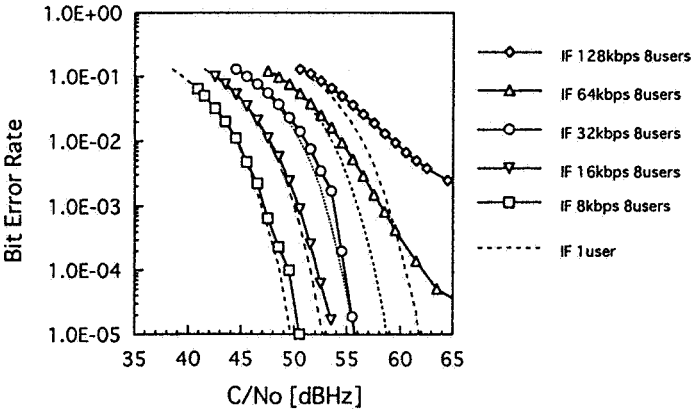


Fig. 3 Bit error rate performance
IF loop back, eight-user CDMA, FEC off.

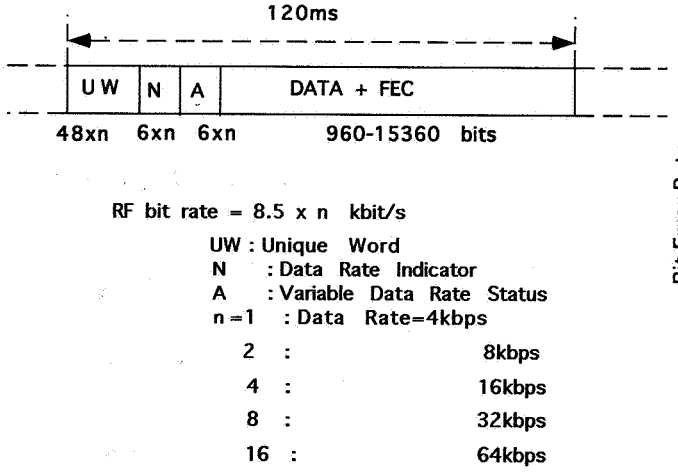


Fig. 1 Signal frame format

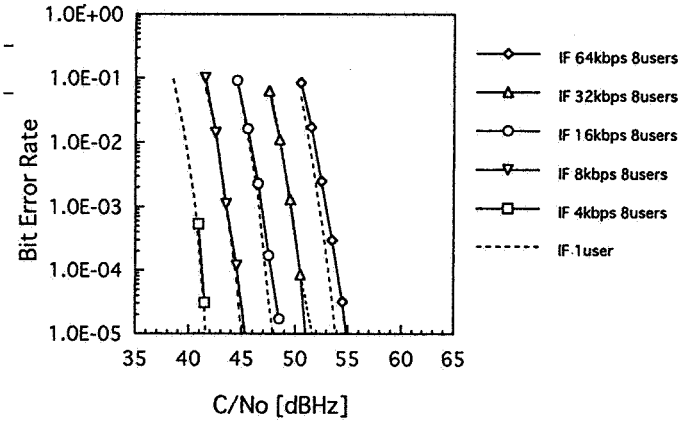


Fig. 4 Bit error rate performance
IF loop back, eight-user CDMA, FEC on.

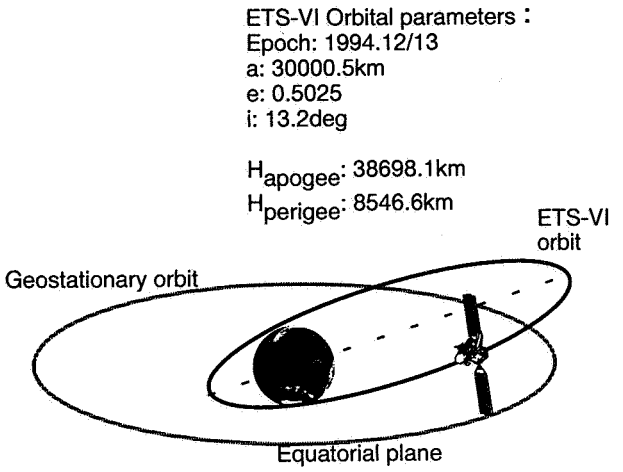


Fig. 5 ETS-VI Orbit

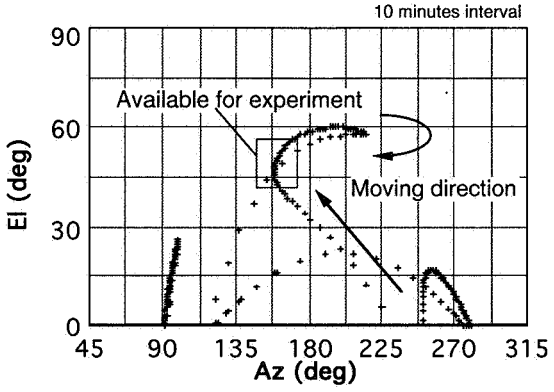


Fig. 6 Loci of ETS-VI at every 10 minutes, viewed from Kashima, Ibaraki, Japan.

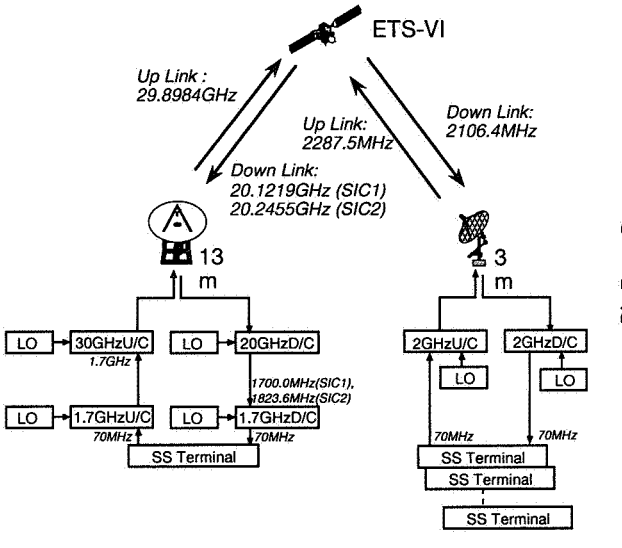


Fig. 7 Experimental configuration

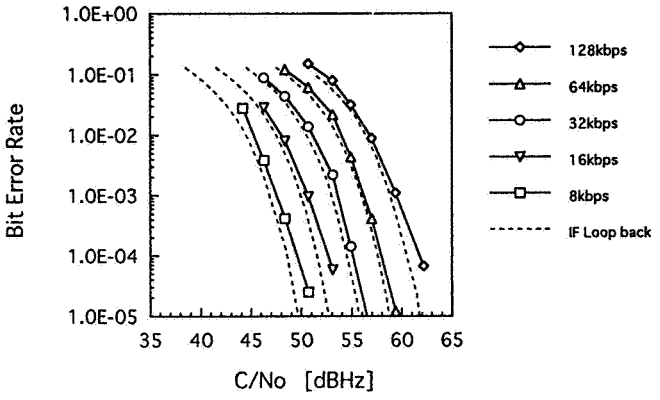


Fig. 8 Bit error rate performance ETS-VI return link (2.3 GHz / 20 GHz), single-user, FEC off.

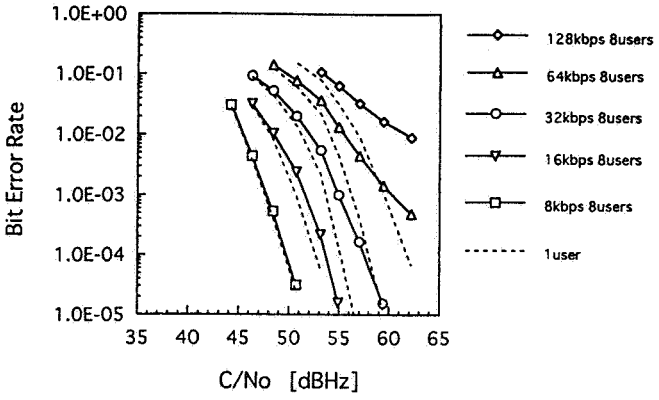


Fig. 9 Bit error rate performance ETS-VI return link (2.3 GHz / 20 GHz), eight-user CDMA, FEC off.

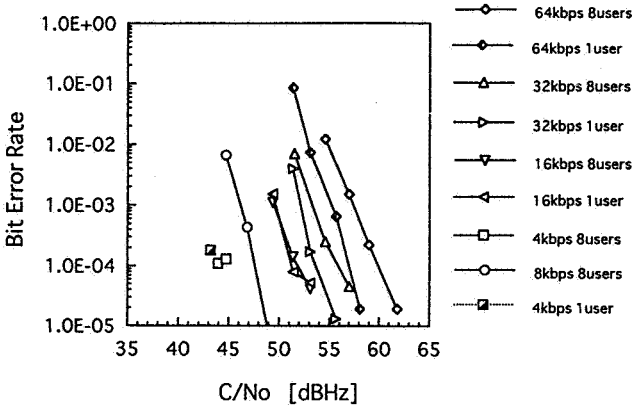


Fig. 10 Bit error rate performance ETS-VI return link (2.3 GHz / 20 GHz), single- and eight-user CDMA, FEC on.

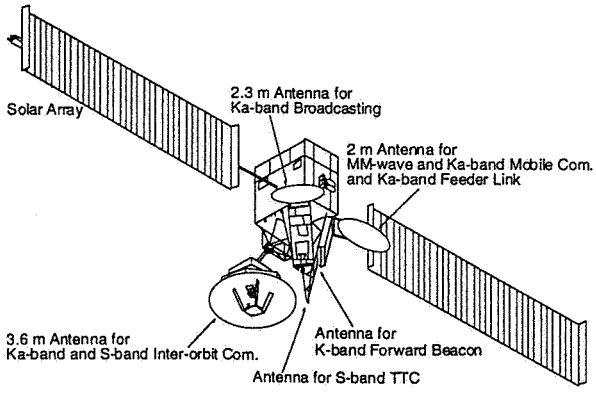


Fig. 11 Conceptual Sketch of COMETS

Table 2. Outline of COMETS advanced mobile satellite communication payload

Antenna	Spot beam antenna shared with Ka- and Q-bands Diameter: 2 m, Circular polarization
Beam	Two Ka-band beams (Tokyo and Nagoya beams) One Q-band beam (Tokyo beam)
Frequency	Ka-band :30.75-30.85 GHz (uplink) :20.98-21.07 GHz (downlink) Q-band :46.87-46.90 GHz (uplink) :43.75-43.78 GHz (downlink)
Transponder	Ka-band :2 (20W and 10W SSPA) Q-band :1 (20W TWTA)
Operation mode	IF repeater : 2x2 matrix interconnection by IF filter bank Wide band (6MHz) and Narrow band (500kHz) Regenerative transponder : BPSK 8 ch. SCPC (uplink) / TDM (downlink) Beam interconnnection by baseband switch

Acquisition Performance of a Digital DS CDMA Receiver in a LEO Satellite Channel

George Efthymoglou and Henry Helmken
Electrical Engineering Department
Florida Atlantic University
Boca Raton, Florida 33431
Phone: 561-367-3452
e-mail: efthymo@volts.eng.fau.edu and helmkenh@acc.fau.edu

I. Abstract

A hardware real-time LEO acquisition system was simulated and tested in the laboratory. A 127 length DS-CDMA sequence is modulated via BPSK onto a Doppler shifted 70 MHz IF carrier at 1 MHz chip rate. A predetermined amount of AWGN is added to model both thermal noise and multiple access interference in the channel. A technique known as Direct Coherent Detection (DCD) is used to recover the I and Q samples with the use of a single A/D. In order to compensate for the Doppler offset a set of candidate carriers is used to cover the carrier uncertainty region. In addition, the linear correlation interval is partitioned into subintervals and the integration results in these subintervals are combined for detection. This parallel acquisition scheme is found to be robust to large Doppler offsets. Its average acquisition time for two different code phase selection criterions is measured as a function of normalized carrier frequency offset and input IF signal-to-noise ratio.

II. Introduction

Direct sequence-code division multiple access (DS-CDMA) systems are currently being designed for Low Earth Orbiting (LEO) Networks that provide global communication services. Such systems suffer from large Doppler offset of a carrier ranging from 30 to 60 kHz, depending on the carrier frequency and satellite altitude. In this paper we evaluate the acquisition performance of a digital DS CDMA receiver in a channel impaired by Doppler shift and additive white Gaussian noise (AWGN).

Code acquisition denotes the coarse alignment of the locally generated PN code with the received signal and is the first operation of a DS correlation receiver. The acquisition performance of a DS system without Doppler offset is studied in [1], [2], among others, while the effects of Doppler and data modulation are studied in [3]-[5]. Doppler offset can introduce inphase (I) and quadrature (Q) phase rotations within the linear "coherent" correlation interval that causes decorrelation between the locally generated PN sequence and the PN sequence in the incoming data [4]. Since a satellite-mobile communication link may start at any time while the satellite is in view, the acquisition performance has to be as little dependent on the initial Doppler offset as possible.

In our acquisition process a set of candidate carriers are employed to reduce the maximum amount of frequency uncertainty, prior to correlation. Carrier spacing is an important design parameter and it depends on the frequency uncertainty range and the correlation interval. To further combat the decorrelation effects of Doppler spreading, the original linear correlation interval is partitioned into subintervals and the integration results in these subintervals are combined to give the test statistic for the code phase that is tested [4]. For this system the acquisition times of the maximum (MAX) and the threshold crossing (TC) phase selection criterions [1], [5] are evaluated experimentally and compared.

The paper is organized as follows. In section II

the experimental setup is described and in section III a technique known as direct coherent sampling (DCD) is described, that allows recovery of the I and Q components of the received signal by undersampling the modulated IF signal. The PN acquisition schemes under investigation are described in section IV. Numerical results of the acquisition performance of the experimental system are presented in section V while some concluding remarks are given in section VI.

III. System description

The received signal at the input to the matched filter is given by

$$r(t) = \sqrt{2P}c(t-\tau)\,d(t-\tau)\,\cos(\omega_0 t+\theta) + n(t) \quad (1)$$

where $c(t)$ is the code sequence and may be expressed as

$$c(t) = \sum_{j=-\infty}^{\infty} a_j\, p_a(t - jT_c), \qquad a_j \in \{-1, 1\}, \quad (2)$$

and $d(t)$ is the data waveform which can be expressed as

$$d(t) = \sum_{j=-\infty}^{\infty} d_j\, p_b(t - jT), \qquad d_j \in \{-1, 1\}. \quad (3)$$

In (1) P is the received power, $n(t)$ is a white Gaussian process representing interference and noise at the synchronizer's input, ω_0 and θ are the carrier frequency and phase, and τ is the propagation delay. In (2) and (3) above, T_c is the chip duration, T is the data bit duration and $p_a(t)$ and $p_b(t)$ are rectangular pulses of unit height and duration T_c and T, respectively. For acquisition of synchronism, the data waveform is normally chosen to have a constant positive value, i.e. +1.

A. Laboratory Setup

The laboratory experimental system is shown in Fig. 1. The I-Q baseband signals representing the desired transmission scenario are developed in software and then loaded through an IEEE-488 bus to an arbitrary waveform generator (AWG) that runs in a continuous mode. In this way, different digital modulation formats, i.e. QPSK, MSK, with desired filtering, that can be generated at baseband, can run in real time. For the particular experiment

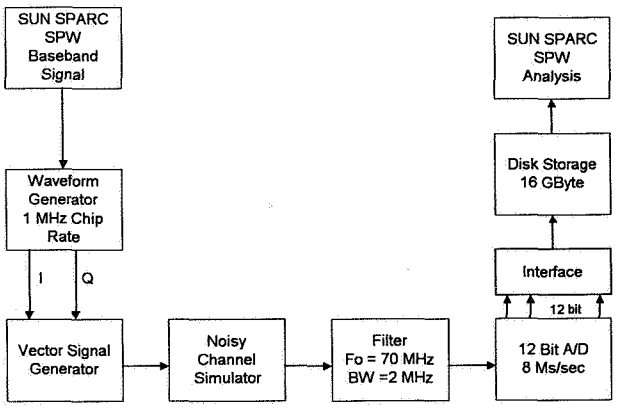


Figure 1 – Laboratoty setup

we create a binary PN sequence with 127 chip period and set the chip rate to 1Mchip/s. The carrier frequency is set to the desired IF frequency of 70 MHz. Doppler offset is introduced by setting the carrier frequency to 70MHz + Doppler offset.

This signal is then passed through a wideband Gaussian noise generator used to simulate the noisy communications channel. A 2 MHz bandpass filter centered at 70 MHz is used to simulate the front end of the receiver. The modulated signal at a given input IF signal-to-noise ratio (SNR) is undersampled by a 12 bit monolithic A/D converter running at 8 MHZ. The samples are stored in a high speed 16GByte RAID disk array. Stored samples are then transferred to the SUN SPARC workstation for further processing.

IV. Sampling strategy

Conventional digital signal processing in radar and communication systems use inphase and quadrature components obtained by two IF-to-baseband converters operated with quadrature LO signals. Each of these components occupies only one-half of the bandwidth of the original signal and can be sampled at one-half the sampling rate required for the original signal. Therefore, quadrature sampling reduces the required sampling rate by a factor of two at the expense of using two phase-locked A/D converters instead of one. Practical problems with DC offset drift, gain changes and quadrature errors limit dynamic range and phase accuracy.

The received bandpass signal $r(t)$ can be expressed in the quadrature form

$$r(t) = I(t)\cos(2\pi f_0 t) \; - \; Q(t)\sin(2\pi f_0 t) \quad (4)$$

where I(t) and Q(t) are the inphase and quadrature components, respectively. The minimum requirements on the sampling rate to allow exact reconstruction are that the sampling rate, f_s, be at least twice the bandwidth W of the information signal. To ensure that spectrum overlap does not occur, when sampling rates are between two times the bandwidth of the bandpass signal and two times the highest frequency in the bandpass signal, the sampling frequency f_s must satisfy [7]

$$f_s = \frac{4f_0}{2D-1} \tag{5}$$

where f_0 is the carrier frequency and D is restricted to integer values. This relation between f_0 and f_s locates f_0 in the center of the band from $(D-1)W$ to DW.

Using direct sampling at the IF we get digitized sampled values $r(t_n)$, where n is an integer index. DCD processing refers to the operation used to obtain the corresponding $I(t_n)$ and $Q(t_n)$ values from the samples $r(t_n)$, where $t_n = \frac{n}{f_s}$. Substituting (5) in (4) we obtain:

$$\begin{aligned} r(t_n) &= (-1)^{n/2} I(t_n) \qquad \text{n even} \\ &= (-1)^{(n-1)/2} Q(t_n) \quad \text{n odd} \end{aligned} \tag{6}$$

Hence for these I_n and Q_n values (I_0,Q_1,I_2,Q_3...) the only computing necessary consists of switching the sampled value r_n to the correct output port with a possible sign change.

In addition, we need accurate phase values for Q_0, I_1, Q_2, I_3 .... This computation requires a digital filtering operation. The method described here and in [7] uses a modification of the sampling theory interpolation function, sometimes called a frequency-window function. By using finite impulse response (FIR) filter weights (determined by the interpolation function) we estimate the value of $r(t_n + \frac{1}{4}f_0)$ from the sample $r(t_n)$. If $f_0 \gg W$ (large D) this time shift of $\frac{1}{4}f_0$ is equivalent to a 90^o phase shift. From (4) and (5) the estimated value $\hat{r}\,(t_n + \frac{1}{4}f_0)$ will then equal

$$\begin{aligned} \hat{r}\,(t_n + \frac{1}{4}f_0) &\approx (-1)^{n/2} Q(t_n) \qquad \text{n even} \\ &\approx (-1)^{(n-1)/2} I(t_n) \quad \text{n odd} \end{aligned} \tag{7}$$

The FIR filter shown in Fig. 2 computes the estimate of $\hat{r}\,(t_n + \frac{1}{4}f_0)$ based on samples of $r(t)$. The

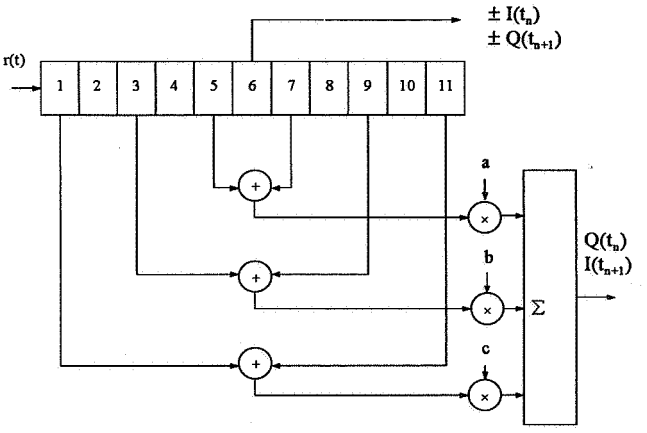


Figure 2 – FIR filter for I & Q interpolation

interpolation function that we used is odd symmetric, with zeros at even valued displacements from its center. Hence only three multiplies are needed in an 11-sample truncation.

V. PN CODE ACQUISITION

The acquisition procedure has two modes of operation, the search mode and the verification mode. The matched filter (MF) synchronization scheme [6] has been adopted in order to achieve fast acquisition of the phase of the code chip sequence. A passive MF provides a fast way of searching over the code phase uncertainty region, i.e. its dwell time is T_c/Δ, where Δ denotes the number of samples per chip.

The code phase uncertainty region is discretized to a collection of $L\Delta$ cells, each of duration ΔT_c, centered upon τ_i, where $i = 1, 2, .., L\Delta$. In doing so any criterion for the selection of the code phase is going to be suboptimal [1]. An accurate estimate $\hat{\tau}$ can be obtained by comparing the test variables from all the cells in the uncertainty region and selecting the maximum. This is known as the MAX criterion. When it is performed in a serial search manner for low complexity applications it may take time to test all cells. In this case a serial search with a threshold crossing criterion can be used, for which each correlation output is compared to a threshold in order to accept or reject the cell under test. A combination of the two schemes can also be used, i.e. the MAX/TC criterion [1]. In this case the code uncertainty region is divided into K sectors with J cells each. Inside a sector a

cell is selected according to the MAX criterion and this cell satisfies the hypothesis for acquired synchronism only if it exceeds a threshold, and if not the search continues with the next sector.

For the TC criterion the event of correct detection corresponds to the fact that the test variable of the correct cell is above the threshold and all other test variables are below the threshold, with probability $P_{d,TC}$, while missed detection occurs when all test variables are below the threshold, with probability $P_{m,TC}$. For the MAX criterion, the event of correct detection occurs when the test variable of the correct cell is above the threshold and is the maximum, with probability $P_{d,MAX}$ and missed detection occurs when all test variables are below the threshold, with probability $P_{m,MAX}$. The probability of false alarm P_f can be obtained as $P_f = 1 - P_d - P_m$ in both cases. The mean acquisition time of different acquisition systems is given as a function of the above probabilities, which can be obtained analytically or experimentally.

Ideally the correlation test statistics can be obtained by performing linear correlation across a fixed dwell time, usually taken to be the code period LT_c. However, the residual Doppler offset causes decorrelation between the incoming data and the local PN code through phase rotations. In order to reduce the maximum amount of frequency uncertainty, we implement the downconversion by candidate carriers prior to correlation. Furthermore, the correlation process is done in two parts: linear "coherent" correlations are performed on subsets of the PN code using I-Q baseband noncoherent matched filters (MFs) and a second "noncoherent" summation of the results of the linear correlations is performed to give the test statistic for the code phase that is tested [4], [5]. It is well known that the peak correlation value reduction due to Doppler offset, L_D, is given by

$$L_D = \frac{\sin(\pi \Delta f T_M)}{\pi \Delta f T_M} \tag{8}$$

where Δf accounts for the residual Doppler offset in the received signal, and T_M is the length of the MF. To combat the effects of decorrelation, the original correlation time, which is equal to the code period LT_c, is partitioned into N subintervals, and the correlation results in these subintervals are noncoherently combined for detection. These correlation results are now affected less by the Doppler offset since the linear correlation interval is $T_M = T/N$. Using the experimental setup we can examine tradeoffs between the number of subintervals N, the number of carriers needed to cover the uncertain carrier-frequency range and the input IF SNR.

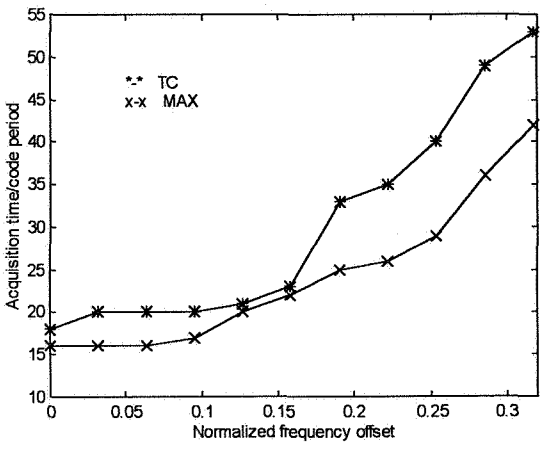


Figure 3 – Average acquisition time versus normalized frequency offset $\Delta f T_M$ with SNR=-15 dB.

VI. Numerical results

The system described above was tested in order to determine the effects of IF input SNR and Doppler offset on the acquisition performance of a direct sequence spread spectrum signal. The processing gain of the system is 21 dB. Furthermore, the A/D converter sampling rate is eight times the PN chip rate and by performing an integration every four samples the acquisition system operates with two samples per chip. The search mode is followed by the verification mode, where 2 out of 4 consecutive correlation peaks have to exceed the threshold, in order to accept synchronism. This stage reduces the false alarm probability considerably.

The input IF SNR is initially set to -15 dB. We select $N = 4$ that results in $\Delta f T_M \simeq .32$ and a peak correlation loss of 1.5 dB for carrier spacing of 20 KHz. Fig. 3 shows the measured average acquisition times as a function of normalized carrier frequency offset $\Delta f T_M$ for the MAX and the TC criterions. The system assumes an initial carrier uncertainty region of ± 30 KHz and uses the three frequencies 0, ± 20 KHz for downconversion prior

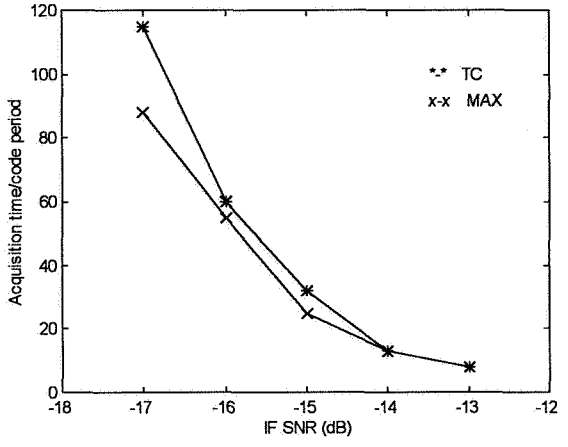


Figure 4 – Average acquisition time versus carrier SNR with $\Delta f T_M = 0.22$.

to coherent correlation. For the MAX criterion, the maximum of $3L\Delta$ MF outputs are obtained and the maximum test variable is selected. If the value of the test variable exceeds a fixed threshold value the verification mode is enabled, otherwise $3L\Delta$ more correlation samples are obtained and the process is repeated. For the TC selection criterion the MF outputs of the three parallel branches are compared every dwell time ΔT_c, and the maximum is compared to the threshold. We observe that the performances of both schemes are similar for $\Delta f T_M \leq .16$, but for larger offsets the MAX criterion performs better.

Fig. 4 shows the measured average acquisition times as a function of the input IF SNR with $\Delta f T_M = 0.22$. The performances of both schemes are similar up to -16 dB. The MAX criterion exhibits less acquisition time as the signal to noise ratio is further decreased.

VII. Conclusions

In conclusion, the acquisition performance of an IF digital receiver that is designed to account for the Doppler shift in the received signal from a LEO/MEO satellite was determined. The selection of design parameters for the DS-CDMA receiver is based on real-time signals and signal acquisition and subsequent processing via SPW software. The results show that the acquisition scheme of a digital receiver that uses a set of carriers for down-conversion and partitions the correlation time into subintervals can tolerate large Doppler offsets of the carrier.

A testbench to simulate a LEO satellite channel has been built and used to enable design trade-offs. For the specific design selection of maximum $\Delta f T_M \simeq .32$ the acquisition performances of the MAX and TC code phase selection schemes were measured from the collected data. The MAX criterion was found to perform better than the TC criterion at large residual Doppler offsets and low IF SNR.

References

[1] G. E. Corazza, "On the MAX/TC Criterion for Code Acquisition and Its Application to DS-SSMA Systems," *IEEE Trans. Commun.*, vol. COM-44, pp. 1173-1182, Sep. 1996.

[2] B. Chung, C. Chien, H. Samueli and Rayeev Jain, "Performance Analysis of an All-Digital BPSK Direct Sequence Spread Spectrum IF Receiver Architecture," *IEEE J. on Sel. Areas in Commun.*, vol. 11, no. 7, pp. 1096-11075, Sep. 1993.

[3] U. Cheng, W. Hurd and J. Statman, "Spread Spectrum Code Acquisition in the Presence of Doppler Shift and Data Modulation," *IEEE Trans. Commun.*, vol. COM-38, pp. 241-250, Feb. 1990.

[4] L. Davisson and P. Flikkema, "Fast Single Element PN Acquisition for the TDRSS MA System," *IEEE Trans. Commun.*, vol. COM-36, pp. 1226-1235, Nov. 1988.

[5] Y. Su, "Rapid Code Acquisition Algorithms Employing PN Matched Filters," *IEEE Trans. Commun.*, vol. COM-36, pp. 724-733, June 1988.

[6] E. Sourour, and S. Gupta, "Direct sequence spread spectrum parallel acquisition in a fading mobile channel", *IEEE Trans. Commun.*, vol. COM-38, pp. 992-998, July 1990.

[7] W. Waters and B. Jarrett, "Bandpass Signal Sampling and Coherent Detection," *IEEE Trans. on Aerospace and Elec. Sys.*, AES-18, Nov 1982.

The L-band Land Mobile (LLM) System:

Access Schemes Trade-off

R.Giubilei, L.Miracapillo

Alenia Aerospazio - Divisione Spazio
Via Saccomuro, 24 - 00131 - Roma Italy

ABSTRACT

The LLM payload is a high performance L-band multibeam payload embarked aboard the ARTEMIS satellite and it is the heart of a telecommunication system offering a land mobile communication service over Europe.
The scope of this paper is to provide an overview of the LLM payload characteristics and to present a trade-off among some candidate access schemes.

1. INTRODUCTION

In the perspective of the next generation of mobile satellite systems and in addition to Inmarsat networks, several regional initiatives have been undertaken to provide L-band capacity over North America (AMSC/TMI) and Europe, the latter by the European Space Agency (ESA's European Mobile Satellite Services - EMSS [1], [2]). ESA has been widely active in promoting the technology for land mobile communications and has established a development plan for having in orbit an adequate capacity. Besides the EMS payload aboard Italsat, launched in 1996, ESA promoted the development of a high capacity payload (LLM, [3]), that will be launched aboard ARTEMIS in 2000. LLM payload key features are:

- the generation of multiple beams,
- the flexible channel-to-beam and beam-to-channel allocation,
- the use of an efficient technique for the utilization of spacecraft power resources through flexible adaptation to traffic demand,
- the possibility of frequency reuse by spatial discrimination between non-adjacent beams, or via orthogonal polarization.

These characteristics led to the implementation of new transponder architectures (*Multimatrix-fed antenna* [4], *Butler matrix* amplification scheme [5], [6]) and to the use of challenging technologies (SAW filters, switch matrix with a large number of interconnections).
The LLM has been originally specified and designed for full-duplex transmission between a Fixed Earth Station (FES) and land-mobile terminals, by using FDMA. However, due to its transparent architecture and a full tunability over the whole mobile services bandwidth, the LLM has an inherent flexibility that can be exploited for enhancing the service capacity. The key area of improvement is certainly the access technique and the best access, modulation and coding scheme to be used for mobile satellite communication is an issue often debated. In particular the trade-off between Code Division Multiple Access (CDMA) and FDMA, and the channel efficiency as a function of the communication system and satellite payload parameters have been several times presented in literature [7], [8] .
In this paper we would reverse the previous approach: the LLM payload characteristics are known. In particular the EIRP, G/T, number of accesses offered, cross-polarization and adjacent beams isolation, C/I are defined and in most of the cases measured during the test campaign , therefore it is worth making the comparison among CDMA, FDMA, modulation and coding schemes, by using the LLM system as a benchmark.

The paper is organized as follows: the LLM payload is shortly described and the basic communication parameters presented, therefore some general findings about the efficiency of various modulation and coding schemes are presented and a trade-off is carried out, based on the actual communication parameters specified for the LLM system.

2. THE COMMUNICATION SYSTEM

The L-band Land Mobile payload will be used for bi-directional communications between terrestrial mobile vehicles and fixed earth stations (FES) both located over the European territory. Up to 445 L-band users may simultaneously access the LLM payload which connects them, via the Ku band feeder link, to the earth stations. At L-band a portion of the spectrum allocated by WARC'92 for mobile satellite service is used, namely: from 1530 to 1559 MHz in the forward (GEO-Mobile) link and from 1631.5 to 1660.5 MHz in the return (Mobile-GEO) link.

At Ku band the frequency allocation is from 14231 to 14250 MHz in the forward (FES-GEO) and from 12731 to 12750 MHz in the return (GEO-FES) link.
The L-band service provides four different coverages: one broad beam covering Europe and North Africa and three overlapped spot beams together covering 90 % of the broad beam coverage (Spot A on northwestern Europe, spot B on central Europe and Spot C on eastern Europe and Asia) (fig. 1). A double circular polarization (LHCP and RHCP) is used in the broad beam and Spot B beam coverages allowing frequency reuse by discrimination of polarization.

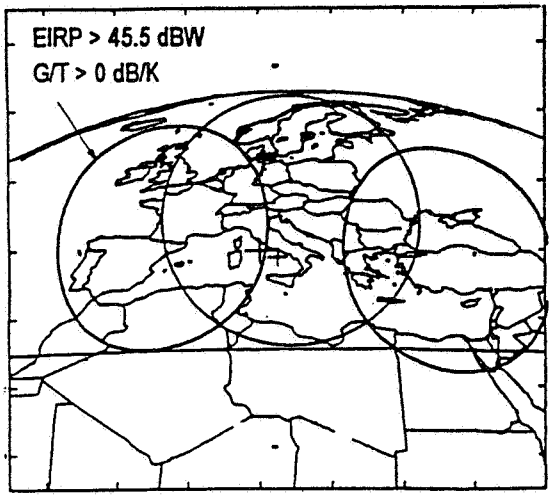


Figure 1: L-band spot beams coverage (from [1])

Spot A and Spot C will be used to experiment with frequency reuse by means of spatial diversity. An European broad beam coverage is provided by the Ku-band feeder link.

Table I shows the RF characteristics of the payload, over the above shown coverages. The overall EIRP shall be 45.5 dBW in the spot beams and 43.5 dBW on the broad beam coverage.

	L Band		Ku Band
	Broad Beam Coverage	Spot Beam Coverage	European Coverage
EIRP/carrier	19 dBW	19 dBW	7 dBW
G/T	-2 dB/K	0 dB/K	-1.4 dB/K
# Users	280	445	445
Polarization	RHCP, LHCP	Spot B: L/ RHCP Spot A, C: LHCP	

Table I: LLM RF characteristics

In order to allow coordination with other systems and organizations active in the same bands and to experiment with frequency reuse by polarization discrimination on the same beam and spatial diversity, the L-band channels are tunable in steps of 0.5 MHz over the whole allocated L-band spectrum. The detailed frequency plan is shown in figure 2. It foresees up to three 1 MHz wide and three 4 MHz wide channels. In the return link the 4 MHz wide channels are clusters of 900 KHz elementary channels. The key feature of the payload is the flexible channel-to-beam and beam-to-channel allocation in forward and return link, so that communication traffic among the different beams be dynamically reconfigurable. This is achieved by a channel-to-beam connection performed by means of commandable switching matrices implemented at intermediate frequency within the payload.

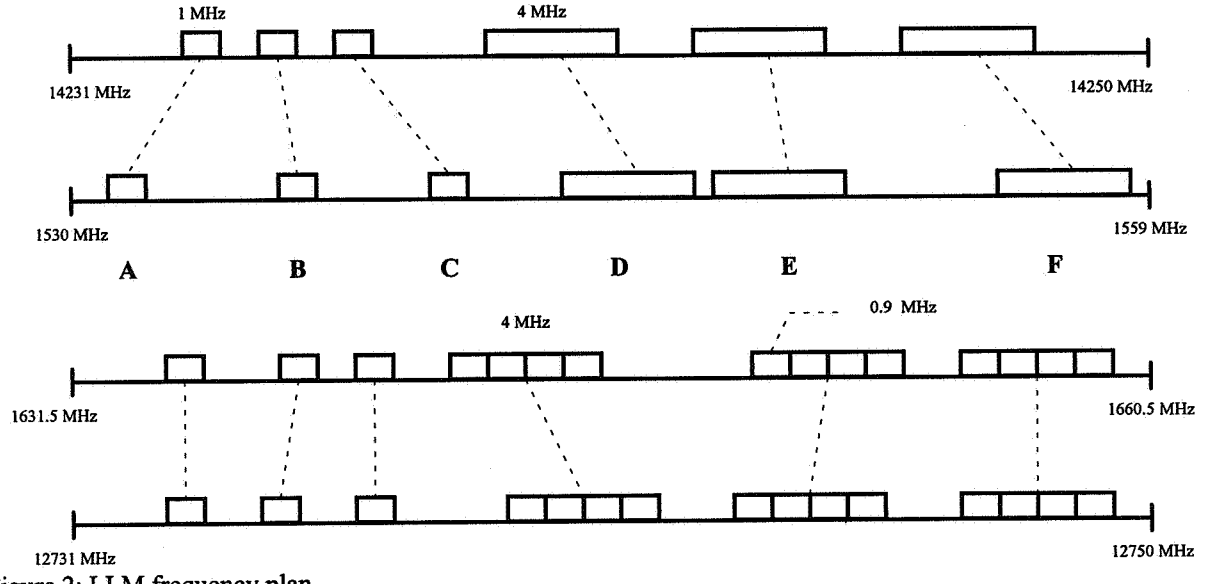


Figure 2: LLM frequency plan

Table II shows the connections that the payload has to guarantee.

Chan	Broad beam Coverage		Spot A	Spot B		Spot C	BW
	LHCP (GL)	RHCP (GR)	LHCP (AL)	LHCP (BL)	RHCP (BR)	LHCP (CL)	MHz
A	•		•				1
B				•	•		1
C		•		•		•	1
D	•	•	•				4
E	•	•		•	•		4
F	•	•				•	4

Table II: Channel-to-beam allocation

Looking at the frequency plan (fig. 2), the channels in both forward and return link are named A, B, C, D, E and F starting from the lowest frequency.
In the return link channels D, E and F represent clusters of 900 KHz sub-channels. Each channel can be connected to the specified beam one at a time, then the payload processes up to 6 channels contemporaneously.

2.1 Communication system parameters

In the following we provide the most significant communication parameters which characterize the system.
The carrier Input Power Flux Density (IPFD), in the 14 GHz feeder link, is -120 dBW/m^2 and the EIRP transmitted per carrier is 19 dBW.
Tables III and IV show some example link budgets derived from [2]. These budgets provide the C/No available for the service at the Mobile Earth Station (MES) and at the Fixed Earth Station (FES).

FES -> Satellite link (14 GHz)	
F.E.S. EIRP @ 1.8 mt antenna, 1 W SSPA	45 dBW
Path Losses	207.4 dB
Atmospheric Loss	3 dB
Additional Losses (Pointing, Polarization,...)	2 dB
Satellite G/T	-1.4 dB/K
Up-link C/No	59.8 dBHz
Satellite -> M.E.S. (1.55 Ghz)	
SAT EIRP/carrier	19 dBW
Path Losses	188.3 dB
Additional Losses	1 dB
M.E.S. G/T	-13 dB/K
Down-link C/No	45.3 dBHz
Overall C/No	44.6 dBHz

Table III: example link budget for FES->MES [2]

MES -> Satellite link (1.65 GHz)	
M.E.S. EIRP	22 dBW
Path Losses	188.8 dB
Additional Losses (Atmospheric, Pointing,...)	1 dB
Satellite G/T @LLM Spot beams	-2 dB/K
Up-link C/No	58.8 dBHz
Satellite -> F.E.S. (12 GHz)	
SAT EIRP/carrier	7 dBW
Path Losses	206.4 dB
Atmospheric Loss	2 dB
Additional Losses (Pointing, Polarization,...)	1.5 dB
F.E.S. G/T	20 dB/K
Down-link C/No	45.7 dBHz
Overall C/No	45.5 dBHz

Table IV: example link budget for MES->FES [2]

3. ACCESS SCHEME TRADE-OFF

3.1 Generalities

The access, modulation and coding schemes to be used for mobile satellite communication is an issue many times debated in literature [7], [8], [9] .
The efficiency is normally expressed in terms of information bits transmitted in 1 sec over a 1 Hz bandwidth (b/s/Hz). In case of a satellite link, it must be taken into account that the transponder power resources are generally limited, therefore a very efficient access scheme (many users) can be limited by the overall transponder available power [9]. This case is referred to as *power limited* condition. The opposite is referred to as *bandwidth limited* condition.
Assuming as applicable the latter condition, it is known that the efficiency of a FDMA access scheme is given by

$$\eta_{FDMA} \equiv \frac{NR_b}{B_A} = r\log_2(M) \qquad (1)$$

where:
- N is the number of users accessing the channel
- B_A is the aggregate available bandwidth
- R_b is the information rate per user
- r is the coding rate
- M is the signal constellation dimension (2 for BPSK)

Code Division Multiple Access (CDMA) is a very attractive technique for implementing a satellite mobile communication service. In CDMA systems all users operate at the same nominal frequency and use simultaneously the same bandwidth. The isolation among

the users signals is provided by the low correlation among the spreading codes. The basic limitation of Asynchronous CDMA (A-CDMA) systems is due to the interference caused on the wanted signal by all the other users, which is approximately seen by the receiver as an increase in the thermal noise level.
The ratio between the energy-per-bit and the thermal noise plus interference spectral density is given by

$$\frac{E_b}{N_o + I_o} \equiv \frac{E_b}{N_o'} = \frac{E_b / N_o}{1 + \frac{(N-1)}{G_p} E_b / N_o} \quad (2)$$

where G_p is the so-called *processing gain*, given by the ratio between the bit duration T_b and the chip duration T_c. This equation is valid in case of *bandlimited signal* [10]. In case of *time limited signal*, a factor 2/3 multiplies the second term into the denominator. The efficiency is given by

$$\eta_{CDMA} \equiv \frac{NR_b}{B_A} = \frac{1}{E_b / N_o'} \left[1 - \frac{E_b / N_o'}{E_b / N_o} \right] \quad (3)$$

where E_b / N_o', given by (2), depends on the bit error rate requested and on the coding scheme used. Taking the limit for $E_b / N_o \to \infty$ we get $\eta_{CDMA} \to \left(E_b / N_o' \right)^{-1}$.
Therefore the efficiency is in general well below 1. The use of coding improves the efficiency, since lower values of E_b / N_o' lead to higher values of efficiency.
The key problem of CDMA systems is the tight users power control needed to keep under control the self-noise and achieve the theoretical performance. In order to solve the problem a new access technique has been proposed in literature [11] by an ESA research team and developed. The concept behind this technique, called Bandlimited Quasi-Synchronous (BLQS) CDMA, is that in a perfectly chip-synchronous system (as seen by the receiver) orthogonal binary sequences can be used to perform the spreading. The orthogonality is maintained in receiving, thanks to a proper synchronization system, thus virtually eliminating the self noise. In practice full chip synchronization cannot be achieved, but it has been shown in [11] that an accuracy of few tenth of chip is reasonable and optimal codes have been found under this hypothesis.
The number of users accessing the channel is the number of orthogonal codes of period $L = rT_b / T_c$ contained in a channel symbol. This number is just L, as in case of *Walsh-Hadamard* sequences. This number decreases with the coding rate r. Eventually the efficiency of a BLQS-CDMA system, in case of BPSK modulation is

$$\eta_{BLQS} \equiv \frac{LR_b}{B_A} = r \quad bit/chip \quad (4)$$

This technique has the main advantage of limiting the effect of the *near-far* phenomenon, typical of A-CDMA. The price to be paid is the complexity of the network synchronization. Figure 3 shows a comparison in terms of efficiency (bit/chip) for coded and uncoded A-CDMA and BLQS-CDMA.

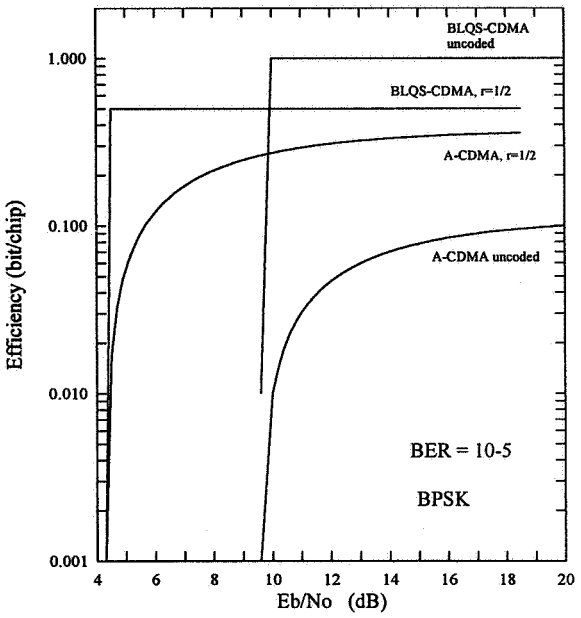


Figure 3: Efficiency of BLQS and A-CDMA

3.2 *FDMA efficiency*

The asymptotic behaviour previously addressed are now investigated to take into account the constraints given by a satellite transponder (e.g. the power), then the results are specialized for the LLM forward link and voice traffic only is taken into account.
The number of channels N that can be allocated for the service are

$$N = \min\left\{ \frac{B_A}{B_S}; \frac{EIRP_A}{V_a EIRP_S} \right\} \quad (5)$$

where:
- B_S is the signal bandwidth
- $EIRP_A$ is the aggregate payload EIRP
- $EIRP_S$ is the EIRP per carrier required
- V_a is the voice activity factor

and the two conditions identify the band limited and the power limited situation, respectively. The signal bandwidth B_S can be expressed by

$$B_S \equiv \frac{R_b(1+\alpha)}{r \log_2(M)} \quad (6)$$

where α is the roll-off factor (40%). Moreover the E_b / N_o requested at the MES receiver is given by

$$\frac{E_b}{N_o} = EIRP_S \frac{Q}{R_b} \quad (7)$$

where

$$Q \equiv G / T_{MES} + 228.6 - L_f \text{ dB} \qquad (8)$$

and L_f are the Free Space Losses. Finally we get

$$\eta_{FDMA} \equiv \frac{NR_b}{B_A} = \begin{cases} \dfrac{r \log_2(M)}{1+\alpha} & \text{(bandlimited)} \\ \dfrac{EIRP_A}{B_A} \dfrac{Q}{V_a} \dfrac{1}{E_b / N_o} & \text{(power limited)} \end{cases} \qquad (9)$$

3.3 CDMA Efficiency

The CDMA efficiency is computed assuming that the available bandwidth (B_A) is partitioned in several (N_{sl}), not necessarily adjacent, slots (e.g. 1 MHz wide), thus

$$\frac{B_A}{N_{sl}} \cong R_c(1+\alpha)$$

and R_c is the chip rate. Starting from (2), after some manipulation we get

$$\eta_{CDMA} \equiv \frac{NR_b}{B_A} = \frac{1}{V_a E_b / N_o'} \frac{\dfrac{EIRP_A}{B_A} Q}{1 + \dfrac{EIRP_A}{B_A}(1+\alpha)Q} \qquad (10)$$

Taking the limit for $EIRP_A \to \infty$ (bandlimited condition) we obtain

$$\eta_{CDMA} \to \frac{1}{V_a E_b / N_o'(1+\alpha)}$$

3.4 LLM Forward link Efficiency trade-off

The efficiency is evaluated in the following for the set of modulation and coding schemes given in Table V and for the typical LLM system forward link parameters given in Table VI.

	BPSK	8-DPSK	BPSK	CDMA
r	1/2, K=7	2/3, TCM 16 state	3/4, K=5	1/3,K=7
M	2	8	2	2
B_S	13.5 KHz	3.4 KHz	9 KHz	1 MHz
$EIRP_S$	16.7 dBW	21.7 dBW	18.2 dBW	16.4 dBW
BER	10^{-3}			
E_b/N_o	3 dB	8 dB	4.5 dB	-
E_b/N'_o	-	-	-	2.7 dB
Margin	4 dB			

Table V: Modulation and coding

The BPSK, r=3/4 scheme has been proposed in [2], for the Mobile Satellite Business Network (MSBN), via EMS and LLM.

Trellis codes (TCM) have been studied and proposed in the frame of NASA's MSAT program for land mobile communications and in particular the scheme here considered is rate 2/3, 16-state, trellis coded 8-DPSK (Differentially coherent 8-PSK), whose performance has been evaluated by computer simulation (see [12]).

The overall margin of 4 dB takes into account the system margin, transponder and implementation losses and the degradation due to fading.

The degradation due to the mobile channel strongly depends on the channel model considered and on the specific modulation, coding and interleaving scheme used. In particular we refer to a Rician channel with K=10, for which the additional loss ranges between 1.0 dB and 2.0 dB ([8], [13]).

B_A	15 MHz
R_b	4.8 Kb/s
V_a	0.4
α	0.4
$EIRP_A$	45.5 dBW
MES G/T	-13 dB/K
L_f	188.5 dB

Table VI: LLM communication system parameters

Figure 4 shows the synthesis of the results. The triangle indicates the actual efficiency, given simply by $445 \cdot R_b (V_a B_A)^{-1} = 0.36$. The straight line at 45.5 dBW shows the actual operating point. At that point the use of high level modulation schemes (as 8-DPSK) is not recommended, because the system in power limited. The threshold over which such schemes are attractive is between 50 dBW and 55 dBW.

The CDMA performance is not better than FDMA, notwithstanding the voice activation factor is included in the efficiency evaluation (10). Similar results have been obtained for the EMS system which has just a single broad beam over Europe [13]. This is mainly due to two major factors not included in the analysis: (i) spatial discrimination given by a multibeam antenna and (ii) cross-polarization frequency reuse

(i) The use of multibeam antennas increases both the power efficiency, by increasing EIRP and G/T, and system capacity through frequency reuse. In FDMA the angular spacing between beams with the same frequency band is determined only by the antenna spatial isolation capability. In CDMA the isolation is increased by the PN codes low correlation and it results in a factor of 3 [8] superior to the best achievable frequency reuse with FDMA.

(ii) PN codes low correlation properties lead to a better tolerance to low polarization isolation of CDMA vs. FDMA systems. Since the CDMA asymptotic efficiency is multiplied by a factor $2/(1+1/\rho)$, where ρ is the polarization isolation, for ρ = 6 dB the CDMA efficiency increases by 60%. Such an isolation value does not provide an adequate C/I for a FDMA system.

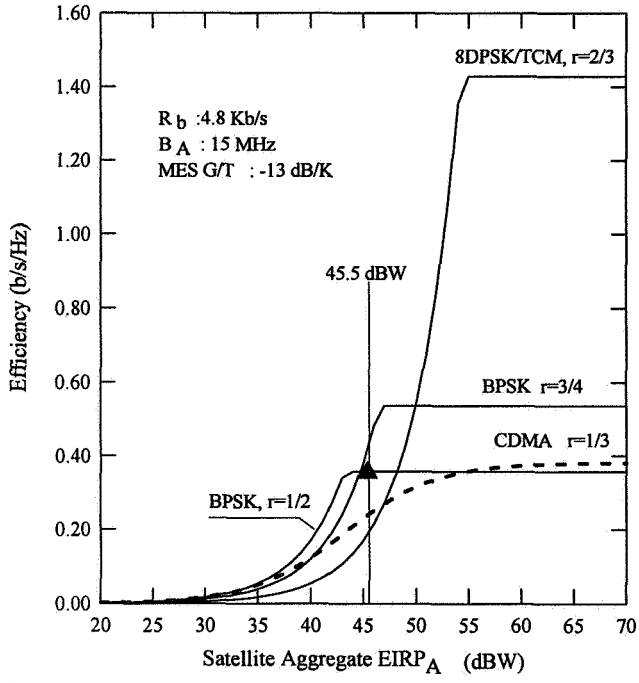


Figure 4: LLM Forward link efficiency analysis

6. CONCLUSIONS

The LLM payload onboard ARTEMIS is a key element of the ESA development plan for having in orbit an adequate capacity at L band to support future Land Mobile Communication Services.

The payload is able to handle 445 simultaneous users and provides an overall EIRP (at L-band) of about 45.5 dBW on the spot beam and 43.5 dBW on a global coverage over Europe. The G/T (at L-band) will be 0 dB/K on spot beams and -2 dB/K over the global coverage.

The main features of the payload are the use of multiple beams, an efficient technique for the utilization of spacecraft resources through flexible adaptation to traffic demand and the possibility of frequency reuse by spatial discrimination between non-adjacent beams, and via orthogonal polarization.

The LLM can support both FDMA and CDMA accesses, and various modulation and coding schemes. The actual aggregate EIRP available on the three spot beams of 45.5 dBW does not allow the use of higher modulation levels (e.g. 8-DPSK/TCM) that are very attractive in case of EIRP increase over 50 dBW.

CDMA can simplify the access procedures to the network and provide, under certain conditions, a better efficiency, however, to exploit its peculiarities, the use of spot beams and double polarization for extensive frequency reuse is mandatory.

Note

The work relevant to the LLM payload has been developed under contract of the European Space Agency. However the views expressed herein, in particular those relevant to access, modulation and coding techniques, do not reflect necessarily official ESA policy.

REFERENCES

[1] C.Loisy, P.Edin and F.J. Benedicto, *European Mobile Satellite Services (EMSS) a Regional System for Europe* in ESA Doc n°ESA/ESTEC/TSM/05/95, May 1985.

[2] R.Rogard, *LMSS: from Low Data Rate to Voice Services*, 15th AIAA - Washington - 1992

[3] L.Miracapillo, T.Sassorossi and R.Giubilei, *The L-Band Land Mobile Payload (LLM) aboard Artemis*, 16th AIAA - Washington Feb.25-29, 1996

[4] A.Roederer and M.Sabadini, *A Semi-Active Multibeam Antenna Concept*, IEEE Int. Symposium on Antennas and Propagation, 1990

[5] W.A.Sadrin, *The Butler Matrix transponder*, COMSAT Tech. Rev., Fall 1974

[6] S.Egami et al., *An Adaptive Multiple Beam System Concept*, IEEE J. Select. Areas Commun., May 1987.

[7] K.G.Johannsen, *CDMA vs. FDMA Channel Capacity in Mobile Satellite Communication*, International Journal of Sat. Comms - Vol.6 - 1988

[8] K.S.Gilhousen et al., *Increased Capacity Using CDMA for Mobile Satellite Communication*, IEEE J.Select. Areas Commun. - May 1990.

[9] A.J.Viterbi, *When not to spread spectrum - A sequel*, IEEE Commun. Mag., April 1985.

[10] A.J.Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison Wesley - 1995.

[11] R.De Gaudenzi et al., *Bandlimited Quasi Synchronous CDMA: A Novel Satellite Access Technique for Mobile and Personal Communications Systems*, IEEE J.Select. Areas Commun. February 1992.

[12] E.Biglieri, D.Divsalar, P.J.McLane and M.K.Simon, *Introduction to Trellis-Coded Modulation with Applications*, Maxwell MacMillan Int. - 1991.

[13] ESA Contract 8692/89/F/RD: *State of the art on advanced access, modulation and coding techniques for satellite land-mobile applications*. Qualcomm, Inc. & Space Engineering - Final presentation - March 1990.

Power Control Performance in Measurement Based S-PCN Channel

P.Taaghol, R. Tafazolli, B. G. Evans
Centre for Communications System Reset (CCSR)
University of Surrey, Guildford,
Surrey GU2 5XH, England
Tel: +44-1483 25 9810, Fax: +44-1483 25 9504
Email: P.Taaghol@ee.surrey.ac.uk

Abstract: *In this paper we highlight the power control issues of the proposed CDMA-based Satellite Personal Communication Networks. We identify the role of both the open and closed loop power control schemes in this environment and investigate their limitations. Correlation models for the shadow fading in such environments are proposed based on channel recordings. We finally evaluate the performance of the closed loop power control in actual recorded channels in different environments.*

1. INTRODUCTION

Recently there has been great interest in the employment of Code Division Multiple Access (CDMA) for Satellite Personal Communications Networks (S-PCN) in search for capacity improvements over the conventional TDMA and FDMA and easier frequency planning schemes. Since the capacity of the CDMA system is interference limited, minimising the effects of Multiple Access Interference (MAI) is of great importance. The use of asynchronous CDMA in the uplink (UT-SAT-LES) gives rise to tight power control requirements as the spreading sequences of different UT's loose their orthagonality with respect to each other to a large extent. It is hence desired that the received signals (from all the UTs within a spotbeam), at the input of the Land Earth Station (LES) receiver to be of the same power level, regardless of the position of the UTs and the characteristics of the channels.

The paper is organised as follows,
In section-2 we discuss the differences between the terrestrial and satellite power control techniques. We then proceed to evaluate the correlation distance of the shadow fading in the satellite environment. And finally we evaluate the performance of the closed loop power control in actual channel recordings.

2. POWER CONTROL IN S-PCN

Power control can be broadly divided into two main schemes, the Open Loop Power Control (OLPC) and the Closed Loop Power Control (CLPC) schemes.

a. Open Loop Power Control

OLPC is designed to track slower variations of the received signal due to shadowing and path losses through the use of a downlink pilot channel, usually transmitted at a few dB's higher than the traffic channel. It is only possible to track slower variations of the received signal with OLPC as it assumes reciprocity between the up and downlink channels. The implementation of the OLPC is not problematic in nature. However, it is quite important to be able to separate the fast variations due to multipath from the slower shadowing fluctuations, accurately in order to ensure correct OLPC operation. The design of the filter responsible for averaging-out the multipath has proven quite challenging [1]. Averaging periods of 20-80λ are common practice, however, due to the movements of the UT (velocity range of 0-150 km/h), any filtering scheme would only be optimised for a particular UT velocity. In addition to this, the statistics of the shadow fading changes significantly for different environments as well as different elevation angles. Consequently, some level of estimation error would always be associated with the OLPC. Nevertheless, knowing the environment and constellation statistics such as the average elevation angle, it would be possible to optimise the filter for the most likely common scenario. In order to do so, second order statistics of the shadow fading such as the effective correlation distance of the shadowing would be required. In section 3, we report some of these values for elevation angles of 60°-80° in two different environments.

b. Closed Loop Power Control

CLPC is designed to track the faster variations of the channel. The fact that a feedback loop exists in the system, imposes an inherent hysteresis phenomenon. That is, by the time the received power is measured (at LES), and the actual power control commands are received by the UT, the fading has changed and the command to increase or decrease the power may no longer be valid.

As far as the accuracy of the CLPC is concerned, the long Round-Trip Delays (RTDs) associated with the satellite

environment (Table-1) are by far the most limiting factor. This significantly limits the Power Control Command Rate (PCCR) which in turn reduces the capability of the Close Loop Power Control (CLPC), in tracking the fast fading.

Table-1 : Typical Round-Trip Delays (UT-LES-UT)

Altitude (km)	Constellation	Min. Elv. Angle	RTD (ms) Min.	Max.
780	LEO	8°	10	33
1414	LEO	10°	19	.47
10355	MEO	10°	138	192

However, CLPC could still prove useful under some circumstances such as a slow moving or a stationary UT which might be in fade for example in the uplink but not in the downlink.

Furthermore, there are indications that in cases were the up and downlink frequencies are significantly separated, the slow shadow variations are not closely correlated. CCSR intends to carry out a simultaneous multi-band measurement campaign shortly, through which the correlation between the slow variations of different frequency bands would be investigated. In the likely event of loosely correlated up and downlink shadowing variations, the CLPC could provide crucial assistance for the OLPC in accurately tracking such variations.

When considering CLPC, one must not also loos sight of the fact that even in the terrestrial environments, were the RTDs are much smaller, the use of very high PCCRs is not common practice due to resource limitations. In fact at higher velocities interleaving becomes an more efficient technique in randomising the errors and hence maintaining the required quality of service. This is quite evident in the field trial results of Qualcomm's IS-95 [2], where the peak (at 10-20 mph) in the Frame Error Rate (FER) distinctively separates the two low FER performance regions to its sides. The FER is kept very low at slow velocities (0-10 mph) as CLPC closely tracks the fast variations. To the other side of the FER peak, interleaving keeps the FER low as it kicks-in from 20 mph. In fact the FER peak falls into a region where neither CLPC nor the interleaving is effective. But as far as the satellite environment is concerned, long interleaving depths are not feasible due to an existing long RTD, which implies lower altitude constellations would be more desirable.

3. CORRELATION MODEL FOR SHADOW FADING IN LAND-MOBILE SATELLITE CHANNEL

Recently there has been a great effort in characterisation of the land-mobile S-PCN channel. Such experiments are very costly and time consuming. Consequently, unlike the terrestrial case, a very limited number of such campaigns have been carried out.

The first-order (time-invariant) statistics of S-PCN channels have been extensively reported [3] [4], however, the second-order (time-variant) statistics are generally much less known. These statistics are of vital importance for realistic evaluation of any power control scheme and have not yet been fully available for satellite channels restricting research to only terrestrial values [5] reported by [6].

For the purpose of power control, we focus attention on correlation distance statistics. As far as the simulation of power control is concerned, correlation distance of shadowing is one of the most important statistics required for both the mathematical and the simulation based analysis of these techniques. Correlation distance of the shadowing simply indicates liklihoodness of two samples separated by the measurement interval. Knowing the correlation distance between to samples, the sampling rate, the standard deviation of the shadow component around it's mean, and the mean of the shadowed signal, a simple shadow fading generator shown in Figure-1 can be developed.

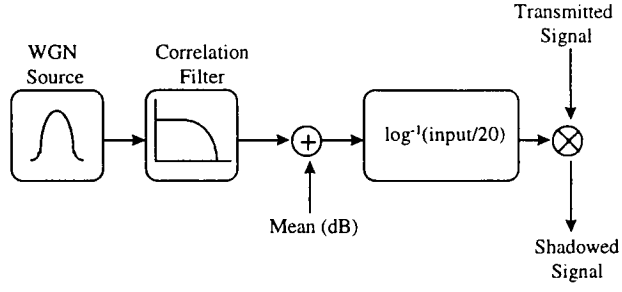


Figure-1: Shadow fading generator

The proposed shadow fading correlation model is based on the narrowband measurement campaign recordings at L-band (1.3 GHz) and S-band (2.3 GHz), carried out by CCSR[1] in spring'1992 [4]. The measurements were carried out using a helicopter emulating a satellite moving in parallel with the mobile unit. We focus attention on suburban and heavily wooded categories as the urban environment is not the main S-PCN operational

[1] Centre for Communication Systems Research (CCSR), University of Surrey

environment, and the open category is not a hostile environment.

a. Log-normal Shadowing

The channel recordings were low-pass filtered in order to average out the multipath variations and hence separate the shadow fading. The low-pass filter was hence designed to be sufficiently narrow in order to average the fast variations, yet wide enough to pass the slower fading[Gold]. Such filters can be implemented using an integrate and dump, averaging the signal over a range of 20-80λ (depending on the environment) as suggested in [Lee] [Gud] and [Gold]. For the purpose of our analysis, in both the heavily wooded and the suburban environments, a length of 20λ (≈ 4.5m at L-band and 2.5m at S-band) was proved sufficient as the multipath component in such environments with relatively high elevation angles are not the dominant element of the received signal. The distribution of the filtered received signal strength (slow variations) was found to be quite log-normal. That is, the signal strength in dB is normally (Gaussian) distributed.

Figure-1 depicts the CDF of the received signal strength in dB, together with the Gaussian fit.

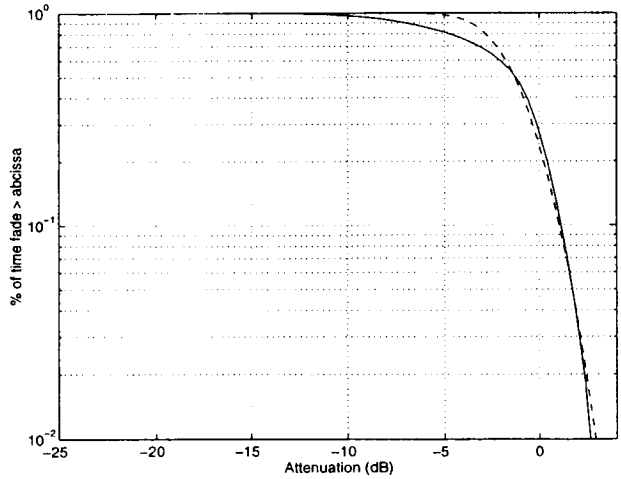


Figure-2 : Cumulative distribution function of the shadow fading in wooded environment, L-band, 60°

b. Correlation Model

In coherence with the approaches of [1] and [5], we propose a simple exponential autocovariance fit,

$$C_\varsigma(\tau) = \mathbf{E}\{\varsigma(t)\varsigma(t+\tau)\} - \mathbf{E}^2\{\varsigma(t)\} = \sigma_\varsigma^2 e^{-v|\tau|/X_c} \quad (1)$$

where X_c is the effective correlation distance of the shadow fading, v is the velocity and τ is the measurement interval. A more appropriate notations of the above expression independent of the velocity can be expressed as,

$$C_\varsigma(\tau) = \sigma_\varsigma^2 e^{-|do-d\,1|/X_c} \quad (2)$$

Where $|d_o - d_1|$ represents the measurement distance. The normalised correlation distance in the following diagrams is defined as the distance at which the correlation falls to e^{-1}. Also note that the results presented here are the normalised autocovariance, so that the covariances at zero lag are identically 1.

Figures-3 and 4 show the normalised autocovariance of the shadow fading in the wooded environment. The route consisted of a 7 km long stretch along a small typical country road of 4 meter width. The route was considered to have high tree density of tall nature and predominantly non-deciduous coniferous trees much taller than those in the suburban environment.

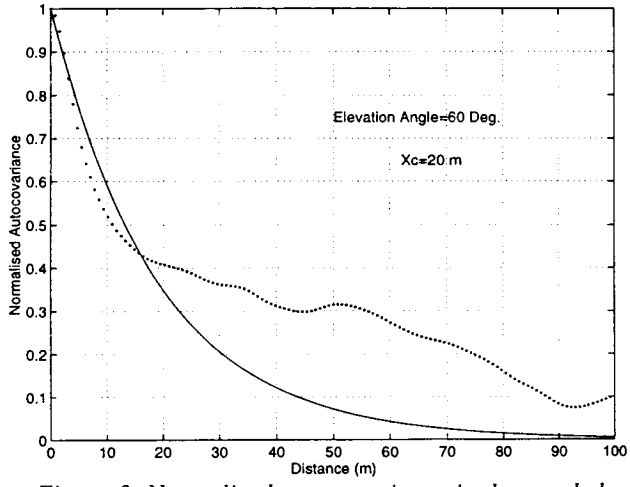


Figure-3: Normalised autocovariance in the wooded environment, L-band, 60°

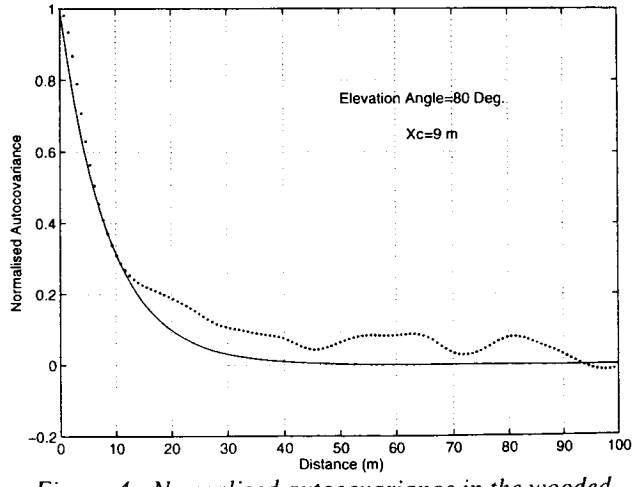


Figure-4 : Normalised autocovariance in the wooded environment, L-band, 80°

From the above it can be seen that the simple correlation model operates satisfactorily up a certain distance. Due to the nature of the tree shadowed environment, generally short correlation distances were measured. It can also be observed that at 80°, the correlation distance of the shadowing is shorter than the 60° case as the obstructing trees are less dense at the top.

Figures-5 and 6 show the normalised autocovariance of the shadow fading in the suburban environment. The suburban route considered can be described as a typical English town suburb. The roadside trees were primarily of tall and deciduous type. The residential buildings along the road were generally two stories high. The distance of trees and buildings along the same section of the road varied from the roadside edge from only a few meters to some tens of meters. However, on average it was estimated to be between 5-10 meters.

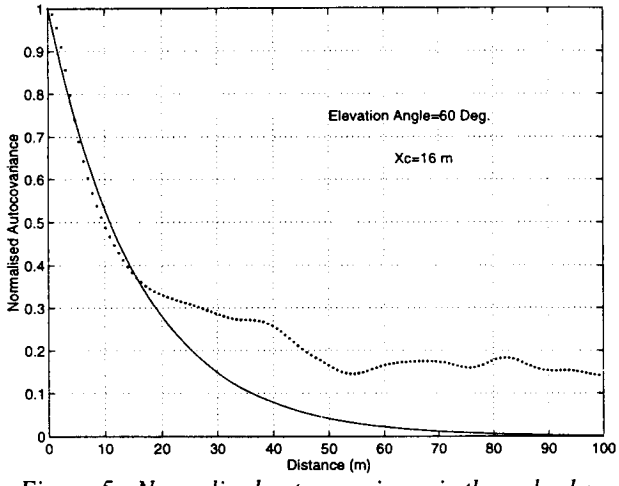


Figure-5 : Normalised autocovariance in the suburban environment, L-band, 60°

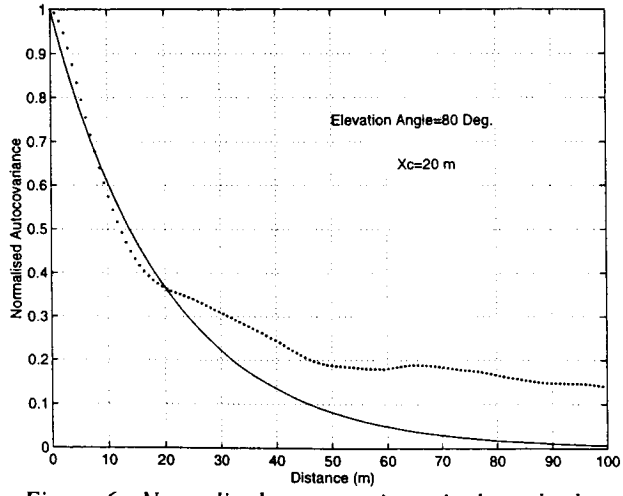


Figure-6 : Normalised autocovariance in the suburban environment, L-band, 80°

From the above, longer correlation distances compared to that of the wooded environment can be seen.

In general, short correlation distances of the order of tens of meters were obtained for elevation angles of 60°-80°, for both the environments. The results were quite expected as at such high elevation angles, any obstruction of the line-of-site will be momentarily, be it tree-tops or building roofs. However, preliminary analysis of the lower elevation angles indicate longer correlation distances. In low elevation angles the situation is expected to get closer to the terrestrial environment as the signal travels through more obstructions of random dimension. Similar results were also obtained for the S-band.

4. CLPC PERFORMANCE

The fixed-step feedback CLPC shown in Figure-7 was set-up in order to evaluate the performance of the closed loop power control using the actual channel recordings.

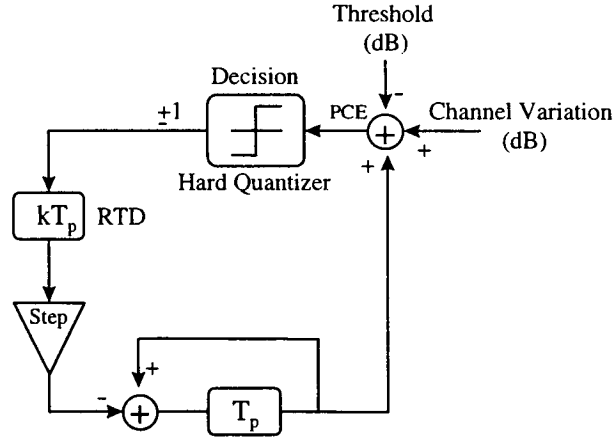


Figure-7: The feedback power control model

where T_p is the power control sampling period, RTD represents the round-trip delay (integer multiple of T_p) and PCE denotes the power control error.

The recordings were pre-processed and the large scale variations (shadowing component) were removed as the performance of the CLPC was to be evaluated. The channel was recorded at a fixed velocity of 8m/s, and a constant sampling rate. It is possible to vary the speed of the data, however, as the speed is lowered, less samples/s will have to processed which would then lead to the loss of resolution. It is, however, possible to overcome this problem by interpolating the lower sampling rate data, preferably using a *sinc-interpolator*.

But for the purpose of this paper simulations were carried out using the actual data recorded at 8m/s, 1024 samples/s. The standard deviation of the PCE was evaluated for a range of step sizes (0.25-1.25 dB). Best results were achieved using a fixed step size of 0.5 dB.

In order to demonstrate the performance degradation of the CLPC in the satellite environment with long RTDs, a range of simulations with different RTDs were carried out.

Figure-8 shows the standard deviation of the PCE for a range of delays covering up to the highest LEO. Longer delays were not simulated as the standard deviation of PCE was approaching the standard deviation of the channel without CLPC. In all cases simulated, the PCE was found to be log-normally distributed.

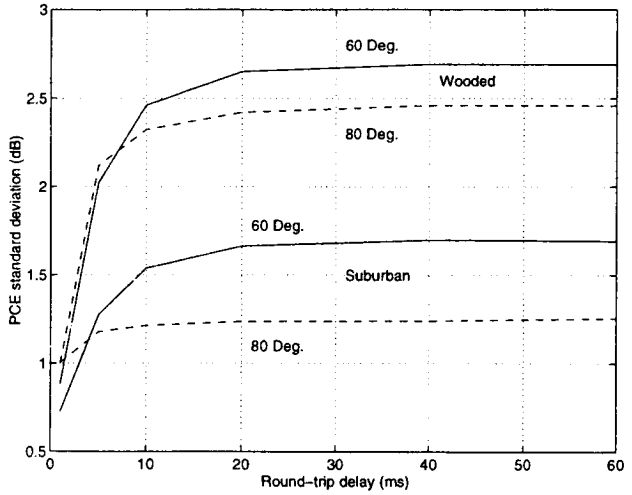


Figure-8: Standard deviation of the PCE for different values of RTD, L-band, 8m/s

It can be seen that for RTDs of more than 20ms, the standard deviation of the PCE reaches its maximum. Note that for lower speeds, longer RTDs can be tolerated. It is also interesting to see that the heavily wooded environment is a more hostile environment as far as CLPC is concerned. It is so, as the small dimension of the tree leaves and stems causes rapid fluctuations in the received signal level. Although these fluctuations could be considered as shadowing, it is neither possible to filter them out (unless averaged over a very short distance which almost certainly result in large estimation errors in other environments), nor to track them with CLPC.

5. CONCLUSIONS

We reviewed the power control techniques applicable to S-PCN and highlighted some of the problems associated both OLPC and CLPC. Simple yet accurate correlation models for the shadow fading in wooded and suburban environments based on channel recordings were proposed. It was demonstrated that the correlation distance for elevation angles of above 60° is in the order of tens of meters in both the environments.

Furthermore, the performance of CLPC in actual channel recordings for different RTDs were evaluated. The distribution of the PCE in all the simulated cases was found to be log-normal. CLPC breaks down as the RTD increases and PCE standard deviations of up to 2.5 dB are not uncommon.

As the PCCR is very limited in the satellite environment, the use of variable step size power control could improve the performance. This would come at no significant overhead despite the fact that each command would comprise of more than one bit, as they are send less frequently compared to the terrestrial case.

6. REFERENCES

[1] J. Goldsmith, L. J. Greenstein and G. J. Foschini, "Error Statistics of Real-Time Power Measurements in Cellular Channels with Multipath and Shadowing", IEEE Transactions on Vehicular Technology, vol.43, No. 3, pp.439-446, Aug. 1994.

[2] QUALCOMM, "Introduction to CDMA and CAI," *IS-95 Standard,* Issue I, pp. 1-58, 1992

[3] Lutz et al, "The Land Mobile Satellite Communication Channel-Recording, Statistics and Channel Model", IEEE Transactions on Vehicular Technology, vol.40, No. 2, pp.375-386, May 1991.

[4] M.A.N. Parks, B.G.Evans and G.Butt, "High elevation angle propagation results applied to a statistical model and an enhances empirical model", Electronics Letters, Vol.29, No.19, pp.1723-1724, 16th Sept. 1993.

[5] M. Monk, L. B. Milstein, "Open-Loop Power Control Error in Land Mobile Satellite System", IEEE Transactions on Selected Areas in Communications, vol.13, No. 2, pp.205-212, Feb. 1995.

[6] Gudmundson, "Correlation model for shadow fading in mobile radio systems", IEE Electronics Letters, vol. 27, No. 23, 7th Nov. 1991.

Other-Cell Interference in Satellite Power-Controlled CDMA Uplink

Erich Lutz
Institute for Communications Technology, DLR, D-82234 Wessling, Germany
Phone: +49-8153-28-2831 Fax: +49-8153-28-1442 e-mail: Erich.Lutz@dlr.de

ABSTRACT

Other-cell interference in satellite systems with CDMA uplink is investigated for different orbit heights (LEO, MEO), for different numbers of spotbeams within a satellite footprint, and for different spotbeam antenna characteristics. For uniformly distributed users, the other-cell interference factor k is evaluated by numerically integrating the other-cell interference power over a satellite footprint. Perfect power control of all users with regard to their respective spotbeam is assumed.

Additionally, for a Gaussian spotbeam antenna characteristic, a lower bound for k is derived analytically, which holds for small nadir angles within the footprint and is independent from orbit height and minimum elevation. A corresponding lower bound for k is numerically evaluated for a tapered-aperture antenna.

The derived values for k indicate that other-cell interference is much more critical for satellite systems than for terrestrial cellular systems.

INTRODUCTION

The capacity of an asynchronous CDMA uplink is limited by co-channel interference (multiple access interference), which is produced by interfering users within the wanted user's own cell, as well as by users in other cells. For perfect power control of user terminals, the uplink capacity (number of active users per cell, N_u) for asynchronous CDMA in cellular networks can be approximated as [1]:

$$N_u \approx \frac{W/R}{\alpha(1+k)}\left[\left(\frac{E_b}{N_{tot}}\right)^{-1} - \left(\frac{E_b}{N_0}\right)^{-1}\right] \qquad (1)$$

with:
W = spread-spectrum bandwidth,
R = data rate,
α = voice activity,
E_b/N_{tot} = required total SNR,
N_0 = thermal noise power density.

k is the other-cell interference factor, defined as interference power produced by users belonging to other cells divided by the interference power produced by other users in the considered cell. It can be seen from (1) that k is an important parameter for the assessment of CDMA in cellular systems.

For terrestrial CDMA, it is usually assumed that the signal power decreases with the fourth power of distance. Here, a typical value for the other-cell interference factor is k = 0.44 [2].

In satellite systems with spotbeams, the interference power is determined by the characteristic of the spotbeam antenna (the terminal antenna is assumed to be hemispherical). The spotbeam antenna is typically a tapered-aperture antenna [3]. Alternatively, the antenna main lobe can be approximated by a Gaussian characteristic, with a quadratic decrease of antenna gain (in dB), [4].

Another parameter determining interference is the definition of the spotbeam contour (edge of spotbeam coverage). Usually, the spotbeam contour is defined by a 3 dB decrease of antenna gain. Higher beam isolation leading to less interference can be achieved by choosing a larger gain decrease at spotbeam contour.

For synchronous CDMA, user signals within a given cell (own cell) may be orthogonal, producing neglectable interference. However, users in other cells normally use different code families and therefore are non-orthogonal with regard to the own cell.

In this paper we consider only the uplink, which is usually more critical with respect to multiple access capacity than the downlink.

OTHER-CELL INTERFERENCE FACTOR

Fig. 1 shows the geometry of the user with respect to the satellite and the characteristics of the satellite spotbeam antenna.

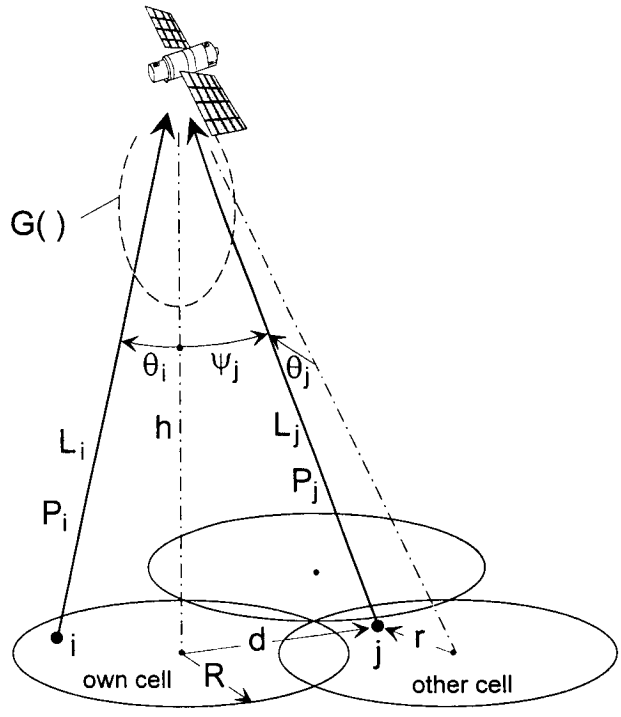


Fig. 1: Own- and other-cell interference scenario.

The other-cell interference factor k is given by the power received from all other-cell interferers j, divided by the power received from interferers i within the considered (own) cell. Assuming hemispherical terminal antennas, we have

$$k = \frac{\sum_{j \in \text{other cells}} P_j/L_j \cdot G(\psi_j)}{\sum_{i \in \text{own cell}} P_i/L_i \cdot G(\theta_i)} \qquad (2)$$

with

P_i, P_j = transmit power of interfering terminal,
L_i, L_j = signal attenuation factor,
$G(\,)$ = antenna characteristic of own spotbeam,
θ_i, θ_j = receive angle of interfering signal, relative to center of respective spotbeam (boresight angle),
ψ_j = receive angle of interfering signal from other spotbeam, relative to center of own spotbeam.

The sum in the numerator has to be taken over all other cells using the same frequency as the own cell. In the following, it is assumed that all cells use the same frequency.

Throughout the paper, the own spotbeam is considered the central spotbeam in nadir direction of the satellite.

For a user at satellite nadir, the received uplink power is designated as $P_{r,0} = P_0/L_0 \cdot G_{max}$ with $G_{max} = G(0)$. For a user anywhere within the own spotbeam the transmit power is assumed to be increased through power control by a factor of

$$\frac{L_i}{L_0} \cdot \frac{G_{max}}{G(\theta_i)} \qquad (3)$$

such that the differences in signal attenuation and antenna gain are compensated, and the received uplink power is maintained as $P_{r,0}$. If ideal user-to-satellite channels are assumed, L_i varies only according to the slant range. However, the analysis also holds for fading channels as far as multipath fading and shadowing can be compensated for. Finally, the power control can also compensate deviations of the terminal antennas from a hemispherical characteristic.

Similar to users in the own cell, the transmit power of terminals in other cells is controlled according to

$$\frac{L_j}{L_0} \cdot \frac{G_{max}}{G(\theta_j)} \qquad (4)$$

such that the received uplink power received at their own spotbeam antenna is $P_{r,0}$, too. It is assumed that the antenna characteristic of the other spotbeams is equivalent to that of the own spotbeam. The interference power caused by any terminal within another cell is therefore given by

$$P_{r,0} \cdot \frac{G(\psi_j)}{G(\theta_j)} \cdot \qquad (5)$$

Assuming uniformly distributed users, the other-cell interference factor k results in

$$k = \frac{1}{A_C} \iint_{\text{other cells}} \frac{G(\psi_j)}{G(\theta_j)}\, dA \qquad (6)$$

(A_C is the cell area). The index j designates the beam which the interferer belongs to.

NUMERICAL EVALUATION OF OTHER-CELL INTERFERENCE

The interferers are assumed to be uniformly distributed over the earth's surface, and are characterized by their longitude L and latitude B. L and B define an interferer's boresight angle θ_j within its own beam, as well as the interference power attenuation $G(\psi_j)$ into the nadir beam.

The spotbeam pattern of the satellite was derived from a flat hexagonal cell pattern: The center of the cell at the origin was taken as satellite nadir. The distances of cell centers from the origin were transformed into nadir angles of the spotbeam centers. The azimuth angles of the cell centers were taken as corresponding angles of the spherical co-ordinates. The numerical evaluation of other-cell interference can, however, be performed for any other spotbeam pattern as well. The half-power beamwidth of all spotbeams is assumed identical. k is evaluated by numerically integrating the other-cell interference power over a satellite footprint, assuming that an interferer belongs to that spotbeam from whose center its angular deviation θ_j is smallest. Perfect power control of all users with regard to their respective spotbeam is assumed.

The following scenarios are investigated:

- MEO system, h = 10 354 km, min. elevation = 10 deg, 169 spotbeams, one-sided beamwidth θ_c = 2.0 deg. The integration extends over 36 most inner beams.
- LEO system, h = 1414 km, min. elevation = 10 deg, 19 spotbeams, θ_c = 14.9 deg.

Two spotbeam antenna patterns are considered:

- Tapered-aperture antenna according to [3,8], with $p = 2$ and $T = 0.9$:

$$G(\theta) = G_{max} \cdot \left(\frac{J_1(u)}{2u} + 36 \frac{J_3(u)}{u^3} \right)^2 \qquad (7)$$

with

$$u = 2.07123 \cdot \frac{\sin\theta}{\sin\theta_{3dB}} , \qquad (8)$$

G_{max} = maximum antenna gain (in boresight direction),
θ_{3dB} = one-sided half-power (-3 dB) beamwidth.

$J_1(u)$ and $J_3(u)$ are the Bessel functions of the first kind and order 1 and 3, respectively.

- According to [4], the main lobe of a spotbeam antenna can be approximated by a Gaussian characteristic, with a quadratic decrease of antenna gain (in dB):

$$G(\theta) = G_{max} \cdot 10^{-0.3\,(\theta/\theta_{3dB})^2}$$
$$G(\theta) = G_{max} - 3\,(\theta/\theta_{3dB})^2 \quad \text{dBi} . \qquad (9)$$

Another parameter determining interference is the definition of the spotbeam contour. Usually, the spotbeam contour θ_c is defined by a 3 dB decrease of antenna gain. Higher beam isolation leading to less interference can be achieved by choosing a larger gain decrease at spotbeam contour.

In order to investigate the influence of spotbeam isolation, the gain decrease at spotbeam contour (identical to the gain decrease at the triple crossing point) is considered as variable c (in dB). Variation of c changes the shape of the antenna gain of a spotbeam and therefore influences the amount of other-cell interference. This is different from the terrestrial case, where the signal attenuation is usually described by a power law of distance. Here, the shape of attenuation does not depend on the separation of the cells.

To account for contour c, (8) is generalized to

$$u = G^{-1}\left(G_{max} \cdot 10^{-c/10}\right) \cdot \frac{\sin\theta}{\sin\theta_c} . \qquad (10)$$

Similarly, (9) generalizes to

$$G(\theta) = G_{max} \cdot 10^{-c/10 \cdot (\theta/\theta_c)^2}$$
$$G(\theta) = G_{max} - c\,(\theta/\theta_c)^2 \quad \text{dBi} \qquad (11)$$

with

$$\theta_c = \theta_{3dB} \cdot \sqrt{\frac{c}{3.0103}} . \qquad (12)$$

The results of the numerical evaluation of (6) are shown in Fig. 2.

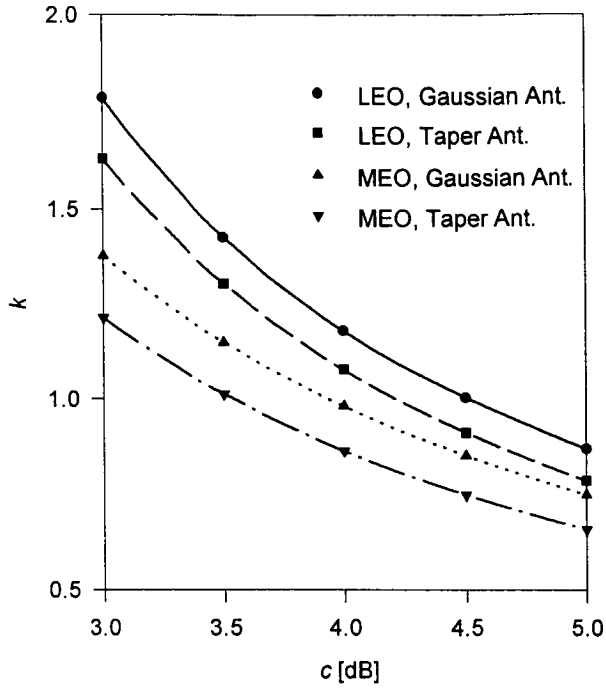


Fig. 2: Other-cell interference factor k versus spotbeam isolation c.

It can be seen that k is much larger than the value $k \approx 0.44$ for terrestrial cellular networks [2]. This shows that other-cell interference in satellite systems is much more critical than in terrestrial cellular networks. Moreover, the values shown in Fig. 2 are substantially larger than the values considered for satellite systems up to now [5]. Some numerical values are compiled in Table 1.

Fig. 2 indicates that other-cell interference can be reduced by increasing the spotbeam isolation. To achieve this, the antenna diameter has to be increased, or the number of spotbeams has to be reduced. For geometrical reasons, k is larger for the LEO scenario than for the MEO scenario.

LOWER BOUND FOR SMALL NADIR ANGLES

For deriving a lower bound of k for small nadir angles, we make the following assumptions:

- Gaussian spot beam antenna characteristic,
- curvature of earth is ignored,
- ideal channels,
- perfect power control of terminals.

For the geometry of the interference scenario we refer to Fig. 1.

Neglecting the curvature of the earth, the receive angle of an interferer, relative to the center of the wanted (nadir) spotbeam is

$$\psi_j = \arctan\, d/h \,. \tag{13}$$

With cell radius R, the contour of the spotbeam is determined by the nadir angle

$$\theta_c = \arctan\, R/h \,. \tag{14}$$

Therefore, the spotbeam antenna gain for an interferer (as a function of distance d of the interferer from the center of the wanted spotbeam) is

$$G(d) = G_{max} - c\left(\frac{\arctan\, d/h}{\arctan\, R/h}\right)^2 \quad \text{dBi} \tag{15}$$

with beam isolation $c = G_{max} - G(R)$. For $h >> d$ and $h >> R$, (15) can be approximated by

$$\begin{aligned} G(d) &\approx G_{max} - c(d/R)^2 \quad \text{dBi} \\ G(d) &\approx G_{max} \cdot 10^{-c/10\cdot(d/R)^2} \,. \end{aligned} \tag{16}$$

This approximation holds for small nadir angles and is quite good for MEO systems with a large number of spotbeams.

(16) describes the attenuation of an interferer, due to the antenna characteristic of the wanted (nadir) beam. Similarly, the power increase due to power control in the interferer's spotbeam can be represented by

$$\frac{1}{G(r)} = G_{max} \cdot 10^{c/10\cdot(r/R)^2} \tag{17}$$

cf. Fig. 1, again taking into account the approximation for small nadir angles.

Assuming uniformly distributed users, the other-cell interference factor can be evaluated correspondingly to (6):

$$k = \frac{1}{A_C} \iint\limits_{\substack{\text{other}\\\text{cells}}} \frac{G(d)}{G(r)}\,dx\,dy$$
$$= \frac{1}{A_C} \iint\limits_{\substack{\text{other}\\\text{cells}}} 10^{c/10\cdot(r^2-d^2)/R^2}\,dx\,dy\ . \qquad (18)$$

Because the earth curvature is neglected, the distance of an interferer from nadir is

$$d = \sqrt{x^2+y^2} \qquad (19)$$

and the distance of an interferer from its own cell center is

$$r = \sqrt{(x-x_0)^2+(y-y_0)^2} \qquad (20)$$

with (x_0,y_0) designating the respective cell center.

Assuming a hexagonal cell pattern and partitioning the integration area according to Fig. 3, the integrals can easily be solved analytically.

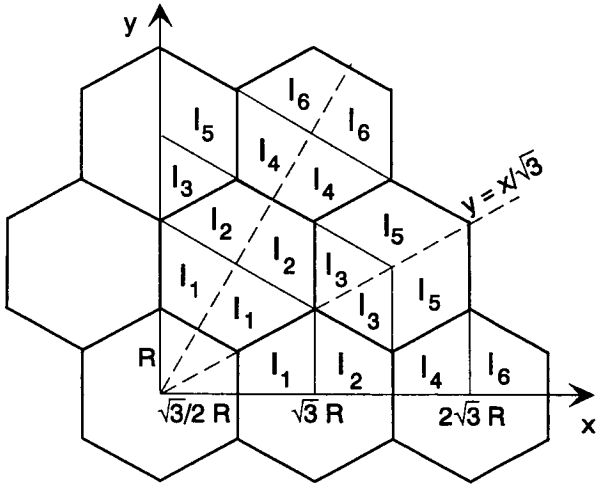


Fig. 3: Partitioning of the integration area for the evaluation of other-cell interference.

For c = 3 dB, the resulting integral values are:

$I_1 = 0.2438 \cdot R^2$
$I_2 = 0.0379 \cdot R^2$
$I_3 = 0.0096 \cdot R^2$
$I_4 = 0.0020 \cdot R^2$
$I_5 = 0.0013 \cdot R^2$
$I_6 \approx 0.$

As expected, the most inner areas yield the main contributions, whereas third-tier cells can be neglected.

With cell area

$$A_C = \tfrac{3}{2}\sqrt{3}R^2 \qquad (21)$$

the normalized other-cell interference is given by

$$k = \frac{12}{A_C} \cdot \sum_{\nu=1}^{6} I_\nu = \frac{8}{\sqrt{3}} \cdot 0.2946 = 1.36\ . \qquad (22)$$

Due to the involved approximations, (18) is a lower bound for other-cell interference, which is independent from satellite height h and spotbeam beamwidth θ_c.

For a tapered-aperture antenna, the evaluation of k according to (18) was performed numerically, resulting in the values given in Table 1.

Table 1 shows that for c = 3 dB interference is more than doubled by the influence of other spotbeams, leading to a more than 50 % decrease of system capacity, cf. (1). Introducing a 3-cell cluster FDM scheme would substantially reduce other-cell interference, but the capacity gain due to reduced interference would be more than outweighed by the capacity loss due to FDM. In principle, interference cancellation taking into account own- and other-cell interferers can reduce interference without decreasing system capacity.

INFLUENCE OF CHANNEL

The above results were derived assuming ideal channels between (mobile) users and satellite. As mentioned before, the results also hold, if multipath fading and shadowing are perfectly compensated by power control. Power control of an interferer impacts its wanted receive signal in the same way as the interference produced by him, because the interferer's signal is received in its own beam and in the victim beam via the same channel. Therefore, perfect power control can eliminate the effect of fading and shadowing on CDMA uplink interference. Power control can, however, increase interference to another satellite if it is in line-of-sight to the user, while the own satellite is shadowed.

As pointed out in [5], power control may compensate for slow shadowing, whereas an additional power increase may compensate for degradation due to fading being too fast for power control. This additional power increase results in an increase of multiple access interference.

Scenario	Antenna	k		
		c = 3 dB	c = 4 dB	c = 5 dB
Lower bound for small nadir angles	tapered-aperture	1.20	0.85	0.65
	Gaussian	1.36	0.97	0.74
MEO-system (h = 10 354 km, 169 spotbeams, ε_{min} = 10 deg)	tapered-aperture	1.22	0.86	0.66
	Gaussian	1.38	0.98	0.75
LEO-system (h = 1414 km, 19 spotbeams, ε_{min} = 10 deg)	tapered-aperture	1.63	1.08	0.78
	Gaussian	1.79	1.18	0.87

Table 1: Numerical values for other-cell interference factor k.

In mobile systems, strong signal shadowing may occur, which realistically can not be compensated by power control. The resulting signal outages decrease the received interference power. However, due to the level variations, the wanted channel requires a higher mean SNR for a given quality of service. Overall, it is expected that the performance degrades.

Usually, the time-share of shadowing increases with decreasing satellite elevation [6]. This means that the nadir beam will receive lower other-cell interference than an outer beam. For a more detailed investigation of this matter an elevation-dependent channel model, e.g. [7], should be used.

For the (less critical) CDMA downlink, the situation is quite different, because the way of an interfering signal from the satellite to the victim terminal is different from its way to the wanted user.

CONCLUSIONS

It has been shown that other-cell interference in satellite CDMA uplinks is much more critical than in terrestrial cellular systems. This should be a helpful information for the comparison of CDMA and TDMA for (mobile) satellite systems.

REFERENCES

[1] **A.M. Viterbi and A.J. Viterbi,** *Erlang capacity of a power controlled CDMA system,* IEEE J. on Sel. Areas in Comm., vol.11, pp. 892-900, Aug. 1993.

[2] **A.J. Viterbi, A.M. Viterbi and E. Zehavi,** *Other-cell interference in cellular power-controlled CDMA,* IEEE Trans. on Commun., vol. 42, pp. 1501-1504, Feb./March/April 1994.

[3] **C. Caini et al.,** *A spectrum- and power-efficient EHF mobile satellite system to be integrated with terrestrial cellular systems,* IEEE J. on Sel. Areas in Commun., vol. 10, pp. 1315-1325, Oct. 1992.

[4] *Satellite antenna patterns in the fixed-satellite service.* CCIR report 558-3 (MOD F), Nov. 1989.

[5] **P. Monsen,** *Multiple-access capacity in mobile user satellite systems.* IEEE J. on Sel. Areas in Comm., vol.13, pp. 222-231, Feb. 1995.

[6] **E. Lutz et al.,** *The land mobile satellite communication channel - recording, statistics, and channel model.* IEEE Trans. on Veh. Techn., vol. 40, pp. 375-386, May 1991.

[7] **H. Bischl, M. Werner and E. Lutz,** *Elevation-dependent channel model and satellite diversity for NGSO S-PCNs.* Proc. of the Vehicular Technology Conf. (VTC '96), Atlanta, 1996, pp. 1038-1042.

[8] **R.E. Collin,** *Antennas and Radiowave Propagation,* New York: McGraw Hill, 1989.

Session 4
Demand, Economics and Technology Issues

Session Chairperson—*Miles Sue*, Jet Propulsion Laboratory, USA
Session Organizer—*Robert Kwan*, Jet Propulsion Laboratory, USA

Prediction of the Potential Demand for Mobile Broadband Satellite Services

Y.F. Hu and R.E. Sheriff
University of Bradford,
Richmond Road, Bradford, West Yorkshire, United Kingdom, BD7 1DP
Phone: +44 1274 384031 +44 1274 384053 Fax: +44 1274 391521
e-mail: y.f.hu@bradford.ac.uk r.e.sheriff@bradford.ac.uk http://manuel.brad.ac.uk/Research/SMG/

Abstract

This paper presents the method used to derive the potential demand for land mobile broadband satellite services. The model considers European demand at national level. In deriving the results, consideration is given to the take-up rate of existing mobile services, the effect of terminal price on the market penetration, as well as the use of databases containing demographic, economic and telecommunications data.

Introduction

A number of new satellite systems have recently been proposed with target operating frequencies in the Ka band or above. At these higher frequency bands, satellites are able to provide services of much wider bandwidth than existing or proposed systems at L or S bands. As a consequence of utilising higher transmit frequencies, a variety of new types of service are possible and new markets for satellite communications will be created. It is envisaged that future mobile-satellite networks will play complementary roles to their terrestrial counterparts. Hence, mobile-satellite services will address market niches where terrestrial networks cannot provide services cost-effectively, or where the terrestrial networks cannot be deployed. As with any other type of service, an initial estimation of the gross potential market and subsequent demand for such services needs to be derived. This paper presents a method for deriving the potential demand for land mobile broadband satellite services. Results are then presented, illustrating the potential market demand in Europe.

Overall Methodology

In order to forecast the market for a mobile broadband satellite system (MBSS), it is necessary to forecast the demand for the different types of terminals and services supported. The different characteristics of these terminals, which support different types of users and services, provides a framework for the methodology of predicting the market demand.

The first step in predicting the potential market is to identify the services and the terminal types upon which the market forecast is based. Two major types of terminals have been identified: portable and mobile. The major distinction between the two types is their level of support in providing user mobility. As far as applications and services are concerned, both types of terminal can support almost all the same services/applications, however, with different quality of services. In this paper, only mobile services will be considered. Taking this into account, the provision of fixed services to complement the terrestrial fixed network will be excluded from the market analysis.

The second step is to identify the gross potential market for such services/terminals. A range of portable and mobile terminals to provide services for individual and group usage is envisaged. The price and, hence, the potential market for each type of terminal will be different.

Once the terminal types and their supported services have been identified, data is then collected on related existing services and products. The data is used to determine the growth and take-up rate of the services. Although it is envisaged that in this multimedia era, numerous new services will be created, broadly speaking, such services can in fact either be regarded as substitutions for, or enhancements to, existing services. For example, colour-fax or fast-fax can be treated as a substitute for an existing fax machine. As far as mobile services are concerned, most of the applications/services are supported by means of cellular terminals. Thus, the market forecast for mobile broadband terminals is based on statistical analysis of the historical trend in cellular mobile phones, rather than being developed for particular type of services. Furthermore, the economic factors that determine the market for each terminal type, such as the price of the terminal, the GDP per capita, the call-charge per minute and the annual subscription fees, are also used.

Finally, the gross potential market and the take-up rate of the terminals are combined to produce a final market forecast.

Mobile Broadband Terminals

For the purpose of market analysis, mobile broadband terminals have been categorised into individual and group types. Table 1. summarises the classification of the

terminals used in deriving the market models, which are broadly based on a subset of those proposed for the European Union's ACTS project SECOMS [1].

For type-A terminals, services of bit rates up to 128 kbit/s are considered; and for type-B, services up to 512 kbit/s are considered.

Terminal	Classification	Use	Type
Portable	PORT-A	Individual	Lap-top
	PORT-B	Individual	Briefcase
Mobile	MOB-A	Individual	Car
	MOB-B	Individual /Group	Trains, Buses, Trucks

Table 1. Classification of Terminals Used in the Market Analyses

FORECAST MODEL

Overview
The potential size of each user group's market depends on several factors:

a) The profitability;

b) The economy of the region. This is often characterised by the GDP or GDP per capita of the region;

c) The market penetration and take-up rate of the service;

d) The tariff of the provided service;

e) The quality of the provided service (QoS).

The last of the above factors (QoS) does not provide a quantified measure and will therefore not be taken into account in the market prediction.

In predicting the market size, several assumptions have been made:

a) The GDP is evenly distributed over a country;

b) The population is evenly distributed over a country;

c) Areas served by existing telecommunications technologies, terrestrial wireless or cellular, were excluded;

d) The size of the satellite market in the remaining areas was assumed to depend upon the same underlying relationship as that for the terrestrial;

e) The satellite market in the remaining areas was assumed to develop at the same pace as the terrestrial wireless/cellular;

f) The terminal usage (in terms of call minutes) will remain static and was estimated at 1000 minutes/year on average.

Profitability
The satellite is expected to have its market in low population density zones, where terrestrial cellular implementation is not profitable. The profitability to provide broadband services depends on the cost of the provision of such services and the penetration of the services, which in turn is determined by the tariff and the GDP per capita. As the cost for the provision of broadband services in the mobile terrestrial network is as yet unknown, it has been assumed that during the initial phase of service deployment, the satellite will aim at providing services in rural areas in developed countries (GDP/Capita > $6000) and in urban areas for other developing countries.

Gross Potential Market
The following gross potential markets (GPM) for mobile broadband satellite services have been identified:

- individual cellular in-fill users: people (pedestrians, car owners, truck drivers) communicating in areas where there is no terrestrial cellular coverage due to geographical or economical reasons or where the terrestrial cellular network has not yet been implemented;
- international travellers: travellers who travel to areas of the world where terrestrial services are poor or non-existent and where broadband satellite type services are available; or business travellers who need to keep in touch by voice, fax, or data or to access e-mail or remote databases;
- public transport users: people who commute via public transport, i.e. buses, trains.

The individual type of terminal has an impact on the GPM as it provides different needs and applications for the users.

Terminals
Portable Terminals: The lap-top type PORT-A terminal has been assumed to support voice, data and e-mail facilities. Potential users include office staff based at a remote location who need to keep in-touch with their headquarters or regional offices. It has been assumed that office staff make up a quarter of the national population.

PORT-B terminals will be catered towards international travellers.

The following table provides a summary of the GPM for each type of portable terminal:

Terminal	Classification	GPM
Portable	PORT-A	Rural office staff
	PORT-B	International travellers

Table 2. Gross Potential Market (GPM) for Individual Portable Terminals

Mobile Terminals: The gross potential market for MOB-A is defined as the number of cars in a country operating outside of terrestrial coverage. The gross potential market for MOB-B is defined as the total number of trains, buses and trucks.

Table 3. provides a summary on the GPM for the mobile terminals:

Terminal	Classification	GPM
Mobile	MOB-A	Cars
	MOB-B	Trains, buses, trucks

Table 3. Gross Potential Market (GPM) for Mobile Terminals

Penetration

Two of the most important steps in market analysis are: i) to determine the growth curve or the take-up rate of the services and; ii) to subsequently determine the penetration of the services. The take-up rate of a service can be characterised by four phases [2]: the introductory phase, the growth phase, the maturity phase and, finally, the decline phase. During the introductory phase, the growth in penetration develops slowly, typically due to: i) the high introductory tariff of the service, and ii) scepticism of potential customers.

Following the introductory phase, the growth in demand tends to rise rapidly due to many factors, including: fall in price, market acceptability, service options, etc. This growth continues until the market reaches maturity.

The length of the maturity phase generally will be sustained before declining due, mainly, to competition from new services.

In [3], several models have been proposed to determine the penetration of the services. Among the proposed models, the Logistic model can most adequately describe the pattern of demand. The equation for the Logistic model is given as follows:

$$P(t) = \frac{M}{1 + \exp(c + dt)} \qquad (1)$$

where P(t) is the penetration at time t,

M is the saturation level

c and d are parameters to determine the growth rate

t is the time variable, e.g. number of years since service introduction

The denominator at the right hand side of equation (1) gives the take-up rate of the new services.

The characteristics of the logistic curve are shown in Figure 1.

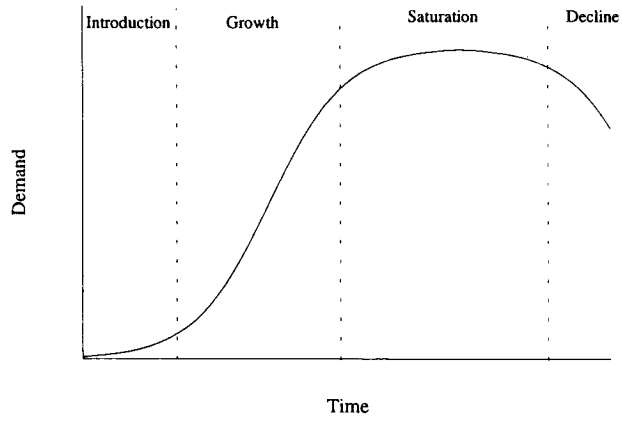


Figure 1. Characteristics of the Logistic Model

One key factor governing the penetration of the market is the GDP (Gross Domestic Product) per capita. This reflects the economic situation of the country. Another factor is the tariff of the service. These two factors determine the affordability of the population in a country to pay for the new services. The affordability to pay, *A*, is defined as

$$A = \frac{GDP\ /\ capita}{tariff} \qquad (2)$$

The terminals offered by a MBSS can be regarded as complementary to the terrestrial cellular mobile terminals. The market penetration of MBSS terminals is determined using historical data for cellular telephones.

In order to provide an estimate for the saturation value, *M*, in equation (1), a relationship between the penetration of the cellular mobile phone and *A* has been derived by gathering data from more than 200 countries for the

income from the cellular services and the number of cellular mobile communications subscribers. Regression analysis is then carried out on the data.

The saturation level is then estimated on a country by country basis. Thus, by expressing M in terms of A, equation (1) can be written as:

$$P(t) = \frac{aA^b}{1+\exp(c+dt)} \quad (3)$$

It has been assumed that the duration between the launch year of the MBSS and year of saturation is 15 years. The price fall of the terminals and the growth in GDP per capita have also been taken into account when forecasting the penetration.

Total Market

From the discussion in the previous sections, the total market can be derived as follows:

$$T(t) = \Sigma[T_{i,port}(t) + T_{i,mobile}(t)] \quad (4)$$

where

$T(t)$ is the total market

$T_{i,port}(t)$ is the market for portable terminals

$T_{i,mobile}(t)$ is the market for mobile terminals

$T_{i,port}$ can be expressed as follows:

$$T_{i,port}(t) = 1/4 \times Pop_i \times rur_i\% \times p_{i,PORT\text{-}A}(t) + Trav_i \times p_{i,PORT\text{-}B}(t)$$

where

Pop_i is the population in country$_i$

rur% is the rural population percentage

$p_{i,PORT\text{-}A}(t)$ is the penetration of PORT-A portable terminals in country$_i$

$p_{i,PORT\text{-}B}(t)$ is the penetration of PORT-B portable terminals in country$_i$

$Trav_i$ is the number of travellers in country$_i$

$T_{i,mobile}(t)$ can be expressed as follows:

$$T_{i,mobile}(t) = cars_i \times rur_i\% \times p_{i,MOB\text{-}A}(t) + (buses_i + trains_i + trucks_i) \times rur_i\% \times p_{i,MOB\text{-}B}(t)$$

where

$cars_i$ is the number of cars in country i

$buses_i$ is the number of buses in country i

$trains_i$ is the number of trains in country i

$trucks_i$ is the number of trucks in country i

rur% is the rural area percentage

$p_{i,MOB\text{-}A}(t)$ is the penetration of MOB-A mobile terminals in country$_i$

$p_{i,MOB\text{-}B}(t)$ is the penetration of MOB-B mobile terminals in country$_i$

Data Sources

Most of the data obtained to date are mainly from the United Nations [4], the ITU [5] as well as from the EU's databases[6]. Table 4. identifies the data used for the prediction of traffic for each user group:

RESULTS

Tariff

This section presents the results of the market prediction analyses. As mentioned previously, the pricing of the terminal and the service tariff can have a significant effect on the potential market for a service. Indeed, the success of a product is closely related to the tariff charged. Tables 5. and 6. illustrate the tariffs used in the analysis. In selecting prices, the market rates for current or near market mobile-satellite receivers were taken into account.

Potential National Markets

The following graphs present the demand for MBSS terminals on a national level in Europe:

- For type-A terminals, the major markets, in order of magnitude, are predicted to be found in: Italy, Germany, Spain, Russia, UK, Austria and France;
- For type-B terminals, again Italy has the greatest predicted market, followed by Germany and Spain.

The graphs show that for type-A terminals, the mobile market (cars) dominates. This situation is reversed when higher data rate type-B terminals are considered, where the portable terminal market is predicted to dominate.

CONCLUSION

The methodology for deriving the market potential for land MBSS type terminals has been presented. Subsequent results show that vehicular type-A terminals command the largest share of the market. Portable terminals dominate the market for type-B. As far as demand at national level is concerned, the main market players are predicted to be Italy, Germany, Spain, UK, and Russia.

Terminal Type	Data for Prediction of the Number of Potential Subscribers	Number of Countries Used
Portable	1) Population	225
	2) Rural area population %	225
	3) Gross domestic product (GDP)	225
	4) Surface area of the region	225
	5) Number of mobile communications subscribers	216
	6) Population of cities over a million	58 (limited to low GDP nations)
	7) Income from telecommunications services	194
	8) International travellers - number of outbound airline travellers per annum	146
Mobile [7]	1) Number of cars	22
	2) Number of buses	22
	3) Number of trucks	22
	4) Number of trains	22

Table 4. Data Used for Market Prediction

	Tariff ($)	
	MOB-A	MOB-B
Handset	1700	1700
Monthly Subscription	100	25.0
Charge/Call Minute	2.4	3.0

Table 5. Mobile Terminal Tariff Rates

	Tariff ($)	
	PORT-A	PORT-B
Handset	3000	3000
Monthly Subscription	100	25.0
Charge/Call Minute	2.4	3.0

Table 6. Portable Terminal Tariff Rates

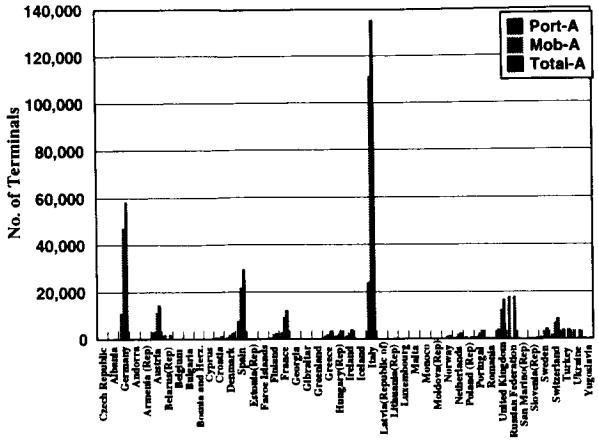


Figure 2. Potential Number of Type-A Terminals per Nation

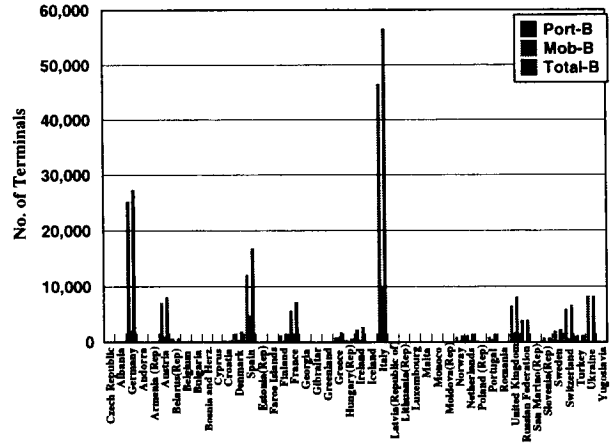


Figure 3. Potential Number of Type-B Terminals per Nation

REFERENCES

[1] http://manuel.brad.ac.uk/Research/SECOMS/

[2] **M.J. Baker**, *Marketing New Industrial Products*, The Macmillan Press Ltd., 1975.

[3] **CCITT**, *Telephone Network and ISDN - Quality of service, Network Management and Traffic Engineering*, Volume II, IXth Plenary Assembly, Melbourne, 14-25 November 1988.

[4] *Demographic Yearbook*, 42nd Issue, UN.

[5] *ITU Statistical Yearbook*, 1994.

[6] *Eurostat Data*, European Commission.

[7] **The UK Dept. of Transport**, "Transport Statistics Report: International Comparison of Transport Statistics 1970 - 1993", Govt. Statistical Service Pub.

Market Perspectives and Business Potentials for Multimedia Satellite Services in Europe

Mario Sforza
Elsacom
Viale Maresciallo Pilsudski, 92 - 00197 Rome, Italy
Phone: +39 6 80902311 FAX: +39 6 8088104
e-mail: sforza@elsacom.finmeccanica.it

ABSTRACT

A large number of regional and global multimedia Ka-band satellite systems has been to date presented in the USA and in Europe. Numerous technical papers have been filed at the level of national administrations and regulatory bodies (e.g. FCC, ITU, Ministries of PTT) to secure orbital positions and spectrum allocation. Multimedia services offered by constellations of satellites orbiting the Earth at different altitudes could soon become the new marketplace frontier and a challenging playground for space industry, satcom operators and service providers around the world. However there are few elemental but key questions which financial entrepreneurs and business decision makers have still difficulties to find an answer for. Where is the multimedia satellite market? How large is it? How this market will be potentially segmented and develop in time? After a preliminary discussion on multimedia and convergence, these basic issues will be addressed with specific emphasis on the European multimedia scenario.

CONVERGENCE AND MULTIMEDIA

Convergence is one of those buzz words which has recently spurred the interest of as many different players as opinion leaders, industry managers, business entrepreneurs, politicians. As a word, convergence has been used in many contexts such as:

- multimedia mass market;
- Information Society or Information Superhighway;
- Global Information Infrastructure (GII);
- interactive television;
- Internet or on-line services.

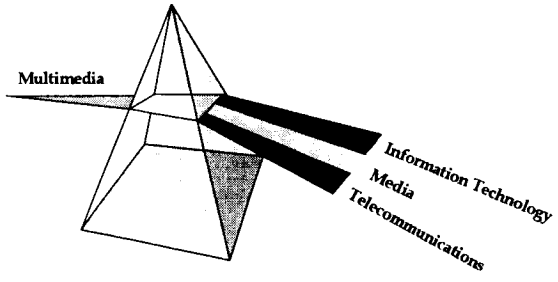


Fig. 1 Convergence and multimedia

Multimedia and convergence will involve many industries, from the telecommunications to the information technology, from the consumer electronics to the media world and to the retail and financial services markets.

The development of the multimedia market and of the convergence will be most likely influenced by some major drivers which can be classified in four categories:

1. technology;
2. market and services;
3. economics and industry structure
4. regulatory aspects

Technology.

Technology is by all means one of the fundamental enabler of the future multimedia market. There are current trends in technology which will have a dramatic impact on the convergence scenario: a) digitisation and compression; b) improved bandwidth; c) two-way, point-to-point switching networks; d) merging of broadcast and server technologies; e) CPU processing power; f) user interface and device; g) encryption. In general, the technology required to support multimedia services and applications will become cheaper and cheaper thereby allowing a broader market to develop. In addition, broadband communication channels will be implemented over a large variety of transportation media (e.g. fibres, coax cables, satellites). For satellite-related technology, time has come for Europe to consolidate a world-wide leadership in some technological key areas of the space industry such as Ka-band terminals and regenerative satellite systems, small satellite design, launchers, optical inter-satellite links, etc. In order to ensure and possibly increase a competitive advantage with respect to other non-EU industries, investments will be needed for specific technological areas of expertise.

Market and services.

Advances in technology will create new services and will enable service providers to offer their existing products at lower prices. Multimedia services will be more easily targeted to specific market segments allowing end-users to interact with the service providers to generate new delivery mechanisms and methods of consumption. There are early but stable indications that people are beginning to change their buying behaviour for audio-visual and multimedia

products and services. We reckon, for instance, that it took WWW (World Wide Web) just four years to reach the 10 million subscribers threshold, much less than other technologies and services (see Fig. 2).

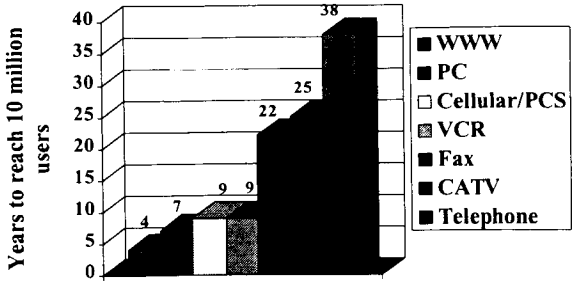


Fig. 2 Pace of development of ICT technologies

Economics and industry structure.
Each of the traditional industries of telecommunications, IT and media will be impacted by convergence and the technology which will eventually support it. Markets are already changing with visible effects on the industry structure. In order to see how the latter may develop, we have considered, though at a very general level, the value chain which could emerge as a result of the convergence process. A simplified version of the multimedia value chain model, based on the ETSI (the European Telecommunications Standardisation Institute) concept [1] is reported in Fig. 3

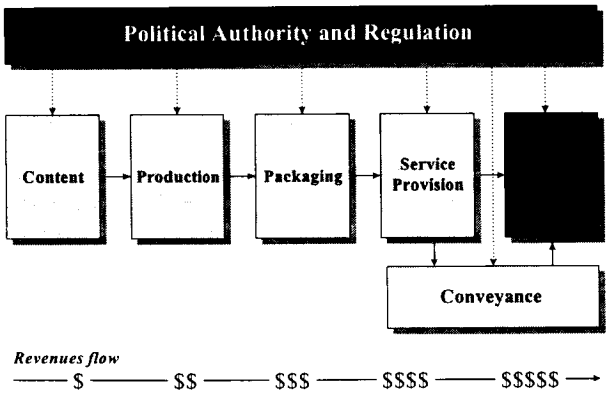


Fig. 3 The multimedia value chain

With respect to similar models well established for the telecommunications and media industry sectors, the conveyance function in the emergent multimedia market could potentially be found not on a straight line between the end-user and the service provider. This basically reflects the fact that how the information and the content is transmitted and conveyed will become more and more irrelevant for the consumer, whose prime (and most valuable) relationship will be with the service provider. Such new element shall be thoroughly understood and evaluated by terrestrial and satellite carriers which could probably see their own margins partly eroded by other players who will have the control of the content and of the consumer base.

Regulatory framework.
Similarly to what already happened for the global mobile satellite initiatives (e.g. Globalstar, Iridium), regulation is not seen by the European market players as a major driver of development. In fact, it is perceived as a barrier to convergence rather than a stimulus. There is a number of issues relating to regulatory aspects which will certainly need specific effort at the level of the European regulatory and standardisation institutional bodies:

- spectrum allocation;
- licensing;
- new services classification;
- public service broadcasting;
- universal service obligations;
- competition;
- standards;
- line of business restrictions;
- accounting separation.

A comprehensive discussion on each single item previously listed is out of the scope of our paper. There are however some basic considerations which deserve specific attention. In general, the development of new industrial activities and service markets will greatly benefit from the definition of a unique European regulatory platform for space-related projects, both for standardisation and RF spectrum management issues. A successful agreement with the US and Asian counterparts, and a full deregulation of the single European telecommunication markets and networks will be other vital elements which could enable the European industry to gain maximum momentum from the new emerging multimedia initiatives. There will be very limited market growth without a co-ordinated deregulation process involving all European countries.

THE EUROPEAN ICT MARKET

Whilst the US market is presently in the midst of a virtual bandwidth feeding frenzy, most important European countries still lag behind a similar pace of expansion. The developments trends witnessed in the ICT and the emerging multimedia US markets will not be immediately followed by Europe as a whole. Unlike the US, most European countries display quite different economical situations, cultural tendencies and buying behaviours. The overall ICT European market turn-over was, at the end of 1996 (see Fig. 4), worth of an overall US$ 365 billion (32% of the global ICT market, compared with the 34% covered by US). Germany, France, Italy and the UK control themselves 70% of such market but with non-homogenous on-line household penetrations and with TV,

PC and telephone density figures quite different from those of the US market.

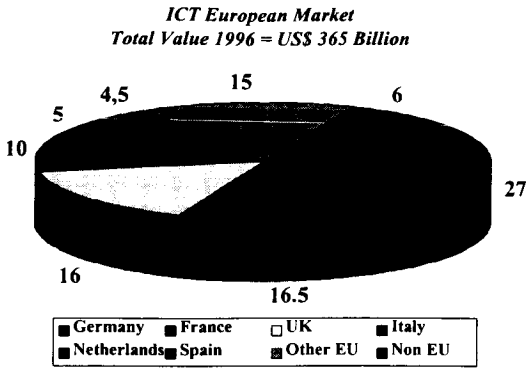


Fig. 4 The ICT European market (%) [EITO]

PC penetration and potentials for on-line services.

As shown in Fig. 5, one striking element which differentiates the US from the four European markets considered is the penetration of PCs. For a thorough evaluation of the future development of on-line consumer

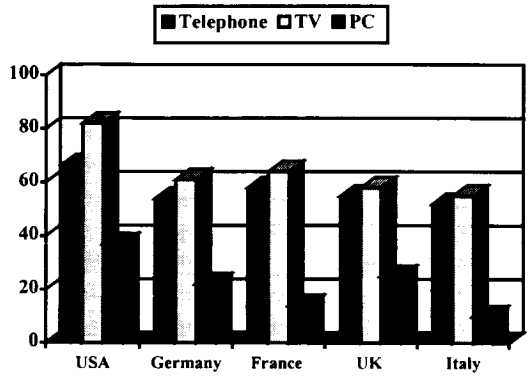


Fig 5 Penetration rates of ICT equipment, 1995 [ITU]

services in Europe, the number of PC households needs to be established and carefully examined. Most analysts agree that the two factors still inhibiting consumer demand for fully specified PCs in Europe are cost and lack of native language software and applications. Though the cost differential between the US and European markets has dropped in the last three years, it is possible, in US computer shops and retail chains, to buy a multimedia PC for about two thirds the price of the same item in Europe. The lack of native language in CD-ROMs of large diffusion and the fact the World Wide Web is still largely dominated by American presence (e.g. nearly 70% of Internet hosts are located in the USA) is the other important factor limiting the penetration of PC in European households. Of the four countries taken into consideration in our paper, it has been estimated that the UK and Germany will be approaching a 45% PC penetration around the year 2005 (similarly to the US market), with Italy and France around 32%. Such take-up rates are also found for the number of on-line households per TV households (Fig. 6), a key indicator to better understand the market and business supporting the convergence between media and Internet-related applications.

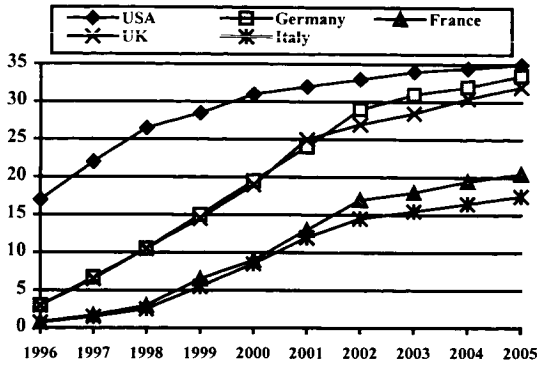


Fig. 7 On-line households per TV households (%) [Kagan World Media]

The digital TV revolution in Europe.

The digital revolution in the European TV market, either via satellite or cable, will be basically driven by the opportunity for service providers to offer pay-per-view services (e.g. movies and sport events). To date and for several reasons not entirely related to technological considerations, cable and satellite networks have shown in most European countries very different penetration curves. The recent progress in the definition of a unique standard platform such as DVB/MPEG should most likely provide new momentum to network operators, terminal manufacturers and service providers. A market could emerge where together with digital TV reception new multimedia services could be delivered to the European household ([2]). The development curves for digital TV households in the four European countries considered are reported in Fig. 7. The Italian market will be probably dominated by digital TV via satellite whilst mixed and maybe integrated networks will take place in the other countries.

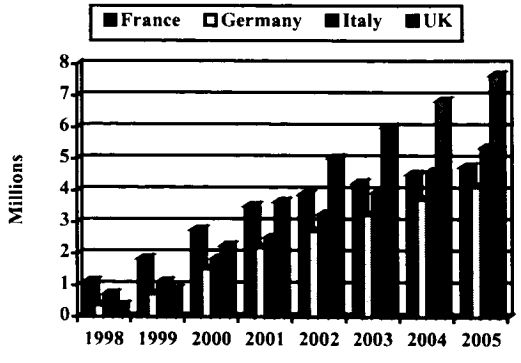


Fig. 6 Western European digital TV households [ITU]

MULTIMEDIA AND SATELLITE DELIVERY

The emerging multimedia applications (e.g. Internet-related on-line services, videoconferencing, E-banking, tele-work, tele-medicine and so forth) and the networks which will deliver these services will be all characterised by a combination of the following distinctive elements:

- broadband
- interactivity
- digital technology
- integration.

The large majority of the IT products and applications have developed so fast in the last ten years to outstrip the conventional bandwidth delivery methods on wireline networks. This trend, clearly shown in Fig. 8, is becoming a major road block for any further progress toward the information society of the next millennium.

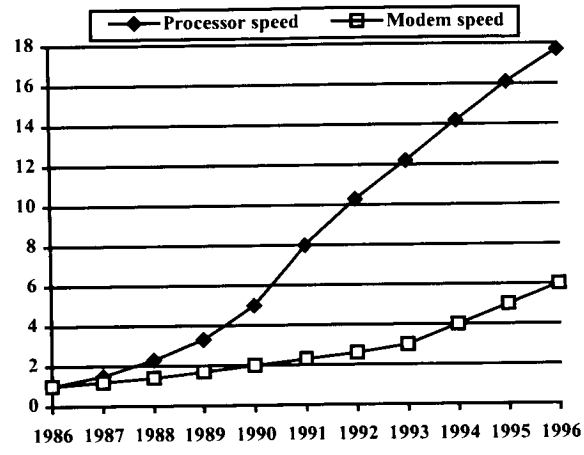


Fig. 8 Relative increment in processor and modem speeds

The digitisation process, began in the computer industry in the early 60s and well-advanced in the telecommunications industry, is now taking place also in the media and broadcasting sector. In the world of convergence, the information bits shall be able to flow from any source to any destination. Such progress will be achieved only with digital technology and some form of intelligent switching capability.

The next generation of satellite systems will be inherently able to respond, in a very cost-effective manner, to the requirements previously presented and discussed. The key technological drivers of the future broadband satellite systems will be:

- use of Ka-band (20-30 GHz) hence the opportunity to deliver information at very high data rates with very small receive antennas (40-60 cm diameter);
- very large bandwidths made available at Ka-band (several Gbytes/satellite);
- on-board processing and multi-beam antennas therefore the possibility to route dynamically the single information packet in any point of the satellite coverage;
- more reliable power amplifiers and more powerful solar cells batteries to increase the average useful life of the satellite (15 years) then optimising the operating costs and reducing the cost per MHz to the end-user.

In the effort to provide an answer to the existing lack of bandwidth, the satellite will however not represent the only available solution. Cable and optical fibre networks will be able to offer even larger bandwidths. The intrinsic satellite features shall be then fully exploited to capture a profitable segment in the future multimedia market. They are basically:

- the capability to deliver services and to broadcast information in very large geographical coverage areas (a truly global network)
- the (large) bandwidth delivered directly at the single end-user premises (the problem of the last mile)
- a very cost-effective and fast service roll-out phase, thus enabling network operators in countries with poor or obsolete telecommunications infrastructure to optimise their investments in a much shorter time than with terrestrial solutions.

EUROPEAN MARKET POTENTIALS FOR A MULTIMEDIA SATELLITE SERVICE

Elsacom, a company of the Finmeccanica/Elsag Bailey group, has carried out a number of in-house and external studies to better understand the potentials of the European multimedia market for a satellite-based initiative. Elsacom is presently working together with Alenia Aerospazio and Elsag Bailey, also of the Finmeccanica group, to define the business idea and the system architecture of the first European geostationary multimedia satellite project, called EuroSkyWay. EuroSkyWay is one of the several Ka-band satellite constellation systems which are presently under study world-wide (e.g. Spaceway, Astrolink, Cyberstar, Teledesic and many others). EuroSkyWay will be a broadband geostationary satellite platform capable to deliver multimedia and interactive contents to the end-users at data rates from 144 kb/s up to 32 Mb/s, with very small size terminals (40-60 cm diameter). The system will allow point-to-point, point-to-multipoint and multipoint-to-multipoint network configurations through state-of-the-art on-board processing and multiplexing matrices. In terms of market segmentation, the potential multimedia services have been divided in three basic sectors:

- the consumer on-line (PC/NC-related)
- the consumer interactive video and video on demand
- the corporate on-line

The consumer on-line services

The European residential PC on-line market segment has grown by over 200% in the last two years; a total of nearly 2.5 million households are connected to on-line services such as Internet, AOL and other similar networks. It is still a relatively small market with the only exception represented by France with its 6 million of Minitel

proprietary terminals. The bulk of the consumer applications (e.g. Internet, video E-mail, on-line gaming, on-line shopping) would certainly benefit from a download speed higher than the conventional 28.8 kb/s modem access, presently used by the large majority of connected residential users. A EuroSkyWay-like satellite platform would therefore be ideally suited to gain a considerable part of this market, expected to grow up to 23 million connected households by 2003. The issue with this kind of applications is that, notwithstanding video E-mail (yet unproved as a killer application), they essentially require not more than a very low data rate for the return path (i.e. the telephone modem up-link would be good enough to cope with the required access speed). Another element to be thoroughly considered and understood is the end-user terminal price. On the basis of elements such as consumer buying behaviour, expected consumer base, multimedia PC price, national average income, competitive technologies and others, the satellite terminal price should not be in excess of 350-500 US$ to have any possible chance of market success. Preliminary conclusions can be therefore drawn:

- the multimedia end-user terminal price will be a major driver
- the high speed return path offered by a satellite multimedia platform might not be perceived as a value-added feature for consumer on-line applications
- digital satellite broadcast services (e.g. Internet multicast, news and papers on-line) using standard platforms such as DVB will reach a large consumer base, when offered in conjunction with digital TV (e.g. Astra and Eutelsat will soon offer this service capability)

The consumer video services

The video and TV broadcasting is a very lucrative market where satellite delivery has already taken a consistent segment and gained large popularity in most European countries. Though with different penetration rates (several countries have fairly large cable networks), digital satellite TV is one of the most promising European market consumer segment. In the early 90s, VOD was set to be the *killer application* for digital TV. After several trial campaigns, most operators have either postponed the start of commercial operations or cancelled the projects entirely. In fact VOD has shown cost/benefit ratios not presently viable in that consumers are not prepared to pay a premium for the extra functionality offered by the service against near-VOD and pay-per-view TV. Another potential application analysed is the video distribution to head-ends. Unfortunately, the majority of the European cable only channels are locally distributed and do not need distribution to cable head-ends via satellite. In addition, and due to a combined effort of rationalisation and optimisation, the number of cable head-ends is set to decrease drastically in the next decade. A limited opportunity could be offered to the future multimedia satellite platform by the film industry. The theatrical distribution and the production process are two areas where the high speed (32 Mb/s and multiples) two-way channel multimedia terminal could represent a valuable working tool. Delivery of high quality movies to cinema screens via digital projectors could become a reality in the next 5-7 years; a satellite platform would drastically cut the distribution costs. As for the production process, satellite broadband up- and return-links could provide higher speed access and improved efficiency in the communication channel between studios, post-production and animation houses. The major problem with both applications is that a relatively slow take-up growth is expected due to the financial investments needed to move from the current terminal analogue technology to the future multimedia digital application. It is therefore considered a niche market.

The corporate on-line services

The corporate market could well exemplify the greatest potential for a future satellite multimedia initiative. The demand for broadband services has grown at a yearly rate of 15% and the corporate data and video communication services market would be probably worth of nearly US $ 60 billion by 2005 in the sole Western Europe. The ability for a satellite multimedia platform to provide a relatively inexpensive high quality, high volume two way communication channel could represent a compelling offer to capture a considerable share of the corporate market. A market where the end-user terminal price is not a big issue due to the fact that such cost is outweighed by the bandwidth usage cost. The corporate market considered in our studies is made of organisations such as: manufacturing, banking and finance, retail, hospitals, transport, utilities, defence, media, governments and others. A sample of the key applications relevant for these corporate sectors is given as follows:

- videoconferencing
- video broadcast
- data broadcast and file transfer
- network interconnect
- Internet and Intranet
- telework
- telemedicine
- disaster recovery

The ability for a future multimedia satellite system to develop its own share in the European corporate market of the next millennium will depend, in our opinion, upon the following elements:

- on-demand (high) bandwidth availability through simple dial-up connection
- cost competitiveness with ISDN and cable/fibre networks but with the inherent capability to offer a 100% market coverage
- a cost-effective and technologically superior return path versus other potential satellite competitors (Ku-band systems, VSAT, DirecPC)
- end-user terminal price and configuration

MARGINS DISTRIBUTION

As already anticipated in a previous section, in the value chain of the future multimedia services business we could witness the raise of new players and the demise of conventional and well-established ones. The distribution of the margins, which are estimated in the range of several billions, could vary according to the type of multimedia application and service considered. Two qualitative examples of estimated margins distribution pie charts are presented in the following figures.

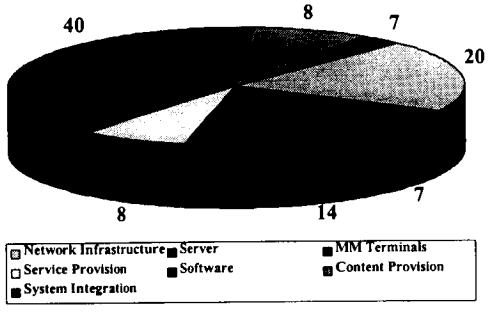


Fig. 9 Margins distribution for on-line corporate multimedia services (%) [Booz-Allen]

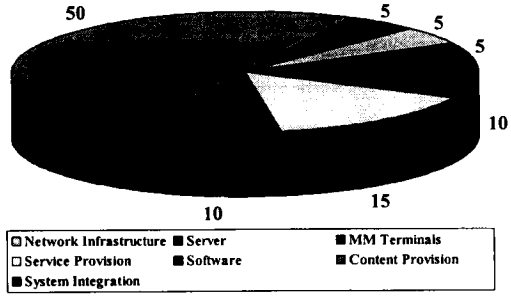


Fig. 10 Margins distribution for on-line/video consumer multimedia services (%) [Booz-Allen]

In the corporate applications and services (est. US 1 billion, in 1996), where the demand for large bandwidth is very high, revenues will most likely concentrate in the technological areas, with the content taking a very minor role with respect to the software needed to run all the different applications. The transport layer (est. 20%), i.e. the network infrastructure segment, could be the second most important strategic area after software (est. 40%), providing satellite initiatives a good background for development.

In the on-line consumer and video multimedia business (est. US 5.5 billion, in 1996), the dominant role could be likely played by those who will create and control the application contents (with an estimated share of 50%). In this sector, the customer relationship with mostly residential consumers will require a quite different service provision structure and strategy (including marketing, billing, etc.) than in the corporate segment. In this multimedia market segment, the network operator will likely have minor chances to play a significant role though, in absolute terms, the overall business could still be significant and worth of further investigations. However the chance for a satellite network to capture a consistent business opportunity could be probably limited, in view of what previously presented.

CONCLUSIONS

Preliminary considerations on the development of a European multimedia market for future satellite systems have been qualitatively presented and discussed in this paper. Emphasis has been primarily given to the understanding of the technological and market trends of the current ICT world. Europe as a whole has been described as a very fragmented scenario where, though very promising, multimedia applications will certainly follow different development curves lagging, in general, behind the US market. Fragmentation will not limit per se the economical success of future multimedia initiatives but will require careful planning and sound marketing strategies.

The future Ka-band satellite-based delivery systems will be inherently able to provide large bandwidths on demand with very small end-user terminals and through flexible and modular network configurations. We believe that, at least in Europe and on the basis of our present forecast and analyses, the principal target for satellite-related multimedia applications will be the corporate market. Such market will be in demand of increasing bandwidth, it will properly value the high speed return path offered by such future systems and it will largely benefit from multinational, possibly global, coverage. The possibility to secure some segments of the on-line consumer market will be strictly linked, in our view, to the possibility to provide a very low cost satellite terminal (less than US $ 500) combined with attractive service offerings bundled in forms and packages which optimally suit each single European culture, language and buying behaviour. The revenues distribution will be likely quite different than in the corporate market (content owners and providers will probably have the leverages to control most of such revenues). A satellite service provider and network operator, such as the newly formed Elsacom of the Italian Finmeccanica/Elsag Bailey group, will have to conceive ad-hoc marketing strategies and establish different partnership to gain a significant market share in both segments.

REFERENCES

[1] *Report of the Sixth Strategic Review Committee on European Information Infrastructure*, ETSI, June 1995

[2] *1995 World Telecommunication Development Report*, ITU, May 1996

Using Mobile Satellites for Fixed or Multimedia Services

Roger J. Rusch, President, TelAstra, Inc.
P.O. Box 4620
Palos Verdes, CA 90274
Telephone: 310-373-1925; Fax: 310-373-5539
Internet: RogerRusch@aol.com

ABSTRACT

The world has a sorry lack of basic telephone service. More than half the people in the world have never talked on a telephone. Now, modern technology has stirred up great interest in universal Internet-like services. This paper addresses the economic, technical, and political factors which will determine the role to be played by mobile and fixed satellite services.

End user cost is a primary consideration in much of the world. The World Bank has estimated that basic telephone services must be provided at end user rates of less than 10¢ per minute. This figure is much lower than the rate anticipated by any of the Mobile Satellite Service (MSS) systems under development, but wider bandwidth, fixed satellite systems should be able to meet this cost target.

The L and S-band frequencies assigned to mobile communications satellites were selected to facilitate communications using relatively small, low gain antennas. These bands are also more tolerant of rain and inclement weather conditions, but, there are technical disadvantages to the lower frequency bands. The narrow bandwidths in L and S-band are limited to low rate voice and data. Use of the higher FSS frequencies at Ku and Ka-band can provide 10 to 100 times more bandwidth, higher capacity, and lower cost services.

Cultural styles and political acceptability will determine the development of these services. Throughout the world there is a wide range of social function and preference that determine usage patterns for communication. Such cultural styles establish the frequency, duration, and content of communications. Some technical and cost decisions depend on these patterns.

The paper describes engineering alternatives and how the design trades determine the service quality and end user costs. The paper quantifies performance factors contributed by the space segment, Earth stations, terrestrial networks and the user terminals. The analysis shows that some satellite constellations offer features which will increase system capacity, reduce airtime service rates, minimize time delay, and reduce the risk of call disconnect.

INTRODUCTION

The ITU Development Sector has sponsored extensive research on the impact of telecommunications on various sectors of the economy and has published several reports that show a strong correlation between economic development and telecommunication investment. Developed countries typically invest a larger percentage of per capita GDP in telecommunications than do developing countries whose focus may be electrical power and roads. Telecommunication in developing countries becomes basic connectivity. Governments (and national telephone service providers) have been compelled to act by the information revolution. Infiltration by the computer culture has had a profound influence on telecommunications.

There is a growing sensitivity to the fact that many people live and work without the benefit of telephones. Efforts to extend terrestrial telecommunications are costly. Investment recovery is uncertain because of light usage in economically disadvantaged regions. Satellites can be a more economical and faster means of providing universal service in rural regions and developing countries.

The developed world has found many benefits from communications satellites. Today we use space based repeaters for television broadcasting, telephone service to stationary terminals, and mobile phone services to ships, aircraft, and land vehicles. Satellites play an important role in expanding the reach of wireless services throughout the world.

Over the past six years, we have witnessed numerous proposals for global communications

networks which could provide needed services directly to end users. A wide range of orbital configurations and frequency assignments have been proposed. Some of the systems are designed primarily to provide MSS, but will also provide FSS. Other systems are designed for fixed service, but may evolve into mobile applications. Recently, a new generation of high capacity "multimedia" satellites has been proposed to provide fixed services using the Ka-band spectrum.

THE ECONOMIC DIMENSION

The direct linkage between economic development and per capita GDP has been understood for a long time. The ITU data [1] on teledensity (Figure 1) clearly shows that more people in affluent countries use telecommunications services. Use of all telecommunications services falls proportionately with the drop in domestic per capita income. The same relationship between telecommunications services and per capita income is also evidenced within countries. Sophisticated communications services are typically employed by the wealthy members of society.

Conventional voice telephony is the most general interactive communications service. The use of mobile communications (cellular) service is unusual in all countries and falls off more rapidly with reduced income levels. The 1994 data shows that access to the Internet is less common than cellular and rare to nonexistent in developing regions.

End user cost is a primary consideration for user acceptance. This is particularly true in the developing world where disposable income is limited. The World Bank has estimated that telecom services must be provided at subscriber rates of less than 10¢ per minute.

Space based services tend to be more expensive than terrestrial services. Satellite communications systems, however, cover huge regions of the world and services can be installed quickly if transponders are in orbit. Ironically, satellites are best suited to serve the people of the world that can least afford service. The task

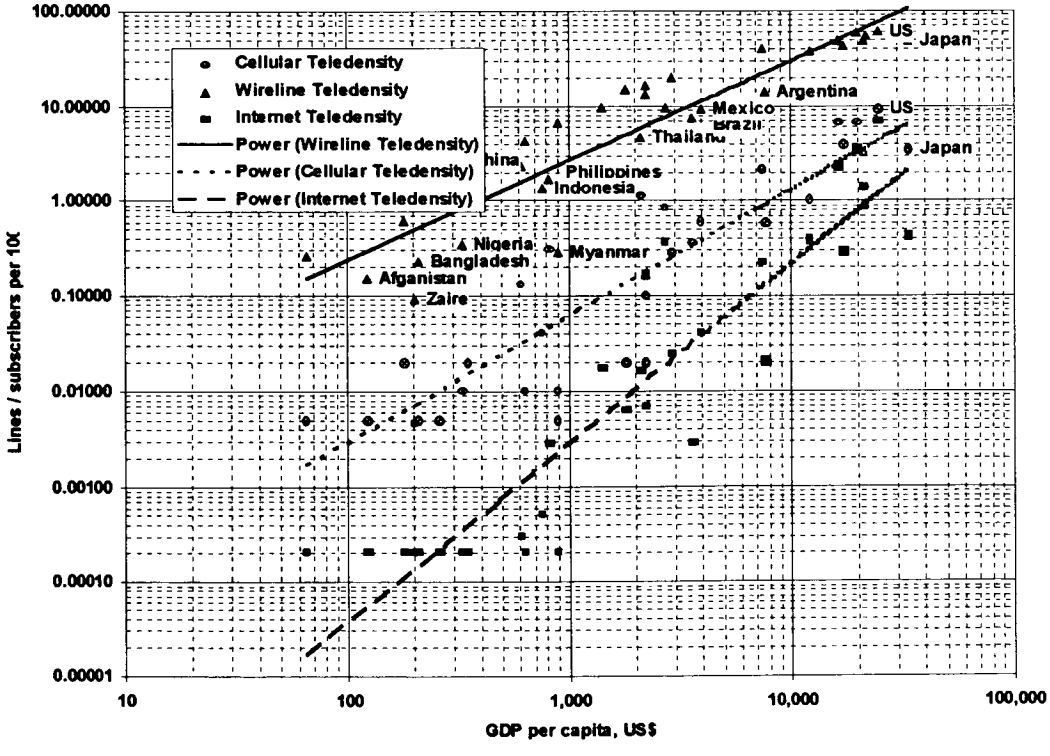


Figure 1. Teledensity and Development

for communications engineers is to drive down the cost of service.

ALTERNATIVES

Figure 1 provides a comparison of MSS and FSS type services based on the best available public information [2]. Capital cost per circuit year is a dominant factor for service costs since investment is made years before revenue is produced. The data shows that there is a basis for lower airtime rates on new FSS systems.

- **Terrestrial Solutions.** The most common communications services are provided by land lines (twisted pairs of copper wire to residences) with local exchanges and switching network centers throughout the world. In the less developed regions of the world, many people cannot afford the service. In other cases, there are vast open regions between users. Providing infrastructure for universal service is extremely costly. Some have estimated that it would cost $6 trillion and require up to 50 years to install at current investment rates.
- **Big LEO or Big GEO "mobile" services** delivered to fixed terminals operating in the L and S-band frequencies. Four systems are under construction to provide mobile services from low or medium Earth orbit. In addition, several satellite systems have proposed to provide mobile service from geostationary orbit. Some of these systems plan to offer fixed services to remote regions. Two of the Big LEO systems, Globalstar and Odyssey plan to provide fixed services at about 50¢ per minute. (Iridium and ICO Global do not propose to provide fixed services.) In order to make effective use of capacity, these systems will use higher gain antennas than the mobile handsets.
- **VSAT delivery using conventional C or Ku-band satellites**. Typical Ku-band VSAT terminals are 1.0 to 1.2 meters (39 to 47 inches) in diameter. These systems provide data access at 64 kbps up to 13.5 Mbps for business customers. Monthly service rates are on the order of $125 to $250 per month including lease of the terminals (In Europe these rates may be 10% to 15% higher). Capital investment for the terminals was initially $15,000 per installation but has dropped to $4000. Currently, however, the total number of terminals are 260,000. In a mass market, production costs would drop. Based on experience in other mass market technologies, a cost of $1000 or less per terminal is within reach if the production volume increases to millions. Cost for the service including lease of the terminal are about 17¢ per minute.
- **Asymmetric satellite services** like DirectPC. Such systems have been called microsats. These systems enable Internet downloading at 384 kbps, but the return link is via telephone modem. DirectPC provides service using a 0.53 meter (21 in.) dish. The satellite antenna terminal and PC card cost about $699 (street price) plus $99 to $200 for installation and $50 for activation.

Program	Capital Cost, US $	Lifetime, years	Fixed voice circuits	Cost per circuit year	Voice rate, bps	User cost per minute	Throughput Gbps
Globalstar	$2,500,000,000	7.5	60,000	$5,556	2,400	~50¢	0.25
Odyssey	$2,700,000,000	12	172,800	$1,302	4,200	~50¢	0.73
INTELSAT 7	$1,215,000,000	12.4	270,000	$363	12,800	17¢*	3.46
INTELSAT 7A	$715,000,000	12.4	112,500	$513	12,800	17¢*	1.44
Astrolink	$3,994,000,000	12	959,375	$347	64,000	~10¢	61.40
M-Star	$6,150,000,000	10	2,591,000	$237	64,000	~10¢	165.82
SpaceWay	$5,171,000,000	15	1,375,000	$251	64,000	~10¢	88.00
VoiceSpan	$4,306,000,000	12	1,125,000	$319	64,000	~10¢	72.00
Teledesic	$9,871,000,000	10	2,000,000	$494	16,000	~10¢	32.00
* includes lease of terminal							

Figure 1. Capital and user costs for Fixed Satellite Service

- **Dedicated interactive satellite services.** The new generation of multimedia satellites [3] has been proposed for service starting in 2000. The Ka-band terminals will be 0.66 m (26 inches) in diameter to enable the higher data rates. The forward transmission rates range from 16 kbps to 6 Mbps with lower return rates of about 384 kbps. Teledesic has said that its satellite terminals will cost about as much as a laptop computer which typically sells for $2000 to $3000 today. Hughes SpaceWay says its terminals will cost $1000. Independent analysis shows that rates can be much less than 10¢ per minute for a telephone circuit

MOBILE & FIXED SERVICE

From a technical perspective, fixed service is vastly different from mobile service. FSS provides nearly ideal conditions for reliable communication. Earth terminals are positioned to avoid blockage by trees and terrain. Only atmospheric conditions must be taken into consideration. In contrast to FSS, MSS must contend with a much more uncertain transmission environment. Terminals, users, and satellites may be moving. Under these conditions the transmission path may be blocked by trees, buildings or terrain. Reflections can cause multipath conditions which can produce fades. Much larger link margins are needed to ensure reliable, continuous communication. Consequently, MSS is more costly than FSS. Hybrid systems may also be more expensive.

FREQUENCY ASPECTS

The L and S-band frequencies assigned to MSS satellites were selected to facilitate communications using relatively small, low gain antennas. These bands also are more tolerant of rain and inclement weather conditions, but, there are cost disadvantages and technical limitations due to the lower frequency bands.

Program	band	bandwidth MHz	satellite MHz
Globalstar	L / S	10	200
Odyssey	L / S	10	300
INTELSAT 7	C & Ku	1,000	2,736
INTELSAT 7A	C & Ku	1,000	3,045
Astrolink	Ka	1,000	9,000
M-Star	Ka	3,000	32,000
SpaceWay	Ku & Ka	1,500	7,080
VoiceSpan	Ka	1,000	8,000
Teledesic	Ka	500	11,905

Figure 2. Satellite spectrum and bandwidth

Bandwidth is a dominant consideration for providing high capacity and high data rate transmissions at low cost. Figure 2 shows the allocated spectrum bandwidth and actual satellite bandwidth with frequency reuse. The narrow bandwidths in L and S-band are limited to low rate voice and data. Use of the higher FSS frequencies at Ku and Ka-band can provide 10 to 100 times more bandwidth, higher capacity, and lower cost services.

TIME DELAY

Long time professionals in the communications industry have been concerned about the quarter second propagation delay from Earth to geostationary orbit and back. This problem is compounded if the call is routed through two geostationary satellites (double hop.) We all recognize that delays cause overtalking and confusion. In fact, designers of the Big LEO satellites selected non-geostationary orbits in order to cut the propagation time delay.

At the same time, the Big LEOs need to expand capacity in order to offer lower cost service. With limited bandwidth, these systems have resorted to extensive speech compression. Even with the most sophisticated digital electronics, there is some degradation in quality due to compression. Furthermore, because these systems operate with vocoders, additional processing time delay is introduced. The time delay for the current generation of vocoders is 80 to 100 milliseconds for compression and expansion on both the forward and return path. When the total processing delay for vocoders is added to terrestrial processing delays, the time delay for the Big LEOs may be as great as for a single hop GEO system.

The next generation of Ka-band "multimedia" satellites have ample bandwidth and could avoid the voice compression. Onboard circuit routing enables a full mesh network and thereby ensures that all calls will be single hop.

USER TERMINALS

Although service rates could be lower for FSS than for MSS, terminals may be more expensive. Today, VSAT terminal and MSS

Program	Terminal type	Antenna	Aperture meters	Terminal cost
Globalstar	cellular	tracking	0.25	$1,000
Odyssey	cellular	tracking	0.25	$1,000
INTELSAT 7	VSAT	stationary	1.20	$4,000
INTELSAT 7A	VSAT	stationary	1.20	$4,000
Astrolink	reflector	stationary	0.66	$1,000
M-Star	array	tracking	0.66	$2,500
SpaceWay	reflector	stationary	0.66	$1,000
VoiceSpan	reflector	stationary	0.66	$1,000
Teledesic	array	tracking	0.66	$2,500

Figure 3. Terminals for direct to user communications

terminals are more costly than cellular handsets. The lower cost for cellular terminals is due primarily to the vast economies of scale that derive from production volumes of millions of units per year. Figure 3 shows a comparison of MSS and FSS terminals. The MSS systems expect to benefit from the large production base of cellular and to offer hand held space telephones at attractive prices.

These systems also hope to build satellite telephone booths. These would be relatively inexpensive stationary FSS terminals which would track the movement of Big LEO satellites. These MSS/FSS terminals are likely to be more expensive than omnidirectional or fixed reflector antennas, however, because of the moving or switching antenna elements. The antennas will only need to be about 0.25 m (9 inches) in diameter in order to provide a gain of 6 dB. Such terminals would enable a much larger number of telephone circuits to be carried by the MSS satellites. Tracking antennas may be more susceptible to failures because of the moving or active elements.

The new generation of Ka-band satellites expects to serve a mass market of FSS terminals for Internet access. These systems will sell relatively small terminals with diameters of 0.66 m (26 in.) which will be relatively easy to install. We presume that these same antennas could be used for receiving television programming.

CULTURAL ISSUES

The whole story does not reside in pure technical tradeoffs, however. There is great divergence in telephone usage patterns. Some cultures are more likely to talk frequently and for long periods on the telephone. Other cultures have not developed this behavior pattern. (One expert notes that the average person in China makes one call per year.) We know cellular phones are less common than voice lines, and we must understand the national demand for data services like the Internet.

Political acceptability is also an important factor which impacts these services. Developing nations insist that their sovereignty be respected and that they be treated as equals by the developed countries that propose to offer global services. Actions taken at the October 1996 World Telecommunications Policy Forum laid the foundation for agreement.

CONCLUSIONS

This paper addresses the effectiveness of providing FSS service using MSS systems. This paper concludes that:

- FSS systems can offer lower airtime rates than MSS systems.
- MSS terminals will be smaller and may be less expensive.

This suggests that the economic decision depends on the amount of usage. If the terminal is used very little, then an MSS solution may be less expensive. If the usage rate is high, then FSS systems will be more economical. Business calls average only 2 minutes in duration and usage is lighter than for residential telephones. Internet users tie up the telephone circuits for extended periods.

The optimum solution will depend on usage patterns and may, in some cases, depend on cultural values. The best answer may vary from country to country. We look forward to the development of all of these intriguing satellite systems and the resolution of the questions discussed in this paper.

REFERENCES

[1] World Telecommunications Development Report, Information Infrastructures, World Telecommunications Indicators, International Telecommunications Union, Geneva, 1995.

[2] Communications Satellite Data Bases, Parts I, II, and III, (12th Edition), published by TelAstra, Inc., 1997.

[3] Financial and Business Evaluation of the New MultiMedia Satellite Systems, published by TelAstra, Inc., 1997.

System Planning Methodology for Satellite Communication Services

Yvonne M. Lazear, Larry H. Capots, Ronald D. Langlais, and Kent M. Price
Space Systems/Loral, Palo Alto, California 94303-4604

Abstract

A systems planning methodology for satcom systems is described and examples are given of possible future geosynchronous satellite designs that incorporate new technologies. Steps in this process include identification of service demand, estimate of economics for services, and selection of technologies including the product development cycle.

1. Introduction

Figure 1 shows a typical system design context, where interactions with management, regulatory agencies, marketing, business planning, and vendors are required in the design of a system. This paper focuses on the interaction among marketing (to identify service demand), business planning (to determine economic viability), and technology (to enable the system design).

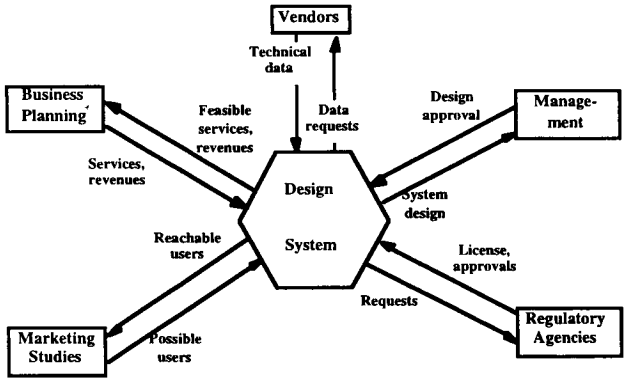


Figure 1: System Design Context

Figure 2 shows the design process flow which has the following steps in order to generate the information needed for the system design. Concurrent engineering techniques are used throughout this process to meet technical requirements for the most economical cost.

1. Write concept of operations
2. Define requirements
3. Perform trades and studies
4. Define System architecture
5. Design segments
6. Build segments
7. Integrate and test
8. Documentation
9. Control System configuration

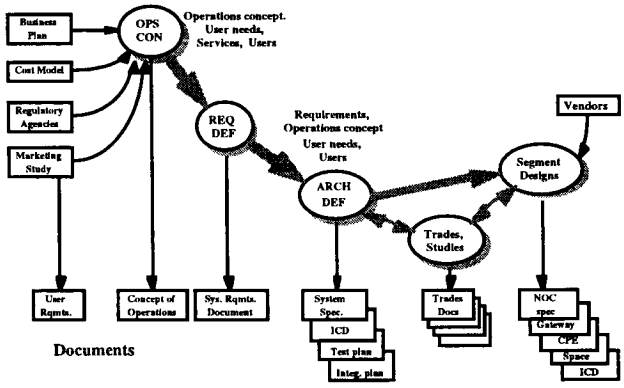


Figure 2: Design Process Flow

2. Service Demand

Market studies are conducted to define customer requirements which consist of service description, quality of service, and price of service. Customers will typically include both buyers of services and sellers of information content. The user requirements are then used to develop a concept of operations, in conjunction with the business plan. **Figure 3** shows the many components of a marketing plan and its interactions. Marketing is dependent on Finance to provide the financial analysis, financial documents and understanding of regulatory requirements to lend credibility to the plan.

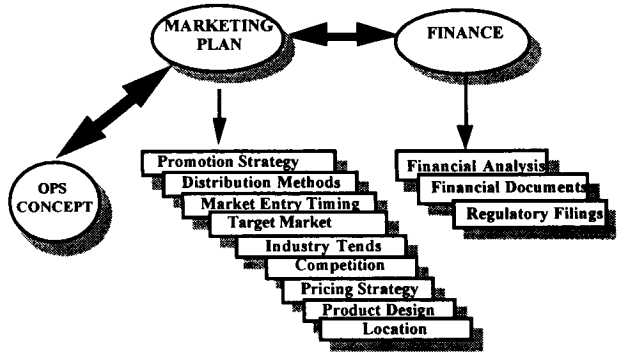


Figure 3: Market Plan Elements

3. Economics of Satellite Communications

A business plan presents information to investors, bankers and other creditors. It also communicates with the people who implement it by setting concrete goals, measurable objectives, and priorities by which a manager can assign tasks, responsibilities, and deadlines, and track results and effectiveness. A business plan also communicates these operations and goals to other interested parties (i.e., employees and suppliers).

Table 1: Business Plan Outline

I. Executive Summary
II. Organizational Plan
III. Marketing Plan
IV. Implementation Plan
V. Financial Analysis
VI. Financial Documents
VII. Supporting Documents

The business plan combines the service demand and the technical feasibility, via a cost model, to produce an estimate of the cost of service. This gives an indicator of business viability by comparing cost with market studies of the price customers would be willing to pay, and price of competing services.

A discussion of the economics of satellite systems is divided into subsections: (1) breakdown of life cycle cost (LCC) by system segment; (2) categories of user costs; and (3) conclusions of satcom economics.

3.1 Breakdown of LCC by System Segment

Table 2 gives the breakdown of LCC for a typical FSS satellite system. The LCC is dominated by the cost of ground terminals and network control (including user billing costs). The total space segment cost (satellite, launch, and spacecraft control) is 36% of the total LCC.

Table 2: Breakdown of Satcom System LCC

System Segment	Percent of Total Cost
Satellite (bus and payload)	18%
Launch (of satellites)	13%
Control (spacecraft)	5%
Control (network)	5%
Ground terminal development and production	35%
System operation and maintenance	24%
Totals	100%

3.2 Categories of User Costs

User costs can be divided into two categories:

- User equipment costs (handset or ground terminal). There may be other costs implicit in the supply of the service (e.g., the existence of electricity, terrestrial phone lines, personal computer, or television receiver).
- System use charges (cost per minute, cost per packet, etc.) for a given type of service. Service could be priced according to data rate, number of packets sent, time of day, quality of service, etc. Each user also has a monthly charge for accounting services.

The system architect tries to minimize the total system cost. This requires a tradeoff between user equipment cost and system use charges. For example, smaller user terminals cost less, but require more of the system resources (bandwidth and/or power) to complete the ccomplete the links. System capacity is thus lowered, and system use charges increase. Ideally, a system supports a mix of small and large terminal services, with higher terminal costs and lower use charges for the larger terminal users.

The system architect tries to minimize both user equipment costs and system use charges. This requires a tradeoff between increasing user equipment size, complexity, and cost in order to increase system capacity and hence lower system use charges. Alternately, the satellite EIRP could be increased, resulting in less system capacity, but reducing user ground terminal size and cost. The existence of large numbers of users pushes the result towards lowering the cost of the ground terminals at the expense of higher system use charges.

3.3 Conclusions of Satcom Economics

The satcom economics conclusions are as follows.

- Technology applications should be judged on a systems level, with impact on user equipment cost and system use charge.
- Ground terminal costs and system operation and maintenance are largest components of system LCC.
- Relatively large, higher power satellites are favored in order to reduce user terminal costs.
- Technologies that increase system capacity or system utilization have a direct impact on system use charge.
- Satellites must be able to interconnect, without adverse cost impacts, with terrestrial wired and wireless networks.

There are a number of ways to reduce user costs, ranging from incremental improvements to complete redesign of system function.

1. Reduce cost and mass of existing hardware by incorporating evolutionary technology advances.

2. Improve the utilization of the system. Systems with flexible architectures (e. g., reconfigurable antennas, on board switching) are more likely to achieve high utilization.
3. Improve the efficiency of transport by minimizing overhead bits. This is unlikely to occur since the trend is to put more intelligence in the network, which results in more signaling and control bits being embedded in the communications traffic. This trend may indirectly improve utilization, but by itself slightly increases user costs.
4. Transmission capacity improvement has the greatest potential to reduce user costs for satellite communication systems. Spacecraft power is the limiting resource, and the user equipment needs a minimum W/m^2 to close the link. If the coverage area (e. g., CONUS) is fixed, the required transmit power is fixed. This is typically the case for broadcast systems with many users. However, for point-to-point communications, power can be reduced by shortening the range (GEO to MEO to LEO) or using a larger gain antenna (larger size or higher frequency).
5. User equipment (ground terminal) cost can be reduced via use of smaller user terminals.

4. Technologies

How technologies are selected is discussed in the product development cycle subsection. What technologies are selected is given in the next subsection.

4.1 Product Development Cycle

The Product Development Cycle (PDC) is a rigorous process used at Space Systems/Loral for the development of new products and technology. The objective of the PDC is to achieve first-time-through developmental success, culminating in full hardware qualification and readiness for full-scale production. When necessary, this process is flexible and remains effective in achieving optimum risk retirement. The ability of the project to endure increased exposure to risk and still attain its risk mitigation objectives, must always be considered when less than full implementation is contemplated.

Figure 4 shows the various developmental phases and stages of the PDC. Specific activities, analyses, and documentation are completed in each stage. Transition between these stages occurs through formal control gate meetings.

PHASES:	PLANNING	DESIGN	QUALIFICATION
STAGES:	STRATEG. MKT. PLAN PRODUCT DEV. PLAN	PRELIM. DES. DETAILED DES.	MFG. READINESS BUILD/TEST
ACTIVITIES:	Requirements Review	Design Analysis	Manufacturing Qual. Test
PRODUCTS:	Design Requirements	Drawings/Specs Tests/Analyses	Qualification Test Report
CONTROL GATES:	Planning Review Bus. Case Valid.	PDR CDR	MRR/TRR QTR
EXIT CRITERIA:	Mktg. Rqmts Met Technic. Feasible	Performance/Cost Sched. Rqmts Met	Qualification Rqmts. Met

Figure 4. Product Development Cycle

Phases and Stages The major phases of the Product Development Cycle are Planning, Design, and Qualification. A "Phase" is the fundamental sequential element of the PDC, and is comprised of a series of developmental stages. A "Stage" is the major subdivision which further characterizes the sequential steps taken within a PDC Phase. It is comprised of a series of PDC Activities and Products which are prepared in support of Exit Criteria for that specific Phase, and culminates in a formal Control Gate review meeting.

The intent of the Planning Phase is twofold. First, it is to identify a requirement for a product that is feasible, both financially and technically, and is consistent with the company's strategic business objectives in the Strategic Market Planning Stage. Secondly, it is to establish the product development and evolution plans, as well as its performance and producibility objectives in the Product Development Planning Stage. Market requirements for this product are also validated with respect to resource constraints at this time. A significant activity in this Phase is a Product Requirements Review.

The object of the Design Phase is to evolve and develop the various aspects of the product within the Preliminary and Detailed Design Stages. This design is based on performance and producibility requirements established in the preceding PDC Planning Phase. These stages culminate in a Preliminary Design Review (PDR) and a Critical Design Review (CDR), respectively.

During the Qualification Phase the product design is evaluated for its manufacturability in the Manufacturing Readiness Stage. The hardware under development is then assembled, integrated, and tested in the Build and Test Stages. The results of qualification testing and feedback from the Build Stage are then analyzed. Subsequently, a determination is made as to the extent of developmental risk mitigated and level of success in achieving performance and producibility objectives.

Activities and Products Each Product Development Cycle stage has a series of operations or tasks, called

Activities, which are performed in support of the Exit Criteria for that stage. The logical yields of these Activities are Products. Products are analyses, reports, drawings, or specifications, which are prepared to satisfy the Exit Criteria for a particular developmental stage. As a design evolves from phase to phase within the PDC, the products associated with each phase evolve and mature as well.

Control Gates and Exit Criteria A formal meeting of all project stakeholders is conducted at the conclusion of each PDC stage. During this time, the results of the PDC activities performed and the PDC products completed are reviewed as to whether the Exit Criteria for that particular stage has been met. Exit Criteria are the development objectives which must be met in order to proceed into the following PDC stage or to conclude product development or qualification.

The control gates for the PDC Planning Phase Strategic Market Planning and the Product Development Planning Stages, are the Product Planning Review and Business Case Validation meetings, respectively. During the Product Planning Review, the company's strategic business plan is reviewed to assure congruence with marketing requirements, engineering and manufacturing feasibility studies, and financial objectives. Among its main objectives are the identification and definition of product evolution trends and design targets.

The Preliminary Design Review (PDR) and the Critical Design Review (CDR) are the control gates for the Preliminary and Detailed Design Stages, as part of the PDC Design Phase. During a PDR the product design concept is evaluated with regard to conformance with predicted performance requirements, based on the results of analyses, modeling, and producibility evaluations. The CDR evaluates the detailed, mature product design with regard to predicted versus measured performance, based on the results of developmental testing, and updated analyses, modeling, and producibility evaluations. Parts, materials, and process qualification are complete by this time. Compliance with higher level architecture requirements, budget, and schedule targets is again reviewed.

The PDC Qualification Phase spans the Manufacturing Readiness, Assembly, and Test Stages. It has as its control gates, the Manufacturing Readiness Review (MRR), Test Readiness Review (TRR), and the Qualification Test Review (QTR), respectively. The MRR verifies that manufacturing drawings, resources, and work instructions are available, and that the design is producible and ready to be manufactured.

The TRR establishes the state of readiness to begin qualification testing. Also, it considers the readiness of the hardware to be tested. The test equipment and software, test facilities, data management arrangements and closure status of design review action items are also reviewed, to determine whether qualification testing may begin.

During the QTR the summarized results of qualification testing, including trend analyses, is presented in the Qualification Test Report which summarizes the results of the actual qualification testing and compares these results with predicted performance values. The results of post test hardware examination, and any deviations from planned test activities, or test criteria are also reviewed at the review meeting. Conclusions obtained from the QTR form a major part of the basis for the final assessment of the full extent of product development risk mitigation achieved.

4.2 Selection of Technologies

A list is given of technologies, roughly in order of impact on economics, that have the potential to influence user costs (user equipment and system use charges). There are three classification; most important, important, and other. Technologies in the "other" category are believed significant, but relative importance was not able to be assessed.

Most important technology areas

- Standards and protocols that are "satellite friendly" and allow seamless, low cost interoperability with terrestrial wired and wireless networks.
- New system architectures to meet user needs via new services. System architectures that lower user costs. Hybrid space–terrestrial communication system architectures.
- On board processing and switching with low mass and power consumption.
- Active antenna technology for spacecraft, gateways, and user equipment. Spacecraft active antennas require high efficiency MMIC HPAs, packaging to reduce cost, and integrated thermal control. Gateways for LEO systems require multiple beams (at least two beams for hand-offs) with wide field-of-view (down to 10° elevation angle). User equipment (at Ka-band) could use a small phased array to track the satellite(s).
- User equipment (ground terminals and handsets) that is compact, transportable, and/or unobtrusive. Lower cost user equipment is required.

Important technology areas

- Higher EIRP and G/T on the satellite to decrease size of user terminal, or to increase user data rate for a given user terminal size. Use of larger satellite

antennas, higher frequency bands, and higher power amplifiers are possible technologies.

- Power generation and storage which is more efficient in terms of mass required for a given capacity
- Launch vehicles with lower cost and higher reliability.
- Network control technologies to manage a number of satellites with interconnected communications services.

Other technology areas

- Intersatellite links with minimum mass and power impact on host satellite.
- Technical support for U. S. positions on international spectrum and orbital assignments. Tools to enable timely analysis and resolution of inter-system interference issues.
- Design for manufacturing of satellites, for systems with many satellites.
- Technologies for efficient orbit and spectrum utilization.

5. Examples

Examples are given of the current Space Systems/Loral (SS/L) GEO satellite technology for 1997, and where GEO satellite technology might be in 1998, 1999, and 2003.

5.1 Current (1997) GEO Satellite Technology

Figure 5 shows the current SS/L GEO satellite product, using the FS-1300 bus. Its features are as follows:

Capabilities

DC power	up to 9.5 kW
Radiated power	2.3 to 2.7 kW
Payload mass	300 to 500 kg
Delivery to orbit	24 months after receipt of order (ARO)

Launch

Atlas 2AS, Ariane IV/V, Proton, Long March, and H-II

Technologies

- Three-axis stabilized
- Graphite central cylindrical structure
- Composite and aluminum structure
- Direct radiating traveling wave tubes
- Lightweight shaped antenna reflectors
- Embedded heat-pipe radiator panels
- Four-panel solar arrays with high efficiency silicon or gallium arsenide cells

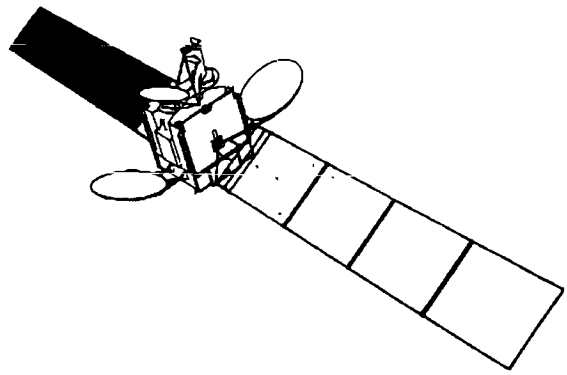


Figure 5: GEO Satellite, 1997

5.2 1998 GEO Satellite Technology

Figure 6 shows the possible 1998 GEO satellite product. Its features are as follows:

Capabilities

DC power	11 kW
Radiated power	3.2 kW
Payload mass	300 to 500 kg
Delivery to orbit	18 months ARO

Launch

Atlas 2AS/2ARS, Ariane IV/V, Proton Long March, H-II and Delta 3

Enabling Technologies

- Medium-temperature (85°C) heat pipes
- High-power solar arrays with improved solar cell efficiency
- Electric propulsion system for north-south stationkeeping
- Advanced attitude control electronics and data handling electronics
- Low cost, lightweight RF electronics, linearized channel amplifiers
- Dual grid-shaped antennas

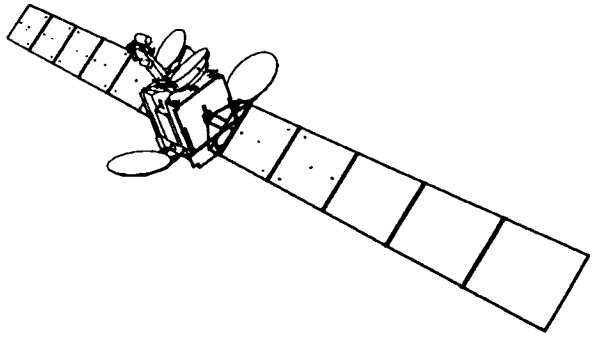


Figure 6: GEO Satellite, 1998

5.3 1999 GEO Satellite Technology

Figure 7 shows the possible 1999 GEO satellite product. Its features are as follows:

Capabilities

DC power	16 to 20 kW
Radiated power	4.6 to 5.8 kW
Payload mass	400 to 1,000 kg
Delivery to orbit	15 months ARO

Launch

Atlas 2ARS, Ariane IV/V, Proton, Long March, H-II, Delta 3 and Sea Launch

Enabling Technologies

- Medium-temperature (85°C) heat pipe
- High-power solar arrays with improved solar cell efficiency and advanced structure.
- Electric propulsion system used for both stationkeeping and orbit raising
- Simplified attitude control subsystem
- Low cost, lightweight RF electronics,
- Linearized channel amplifiers
- Bipolar NiH2 batteries for high energy density
- Phased array antennas
- Reconfigurable and shaped antennas
- Lightweight bipolar NiH2 batteries
- Deployable heat rejection radiators

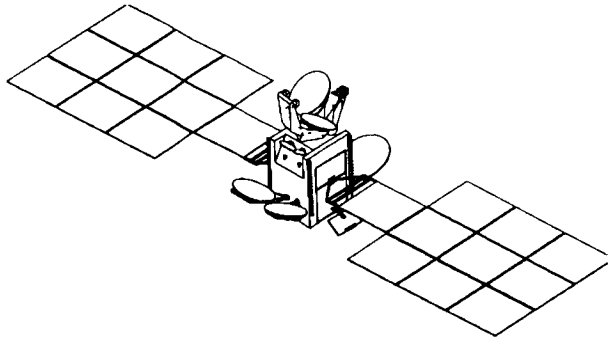


Figure 7: GEO Satellite, 1999

5.4 2003 GEO Satellite Technology

Figure 8 shows the possible 2003 GEO satellite product. Its features are as follows:

Capabilities

DC power	>30 kW
Radiated power	>9 kW
Payload mass	450 to 1,200 kg
Delivery to orbit	12 months ARO

Launch

Atlas 2ARS, Ariane IV/V, Proton, Long March, H-II, Delta 3, Sea Launch and Recoverable Launch Vehicle (RLV)

Enabling Technologies

- GPS tensor for attitude determination and orbit control
- Very high efficiency multi-junction solar cells
- Advanced deployable radiators
- Very high energy density chemical storage devices
- Physical energy storage devices
- High performance filters and computers
- Autonomous satellite control and command

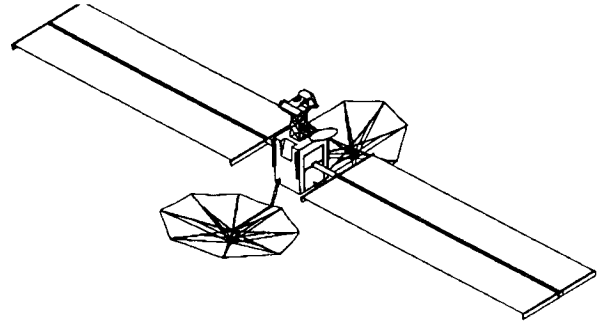


Figure 8: GEO Satellite, 2003

Business Model for Broadband Satellite Communication Services

Dawn McCraw
Motorola Satellite Communications Group
2501 S. Price Rd., Chandler, AZ, 85248-2899 USA
Phone: 602-732-2757 FAX: 602-732-2332
dawn_mccraw@sat.mot.com

ABSTRACT

Global broadband satellite systems are on the horizon and are opening a window of tremendous opportunity for a wide spectrum of service and equipment providers in this new millennium of high bandwidth, global, interactive, multimedia-based information sharing. With emerging commercial technologies such as phased array antennas, baseband processing payloads, and low earth orbit (low-delay) satellites, traditional models of the markets effectively addressed by satellites are being challenged.

As with any new technology, an extensive market assessment must be performed to determine technical design requirements and business viability. This paper examines the many technical and business trade-offs involved in determining the markets and services which can be competitively provided by broadband satellites. The business model is market-driven and allows iteration scenarios of various usage patterns, service willingness-to-pay assumptions, application and market mixes by region of the world, data rates and subscriber multiplexing factors. The subscribers served by the system are checked against a total available market size by region and application. Satellite and constellation requirements are derived from business and market scenarios and financial parametrics such as cash flow, IRR, and NPV are determined, thus allowing evaluation of technical requirements trades against business performance.

INTRODUCTION

The last several years have seen a tremendous amount of activity surrounding the potential of broadband satellite systems, especially global ones. The US FCC Ka-band filings are such an example. A single LEO (Low Earth Orbit) filing and fourteen GEO (Geosynchronous Earth Orbit) filings are fighting for Ka-band US spectrum and many of those systems are global. Additional systems are being filed outside the US. Why? Technology advances in low power baseband processing, flexible phased array antennas, broadband (>50 Mbps) Ka-band and millimeter wave (inter-satellite crosslink) technology and low cost user terminal technology have enabled a new generation of space systems that offer competitive advantages over terrestrial systems. The most compelling advantage is the tremendous flexibility that can be offered. A LEO system can meet all ATM quality of service requirements and seamlessly interwork with existing networks, yet capacity can be moved around the world instantaneously. A small 18" phased array can be placed anywhere and it can automatically track the satellite and provide a two-way connection of tens of megabits per second. Larger antennas can accommodate OC-3 (155.52 Mbps) rates. In addition, because of baseband processing and on-board packet switch technology, tens of millions of simultaneous users can now be supported by such a global system.

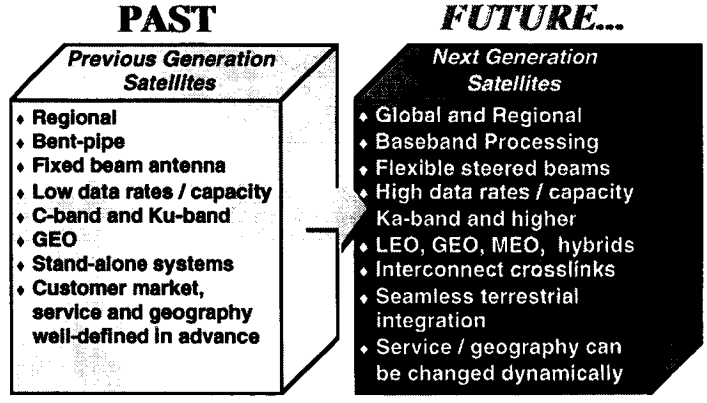


Figure 1 - Space System Evolution

These systems are viable as the new multimedia age is still in its infancy. We have watched one telecom carrier after another announce mega-projects promising fiber to our homes, video-on-demand, videoconferencing, telecomuting, high speed internet access, a single phone that can roam between work, home and car... only to back off and scale down. Why is this happening? Because the technology upgrades to existing plant investment and the new infrastructure required to enable these services will cost billions (possibly hundreds of

billions of dollars) and the government will not be subsidizing it.

Newly privatized and de-regulated carriers have been in the safe, traditional voice business for over sixty years and such a sizable investment on a market that does not yet exist could be too much to hope for in a time when carriers are focused on retaining their existing market share. Even optimistic projections show that the technologically advanced US will fall far short of providing higher bandwidth (128 kbps-ISDN to 10 Mbps-Cable Modem) to homes in the next decade.

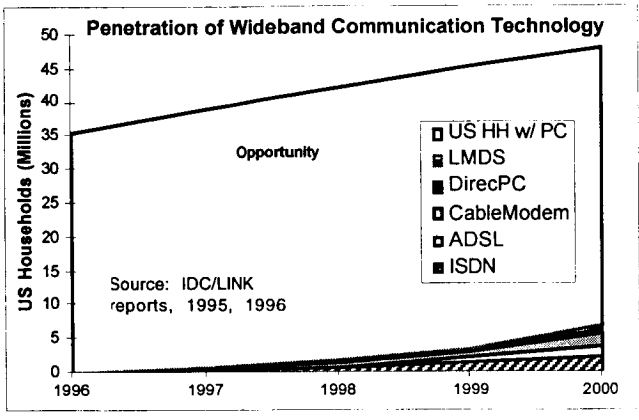


Figure 2 - Projected Penetration of Residential Wideband Technology* in US

* LMDS: Local Multipoint Distribution System; Terrestrial wireless 10-45 Mbps to the home and 2.048 Mbps from the home

DirectPC: Hughes broadcast-only satellite delivered data service at 5 Mbps to the home, phone line from the home

Cable Modem: Electronics upgrade to existing cable infrastructure to deliver 10-30 Mbps to the home and 768 kbps from home

ADSL: Asymmetric Digital Subscriber Line; the Telco carrier's electronic upgrade to existing copper twisted pair. Delivers 1.5-6.1 Mbps to the home, <2.048 Mbps from the home.

ISDN: Integrated Services Digital Network; 128 kbps two way

Understanding the demand, willingness to pay and take-up rates for services is crucial to justifying the investment in broadband systems, terrestrial or otherwise. No one can accurately predict telecommunications services and demand into the next millennium. But the phenomenon of the internet and world-wide web attests to the desire of people to become more interactive, knowledgeable, and global. They want the power to quickly find and process any information which can make their life easier. Estimates of Internet users range from 10 to 35 million growing at 100% per year with the internet services market as much as \$36 B in 2000[1]. Demand growing at this rate will cripple the existing circuit-switched phone infrastructure (as many statistics are already demonstrating). Carriers must develop creative ways to serve the growing data market effectively. The graph below (in log scale) demonstrates that this type of usage is undergoing phenomenal growth while traditional circuit-switched types of communications is leveling off.

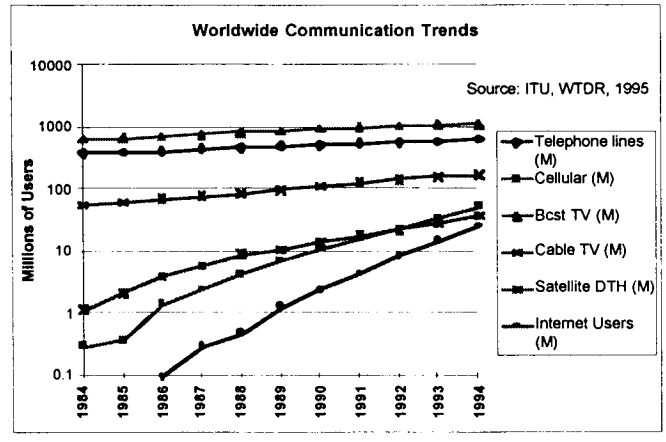


Figure 3 - Trends in Communications

Many contend that once the bandwidth is made available, pent-up demand for broadband services will explode. Developers will create services that we haven't yet dreamed of and the "killer app" just hasn't been invented yet. This answer, however, is not sufficient for potential investors in multi-billion dollar global systems.

Previous generation networks (both satellite and terrestrial) which were optimized for a single service type, such as voice or video, or single service modes, such as point-to-point or broadcast, could be analyzed more easily than the large variety of applications on the horizon today. This business model allows analysis of infinite combinations of applications and service types over a single assumed infrastructure.

[1] Sources of estimates: Harris Poll (11/96), IDC (5/96), Computer Intelligence Infocorp (5/96), Wall Street Journal (3/96), Intervid (3/96), Morgan Stanley (2/96), Nielsen (10/95)

This business model has been created to run various scenarios of regional market penetration for multiple simultaneous services and aid potential network operators and/or investors in determining the business viability of this multi-billion dollar venture. It could also be used to facilitate the allocation of bandwidth among various investors and service providers within particular regions. In addition, it will derive satellite uplink and downlink channel requirements and regional coverage requirements (constellation design) based on market demand. The methodology behind the model can also be useful in the analysis of terrestrial networks and services.

MODEL METHODOLOGY

The business model is market-driven. Market indicators, business goals and technical parameters (including cost) make up the three major components of the model (as shown in Figure 4).

In the *business component* there are essentially four business "toggles" or variables that can be tuned to different scenarios to observe business parametrics. These toggles are:

- *% of Satellite Capacity to Each Region of the World:* An assumption is made as to how much world-wide capacity, C_T might be sold. The model assumes that revenue can only be generated from a full-duplex connection. In other words, charges are either billed at the origination *or* termination point and at each end the user is "listening" on one channel and "talking" on another channel. This connection uses four simplex satellite channels. C_T is then split into the eight regions described in the Regional Demand Section.
- *% Satellite Capacity for Each Market Segment:* A further refinement in understanding the capacity allocated to each region is the assessment of how that capacity is addressing the five market segments - Large Business, Small/Medium Business, Direct to Home (DTH), Wireless Backhaul and Gateway (carrier) Services.
- *% Satellite Capacity for Applications within a Market Segment:* Within each market segment certain applications may be more interesting than others. For example, in the Large Business market, one may have less confidence in the videoconferencing market and more in the LAN interconnect/ intranet market. A choice can be made as to where system resources should be focused.
- Loading of Satellite Capacity per Year: Due to the nature of satellite systems (especially LEO's), solutions are more of a "big bang" approach than an incremental one. But even though large amounts of capacity are available instantly, it does not mean it can be sold into the market instantly. This toggle allows discretion of how much capacity (and how many terminals) can be realistically sold each year. The model does assume that the investor will have a higher 'per subscriber' cost of capacity when the system is not fully loaded. For example, when the system is only 20% loaded, a bit will cost the operator five times more. Whether that increased cost is recouped is a function of what the market will bear, or the 'willingness to pay'.

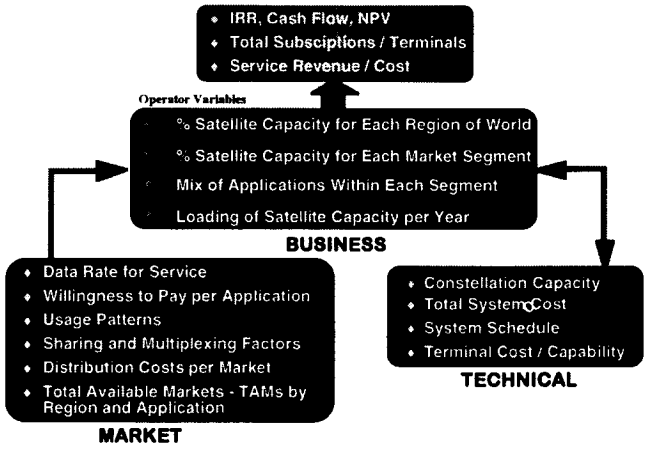


Figure 4 - Business Model Methodology

The *market component* is comprised of assumptions relating to:

- *Market Segments:* Broken into Large Business, Small/Medium Business, Direct to Home, Wireless Backhaul and Gateway Services
- *Service Types:* These characterize the various services by network parameters such as peak to average data rate, phase jitter, error tolerance, etc. and range from constant bit rate service to broadcast.
- *Applications:* Many distinct end-user services (or applications) can fall into the same Service Type. Applications are

differentiated by assumptions such as willingness to pay, usage patterns, sharing factors, etc.

Because the model uses the end-user willingness to pay, factors must be included to move from the network operator wholesale business (where cost only includes the satellite system and network operations) along the service provider, distribution, and retail value chain out to the end-user. These factors are called *cost of sales (COS)* and include such things as advertising, marketing, distribution facilities, and profit along the chain. At this point the COS is simply a function of revenue by market segment. Markets such as Gateway Services with 10% COS would not incur the expenses associated with such markets as Direct to Home where it is assumed that 50% of each revenue dollar is used in costs over and above the network operator costs. The COS are added into the cost of providing bits to each market and shown on a per minute basis at each application level. This allows a more realistic analysis of the end-user willingness to pay against the end-to-end per subscriber cost of providing service for each application.

Satellite service is never assumed to receive a premium rate. Conservative assumptions are made on today's terrestrial rates for specific services. Higher data rates are usually specified over those being provided today and then prices are decreased at variable rates per year (for most applications 5% per year has been assumed - this is approximately a 50% reduction by 2010).

For each application, a total available market (TAM) demand of subscribers has been derived through a multitude of research sources, both primary and secondary. Many regional statistics have been used to determine these markets by the eight regions identified. Once the business toggles have been turned and subscriptions or sites calculated for each application, there is a comparison made to the specific market demand for that application. The sanity check is that the assumed capacity and usage characteristics do not create a subscriber base by region that is not supported by market demand.

Once market assumptions have been entered, many business parameters can be analyzed. Outputs include net cash flow, Internal Rate of Return, Net Present Value, number of subscriptions as well as number of terminals sold (service bundling assumptions create fewer terminals -or users- than 'subscriptions').

In the *technical component* of the model, the simultaneous "on air" users of each service type by region will be derived based on market inputs. This will then determine satellite channel requirements leading to cost, size, weight and power estimates. The business and technical parameters can be iterated and traded to determine the optimal solution space.

There is also a feature of the model which allows cross-checking of a particular satellite system point design against business assumptions to give a measure of viability within assumed cost estimates. This is explained in further detail below.

REGIONAL DEMAND

In order to derive system requirements such as capacity and constellation design, we must first understand regional traffic demand around the world. For this purpose, the world has been divided into the eight arbitrary regions shown in Figure 6.

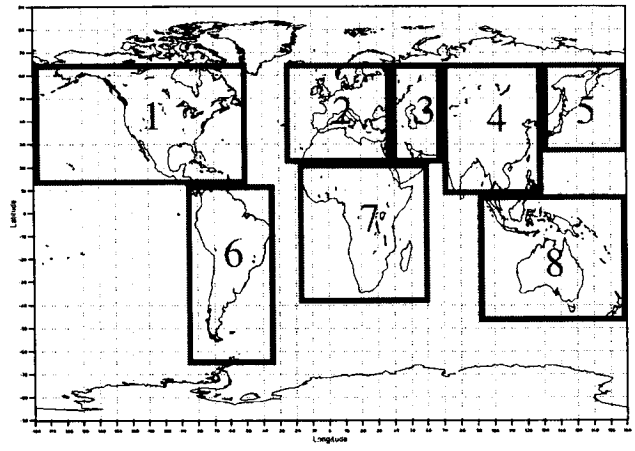


Figure 6 - Regional Segmentation for Business Model

The total assumed constellation capacity, C_T can be divided into the eight regions of the world and split across the five market segments. This allows the flexibility of targeting specific market segments by region as illustrated in Table 1. In the case shown below in Table 1, a maximum of 37% of the selected worldwide capacity can peak over North America with the assumed system constellation and satellite design. Increasing the capacity above the current design will yield new satellite channel requirements that much be traded against higher costs for power, size and weight increases.

When considering a LEO system, there is a unique flexibility in being able to peak capacity over specific areas of the world. A smaller satellite can be designed because the low earth constellation allows for battery charging time over water. Batteries and solar panels can

be much reduced over a GEO satellite with similar capacity. Global coverage requires many LEO satellites (hence a higher cost), but this also comes with the advantage of significantly higher system capacity and global flexibility (as well as fiber-like signal delay).

Region	Regional Split of Capacity	Large Business	Direct to Home	Wireless Backhaul	Small/ Med Business	Gateway Services
World-wide Market Split	NA	11%	44%	3%	36%	6%
North America	37%	10%	40%	5%	40%	5%
Europe	23%	15%	40%	5%	35%	5%
Russia	0%		90%		10%	
China	10%	10%	50%	5%	30%	5%
Japan	10%	20%	40%		35%	5%
South America	13%		55%	5%	30%	10%
South Africa	2%		80%		15%	5%
Australia	5%	10%	45%		40%	5%

Table 1 - Example Allocation of Global Capacity into Regions and Market Segments

The result of the regional allocations above of C_T is a matrix of the desired Gbps (gigabits per second) allotted to each region by market segment. We next examine how those regional bits are utilized.

MARKETS, SERVICES, AND APPLICATIONS

There are five different service types supported by the satellite systems and analyzed in the business model. These are:

- Constant Bit Rate Service : This channel type cannot be shared with other users and is essentially today's circuit-switched model.

  Applications modeled: Videoconferencing, voice, videophone, video backhaul (broadcasters)

- Available Bit Rate Service: The transmit and receive data rates for such applications are assumed symmetric (or of equal speed). ABR service is consistent with ATM Forum definitions of bursty applications with a peak to average ratio >1. These applications are very inefficient in a circuit-switched system where there is significant "dead time". A packet-switched system, however, with baseband processing can take advantage of the spaces between bursts and multiplex in many additional geographically dispersed users driving down the cost per bit.

  Applications modeled: LAN/WAN interconnect, intranet, email, branch office remote access, video games, Electronic Data Interchange, Point of Sale

- Broadcast / Multicast Service: This service can be provided from a single location and duplicated and re-transmitted to selected beams throughout the world. It is a one-way (downlink) service and the number of subscribers is purely determined by the market demand. Broadcast service is one that satellites (especially GEO) perform much more efficiently than terrestrial networks. Either LEO or GEO satellites which utilize multiple beam antenna systems have increased flexibility in targeting multicast signals to a subset of beams. The model does take into account the capacity lost in re-generating the signals into multiple beams.

  Applications modeled: Multicast service (newspapers, video magazines, financial services, training), entertainment broadcast video, digital audio

- Asymmetric Service: The primary example of this service is on-line access. In the model a 100:1 ratio is assumed between the user receive and transmit rates (based on traffic analysis data for web access). For example, the user receives a 2.048 Mbps channel from the server via satellite and transmits requests to the server at 20 kbps. Assuming an asymmetric model frees up a lot of satellite capacity that can be sold to other users.

  Applications modeled: Internet and on-line access

- Broadcast Interactive: This is a very interesting service and is most closely modeled by a home shopping scenario where video is broadcast to all subscribers but they have a low bit rate upstream order wire. The on-air subscriber base is determined by the uplink data rate. The downlink service data rate is fixed at 6.312 Mbps for each channel and the uplink channel rate is as low as 1 bps. Once the number of channels available for service has been selected, the remaining capacity must

be utilized for uplink data, thus limiting subscribers.

Applications modeled: Entertainment video with an order wire (home-shopping, games, political voting...)

Within each market segment, capacity can be allocated to various applications as shown in the example in Table 2. The result is an instantaneous allocation of channels to specific applications by region and market segment.

Application	Large Business	Direct to Home	Wireless Backhaul	Small/ Med Business	Gateway Services
Desktop videoconferencing	10%	5%		10%	
Videophone		5%			1%
LAN Interconn, intranet (Global)	20%				
LAN Interconn, intranet (Regional)	30%				
LAN Interconn, intranet (National)	20%				
Internet backbone					49%
LAN access		7%			
Voice (one-to-one)		13%		25%	
Voice backhaul			100%		50%
Direct online access		25%		35%	
Branch office remote access	20%				
Electronic transactions (POS)				20%	
Digital Audio		5%			
Multicast Service		5%		10%	
Video Broadcast (one-way)		5%			
Video Broadcast (interactive)		25%			
Games		5%			

Table 2 - Example Allocations of Market Segment Capacity into Applications

We now examine the assumptions and calculations that make use of these channels assignments for each application.

APPLICATIONS ANALYSIS

The revenue generated by each application must be determined to derive business parameters. For each application there is first a set of assumptions:

- Data Rate: The service rates modeled are 64 kbps, 128 kbps, 384 kbps, 2.048 Mbps, 6.312 Mbps, 10 Mbps and 51.84 Mbps. These service channels are multiplexed onto satellite uplink and downlink channels: 2.048 Mbps (E1), 10 Mbps (Ethernet), 51.84 Mbps (OC-1).
- Willingness to Pay: In most cases a 1995 terrestrial rate is the starting point. That price is then decreased by a variable percent per year (usually assumed 5%). For example AT&T WorldWorx rates for 128 kbps business videoconferencing is $2.60 per minute. The model assumes that the service will be provided at 384 kbps and the price will drop 5% per year to $1.20 per minute in 2010.
- Sharing Factor: This represents the number of users that can share a channel. For constant bit rate applications, this is 100%. For bursty applications the sharing varies; residential on-line access for example, assumes 6 users can share an average 2.048 Mbps connection (1/6 = 17%).
- Occupancy: This takes into account the subscriber usage. It is the time the user is receiving service divided by the total time the service is available. In all markets except Gateway Services (where 24 hour usage is assumed), there are different assumptions of the available service hours. Thus, time of day is taken into account and it is not assumed that revenue is being generated 24 hours a day. An interesting point to note is that in today's circuit-switched model, when a user is on-line for 3 hours, they occupy bandwidth the entire time, even when there is no data exchange taking place. In a packet-switched model, communication bandwidth would only be consumed when actual data exchange occurs (not just the "off hook" time
- Advertising Factor: For applications such as on-line service, digital audio, multicast service (magazines), video broadcast and video broadcast interactive, there is additional revenue that must be accounted for (just as there were additional costs in the cost of sales). The factor does not affect the willingness to pay, but increases the regional revenue 1.5 to 2.5 times depending on the application.

Cost per Minute:

The first parameter which must be calculated is the cost per minute of service. The cost per year of the system is estimated and includes space vehicles, launch, ground control equipment, space vehicle and launch insurance, spares, replenishment, operations costs, and COS as described in Section 2. The cost is then divided by the Gigabits delivered in the study period (2001 to 2010 in this case) assuming a 24 hour day or

25,228,800,000 Gigabits. This Cost per Gigabit can then be converted into a Cost per Minute for various channel sizes (64 kbps through 51 Mbps). It must then be multiplied by the relevant sharing factors for each application since the cost of a single channel may be shared by several users.

Number of Subscribers and Revenue:

When considering CBR and ABR services, the number of subscribers is determined by taking into account the allocation of channels for a specific service, the sharing factor, the occupancy and the loading of capacity. For connection-oriented applications which are subject to tail charges, there are additional multipliers that take into account assumptions on the percent of traffic which terminates or originates outside of the satellite network. Revenue per year is simply the subscribers per year times the price or 'willingness to pay' per year (taking into account occupancy and per minute rates).

For Asymmetric services, full duplex channels (not connections) must be divided up between the asymmetric data rates. Thus, the subscriber receive rate and transmit rate must be taken into account and an uplink efficiency (or capacity) degradation must be considered.

The number of subscribers supported for Broadcast Interactive service is determined by the subscriber transmit -i.e. order wire- data rate (and its uplink efficiency factor). Another factor which must be taken into account is the capacity used to re-transmit the broadcast signals throughout the region among interconnected satellites (via intersatellite crosslinks). These factors are the number of broadcast beams on board each satellite and the minimum number of satellite in view of a region. Because these factors do not allow for a fine adjustment in subscribers, there is a also "take-up rate" consideration which can lower the number of subscribers (and revenue) to a reasonable percent of the total available market.

For the Broadcast (one-way) service, the number of subscribers supported is purely a function of the market. Technically, an infinite number can be supported. All that is required is a user terminal which can tune in to the channel. In the model, the number of assumed subscribers is kept to less than 2% of the total available market.

SUMMARY

Satellite systems, while expensive, can achieve ubiquitous global broadband penetration into target markets faster than any terrestrial infrastructure and have unequaled geographic and service flexibility. All assessments from the business model to date demonstrate that LEO satellite service in particular is extremely competitive with terrestrial service (much cheaper per subscriber in most cases). New satellite technology can and must seamlessly integrate into terrestrial networks; stand-alone systems will not work. Satellite systems are a potential solution in developing the broadband market and services prior to fiber investment. Because of its tremendous capacity, wide range of service provision, high data rates, and multiplexing advantages of supporting so many data users, it is no longer just a niche market solution.

The broadband satellite business model can be a valuable tool in assessing the viability and profitability of both market and technical trades.

Session 5
Current and Planned Systems

Session Chairperson—*Barry Evans*, University of Surrey, UK
Session Organizer—*Jack Rigley*, Communications Research Centre, Canada

Inmarsat Aero-I Flight Trials

Richard Locke
Inmarsat
99 City Road, London EC1Y 1AX UK
Ph. +44 171 728 1134
Facs: +44 171 728 1190
Richard_Locke@inmarsat.org

John Sydor
Communications Research Centre
3701 Carling Avenue PO Box 11490
Ottawa, Ontario, Canada K2H 8S2
Ph.1- 613-998-2388/ Facs. 1-613-998-1686
john.sydor@crc.doc.ca

Introduction

Since 1990 , Inmarsat has been providing the aviation community with global telephone, facsimile, and data services via satellite. The current Inmarsat aeronautical products, especially "Aero-H", have been very successful at addressing the communications needs of long-haul commercial aircraft. However, for smaller aircraft such as medium-haul jets, or business and commuter aircraft, the physical characteristics and cost of the equipment have traditionally made the prospect of on-board satellite communications impractical.

In 1994 and 1995, Inmarsat held a series of meetings with the aeronautical satcoms industry to define the specifications for an evolutionary and practical satellite communications product which would satisfy the requirements of many segments of the aviation community. This study group, with the aid of extensive market research, determined that the new system should have the following features (compared to the current Aero-H product):

- reduced size and weight of both the internal and external equipment,
- reduced equipment cost,
- identical services as the current product (voice, fax, and data) but achieved with power and bandwidth savings,
- lower end user charges (per minute cost),
- backward compatibility with the existing system (in terms of network management and access protocols),
- and a clearly defined upgrade path which would allow users of the current system to take advantage of the new technologies.

The result of this collaboration between Inmarsat and industry was the "Aero-I" specification.

Aero-I Features

Aero-H was designed for use with the Inmarsat-2 series of satellites which have L-band global beams for communications with the mobile Aircraft Earth Station (AES) terminals.

The Aero-I design is feasible largely because of the L-band spot beam capabilities of the new Inmarsat-3 series of satellites. The first satellite successfully entered into service in May 1996, and the fourth satellite which will complete the geostationary constellation will be launched in 1997.

Spot beams on the Inmarsat-3 satellite have allowed the minimum antenna gain to be reduced from 12 dBic (as in the global beam Aero-H system) to 6 dBic. This in turn will permit antenna manufacturers to produce Aero-I equipment specifically targeted for use by smaller aircraft. The use of spot beams has also reduced the per-channel EIRP required to sustain voice communications. AES manufacturers may therefore incorporate a smaller and cheaper HPA into their Aero-I products.

For telephony, the current Aero-H system encodes the output of a 9.6 kb/s LPC voice codec within a 21 kb/s RF channel (SCPC). The Aero-I design incorporates the Advanced Multiband Excitation (AMBE) codec which has been developed by Digital Voice Systems, Inc. This codec provides the same (or better) voice quality compared to LPC but at a rate of 4.8 kb/s. By implementing advanced voice coding techniques, the same circuit-mode voice service is compressed onto a 8.4 kb/s RF channel ($1x10^{-3}$ nominal BER). The reduction in per-carrier bandwidth and EIRP requirements will translate into reduced per-minute charges for the end user.

The 8.4 kb/s voice channel can also be used to provide Group 3 fax and data services at a maximum rate of 2.4 kb/s. Since the voice channel operates at a $1x10^{-3}$ BER which would be unsuitable for fax and data communications, half of the codec rate is used for additional coding. While this reduction in the maximum user rate could be seen as a disadvantage, it does however permit the user to send fax and data without the need to transmit on a higher power (and therefore higher cost) circuit mode channel. Future

enhancements to Aero-I are likely to include the use of compression techniques to make the full codec rate available to the user for fax and data.

The Aero-I specification also provides for packet mode data communications up to a maximum rate of 4.8 kb/s via either a slotted-Aloha random access channel or a reservation TDMA channel.

Existing Aero-H terminal owners may also take advantage of the new technologies incorporated in the Aero-I system design. The "Aero-H+" specification takes an existing Aero-H terminal, replaces the 21 kb/s circuit mode channel modem with its new 8.4 kb/s counterpart and updates the terminal's software to allow for spot beam operation. In essence, the only difference between Aero-H+ and Aero-I is the minimum antenna gain (Aero H+ terminals retain their original 12 dBic antennas). Users of Aero-H+ systems would be charged at a lower per-minute rate because of the power and bandwidth savings associated with the new channel design.

The Aero-I Trials Program

In order to test the viability of the Aero-I system, Inmarsat and the Communications Research Centre (CRC) conducted an in-flight trials program in late February 1996. A prototype terminal was installed on an Ontario Air Ambulance Cessna Citation and a companion terminal was installed at an earth station which is part of the Inmarsat Aeronautical Network ground segment. Since the first Inmarsat-3 satellite had not been launched at the time of the trials, the expected Aero-I link budget conditions were simulated using an Inmarsat-2 satellite.

The prototype terminal (Figure 1) that was installed on the Cessna Citation incorporated a 4.8 kb/s DVSI AMBE voice codec, a 8.4 kb/s Aero-I (C-channel) modem, a CAL Corporation AMT-100 Aero-H Satcom Terminal (for RF functions), and an Aero-I antenna with a Beam Steering Unit (BSU).

Monitoring equipment on the ground and in the air recorded the parameters necessary for evaluating the terminal's performance, including modem and codec performance data, received BER, received C/No, Doppler offset, and antenna pointing data. Audio (used later for subjective evaluation of the received voice quality) was recorded on a digital audio tape system, synchronized with the QMS.

One specialised piece of monitoring equipment on board the aircraft was the Quality Monitoring System (QMS), developed by CRC. The QMS has a number of features which allow the statistical analysis of the received satellite signal as a function of time. With the QMS it is possible to display the signal's C/No and mean signal C/No as a function of time, as well as the C/No cumulative distribution function and variance. The QMS is also capable of recording the antenna position in azimuth and elevation with respect to the aircraft and the orientation of the geomagnetic field. CRC used this information to analyze the spatial relationship between the satellite, aircraft, and antenna pointing, and correlate this with the measured signal quality and Aero-I terminal performance.

The Antenna and Autonomous Beam Steering Unit

A mechanically steered single element Highly Shortened Conical Helix antenna (Ref.1) conforming to all Aero-I requirements was designed and built specifically for these trials. The antenna measures about 18 inches (45.7 cm.) in length, 6 inches (15.2 cm.) in height, and 6.5 inches (16.5 cm.) in width. It was mounted on top of the aircraft fuselage midway between the nose and the tail. The antenna contained an integral diplexor, LNA, GPS antenna, and three dimensional magnetic field sensor (Figure 2).

The Beam Steering Unit (BSU) used a combination of GPS and geomagnetic field information to determine satellite position and aircraft orientation change. The unit steered the antenna toward the satellite autonomously, not relying on any aircraft derived navigational information. (Ref. 2,3)

The tracking algorithms used in these trials were experimental and by no means optimized. Crosswind correction and signal strength peaking algorithms were not fully implemented in the BSU. As a consequence, tracking for the majority of tests was carried out in an open-loop manner with no input from the terminal.

Test Plan

The objective of the flight trials was to test the operation of the Aero-I terminal at a number of different geographical locations, over land and water, and at different satellite elevation angles. A second objective was to observe the performance of autonomous tracking system, especially at latitudes where the geomagnetic field tended to the horizontal. These objectives determined the flight plan - a 12000 kilometre course along the eastern seaboard of North America down to the Caribbean Sea. This course provided the opportunity to evaluate the Aero-I system at elevation angles between 3 and 58 degrees. At various locations along the course, the performance of the Aero-I terminal was tested while the aircraft was subjected to maneuvers typical of an aircraft in flight.

Data Analysis and The use of Aircraft Dynamics as Reference Conditions

The Aero-I communications link is affected by a number of signal stability and propagation phenomena:

-Satellite signal power stability,
-aircraft induced multipath,
-ground induced multipath,
- shadowing,
-antenna gain changes due to aircraft fuselage curvature, and
-autonomous tracking system error.

Evaluating the effect that each of these has in isolation is difficult, and is not particularly meaningful. In an effort to simplify the analysis the tests were categorized in terms of aircraft flight dynamics and the effect these had on Aero-I performance. Quantifying communications performance in terms flight states was attractive because the occurrences of such states have tangible statistical meaning to aeronautical operators.

Reference Pilot Power Stability and Correlation with the Communications Channel

Two CW pilot signals were used as references for the QMS system. The Inmarsat AORE and AORW satellites generated pilot signals at respectively 1541.000 MHz and 1537.525 MHz. The communications channel frequencies used for the trials were 1547.5625 MHz for receive and 1645.0625 for transmit. Because of the close correlation between the reference pilot signal and the receive channel, it was assumed that whatever propagation phenomenon affected the pilot channel would similarly affect the receive channel. Therefore the QMS results, which were actually for the reference pilot signal, were applied to the communications channel.

It was noted over a number of tests that reference pilot signals displayed a stability of +/- 1.0 to +/- 1.25 dB (of the mean) for 99% of the time. This range of pilot signal fluctuation is about 0.5 dB greater than noted in other published results where a wide bodied aircraft and high gain (12 dBic) antenna were used (Ref.4). The poorer stability in our case was most likely due to the degraded ground multipath discrimination that is present with the Aero-I antenna and narrower fuselage of the Cessna aircraft.

The above stability statistics for the reference pilots quantify the best possible signal characteristics that could be expected with the Aero-I terminal. These statistics were used as reference values against which other (dynamic aircraft) performance statistics were compared.

Stationary Aircraft Conditions

Understanding the nature of satellite signal propagation around a stationary aircraft is important because this is a common state for the aircraft to be in during communications. Additionally, other than for the presence of the ground reflected multipath, the stationary propagation environment is similar to that of an aircraft flying in straight, level flight.

Stationary tests were performed by situating the aircraft on the runway in an area far from buildings and other possible sources of multipath. Only reflections due to the ground and aircraft were anticipated. The aircraft was turned to specific points on a circle (usually every 45 degrees) and measurements were undertaken with the QMS and Aero-I system operating. By taking such measurements at different geographical locations, the effects of different elevation angles could be recorded.

These tests demonstrated the repeatable nature of multipath nulls at specific positions around the aircraft. One typical example is shown in Figure 3. Here the antenna is at 305 degrees azimuth with respect to the aircraft nose, pointing at a 37 degree elevation angle. The antenna is looking essentially into free space with the port wing far behind the main lobe; only the forward part of the fuselage is visible to the antenna. Nevertheless, a pronounced null exists, causing about a 1.75-2.0 dB drop in signal strength. This null manifested itself throughout the trials during dynamic testing, with its effects only disappearing at higher satellite elevation angles.

Such nulls, which are not attributable to antenna beam scalloping or other electromagnetic effects, were found all around the aircraft. Generally, the aft portion of the aircraft at low satellite elevation angles showed the greatest occurrence of such nulls. It was difficult to determine how wide these nulls were. In many cases they would disappear with several degrees change in the aircraft azimuth position; at other times they were quite broad, possibly being a conglomeration of a multitude of reflections.

Ground Maneuvers

A number of tests were conducted to determine if ground motion had an effect on signal reception. Tests were conducted on runways with the aircraft moving in a straight lines on the runway. In other tests the aircraft was forced into sharp turns. In all cases the satellite was tracked by the terminal, with less than

0.5 dB of signal loss being attributable to tracking error.

Motion induced multipath had a negligible effect on signal quality in these tests. During ground movement, the received signal usually maintained a stability of +/-1 dB within the mean level for 99% of the time; very similar to the reference pilot statistics, providing the satellite was clearly visible by the antenna and was at a reasonable elevation angle (>10 degrees).

There was no perceptible degradation in communications attributable to the Aero-I terminal while the aircraft was undertaking ground based maneuvers. However, shadowing by hangars and airport buildings did cause deep outages at times, especially at low satellite elevations.

Level Flight at Low Satellite Elevation Angles (3-5 degrees)

A number of tests were conducted with the satellite at a low elevation angles. Figure 4, typical of such tests, shows the received C/No while the aircraft was mostly in straight, level flight. The signal was received between 200 and 260 degrees azimuth with respect to the nose of the aircraft. The tail and wing of the aircraft were visible to the antenna.

For this particular flight condition and satellite elevation, the received signal was within +/-4.5 dB of the mean for 99% of the time. The large variance can be attributable to both ground and aircraft multipath. At such low elevation angles the antenna has virtually no discrimination of direct and multipath signals. The communications link is maintained because of the robustness of the modem, which could tolerate up to about 2 dB of smoothly changing signal loss. The BER would fluctuate significantly during such instances, but as long as the modem/codec maintained lock and synchronization, voice was intelligible with only occasional garbling. When voice loss did occur, recovery was usually quite quick.

One experiment that was done at low elevation was signal tracking: an antenna tracking algorithm was activated in which a search for the peak signal was attempted. A peak could not be found within the 200 to 260 degree azimuth span (Figure 5). It became apparent that in a high multipath environment, the variance of the received signal swamped any meaningful signal strength optimization attempted by peaking of antenna gain. In fact, it was noted that the dithering caused by the antenna worsened the reception statistics.

Level Flight at Moderate Satellite Elevation Angles (10-20 degrees)

Analysis of 18 straight and level flights, where the elevation was between 7 and 20 degrees, indicated that the signal variance is about +/-2.3 dB for 99% of the time for signals received from the aft end of the aircraft while only +/- 1.1 dB for signals received from the forward end. Forward signals approach the reference pilot signal stability limit of about +/1 dB for 99% of the time, and as a consequence, are probably quite free of aircraft induced multipath. The aft signals, by their higher variance, indicated that aircraft induced multipath is pronounced.

In straight, level flight at moderate elevation angles there were only rare instances of voice loss with the Aero-I terminal.

Level Flight with High Satellite Elevation Angles (>35 degrees)

At high elevation angles (>35 degrees) the antenna main lobe begins to display significant directivity because it is pointing upwards and away from the metallic surface of the fuselage. Multipath rejection by the antenna becomes noticeable. Typically the variance is in the order of +/- 1 dB for 99% of the time. Multipath nulls due to the tail would, on occasion, still be noticed.

Under such conditions there were never any instances of voice loss with the Aero-I terminal.

Tracking through Banks and Turns

In straight and level flight the BSU maintained satellite tracking to within 3 degrees, at worst. However, it is during turns that tracking efficacy is most important. During turns having banks of 15-25 degrees azimuth tracking was within several degrees of the expected values. There is more error with elevation tracking, but it did not exceed +/-7 degrees, resulting in negligible gain change (+/-0.5 dB) due to antenna mispointing . For faster turns, such as those having peak bank angles of 45-60 degrees, the elevation error remained at +/-7 degrees. Azimuth tracking error however increased to +/-12 degrees. Such tracking error results in an antenna gain loss of about of 1 dB, at maximum.

A typical 360° turn is shown in Figure 6. Here the aircraft had peak bank angles of 30°. As can be seen the C/No is well maintained. Severe multipath/shadowing nulls are seen at 330°, 250°-200°, and 120° azimuth. Note the existence of the null at 305°-310° described earlier.

The autonomous tracking system used in these trials was not fully optimized. Some of the errors noted above may have been due to cross wind. Commercial versions of the autonomous tracking system have mitigated such problems, resulting in better tracking performance.

Modem and Codec: Combined Performance

The quality of the 4.8 kb/s AMBE voice codec was very good and allowed for natural conversation. Usually it would take the called party a number of seconds to become acquainted with the synthesized voice and satellite delay. After this, the conversations would proceed normally.

Background noise, which was significant in the confines of the small twin jet aircraft, did not have a degrading effect on the quality of voice communications. Quality seemed to be more dependent on the distance the microphone was from the lips of the speaker.

Within environments having significant sporadic multipath activity, where the corrected BER reached about 0.03-0.06, speech losses were noted. With such sporadic bursts of error, recovery of the modem/codec was very quick. If the sporadicity persisted, then the power control system would increase the link power appropriately until the BER was restored to its nominal operating point. Such system behavior adds to the resiliency of the channel and is a feature in both the Aero-H and Aero-I specifications.

The modem and codec combination are tolerant of smoothly dropping signal levels. It was noted on occasion that the signal level could drop by 2-3 dB over a period of 3 seconds without any loss in voice continuity. Such a drop, if followed by a consistently low signal (and high BER) would cause the Aero-I power control system to dynamically increase the power on the affected link.

Similar resiliency was found in situations where the signal would fluctuate in a quasi-random manner over a wide dynamic range. This was typical of a low satellite elevation environment. Even with such poor signal stability, intelligible communications were maintained.

Surprisingly, the modem/codec seemed to be most vulnerable to signal fluctuations that were rapid and sinusoidal. Even when the mean C/No was above threshold, fluctuations of several dB at a 3-5 Hz rate would cause the modem to lose lock for inordinately long periods of time (>10 secs).

Conclusion

The Inmarsat Aero-I flight trials program demonstrated for the first time the performance of an aeronautical terminal designed around a 4.8 kb/s digital architecture. This system incorporated new types of modems and codecs that show significant resiliency in face of sometimes severe multipath conditions.

The modem/codec combination demonstrated operating margins that could tolerate 2 dB loss in signal power without inordinate loss of communications. The Aero-I power control system responded quickly to changes in the BER and signal power: a factor of some importance with the severe multipath environment that occurs with small aircraft and larger beamed, low gain antennas that will typify Aero-I systems.

These tests demonstrated the performance of one particular autonomous tracking system. Such systems can and do function with a high degree of pointing accuracy during severe aircraft maneuvers. The resiliency demonstrated with the modem/codec system and the ability to dynamically control link power indicate that the small gain loss due to antenna tracking error will not impair the performance of the Aero-I system. These trials also indicate that the direction of multipath and nulls around a small aircraft are repeatable, and as such, can probably be countered by sophisticated signal processing and antenna steering or beamforming techniques.

References

[1] Short Conical Antenna U.S Patent Number 5479182 Dec 26,1995

[2] John Sydor and Martial Dufour, Aeronautical Satellite Antenna Steering Using Magnetic Field Sensors. Proceedings of the International Mobile Satellite Conference June 16-18, 1993 Pasadena, Calif. JPL Pub No. 93-009

[3] C.A. Sutherland and J.T. Sydor, MSAT Aeronautical Mobile Satellite Communications Terminal Development. Proceedings of the Fourth International Mobile Satellite Conference June 6-8, 1995 Ottawa Canada. JPL Pub No. 95-12 / CRC Pub No. CRC-CP-95-001

[4] T.Ide et al, Propagation Results of Aeronautical Satellite Communication Experiements Using Inmarsat Satellites. IEEE Transactions on AeroSpace and Electronic Systems Volume 28, No. 4 Oct 1992

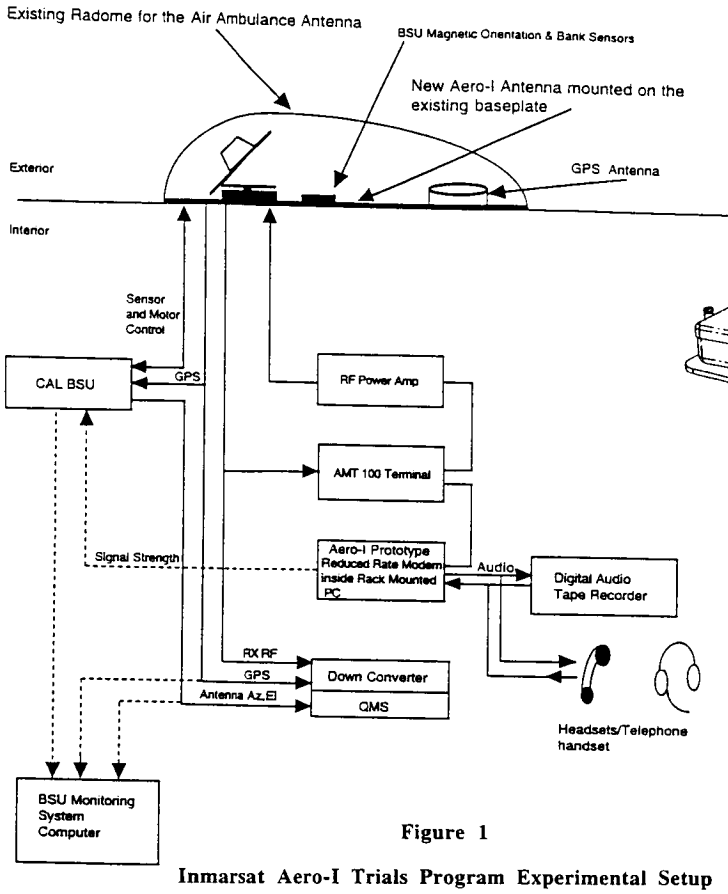


Figure 1

Inmarsat Aero-I Trials Program Experimental Setup

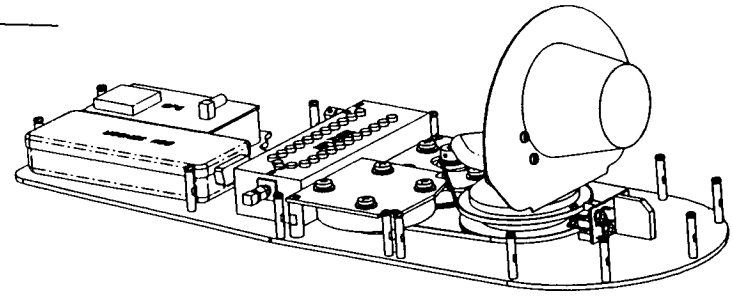


Figure 2

AERO-I Antenna without radome or shielding

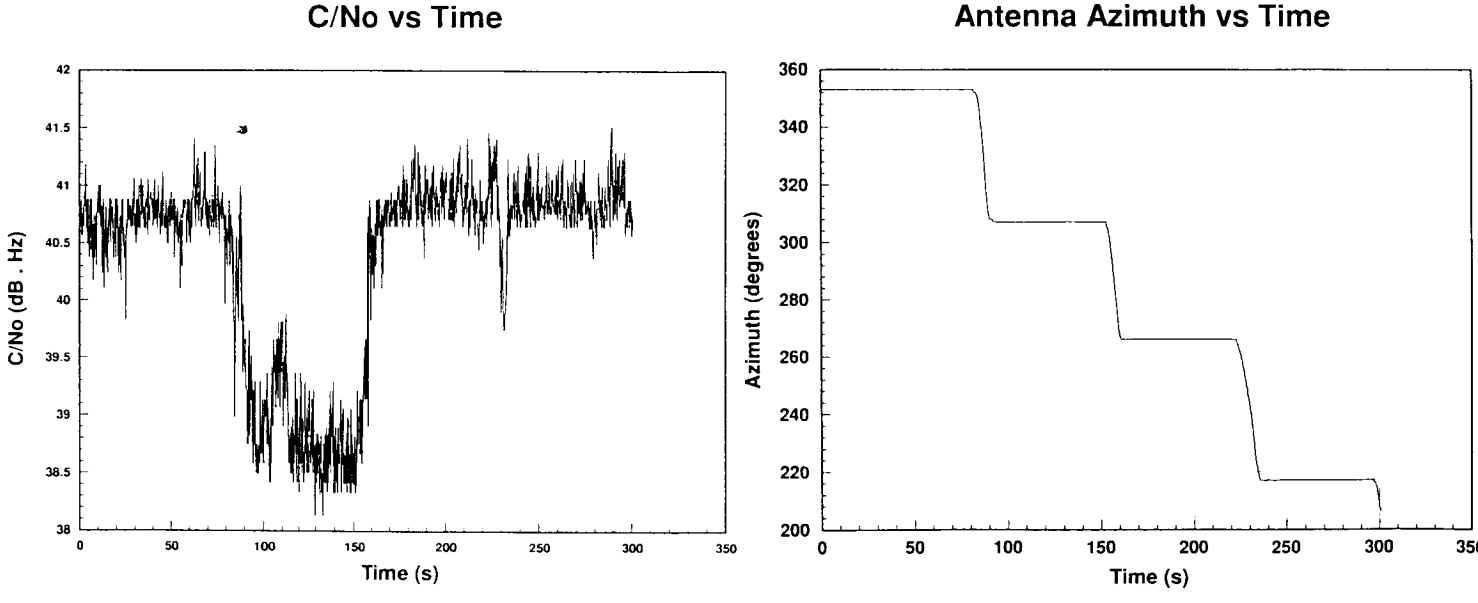


Figure 3

305° Multipath Null

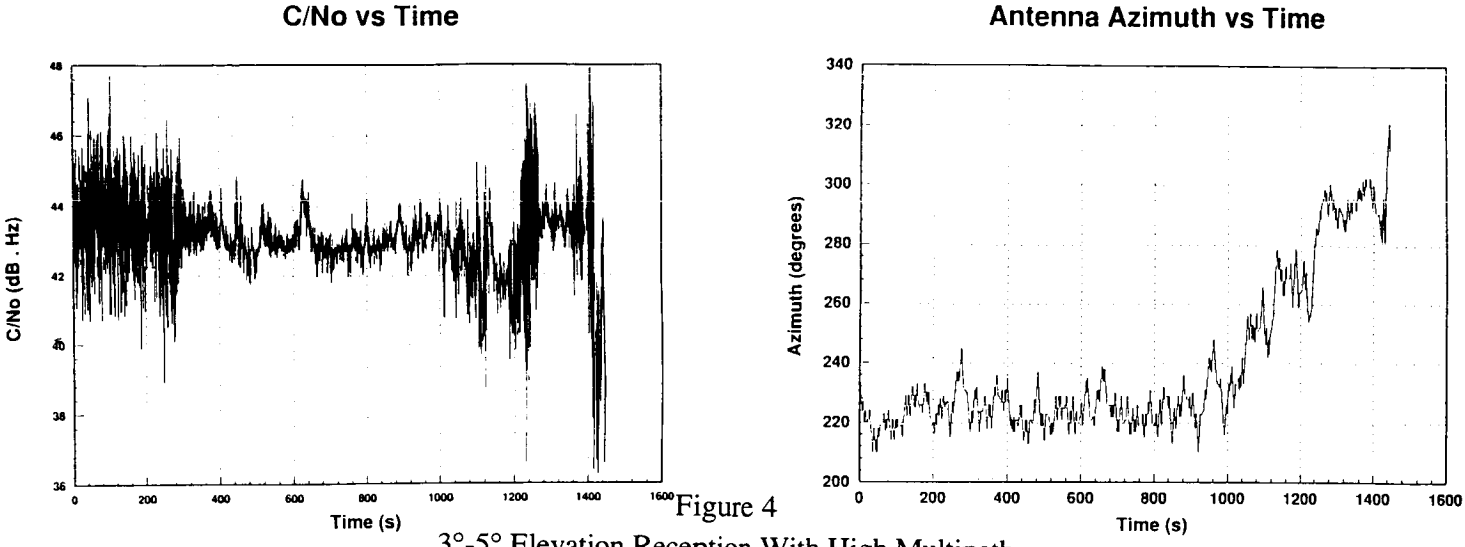


Figure 4
3°-5° Elevation Reception With High Multipath

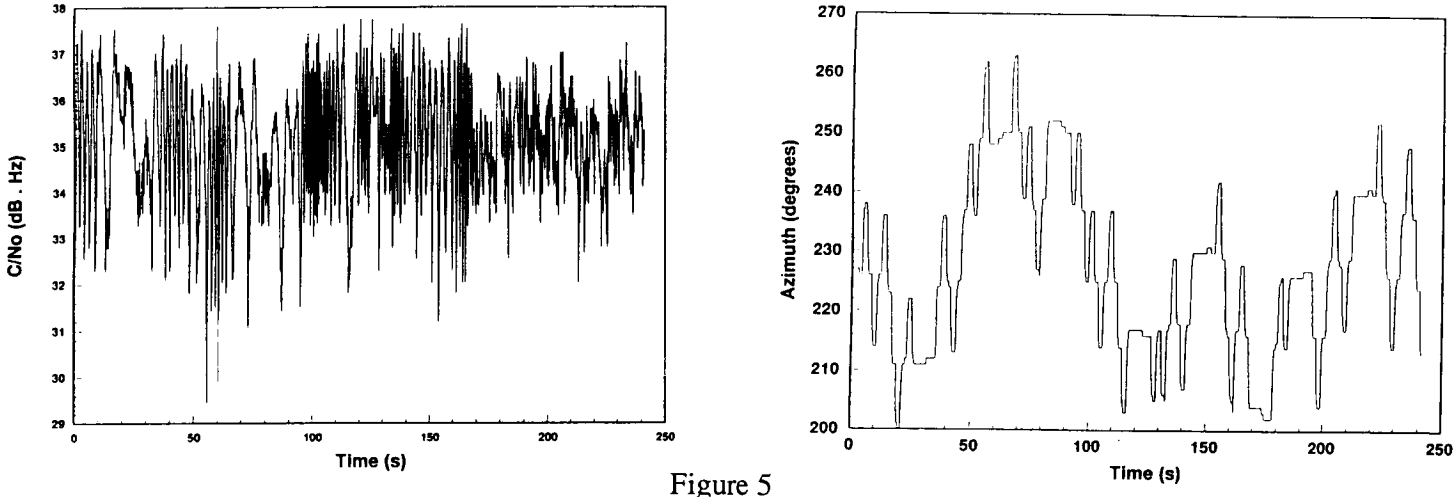


Figure 5
Effects of Antenna Dithering In High Multipath

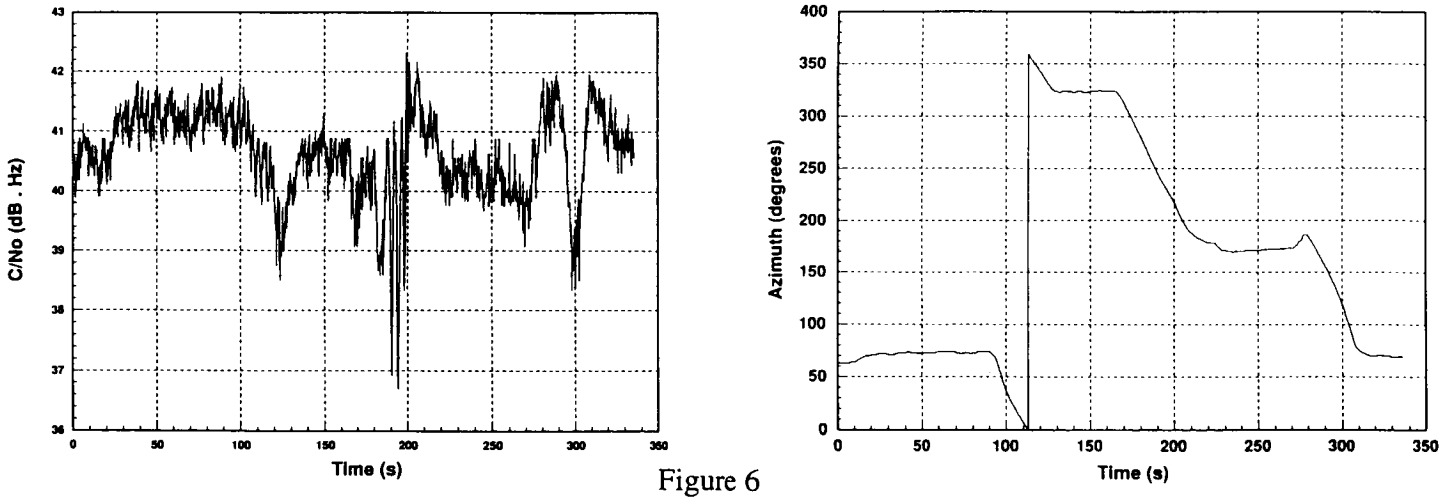


Figure 6
360° Turn With Peak Bank Angles of 30°
Nominal Satellite Elevation is 35°

EAST: A Distributed Architecture Enhancing a Multi-operator Business Approach

Eric Copros
Matra Marconi Space
31, rue des Cosmonautes - Z.I. du Palays
31077 Toulouse Cedex - France
Phone: +33 562 19 73 70 FAX: +33 562 19 76 52
email: eric.copros@tls.mms.fr

Gary A. Johanson
Westinghouse Wireless Solutions Company
930 International Drive
Linthicum, MD 21093 USA
Phone: +1 410-765-9045 FAX: +1 410-765-9745
email: johanson.g.a@wireless.westinghouse.com

ABSTRACT

Sponsored by Matra Marconi Space, EAST (for Euro-African Satellite Telecommunication) is a new telecommunication system for Africa, the Middle-East and Europe.
EAST offers highly competitive cost per circuit and rapid deployment of the network. It provides a wide range of services to mobile users, focusing its capacity on areas seeing a large peak in demand. It provides flexible interconnection with terrestrial networks via distributed gateways.
EAST satellite services will enable licensed national service providers to offer high quality fixed and mobile services over 100% of their territory. Fully compatible with established terrestrial and emerging satellite standards, EAST's flexible interconnection architecture will allow subscribing regional operators to optimise their future networks.
This paper describes how the EAST architecture and operational structure guarantee a wide autonomy to regional operators and ensure a high grade of service availability. EAST, acting as a carrier, will have access to the orbital slots. EAST provides the satellite resource to the operators and coordinates, when required, inter-operator exchanges. It provides the infrastructure for a soft sharing of resources between the regional players.
Regulatory and licensing permissions, necessary to operate the system, are fully under regional control. The regional operators are wholly responsible for the management of their subscribers and services. In particular, the call establishment process is handled by them. Regional operators manage their own set of satellite resources according to their network planning policy and to their expected market growth.
The EAST architecture gives the opportunity to subscribing operators to manage their own satellite sub-network. Moreover, such a concept prevents a fully centralised Network Control Centre from becoming an over-dimensioned mandatory thoroughfare and a major single point failure of the whole system.

INTRODUCTION

EAST is a recent initiative company proposing a satellite communication system addressing the market of mobile and rural telephony over Europe, Africa and the Middle-East.
Relying on geostationary satellites to better focus its capacity on highly traffic-demanding areas, EAST will provide a wide range of services along with GSM (Global System for Mobiles) interoperability by the year 2000.
The system, currently under development, is based on a distributed architecture offering a wide range of autonomy to the local players in the world of communications: the national (and/or regional) operators.

EAST SYSTEM OVERVIEW

EAST Services

EAST provides high capacity hand-held and fixed services for remote areas which are not sufficiently covered by existing telecommunications infrastructure.
Primarily dedicated to voice telephony, EAST also offers a range of teleservices (facsimile, messaging, transparent data services as well as typical supplementary and alternate services provided by a classical GSM network) and low data rate bearer services.
The EAST service is connected to the Public Switched Telephone Network (PSTN) and to the Integrated Switched Data Network (ISDN). The system provides interoperability and roaming between the satellite network and the local Public Land Mobile Networks (PLMN), using established terrestrial standards (such as GSM or Digital Cellular System, DCS) for users equipped with dual-mode terminals.
Thanks to advanced satellite payloads, 10,000 to 30,000 Erlangs of mobile and fixed traffic will be supported per satellite, assuming full rate vocoders.

On-board digital processing allows direct, single-hop communications between EAST terminals (either mobile or fixed).

EAST system features

The EAST system operation is based on the following general concepts:

- Interoperability with PLMN,
- Distributed management by regional operators,
- Centralised, high-level management for security and network co-ordination,
- Networking functions to support least cost routing,
- Flexible numbering scheme and unique subscriber number,
- Mixed centralised/distributed resource management.

Interoperability with PLMN: EAST provides interoperability with the PLMN to dual-mode terminals equipped with GSM and/or DCS standards. The EAST network design is based on a maximum reuse of the GSM network functions and standards. EAST provides transparent roaming functions between EAST and the PLMN. Mobility management is performed by the standard GSM functions of the MSC (Mobile services Switching Centre) and associated data bases (Home Location Register, HLR and Visitor Location Register, VLR).

Distributed management: Regional Operators manage their own subscribers and services.

High level management: a centralised Network Control Centre (NCC) ensures security at the overall network level (in particular it monitors and controls satellite health and configuration) and guarantees an efficient sharing of the resource, particularly for operators sharing the same beams.

Least Cost Routing: least cost routing is one of the key features of the EAST system; allowing to utilise in the most efficient way, and in accordance with the telecommunication regulations, the connectivity provided by the satellite segment and to optimise terrestrial connections. For outgoing calls, least cost routing ensures the selection of the optimum access point to the satellite to minimise the overall connection cost and to account for the actual ground segment configuration. For incoming calls, least cost routing allows to accept a call at any gateway in the network.

Numbering scheme: providing the terrestrial network owns enough functionalities to recognise an EAST subscriber number (for instance if an EAST country code is used or if the PSTN owns Intelligent Network functions), least cost routing can be easily implemented. A unique subscriber number gives EAST users access to international roaming inside the whole EAST service area.

Mixed resource management: satellite resource management is performed on a mixed distributed/centralised basis, ensuring an efficient use of the satellite resource in terms of power and spectrum and allowing regional operators to autonomously perform their own network planning.

EAST SYSTEM ARCHITECTURE

Architecture Overview

The architecture of the EAST system relies on a flexible and powerful space segment, advanced, state-of-the-art terminal technology and a ground segment infrastructure taking advantage to the maximum extent possible of existing GSM products. This architecture is illustrated in figure 1.

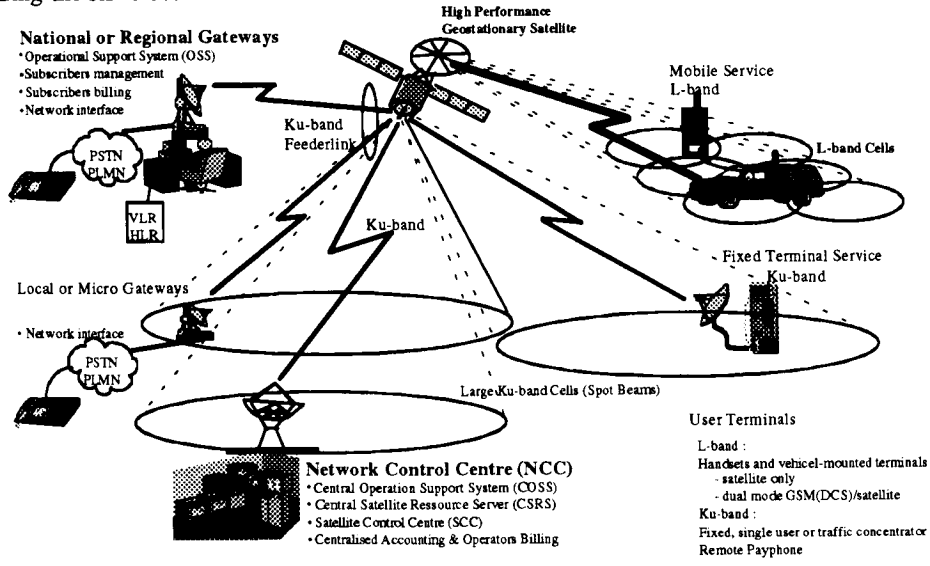


Figure 1 - EAST Architecture overview

The ground segment is composed of a Network Control Centre (NCC), several National Gateways (NGW) and some smaller gateways (Local Gateways, LGW or Micro Gateways, μGW).

The Network Control Center
The NCC provides general network control (including payload control) and management. It co-ordinates the use of the resource across the entire system and allocates the resource to the gateways. The NCC performs the commissioning of new equipments and manages the software and hardware configuration of the different Network Entities. The NCC is in charge of controlling the access to the network and of broadcasting system-wide information in all satellite beams (through the Broadcast Control CHannel). It also synchronises the gateways and the user terminals.
The system supports a redundant NCC in hot stand-by for emergency operations or for any other operational reasons.

National Gateways
A NGW serves a country or a region It contains a full complement of mobility management facilities (HLR, VLR, MSC, Authentication Centre, AuC, Equipment Identity Register, EIR), as well as the accounting facilities and customer care centre attached to the provision of the EAST services over their registration area.
The NGW handles subscriber affiliation, data base management, least cost call routing, real time switching and equipment management. A NGW manages its own pool of satellite resource and reports to the NCC for any modification of this pool (that may potentially involve a payload reconfiguration).
The capacity of a NGW is between 200 and 2000 circuits.
A NGW is able to control one or several local gateways of smaller capacity.

Local gateways
A local gateway is a smaller gateway designed to provide a lower capacity terrestrial connection (from 30, for a μGW to 90 circuits, for a LGW) within a region in order to reach locations in a more efficient manner than terrestrial trunking from the NGW. Local gateways enable efficient least cost routing, avoid communication bottlenecks on the terrestrial connections and provide low cost remote satellite access points to a NGW.
A local gateway is a slave entity attached to a given National Gateway. It contains minimised functionalities compared to the NGW: simplified call and protocol management during the call, channel processing and interface to the terrestrial networks. A LGW does not manage the call set-up, nor is it in charge of managing subscribers or services. A LGW only has temporary data bases to store in-call users profiles.

EAST FUNCTIONAL DISTRIBUTION

Operational structure
The EAST system operation involves a multi-tier business structure. Each of the actors shall be considered as legally independent companies tied by contractual agreements and making use of the same system within an agreed set of rules.

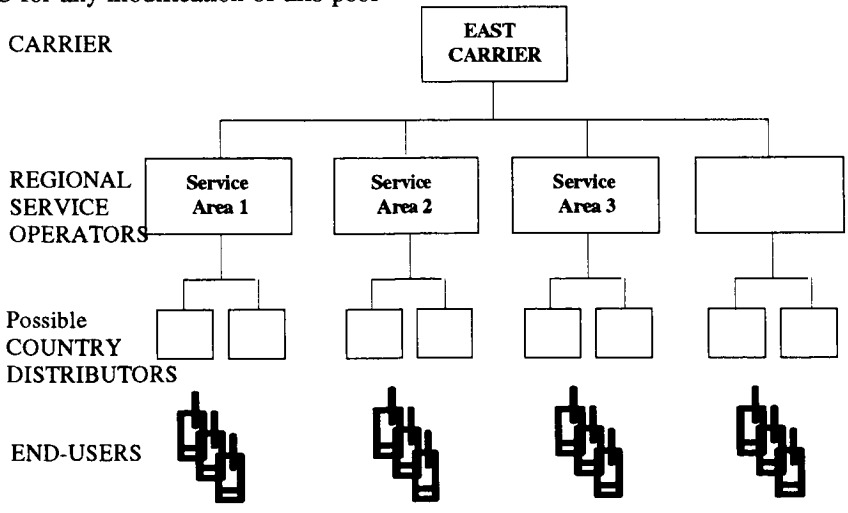


Figure 2: EAST operational Structure

The EAST Carrier is in charge of the overall system management and makes available system resources to the Regional Service Operators (RSO), who provide a local or regional infrastructure for the Services Providers (SP) or directly to the end-users. This business structure is illustrated in figure 2.

EAST Carrier: The basic role of EAST carrier is to design, procure and manage the communication system. It supplies marketing expertise for commercial services and organises the evolution of services to meet user needs. More generally, the EAST carrier has a major co-ordinating role in the network and between the different RSOs.

Regional Service Operators: The Regional Operators purchase capacity from EAST and manage their customers. They also retail subscriber units. They also handle all the local regulatory matters and manage the interfaces to the terrestrial infrastructure. They perform their own network planning.

Service Providers: Service Providers (or country distributors) buy airtime from the RSOs. They market the EAST services within their country and bill their own subscribers. SP may also offer value added services. On some regions, RSOs may act as country distributors.

This business structure matches perfectly the physical architecture of the EAST system, through an appropriate distribution of functions.

Functional distribution

The functionalities required for the *EAST Carrier* to fulfill its mission are concentrated in the NCC: the NCC hosts the Central OSS (Operational Support System), facilities to manage the RSO (Regional Operators Administration Centre) and to control the satellite (Satellite Control Centre). It also hosts entities dedicated to the management of resource (the Central Satellite Resource Server and the Satellite Resource Management System), as well as functions specific to network synchronisation and access control (the Network Signalling and Synchronisation Access).

The Regional Operators own the National Gateways and a set of Local and/or Micro Gateways, depending on their business strategy and on the configuration of the local terrestrial infrastructure. As mentioned previously, the NGWs own all the facilities to manage user mobility and profile and to establish calls by means of their pool of resource.

The Service Providers mainly deal with the management of services. They may handle Local gateways but management of operational infrastructure is not mandatory for them.

This distribution of roles between physical entities is illustrated in figure 3. This distribution fits with the role of each player, the EAST carrier as a resource provider , local partners as telecommunication operators and service providers.

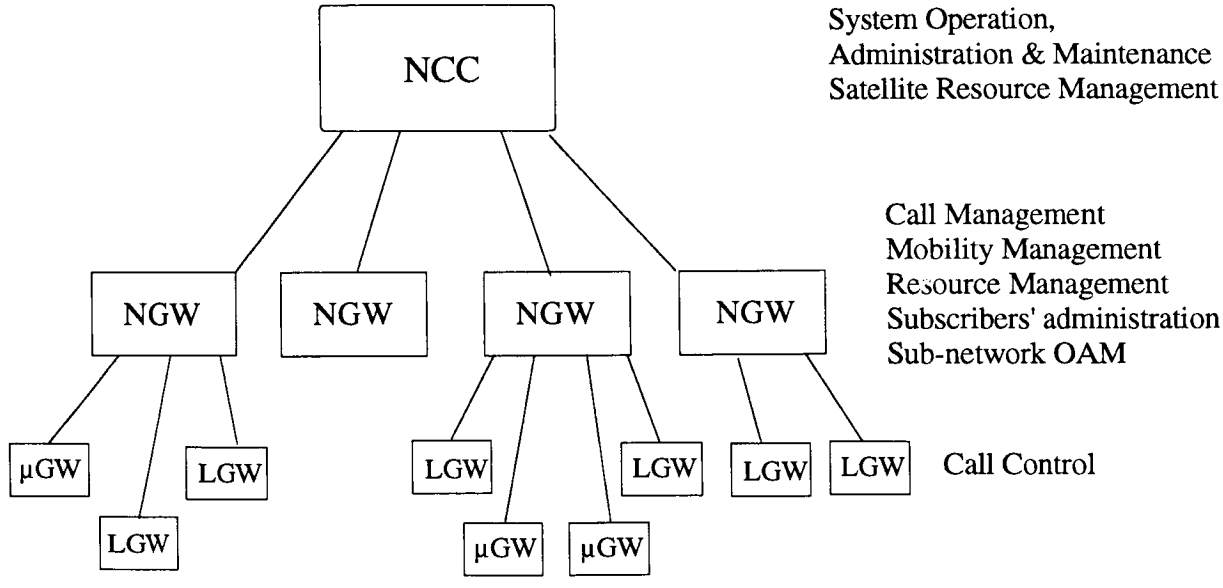


Figure 3: EAST system hierarchy and functional distribution

BENEFIT OF THE EAST APPROACH

In the approach selected for the EAST system, the centralised functionalities are strictly limited to the minimum (non real-time) functions:

- satellite control,
- management of common resource,
- long term, high-level network management,
- network synchronisation,
- assistance to regional operators,
- system deployment and commissioning.

The management of real-time operation (call set-up, location updates) is left to the regional operators, leading to a wide autonomy of the operator in their management of subscribers, network and services, but also significantly improving such network performances as call set-up delays and congestion rate:

- customer care is left to local operators, solving problems such as the "language fences" in recording machines (what about a grand-father of southern Spain trying to leave a voice mail to his son and facing an answer in English?) or problems like subscription or billing recovery,
- local regulatory issues are left to those best acquainted with these problems: the local operators,
- thank to the variety of ground segment products, the RSO sub-network topology may fit the actual configuration of the terrestrial PSTN and PLMN (number of lines, size of the trunks,...),
- a wide autonomy is left to regional operators regarding the way they perform their network planning (distribution of gateways and capacity); in addition they may adapt their strategy of resource reservation to their actual grade of service requirements (they manage their own pool of resource).

Compared to the current generation of centralised systems, there is no risk of traffic bottle-neck at the NCC. Each gateway MSC is easily derived from standard GSM products and capacity. The overall system processing capacity is even significantly larger and the signalling load is reduced.

Another important aspect of the network architecture is reliability and, consequently, service availability. A centralised point for all on-going (or even incoming) calls is either a major risk to availability or a head-ache for reliability engineers. The EAST architecture ensures that calls are not dropped in the event of a NCC failure, even if switching to the back-up NCC take some time. In most cases, such an event does not even prohibit new call set-up.

The EAST architecture has been thought as a major step towards the new generation of UMTS-compliant S-PCNs. All the components of the system: the gateways and the NCC, derived from standard GSM products, the advanced digital processing payload, the efficient hand-held terminals and the adaptive, satellite-and-GSM-compatible air interface contribute to the achievement of this goal.

The Globalstar Satellite Cellular Communication System Design and Status

Dr. Fred J. Dietrich
Globalstar L.P.
3200 Zanker Road, San Jose, CA 95164-0670
phone: (408) 473-7188 FAX: (408) 473-7809
e-mail: fred.dietrich@globalstar.com

Abstract-The Globalstar cellular communication satellite system is described, including its use of Code Division Multiple Access (CDMA) as the basic modulation scheme, use of diversity for signal quality and power control, and complex phased arrays on the satellite. In addition, the financial and business development status of Globalstar are discussed.

1.0 INTRODUCTION

Globalstar is a satellite-based cellular telephone system that allows users to communicate from anyplace in the world between 70° North and South latitudes. It provides clear communication thanks to Code Division Multiple Access (CDMA) transmission, and avoids outages caused by blockage of signals by using diversity signals from two satellites[1].

Globalstar is a partnership of a number of companies. Loral and Qualcomm, Inc. are general partners, and 10 others serve as limited partners (Figure 1).

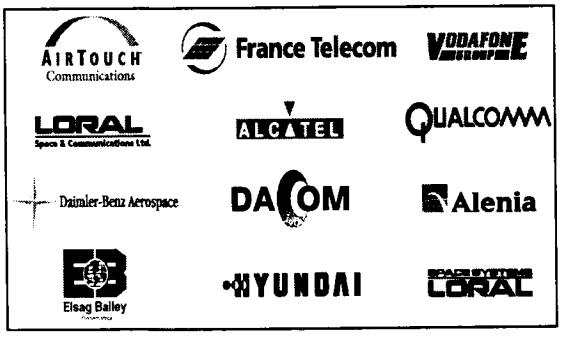


Figure 1. Globalstar Strategic Partners

2.0 SYSTEM DESCRIPTION

2.1 Overall Description

The Globalstar system consists of a Walker 48-8-1 constellation, that is, 48 low-orbiting (1400 km high) satellites in eight orbits, inclined 52° with respect to the equator, with six satellites in each orbital plane. They contact users on the 1.6 gigahertz (GHz) L-band and 2.5 GHz S-band via a hand-held radiotelephone (user terminal or UT) and communicate with the large Gateway ground antennas on the 5 and 7 GHz C-bands. Ground stations (Gateways) use 5.5 meter antennas and connect to the public switched telephone network (PSTN). The UTs are the same size as a conventional cellular telephone, but with a thicker and longer antenna. Figure 2 shows the elements of the system.

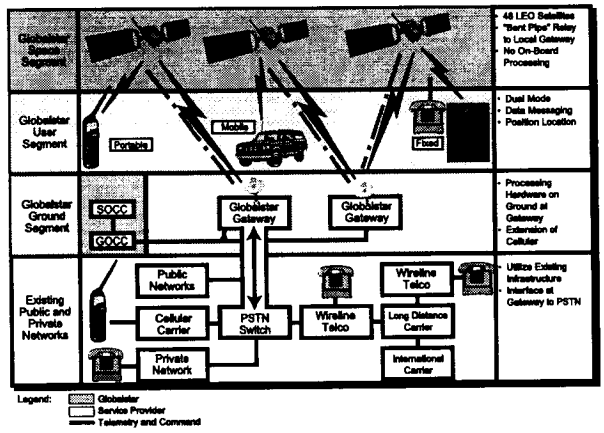


Figure 2. The Globalstar System

2.2 CDMA Operation

Code division multiple access (CDMA) is a spread-spectrum technique which was developed for cellular applications and has now been standardized by the Telecommunications Industry Association (TIA) as IS-95. The basic concept of CDMA is shown in Figure 3. With minor modifications the technique is well-suited to the mobile satellite applications [2].

Globalstar contains a forward link, which consists of an uplink from a Gateway ground

station and a downlink to a UT, and a return link, which consists of an uplink from the UT and a downlink to the Gateway.

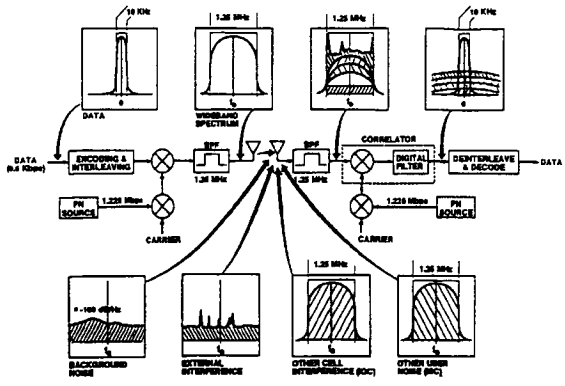


Figure 3. The CDMA Concept

Globalstar makes extensive use of diversity, where a particular call circuit is put through two or more satellites simultaneously. This allows a mobile user, for example, to drive past a building or row of trees and lose one signal completely, but maintain contact through the second and/or third signals, preventing signal dropouts. A RAKE (multi-channel digital) receiver is used to combine these signals at both ends of the link. A typical forward link for a telephone conversation would have an EIRP from the Gateway of 41 dBW and the signal would be received at the UT with an Eb/No of 3.9 dB with two circuits operating. For the return link, the UT typically has an EIRP of -7.3 dBW, and the return link Eb/No is 5.7 dB, assuming satellite diversity. These values are good enough for high-quality audio reception

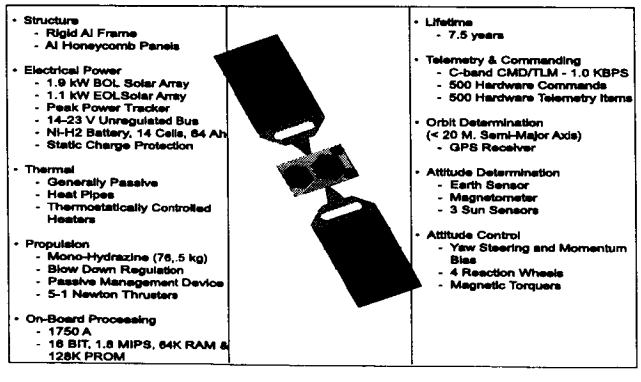


Figure 4. Satellite Characteristics

The Globalstar Air Interface (the specification for the link operation) specifies a forward link CDMA waveform that uses a combination of frequency division, pseudo-random code division (Welsh code), and orthogonal signal multiple access techniques. Frequency division is employed by dividing the available spectrum into nominal 1.23 MHz bandwidth channels. Normally, an MSS Gateway would be implemented in a beam service area with a single radio channel until demand requires additional radio channels. One Walsh circuit has a maximum rate of 4.8 kb/s.

The Globalstar Return CDMA Channel also employs PN spreading using a quadrature spreading code of length 2^{15}. Here, however, a fixed code time offset is used. Signals from different handsets (UTs) are distinguished by the use of a very long (2^{42}-1) PN sequence whose time offset is determined by the user address. Because every possible time offset is a valid address, an extremely large address space is provided. This also provides a high level of privacy.

To maximize capacity, CDMA requires that the Eb/No received from all UTs is at a similar level at the Gateway. UTs transmitting more power than normal create more interference which reduces the system capacity. To resolve this problem, Globalstar uses dynamic power control on the Return Traffic Channel.

Closed loop power control is also used. For closed loop control, the Gateway compares the signal-to-noise ratio from the UT to a threshold and sends out a command to have it increase or decrease its transmitted power until the ratio corresponds to the threshold in the Gateway. Its time period is typically 50 milliseconds, meaning it does not fix fast fades.

3.0 Hardware Description

3.1 Satellite Design

Fig. 4 shows a drawing of the Globalstar satellite with callouts. The satellite is 3-axis stabilized with the earth-facing panel always parallel to the orbit tangent. A global positioning system (GPS) receiver is used to accurately determine

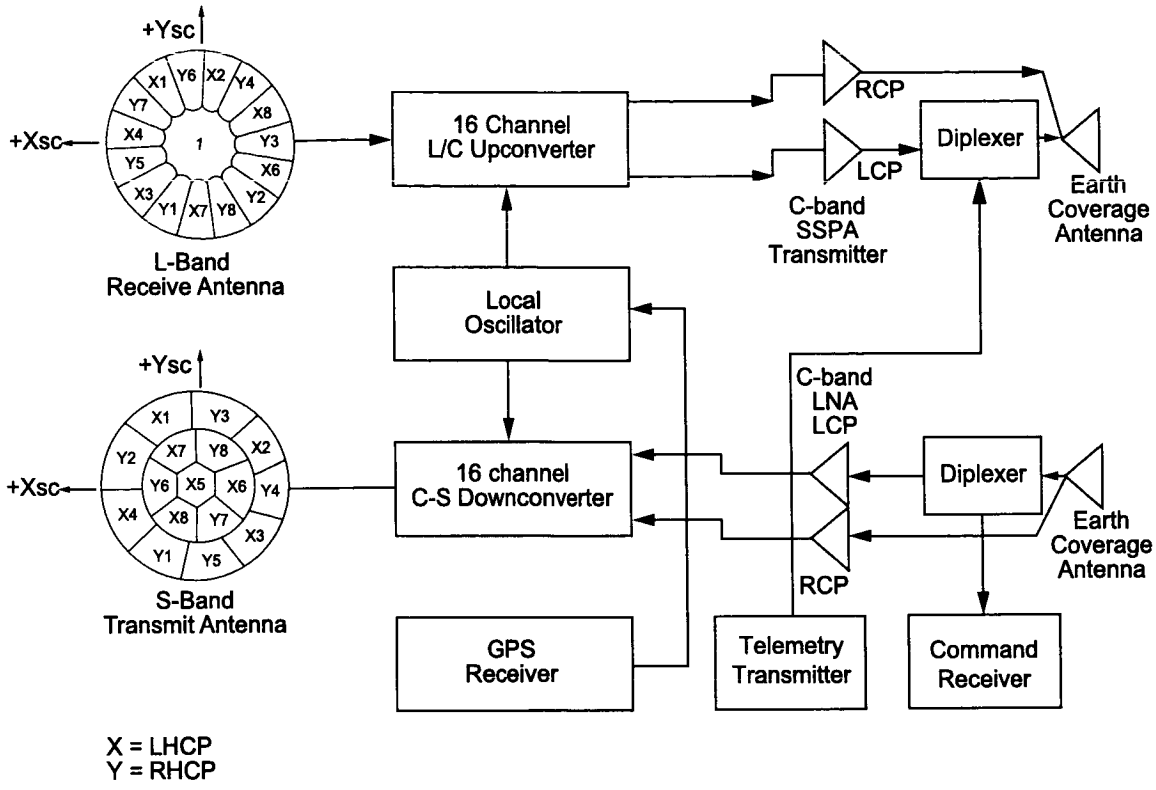


Figure 5. Globalstar Satellite Payload

the orbit parameters, and also to supply accurate time and frequency to the satellite systems. Solar panels and a large nickel-hydrogen battery provide power for all phases of the mission.

Battery recharge takes place over the oceans, where there is less traffic. The attitude control system uses small (one Newton) thrusters for attitude control. Yaw steering is employed to provide sufficient solar array power during all phases of the mission.

3.1.1 Payload Design Figure 5 shows a block diagram of the communication subsystem payload As you can see, it is a "bent-pipe" transponder. Uplink signals from the Gateway at 5 GHz are downconverted to 2.5 GHz and transmitted to the UT. The UT responds by transmitting at 1.6 GHz, which is upconverted in the satellite to a 7 GHz downlink. Switchable gain control and filtering are provided in each of the 16 channels for circuit control.

3.1.2 L- and S Band Antennas The objective of the L- and S-band array antennas for Globalstar is to form 16 independent, fixed beams each for receive and transmit that cover the visible earth. They provide frequency reuse on the beams to allow reusing the narrow spectrum to a separate transponder in the satellite. The approach taken for implementing these beams is with a separate power divider for each beam, which shapes and steers the beam, and combiners/dividers that share these beam dividers among the 16 channels. The arrays contain active circuits—power amplifiers (PAs) and low noise amplifiers-(LNAs) [3].

Figure 6 shows a block diagram of the S-band array. It shows single signal paths to two elements. The operation of the entire array can be deduced from this diagram. The transmit array forms 16 simultaneous circularly-polarized beams, a central beam and two rings of outer beams, as shown in Figure 7a (predicted) and 7b (measured). It can be seen that the desired pattern shape was achieved quite accurately. Figure 8 shows a photograph of this array. The

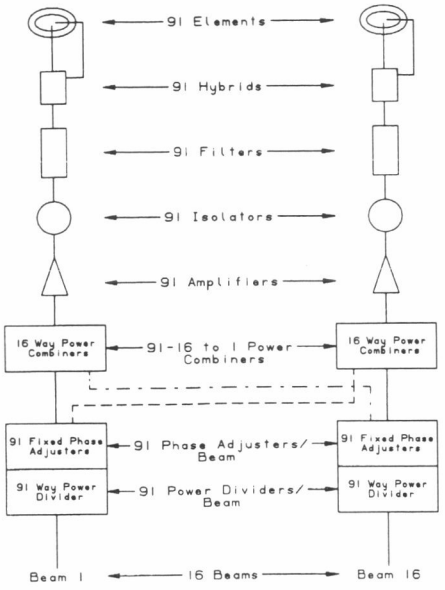


Figure 6. Diagram of S-Band Phase Array

hexagonal shape of the antenna results from the use of an equally-spaced triangle lattice array of 91 radiating elements. The array uses phase only, not amplitude, to form the beams.

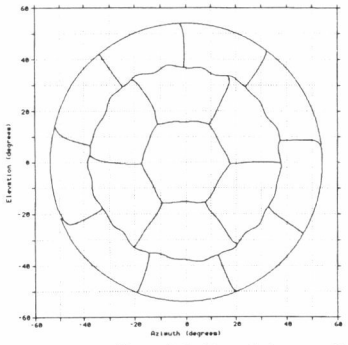


Figure 7a. Predicted S-Band Array Pattern

The block diagram for the L-band receive array is identical to that for the S-band array except that the amplifier is turned around; that is, it receives a signal from the radiating element and sends it to the power divider network, and it is a low noise amplifier (LNA). Also, there are only 61 instead of 91 elements/amplifiers. The beam shapes are different from those of the S-band array, consisting of a central beam and one ring of 15 outer beams.

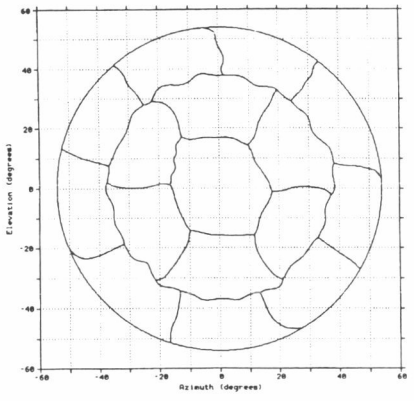


Figure 7b. Measured S-Band Array Pattern

3.1.3 C-Band Isoflux Horn Antenna The C-band antennas are designed to have a pattern that compensates for the space loss variation that occurs on paths from the edge of the earth (10° elevation, maximum path length/loss) to straight down, which is the minimum length/loss. Figure 9 shows measured patterns for both transmit and receive.

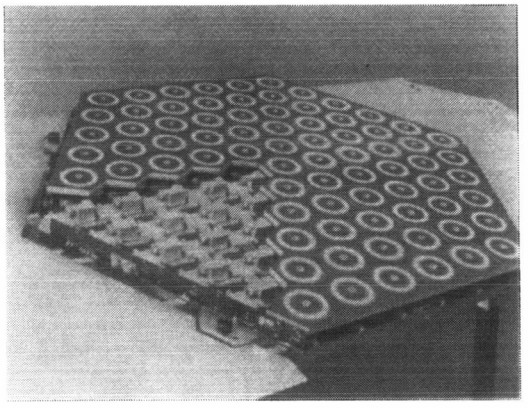


Figure 8. Photograph of S-Band Array

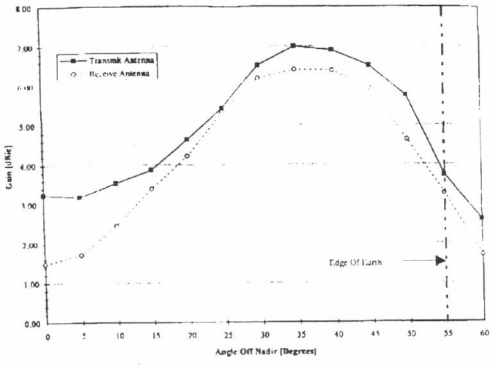


Figure 9. Measured Pattern of C-Band Horn Patterns

3.2. Gateway Design

A block diagram of the Gateway can be seen in Figure 10. The antenna is approximately 6 meters in diameter. It contains all the electronics necessary to perform the CDMA communication, including RAKE receivers (multi-channel digital receivers). It also connects to the PSTN through a switch, and also provides a GSM interface (GSM, or Global System for Communications, is the European cellular standard).

3.3. User Terminal (UT)

The UT will typically be a hand-held dual-mode unit, although a variety of single-, dual- and tri-mode units will be available, operating on both the Globalstar system and one or more of several terrestrial cellular systems. In Globalstar operation the UT will transmit an average EIRP of about -10 dBW (maximum -4 dBW), and contains a 3-channel RAKE receiver so that it

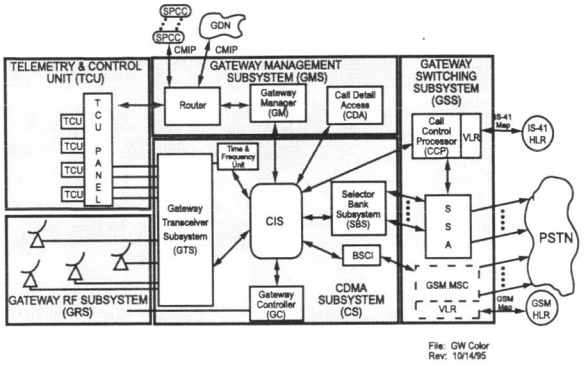


Figure 10. Gateway Block Diagram

can receive signals from more than one satellite simultaneously. Globalstar will also produce fixed user terminals, which will typically be a solar-powered phone booth in a village.

User Terminals will be manufactured by Qualcomm, Telital, and Ericsson. A photograph of a UT can be seen in Figure 11.

3.4. World-Wide Network

The Globalstar Data Network (GDN) provides interconnection of the Gateways to the Globalstar Operations Control Center (GOCC), and to the Satellite Operations Control Center (SOCC). This will be a network of links connecting the various sites, including all the Gateways. The GOCC and SOCC are located in the Globalstar

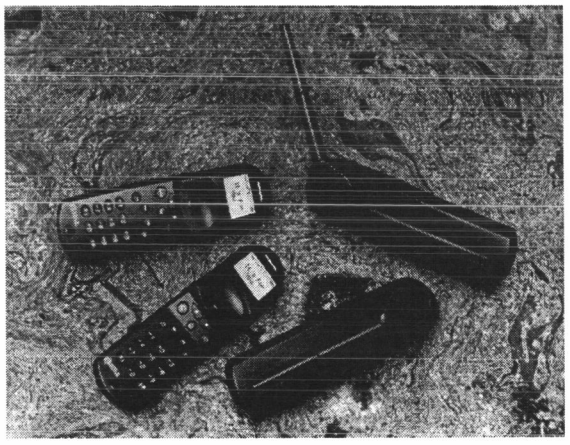


Figure 11. Photograph of User Terminals

facility in San Jose, with an alternate site near Sacramento.

4.0 PROGRAM STATUS

Globalstar is well on its way toward the goal of full service in January 1999. Satellites are being integrated at Alenia in Rome, Gateways are being installed in several locations around the world, and UTs are being manufactured by at least two companies.

Figure 12 shows the qualification model payload undergoing testing at Alcatel. It has passed all of the communication tests imposed on it. Figure 13 shows the Alcatel Gateway antenna. It uses an altitude-azimuth-cross elevation mount. Figure 14 shows the near-field antenna range at Alenia. Figure 15 shows the GOCC/SOCC Operations Center in San Jose.

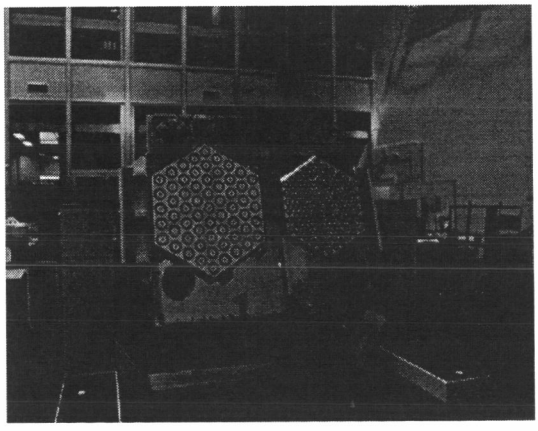


Figure 12. Qual Model Satellite Payload

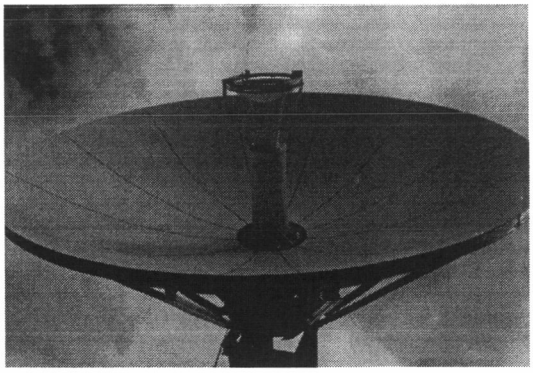


Figure 13. Gateway Antenna Photograph

5.0 SUMMARY

Globalstar is a technologically superior mobile satellite communication system that is on schedule and only a year away from beginning preliminary service. With over 80% of the required $2.5 billion funded, and with service provider agreements in 104 countries, Globalstar will soon have the world talking.

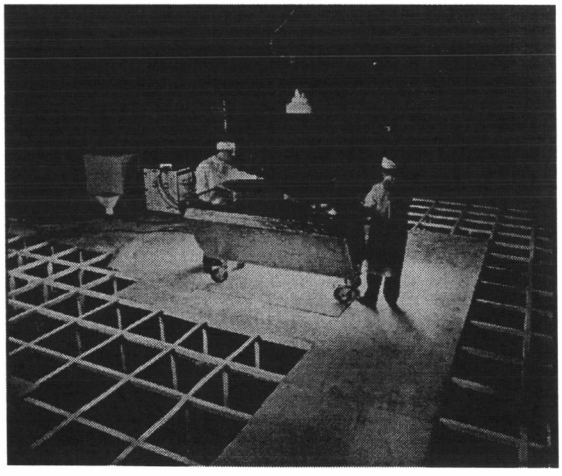


Figure 14. Near Field Antenna Range at Alenia

Figure 15. GOCC/SOCC Operations Center

REFERENCES

[1] **F. Dietrich**, *Globalstar Update*, presented at WESCON96, Anaheim, CA, October 22-24, 1996.

[2] **P. Monte and S. Carter**, *The Globalstar Air Interface: Modulation and Access*, presented at the AIAA 15th International Communication Satellite Systems Conference, San Diego, CA, February 28-March 3, 1994.

[3] **P. Metzen**, *Satellite Communications Antennas for Globalstar,* presented at the JINA International Symposium on Antennas, Nice, France, November 12-14, 1996.

The Asia Cellular Satellite System

Nam P. Nguyen, Pedro A. Buhion Jr., Adi R. Adiwoso

P.T. Asia Cellular Satellite
Jakarta, Indonesia
Phone: +62-21-522-2050
Fax: +62-21-522-2044

ABSTRACT

The Asia Cellular Satellite System (ACeS) is being developed to serve a market of three billion people, many of whom has no access to telecommunications services. The ACeS system is a mobile-satellite communications system providing digital voice, facsimile, data transmission and paging services to handheld, mobile and fixed user terminals. The system design is based on the Global System for Mobile (GSM) standards to take advantage of proven technologies and minimize development risks. In addition, the system will integrate the operation of satellite-mobile communications with terrestrial cellular networks, allowing users to roam seamlessly between ACeS and terrestrial celluar networks by automatically routing calls to a single number. The first ACeS satellite, the Garuda-1 satellite, is scheduled for launch in September 1998. Following the completion of in-orbit tests, the ACeS system will become operational in early 1999.

1.0 INTRODUCTION

This paper provides an overview of the ACeS system. It describes the three segments of the system: space, ground and user. The space segment comprises of one or more Garuda satellites and a Satellite Control Facility. The Garuda satellite is being built on the Lockheed Martin's state-of-the-art A2100AX bus. It is equipped with a digital channelizer, a beam forming network and multiport power amplifiers which allows switching of satellite power and bandwidth resources to the spot beams with high traffic demands.

The ground segment comprises of one Network Control Center (NCC) and a number of Gateways. The NCC performs the overall control and management of the ACeS system. The gateways, located anywhere the service area, provide the primary interface between ACeS and other telecommunications networks including Public Switched Telephone Networks (PSTN), Public Land Mobile Networks (PLMN) and Private Networks (PN).

The ACeS user segment encompasses three types of user terminals (UT): handheld, mobile and fixed. The ACeS handheld UT is dual-mode -- satellite/terrestrial cellular -- with automatic or alternate methods of mode selection under user control. The handheld UT is comparable in size and weight with existing terrestrial cellular phones. The mobile and fixed UTs are essentially the same as the handheld UT except that these terminals have larger antennas with higher directivity and gain.

2.0 ACeS SERVICES

ACeS will offer standard GSM service features plus other features specifically designed for the ACeS system, including:

- High quality digital voice.
- Group 3 facsimile.
- 2.4 kbps full duplex data.
- DTMF signaling.
- Standard GSM features for call transfer, call waiting, call holding, conference calls, three-party service, call barring, operator intervention, operator assistance and operator call trace.
- A special feature called "High Penetration Alerting" informs the subscriber that a call is being attempted but cannot be connected due to signal blockage. The subscriber can then move to a line-of-sight location to receive the incoming call.
- Seamless roaming throughout the service area and into terrestrial cellular networks by ACeS subscribers, all with a single telephone number.
- Roaming of terrestrial cellular subscribers into ACeS system with the use of an ACeS dual-mode handheld user terminal and the subscriber identity module (SIM).

Optional features including short message service, voice mail, store and forward fax and high power

paging (which is similar to the high penetration alerting feature) will be offered in a future date.

3.0 SERVICE AREA

ACeS focuses specifically on the unique needs of the Asian market that has a combined population of three billion people, many of whom have no access to telecommunications services. ACeS implements key services and features and provides low cost handheld user terminals to satisfy the potential large demand for wireless telecommunications within ACeS' coverage area. Figure 1 shows the ACeS' coverage area which encompasses China and Japan to the North, Indonesia to the South, the Philippines to the East, India and Pakistan to the West, as well as some twenty other countries in Asia.

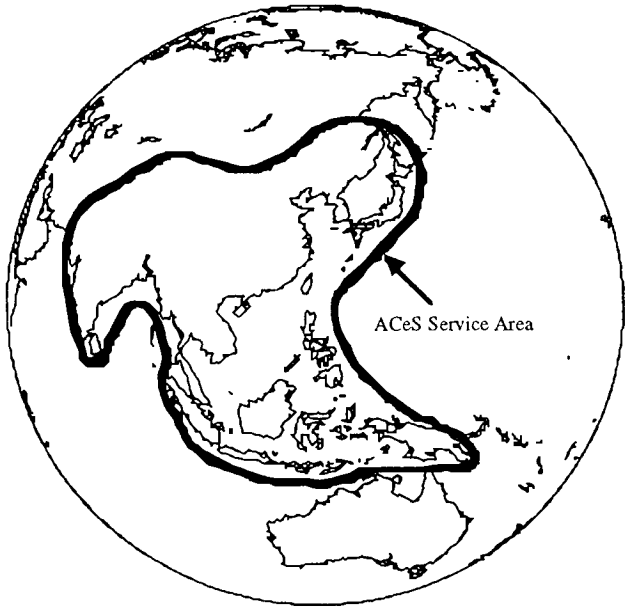


Figure 1: ACeS Service Area

4.0 SYSTEM OVERVIEW

The ACeS system is comprised of several innovative design features not available in existing mobile satellite systems, which enable ACeS to provide high quality, low cost mobile communications services to handheld user terminals. The communications link is established via the Garuda satellites which incorporates state-of-the-art technologies and production techniques to maximize the communications capability. Figure 2 shows a block diagram of the ACeS system architecture. The ACeS system consists of three segments - space segment, ground segment and user segment. The space segment includes one or more geosynchronous satellites and a Satellite Control Facility (SCF).

The ground segment consists of a Network Control Center (NCC) which provides overall management and control of the ACeS system. The NCC and SCF will be located in Batam Island, Indonesia. The ground segment also includes one or more Gateway stations which provide connection to terrestrial networks and management of the ACeS subscribers. ACeS plans to commence its operation with three gateways, one in each of the founding partner's countries -- Indonesia, the Philippines and Thailand. The ground segment has two types of customer management databases: ACeS Customer Management Information System (ACMIS), which is located in the NCC, and Gateway Customer Management Information System (GCMIS) which is located at each of the gateways.

The user segment consists of three types of user terminal -- fixed, mobile and handheld terminals. The handheld terminal is dual mode - satellite mode and cellular terrestrial mode.

The different elements of the ACeS system are connected together via the radio links using the TDMA/FDMA signaling protocols defined in the ACeS Satellite Air Interface Specifications. This proprietary satellite air interface is based on GSM signalling. There are four different connectivities within the system:

- The gateways and NCC transmit with the user terminals using the satellite C-band to L-band link. This is called forward link.
- User terminals transmit to gateways and NCC using the satellite L-band to C-band link. This is called return link.
- ACeS also provides user terminal to user terminal (i.e., mobile-to-mobile) connection via single-hop L-band to L-band link. This is accomplished by making the L-to-L cross-strapping within the digital channelizer on board the satellite.
- The system control and management signaling between the gateways and the NCC is performed using the satellite C-band to C-band communication links. Alternatively, this link can also be carried by terrestrial facilities, if necessary.

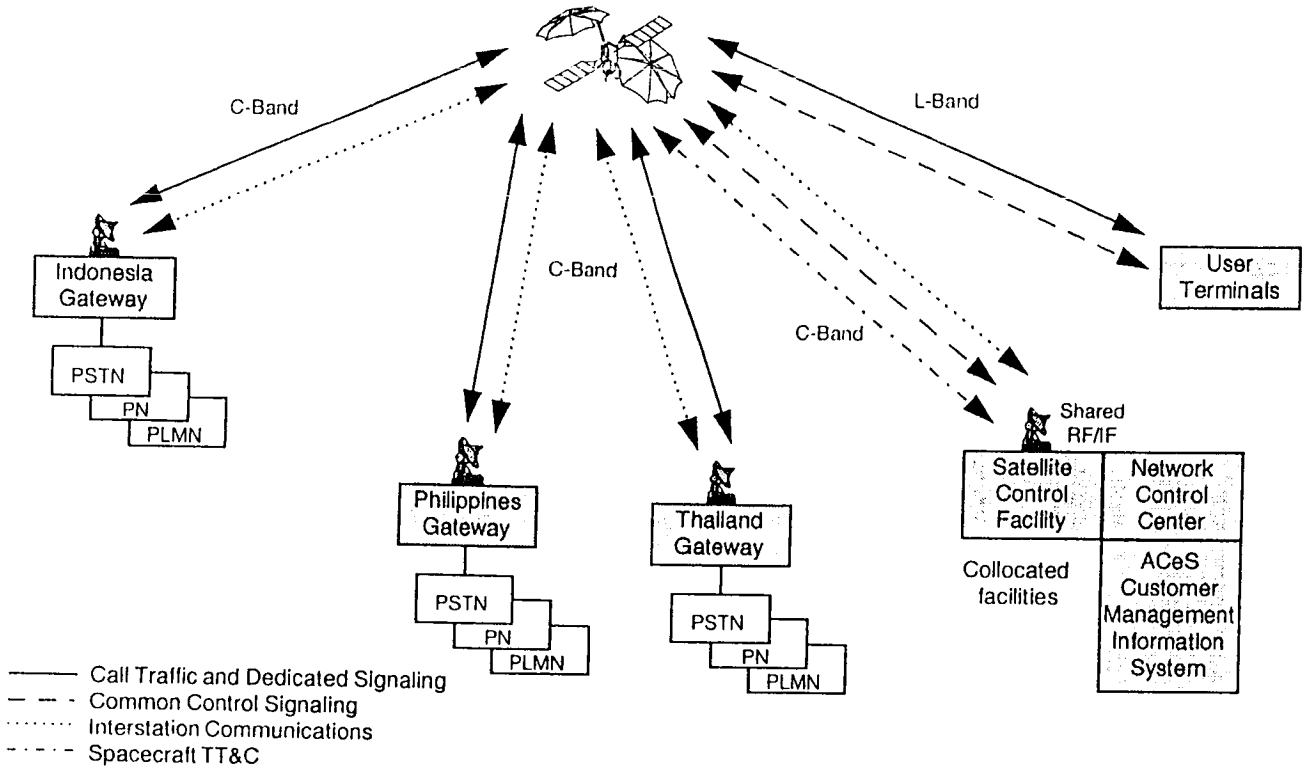


Figure 2: ACeS System Architecture

5.0 THE GARUDA SATELLITE

The first ACeS satellite, Garuda-1, utilizes the state-of-the-art Lockheed Martin's A2100AX spacecraft bus and accommodates a C and L-band communications payload. The satellite is designed for a twelve year service life. Figure 3 shows the deployed configuration of the Garuda-1 satellite.

The Garuda-1 satellite receives signals from the NCC and gateways in the extended C-band 6425-6725 MHz and transmits to back these earth stations in the 3400-3700 MHz frequency band. The communications links to the user terminals are in the conventional L-band, 1626.5-1660.5 MHz receive and 1525-1559 MHz transmit.

The ACeS C-band coverage area is provided by a single C-band beam produced by an antenna subsystem with a dual-surface shaped reflector. The L-band coverage has 140 spot beams which are formed by separate transmit and receive antenna subsystems. Both of the L-band transmit and receive antenna reflectors are 12 meter in diameter. This multi-beam configuration can support a frequency reuse factor of 20 using a 7-cell frequency reuse pattern.

The payload provides an L-band G/T of more than 15 dB/K and an L-band total EIRP of at least 73 dBW. The satellite is capable of supporting 11,000 simultaneous voice communication links to handheld terminals with a 10 dB link margin. However, there are enough transponders on the satellite to accommodate more than 28,000 voice circuits with lower link margins.

The digital channelizer on board the satellite dynamically routes signals between the C-band beam and L-band beams and vice versa. The channelizer also routes L-band frequency subbands between L-band beams to support single-hop mobile-to-mobile communications.

Figure 4 shows a simplified block diagram of the communications payload of the Garuda-1 satellite. The signal from the gateway is received by the C-band antenna, downconverted and passed through the digital channelizer where the channelization and signal

switching take place. Depending on the configuration of the channelizer, the signal may be routed to the C-band upconverter for C-to-C communications, or to the L-band upconverter and transmit beam forming network for the C-to-L forward link. On the return link, the L-band signal from the user terminal is received through the L-band antenna and passed through the receive beam forming network before being processed by the channelizer. The signal then may be routed through the L-band upconverter and transmit beam forming network for single-hop mobile-to-mobile communications, or through the C-band upconverter for L-to-C communications.

ACeS employs separate L-band transmit and receive antenna subsystems to reduce the risk of passive intermodulation (PIM) problems. The L-band feed arrays and reflectors are identical for both transmit and receive. Each feed array is machined from a single block of aluminum alloy. The reflectors are 12 meters in diameter and are based on the flight proven TDRSS deployable mesh reflector technology. Edge mounting of the reflector is used to reduce mass and folding ribs are used to minimize the stowed envelope. The reflectors are folded and stowed against the East and West side panels of the spacecraft for launch and are deployed in orbit in a controlled, low energy sequence. The reflector surface is made of a gold-plated molybdenum wire mesh.

5.1 L-band Transmit Antenna Subsystem:

The L-band transmit antenna subsystem forms the 140 L-band transmit beams over the ACeS service area. Figure 5 shows the block diagram of the L-band transmit subsystem. The low-level beam forming network provides the amplitude and phase weighting for each beam and the multiport power amplifiers provide the power needed to give the required EIRP per beam. With the butler matrix input and output networks, the feed elements are shared between beams and the power is shared between the multiport power amplifiers. This enables a high degree of power distribution flexibility between beams to accomodate traffic variation among the beams while minimizing the number of feed elements and power amplifiers. In addition, the use large number of feed elements also allows efficient shaping of the beam pattern to reduce the co-channel interference.

5.2 L-band Receive Antenna Subsystem:

The basic structure of the L-band receive antenna is shown in Figure 6. Each beam is formed by combining the amplitude and phase of the output signal of a number of feed elements. Similar to the transmit antenna, feed elements are shared between the beams to provide efficient beam pattern shaping to minimize the co-channel interference.

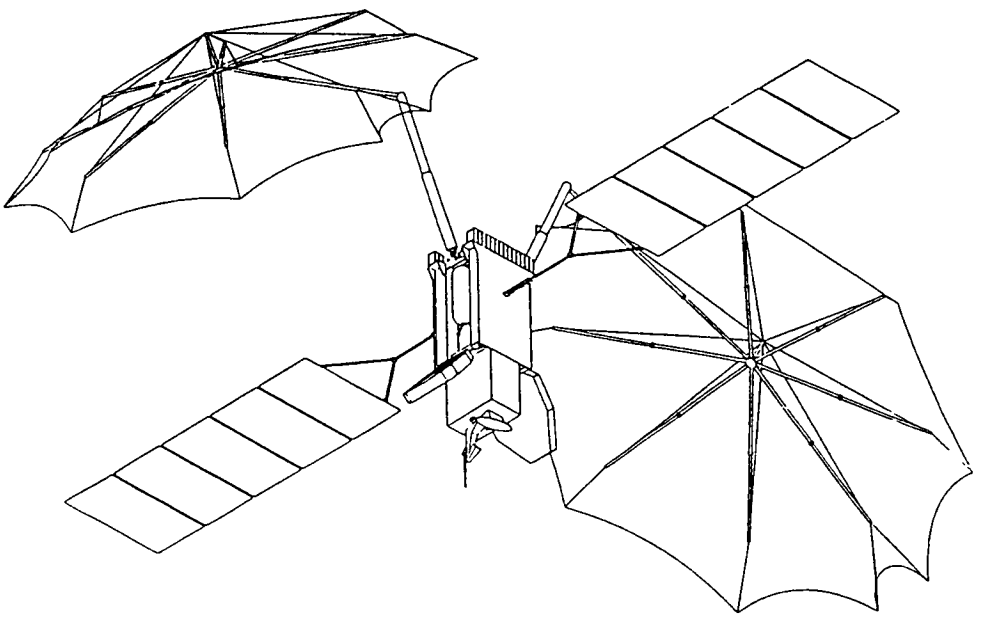


Figure 3: The Garuda-1 Satellite

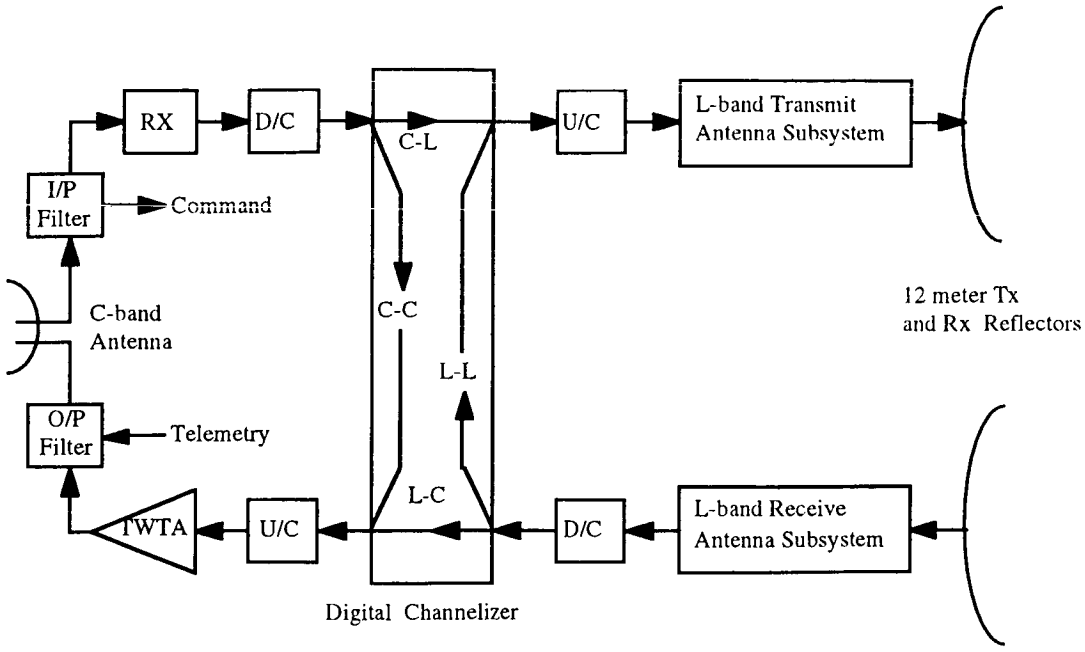


Figure 4: The Garuda-1 Payload Block Diagram

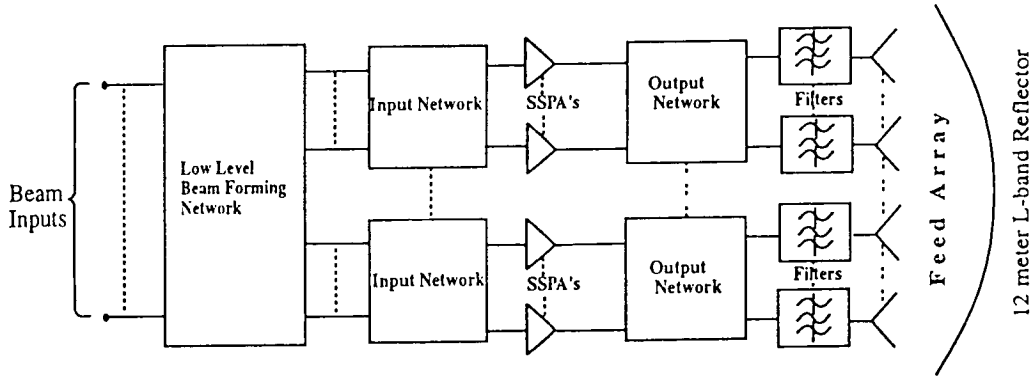


Figure 5: The L-band Transmit Antenna Functional Block Diagram

5.3 Digital Channelizer:

A digital channelizer on board the Garuda-1 satellite is utilized for the channelization and routing of the frequency subbands between the C-band and L-band beams. The channelizer can route a number of forward C-band frequency subbands to the same number of return C-band frequency subbands for signaling and control communications between the NCC and the gateway stations. Other forward C-band frequency subbands are routed to the forward L-band frequency subbands in individual L-band spot beams for communications between the NCC or gateways and the user terminals. The return L-band frequency subbands from each L-band spot beams are routed by the channelizer to either the return C-band for communications between user terminals and gateways or NCC, or to the forward L-band for single-hop mobile-to-mobile communications within the same beam or between any two L-band beams. The routing table in the channelizer, which determines the routing of each frequency subband is fully configurable through ground commands from the NCC via the SCF.

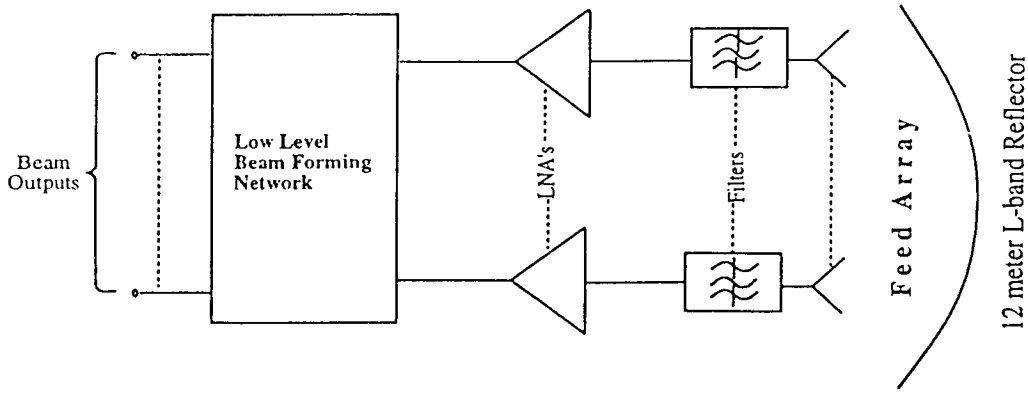


Figure 6: The L-band Receive Antenna Functional Block Diagram

6.0 THE SATELLITE CONTROL FACILITY

The ACeS Satellite Control Facility (SCF) will be located in Batam Island, Indonesia. The SCF is consist of the hardware and software necessary for the control of the Garuda satellites. In addition to the functions of a conventional satellite control facility, the SCF also generates commands to control and configure the digital channelizer. The channelizer configuration commands are generated according to the instructions from the NCC. The SCF also passes telemetry data related to channelizer status and configurations to the NCC.

7.0 ACeS NETWORK CONTROL CENTER

The ACeS Network Control Center is co-located with the SCF in Batam Island, Indonesia. The key functions of the NCC include monitoring and configuring the channelizer routing table; control of frequency reuse pattern; allocation of network resources to the gateways; generation of network timing and synchronization; generation of system broadcast and paging signals; support of call set-up process; monitoring of operational status of all gateways; and monitoring of network loading. Associated with the NCC is the ACeS Customer Management Information System (ACMIS). The primary function of ACMIS is to collect satellite resource usage information and perform related accounting functions such as generating bills to ACeS' customers, i.e., the gateways.

Figure 7 shows the block diagram of the NCC which consists of the following subsystems: Antenna IF/RF, Signaling Channel Equipment (SCE), Access Control Equipment (ACE), Network Synchronization Subsystem (NSS), Interstation Communications Subsystem (ICS) and ACMIS, all connected together by an ethernet local area network (LAN).

The Antenna RF Subsystem is used for transmission and reception of signals to and from the Garuda satellite at the extended C-band. It shares a 15 meter tracking antenna with the SCF.

The SCE's basic function is to provide the modulation and demodulation of signaling messages between the NCC and the user terminals via the Garuda satellite. The signaling includes system broadcast, access and paging channels.

The ACE consists of two computers. The primary function of one computer is to perform call set-up control. The other performs network management and network resource allocation. Both computers are fully redundant for high system availability.

The NSS provides the master reference for network TDMA synchronization. The NSS consists of satellite ranging equipment, timing pulse generators and GPS equipment.

The ICS consists of satellite modems that facilitate communications between the NCC and the gateways via the C-band to C-band satellite connection.

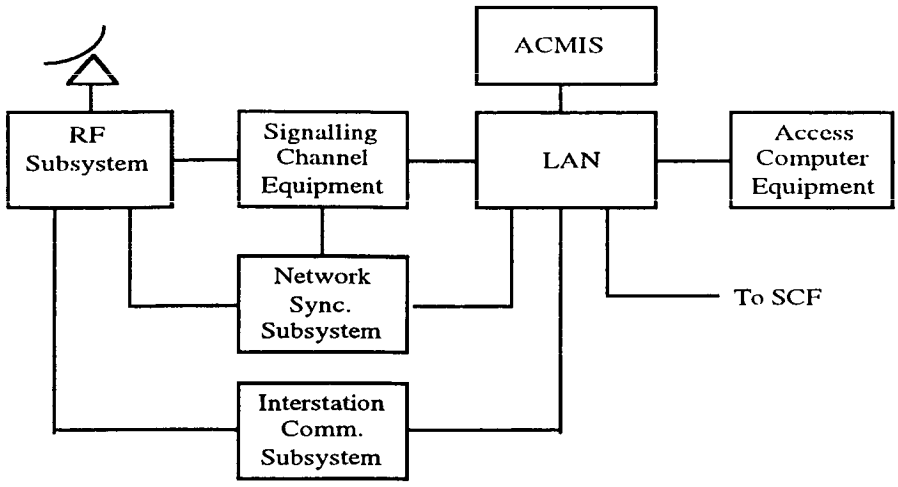


Figure 7: ACeS Network Control Center Functional Block Diagram

8.0 ACeS GATEWAY STATION

The ACeS gateway provides connectivity to terrestrial networks. It also performs call set-up in conjunction with the NCC. The gateway manages its own resources and assigns frequency subbands and TDMA time slots, which are allocated to it by the NCC, to users on a call-by-call basis. The gateway also performs frequency and timing offset control to remove frequency shift due to doppler effect and the effects of propagation delay variation among the user terminals. Other functions performed in the gateway mobile switching center (MSC) include mobility management and fraud protection through user authentication processes.

Initially, ACeS will start its operation with three gateways located in Jakarta, Manila and Bangkok. Figure 8 shows the functional block diagram of the gateway.

The Antenna RF Subsystem is used for transmission and reception of signals to and from the Garuda satellite at the extended C-band. It has a 13 meter tracking antenna.

The Traffic Channel Equipment's (TCE) main function is to provide the modulation, demodulation and vocoder processing of the voice signal. Data and fax traffic will bypass the vocoder function in the TCE. The TCE also performs power control function that would save satellite power as well as the user terminal power.

The Gateway Station Controller (GSC) performs call set-up and tear-down process in conjuction with the NCC. It assigns the gateway resources to calls being set-up and tears down calls being terminated.

The Mobile Switching Center (MSC) provides the connectivity between terrestrial networks and the ACeS network. The MSC handles all of the SS7 signaling capabilities with the fixed network for the process of connecting calls. The Home Location Register (HLR), Visitor Location Register (VLR), and Equipment Identity Register (EIR) are used for mobility management and authentication of users.

The NSS and ICS were described earlier in Section 7.0. The gateway NSS is slaved to the NSS at the NCC.

The Gateway Customer Management Information System (GCMIS) is a customer management information system located at each of the gateways. GCMIS will be responsible for maintaining a detailed database of subscribers and user terminals and subscribers' accounting management.

9.0 ACeS USER TERMINALS

There are three different types of ACeS user terminals (UT): handheld, mobile and fixed. Handheld UTs will be similar in size and weight to those used in terrestrial cellular networks. They offer all ACeS service features including a port for facsimile and data services. Both the ACeS handheld and mobile UTs will be dual-mode. This dual-mode capability will allow ACeS users to roam between the ACeS and terrestrial cellular networks. The handheld UTs' antenna has a hemispheric pattern whereas the mobile

UTs have higher antenna directivity to reduce satellite power requirement.

Fixed UTs are to be installed in locations where there is limited, or no access to terrestrial telecommunications. They utilize higher antenna directivity allowing efficient use of satellite power.

10.0 ACeS BUSINESS MILESTONES

The ACeS system concept was originated in 1993 by P.T. Pasifik Satelit Nusantara (PSN) of Indonesia and was developed throughout 1994 by PSN along with Philippine Long Distance Telephone Co. (PLDT) of the Philippines and Jasmine International Overseas Co., Ltd. of Thailand. The three companies completed a joint venture agreement and P.T. Asia Cellular Satellite was incorporated in June 1995 in Indonesia.

On July 5, 1995, Lockheed Martin Corporation of Bethesda, Maryland, was awarded the contract for the construction of the Garuda satellites, the Satellite Control Facility, the Network Control Center and the three gateways.

September 24, 1996 marked the completion of three significant program milestones with the signing of the user terminal, launch vehicle, and financial contracts in Jakarta. ACeS signed the user terminal contract with L.M. Ericsson of Sweden for the design, development and manufacturing of the ACeS dual-mode handheld user terminals. The second key milestone was the signing of the launch vehicle contract with Lockheed-Khrunichev-Energia International, Inc. (LKE) for a Proton launch in the third quarter of 1998. LKE is a joint venture between Lockheed Martin, Krunichev and Energia of Moscow, Russia. The third key program milestone was the signing of an underwriting agreement with PT Danareksa Finance for US$250 millions.

On March 24, 1997, ACeS signed a loan agreement with a consortium of Indonesian banks, led by P.T. Danareksa Finance, completing the financing requirements for the entire ACeS program.

11.0 CONCLUSIONS

The ACeS system focuses specifically on the unique needs of the Asian market. The system will provide voice, data and facsimile services to handheld, mobile and fixed user terminals. ACeS service area includes over twenty countries with a combined population of approximately three billion people, where the majority has no access to telecommunications services. With its low infrastructure cost per square kilometer of coverage, ACeS believes that it will provide the best answer to the ever growing telecommunications demand in the Asian region.

12.0 ACKNOWLEDGEMENT

The authors would like to acknowledge and thank Vern Trail of TMC Consulting, Mehran Shariatmadar of Telesat Canada and Mike Ward of Wardcom for their valuable comments.

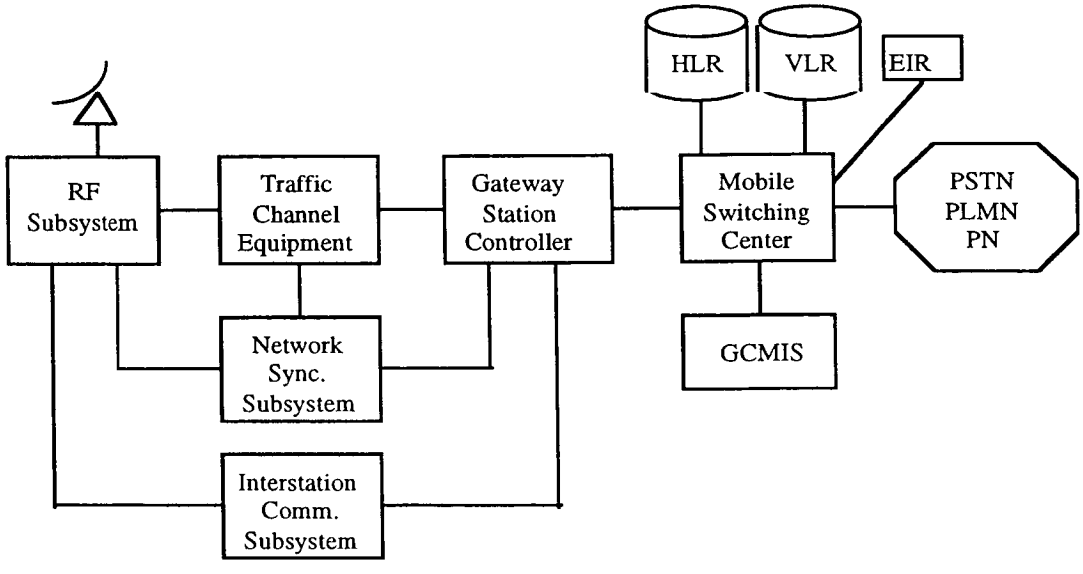


Figure 8: ACeS Gateway Functional Block Diagram

ELLIPSO™ - An Affordable Global, Mobile Personal Communications System

John E. Draim
Advisor, Constellation Design
Mobile Communications Holdings, Inc.
1120 19th St. NW, Suite 460
Washington, DC 20036
Phone: (202) 466-4488 Fax: (202) 466-4493

David Castiel
President and CEO
Mobile Communications Holdings, Inc.
1120 19th St. NW, Suite 460
Washington, DC 20036
Phone: (202) 466-4488 Fax: (202) 466-4493

W. J. Brosius
Chief Scientist
Mobile Communications Holdings, Inc.
1120 19th St. NW, Suite 460
Washington, DC 20036
Phone: (202) 466-4488 Fax: (202) 466-4493

Amb. Gerald Helman
V.P. Policy and International Programs
Mobile Communications Holdings, Inc.
1120 19th St. NW, Suite 460
Washington, DC 20036
Phone: (202) 466-4488 Fax: (202) 466-4493

ABSTRACT

MCHI is developing the Big-LEO ELLIPSO™ Global Mobile Personal Communication System (GMPCS) to provide affordable personal telephonic and data communications using hand-held (cellular-type), mobile, or fixed user terminals. These terminals will operate through a unique communications satellite constellation that employs both elliptical and circular orbits. The satellites use simple bent-pipe transponders and employ CDMA technology that permits sharing the assigned frequency bands with the other Big-LEO CDMA systems (Globalstar, Odyssey and Constellation). The use of elliptic orbits provides the major cost reductions realized by the ELLIPSO™ system, because fewer satellites (and also fewer launch vehicles) are required, as compared with circular systems. Using elliptical orbits, earth coverage can be biased by both latitude and time of day so as to match the market needs for particular geographical regions and their populations' daily work schedules. Several independent studies have concluded that the ELLIPSO™ system provides the lowest cost-per-billable-minute of any of the Big-LEO systems.

The system's basic elements are described (space segment, ground segment and user terminals). Major program milestones are listed, along with process involved in obtaining an FCC license for the system.

TARGETING THE MARKET

The implementation of satellite personal mobile communications systems appears at first blush to be directly in conflict with terrestrial based wireless phone systems (cell phones). Indeed, it is difficult to see how a satellite-based system could compete economically with a small, regional terrestrial cell-phone system located in a highly populated area. The large investment in satellites, satellite launch vehicles, and satellite gateways may not be at all justified when the cost of these assets are compared with a number of terrestrial towers located at nodes within the limited area. It begins to make much more sense, however, when the total global scenario is considered. There are, (even in developed countries such as the U.S. and Canada, and particularly in the developing nations) vast, sparsely populated regions of the world where it would be uneconomical to emplace a large number of terrestrial stations to service a limited base of subscribers. Thus, the satellite personal communication system should be viewed as a supplement to existing wireless and wireline terrestrial systems, and not as a replacement system. It has been estimated that half the world's population lives more than two hours from a telephone, and that there are 40 million applicants on the waiting lists for telephone service. Present unserved potential users for such services, globally, have created an attractive market that a satellite system can satisfy if properly implemented. It is expected that the satellite system penetration of the total market will be only 5-15%; yet, even this figure is predicated on a roughly comparable terrestrial

cellular system, with affordable cost-per-minute charges to the telephone user.

Another important factor to be considered is the geographical distribution of the market. It is common knowledge that the vast majority of the world's population resides in the Northern Hemisphere, along with most of the world's commerce and industry. See Figure 1. Thus, it makes eminently good sense to design a satellite system that correlates with this uneven global distribution pattern. [1] Elliptical orbits provide this capability by concentrating satellites near the orbit's apogee (point of greatest altitude) which is placed over the Northern Hemisphere.

ELLIPSO's elliptic orbits and subconstellations enable a staged geographic implementation. First priority is being given to the equatorial ring of satellites (CONCORDIA sub-constellation). This ring provides continuous service in temperate regions (with the capability of extending service to any user between approximately 55° North and South latitudes). Later, the two inclined planes (BOREALIS sub-constellation) will be placed into orbit, completing the highly biased coverage of the Northern Hemisphere; they continuously cover the region from 25° North latitude to the North Pole.

In bringing a new system into being, it is reasonable to adapt a phased sequence of introduction. The ELLIPSO™ system is well adapted to this approach. The variety and complexity of services will be implemented in roughly three stages. The first stage will address basic voice services (POTS), basic data services (paging) and G3 facsimile. The second stage will introduce GSM features, cellular interoperability, push-to-talk trunked radio and geo-location services. The third stage will include enhanced data services, with cellular digital packet data (CDPD) and internet gateways. The flexibility and economy of the three-plane ELLIPSO™ system[2] have been verified by subsequent studies and analyses, with improvements being made without changing the basic concept.

ELLIPSAT - A SERVICE PROVIDER

MCHI is in the process of forming value-added partnerships (VAPs) for the purpose of distributing communications services in various regions. In many cases, this will involve working relationships with the PTT's, or equivalent, in the various countries. These in-country agencies will market the service locally and participate in the billing and records-keeping functions necessary to service subscribers. In most cases, they will be the primary interface between the system operator and the end users. The ELLIPSO™ system thus provides wholesale voice and data services to the VAPs, which in turn provide retail telephony data services and bundled solutions to in-country end users (business and individuals).

FCC LICENSING HISTORY

Licensing Applications:
MCHI applied to the FCC for a license on November 5, 1990; it was the first of the Big-LEO organizations to file. Subsequent to MCHI's application, filings were made on behalf of Motorola (Iridium), Loral (Globalstar), TRW (Odyssey), Constellation and AMSC.

Spectrum Issues:
In February 1992, MCHI participated in the International Telecommunication Union's World Administrative Radio Conference in Geneva, and assisted the US delegation in persuading other governments to allocate for global use a portion of the radio spectrum for the Big-LEOs. The above Big-LEO applicants all subsequently participated in the negotiated rule-making process established by the FCC, which then made recommendations to the FCC on means for sharing the available spectrum.

Licensing Process:
Largely based on the results of the negotiated rule-making process, the FCC issued a Report and Order in October 1994 which established the rules under which licenses would be granted. These rules: (a) set a limit on the number of licenses to be granted at five, (b) required essentially global coverage by all systems (55°S through 70°N latitude), and (c) established "strict" financial standards which required applicants to demonstrate 'irrevocable commitments' to pay for the construction and launch of the entire constellation of satellites.

FCC's Financial Requirements Issue:
In November 1994 MCHI put together an extensive financial package with strong support from Westinghouse, Israel Aircraft Industries, Barclays Bank and others, believing that it had met the FCC's strict financial requirements. On January 31 of 1995, the FCC issued separate Report and Order's granting licenses to Motorola, Loral and TRW the right to proceed with construction of their systems. Also on

that date, the FCC deferred MCHI's application on grounds that it had failed to meet the FCC's strict financial standards. Having used the financial standards to determine the license grantees, the FCC stated that the technical viability and qualifications of the applicants were not considered in awarding the three licenses.

Mutual Exclusivity Issue:
Another point raised by the FCC was the 'mutual exclusivity' issue resulting from the existence of six applicants when a determination that only five licensees could be accommodated within the available spectrum. This had been agreed to by the applicants in the negotiated rulemaking process. Later in 1996 this condition no longer prevailed, due to the withdrawal of AMSC from the applicant group seeking licenses.

MCHI Lawsuit:
In 1995, MCHI filed a lawsuit against the FCC with the Federal District Court on the grounds that the negotiated rulemaking process was seriously flawed and arbitrary. In later separate petitions, MCHI held that FCC had failed, in its Report and Order, to properly take into account small business issues. The Telecommunications Act of 1996 specifically directs that the FCC identify and eliminate "market entry barriers for entrepreneurs and small businesses in ... telecommunications services and information services." [3] This MCHI lawsuit against the FCC is still pending.

Current Status:
Following the license denial of January 1995, resulting from decisions by FCC's International Bureau, MCHI then appealed to the full FCC Commission. On June 24, 1996, the full FCC Commission again denied MCHI's appeal and directed that MCHI file a revised financial showing. This was done on September 16, 1996 with further supplements submitted on November 13, 1996, and on January 27, 1997. At the time of this writing, MCHI is still awaiting licensing action by the FCC.

LOW COST TO THE USER - A NECESSITY

The key to success in the mobile satellite services arena, in the view of the developers of ELLIPSO™, is to provide communications at a low and affordable cost. It should be an 'everyman's system', not just a system for the corporate businessman, or to suit the 'lifestyle of the rich and famous' individual. At every decision stage, the ELLIPSO™ system designers confront the cost/performance question, and endeavor to ensure that the ultimate system maintains a high level of 'affordability'. This approach has led to: (a) the adoption of the technically sophisticated, geographically tailored constellation, (b) a highly automated orbit determination and station-keeping process, (c) a simplified, bent-pipe transponder satellite, (d) a standardized, 'turn-key' ground-control station capable of operation by native in-country communications technicians with a minimum of training., and (e) a dual-mode mobile transceiver (hand-held user terminal) allowing use of nearby terrestrial facilities wherever available and less expensive than going through satellites.

SYSTEM DESCRIPTION

Space Segment:
The space segment of the ELLIPSO™ system is composed of three planes of medium-altitude satellites. See Figure 2. All three planes, by design, remain approximately edge-on to the sun.[4] This simplifies the design of the solar array mechanism as well as the that of the thermal control subsystem. Since the apogee altitudes of the inclined elliptic plane satellites and the mean altitudes of the equatorial ring of circular orbit satellites are approximately equal, the same design can be used for all satellites in the system. The early baseline design has undergone several improvements to decrease the long-term perturbations (that translate into reduced station-keeping fuel requirements).[5,6] Other orbital studies have led to finding the most efficient methods for orbital insertion into the elliptical orbits.[7]

Satellites: The ELLIPSO™ satellites are of conventional design, with one large receive and one large transmit planar array antenna located on the earth-pointing face of the satellite. The satellites always face directly at earth center. Sixty-one spot beams are formed to cover the visible surface of the earth; these beams are of fixed geometry. The transmitted power level is automatically adjusted to maintain a constant power flux density on the earth's surface. A two-wing solar array is deployed for prime satellite power; this array employs gallium-arsenide technology giving extended lifetime (and minimizing degradation). All electronic devices, microprocessors, and solar arrays are designed to be highly radiation resistant. The mass of each satellite is approximately 700 kg.

Constellation Design: The design of the ELLIPSO™ constellation is unique among the Big-LEO systems in that it uses elliptical orbits. Considerable technical effort has gone into the design and refinement of these orbits to maximize the efficiency of coverage using the minimum number of satellites. The tailoring, or biasing, of earth coverage to match communications needs is only possible using an elliptical orbit constellation. Proper orientation of these elliptical orbits allows us to adjust to changing demands by latitude, longitude, and even time of day. Circular orbit constellations are inherently inefficient and wasteful, creating the need for large numbers of satellites when perhaps less than half that number would suffice using more efficient elliptical orbits.
The ELLIPSO™ constellation design is protected by U.S. Patent #5,582,367 issued on December 10, 1996. Further patents are pending.

The present baseline ELLIPSO™ configuration consists of three planes of satellites. The equatorial ring is presently configured with seven equally spaced circular orbit satellites at an altitude of approximately 8060 km (4.8 hr period). This ring essentially provides coverage for users between 55° South and 25° North latitudes. An optional orbital pattern for these equatorial satellites employs slightly elliptical orbits in which the apogees of all satellites always point towards the sun (APTS), thereby biasing coverage towards daytime hours. In essence, we have stolen some coverage from the nighttime in order to augment the daytime coverage. Of course, this correlates well with the normal communication patterns of fewer nighttime calls and an increased number of calls placed during the daytime (working) hours. In addition to the equatorial ring, there are two inclined (i= 116.58°) elliptical orbit planes containing five satellites each; the satellites in each inclined plane are equally spaced in time (or mean anomaly), so that they appear to bunch together and are much higher in the Northern Hemisphere, giving the enhanced performance in the Northern latitudes. These inclined orbits are all sun-synchronous with periods of approximately 3 hours, with one plane having a noon ascending node, and the other a midnight ascending node. This places the inclined satellite planes always edge on to the sun- year in year out. These inclined planes can also be biased in favor of daytime coverage by adjusting the argument of perigee, (this amounts to merely leaning the ellipse towards the sun in the plane of the orbit).

The ELLIPSO™ orbits in all three planes have been found to have very stable orbits, requiring very small amounts of fuel (less than 10 kg) to correct perturbations and maintain the intersatellite spacing during the 5-7 year lifetime of the satellite.

Satellite Operations: The operation and housekeeping of the ELLIPSO™ satellites will be carried out from a primary and a secondary TT&C station. Use of two stations at a mid-latitude that are spaced approximately 90° apart will provide down- and up-loading opportunities at least once per orbit for each satellite. These TT&C stations will be located on U.S. territory. Precision orbit determination will be accomplished through use of GPS receivers on board each satellite. Necessary station-keeping maneuvers will be calculated and implemented using conventional methods.

Ground Segment:
The ground segment of the system is composed of the stations and facilities necessary for maintaining the satellites, along with the stations necessary to serve the communications functions of the system both at the user and terminal ends, including the user terminals themselves.

System and Network Control: The System Coordination Center (SCC) is the headquarters and operations center for the entire system; it will be located on US territory. It functions to perform system planning and development, as well as to oversee the overall constellation control and maintenance functions. Links are provided to the TT&C stations to collect and disseminate satellite ephemeris, satellite housekeeping, and satellite corrective maneuver data to and from the satellites. The Regional Network Control Centers (RNCCs) handle the billing, configuration, data base and system management center for the communications functions; they may be segmented into several regional stations for political, geographic, or commercial reasons. The ELLIPSO™ Switching Offices (ESOs) handle real time call management and network interface functions within the system, connecting with the ELLIPSO™ Ground Control Stations (GCSs) and RNCCs. They also interface with other external systems such as PSTNs, or private networks (in either direction). The GCSs function as gateways- ground entry points for user signals- providing baseband active circuits to the ESOs, as well as being in direct contact with the TT&C stations and the controlling RNCC.

User Terminals:
The User Terminals (UTs) can be of several types (multiple configuration options). There is a dual mode option allowing the user to communicate either through a satellite or directly through an ELLIPSO™ GCS, for subsequent alternative call routing through another server such as cellular or a PSTN, whichever path is determined to be more effective and less costly. In addition, the user has the option of using a hand-held transceiver, or a mobile (vehicle-mounted) version, or in some cases a fixed remote phone booth (the "village phone").

CONCLUSIONS

The ELLIPSO™ mobile communications satellite system is an efficient, affordable system in which costs have been aggressively held down through use of elliptic constellation design, simplicity in the construction of satellites and ground terminals, and extensive labor-saving automation. These facts are borne out by the fact that the total cost of the ELLIPSO™ system is significantly less than the other, already licensed Big-LEO systems. In addition, independent studies comparing the various Big-LEO systems approaches have concluded that the ELLIPSO™ system provides superior design features. [8,9] One of these independent studies (from which Fig. 5 was taken) concludes that ELLIPSO™ indeed succeeds in providing the lowest billable cost-per-minute achievable of all the Big-LEO communications systems.[9]

REFERENCES

1. Draim, John E., and Cooper, James K., *Satellite Constellations for Optimal Northern Hemisphere Coverage (An Annotated Briefing)* ANSER C^3 and Surveillance Division Note CSDN 83-1, June 1983

2. Castiel, D., Brosius, W. J., and Draim, J.E., *ELLIPSO: The ELLIPSO Mobile Satellite System.* AIAA-94-1140-CP, Presented at the 15th AIAA International Communications Satellite Systems Conference, San Diego, CA, Feb. 28- Mar. 3, 1994

3. 104th Congress, January 3, 1996, *Telecommunications Act of 1996,* Sec. 257.

4. Castiel, D., Brosius, W. J., and Draim, J.E., *ELLIPSO: Coverage Optimization Using Elliptic Orbits* AIAA-94-1098-CP, Presented at the 15th AIAA International Communications Satellite Systems Conference, San Diego, CA, Feb. 28- Mar. 3, 1994

5. Sabol, C., Draim, J.E., and Cefola, P.J., *Refinement of a Sun-Synchronous Critically Inclined Orbit for the ELLIPSO™ Personal Communication System,* AAS Paper 95-340, AAS/AIAA Astrodynamics Specialist Conference, Halifax, Nova Scotia, 14-17 August, 1995.

6. Draim, J. E., and Castiel, D., *Optimization of the Borealis and Concordia Sub-Constellations of the ELLIPSO Mobile Communications System,* Published Paper IAF-96-A.1.01, 47th International Astronautical Congress, Beijing, China, October 7-11, 1996,

7. Sabol, C., Draim, J.E., Cefola, P.J., *Optimal Orbit Transfer to a Sun-Synchronous Critically Inclined Orbit for the ELLIPSO™ Personal Communication System,* Paper AIAA-96-3615, presented at the AIAA/AAS Astrodynamics Conference, San Diego, CA, July 29-31, 1996

8. Gaffney, L.M., Hulkower, N.D., Klein, L., Lam, D.N., *A Reevaluation of Selected Mobile Satellite Communications Systems: Ellipso, Globalstar, Iridium, and Odyssey,* Mitre Corporation MTR 93B0000157, May 1994

9. Violet, M.D., *Extension of the Cost per Billable Minute Model to Address Systems Using Highly Elliptical Orbits.* Private Correspondence with Dr. Paul Cefola, MIT, April 30, 1996

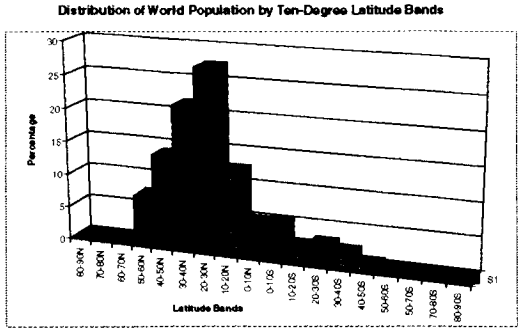


Fig. 1
Distribution of World's Population by Ten-Degree Latitude Bands

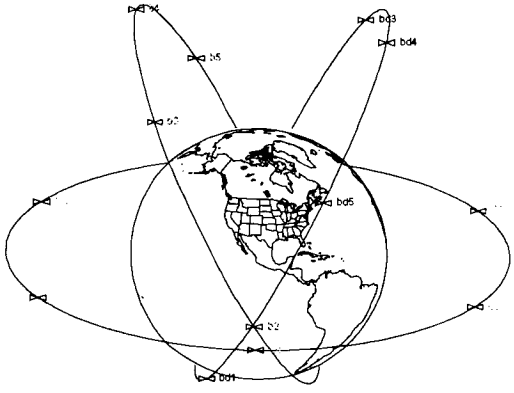


Fig. 2
ELLIPSO™ Constellation

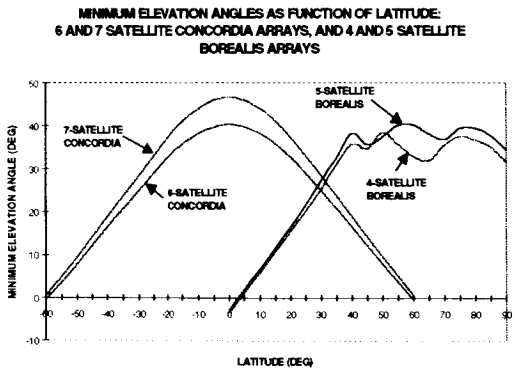


Fig. 3
Minimum Elevation Angles, 4- and 5- Satellite Inclined and 6- and 7- Satellite Equatorial Rings

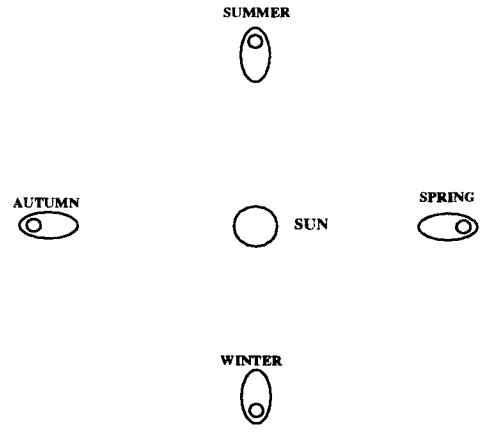


Fig. 4
Apogee Pointing to the Sun (APTS) Orbit

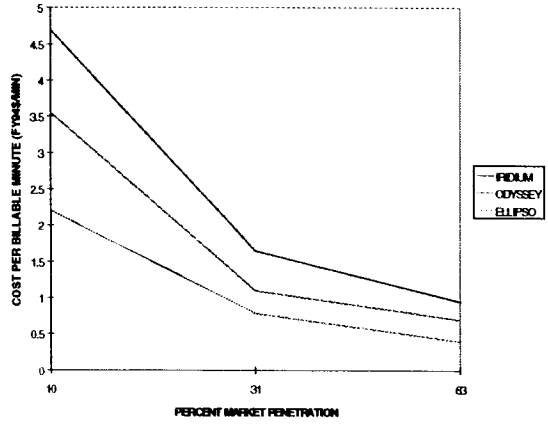


Fig. 5
Billable Cost per Minute, for Three Big-LEOs[5]

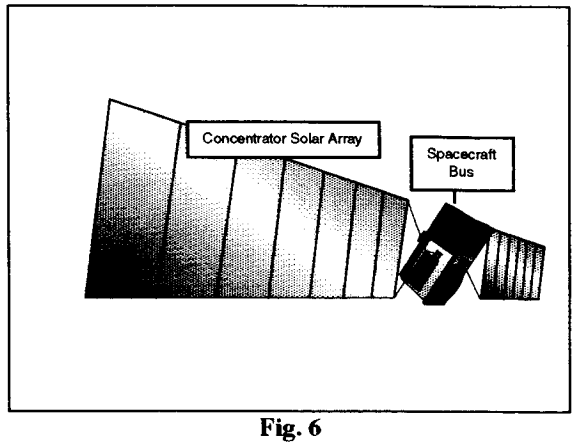


Fig. 6
ELLIPSO Satellite

The Hughes Geo-Mobile Satellite System

J. Alexovich, L. Watson; A. Noerpel*; and D. Roos*
Hughes Space and Communications Company
P.O. Box 92919
Los Angeles, CA 90009-2919
Tel: 310-662-6096 – Fax: 310-662-7060
e-mail: jralexovich@ccgate.hac.com

*Hughes Network Systems, Inc.
11717 Exploration Lane
Germantown, MD 20876
Tel: 301-428-5500 – Fax: 301-428-5707

ABSTRACT

A new generation of high performance geosynchronous satellites provides high quality voice and data links directly to handheld terminals. The ability of Mobile Satellite Systems (MSSs) to fill in terrestrial cellular networks at low cost, and also provide better roamer services has created a significant worldwide demand. These MSSs provide a cost effective and expedient solution to the telecommunication infrastructure needs in remote locations where it is not feasible to build terrestrial systems. In response to the demand for MSS services, Hughes Space and Communications Company is teamed with Hughes Network Systems to develop the Hughes Geo-Mobile (GEM) Satellite System (hereafter referred to as the GEM system). The GEM system is a fully integrated mobile communications network modeled after the Global System for Mobile Communications (GSM), and provides GSM type services to large geographic regions. This paper provides an overview of the GEM system.

1. INTRODUCTION

The GEM system, depicted in Figure 1, is an MSS that integrates a Hughes geosynchronous satellite with a

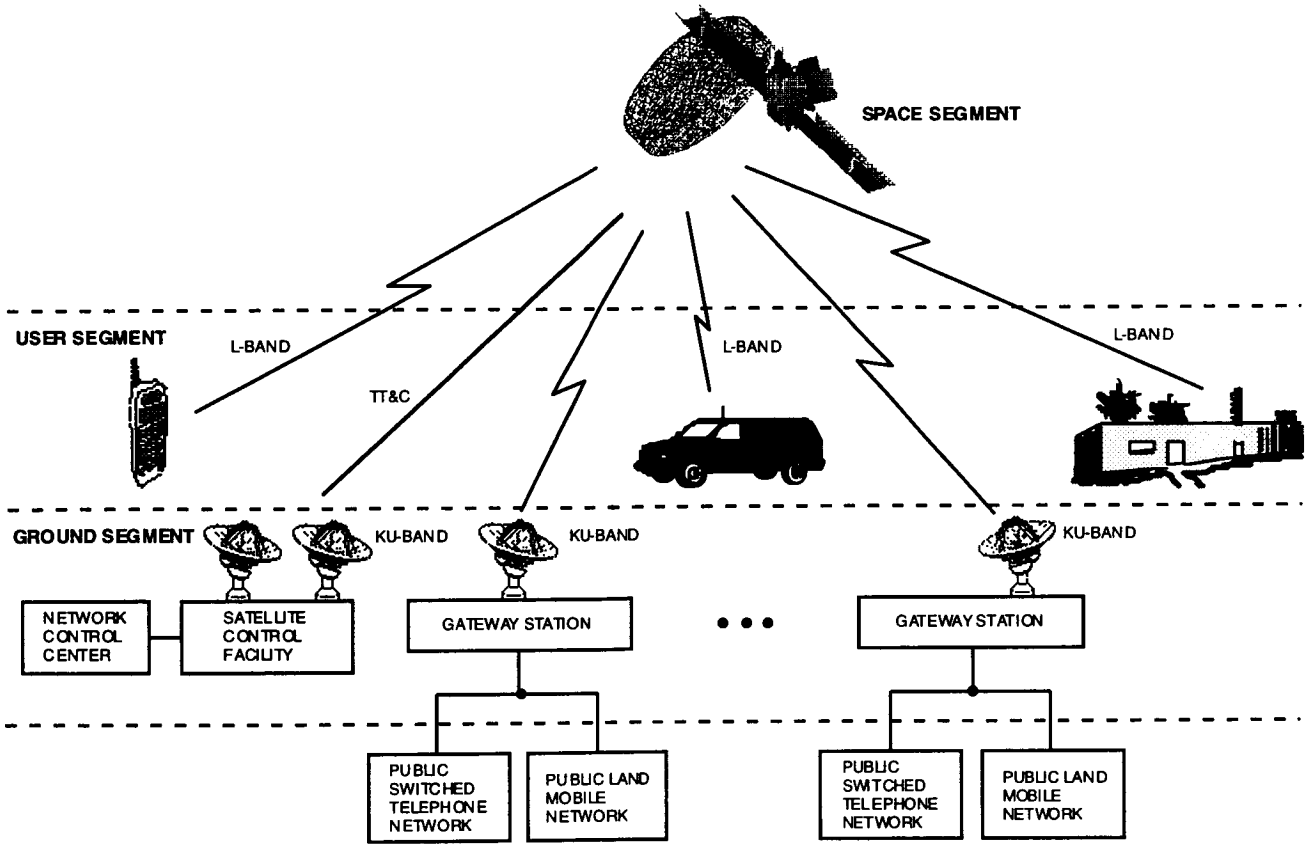


Figure 1. Hughes Geo-Mobile Satellite System

ground segment and user terminal segment to provide communications services similar to those provided by terrestrial cellular systems (e.g., voice, data, facsimile, and other supplementary services). The ground segment includes gateway stations that are interfaced with the Public Switched Telephone Network/Public Land Mobile Network (PSTN/PLMN) so that mobile subscribers can access users of the PSTN/PLMN as well as other GEM mobile subscribers from anywhere within the satellite service region. Users access the system at L-band via terminals that are similar to those used with ground based cellular systems (e.g., handsets or vehicular terminals).

The GEM common air interface (i.e., waveform and access control design for the user terminal to GEM system interface) is a modified version of the terrestrial GSM design. Modifications were made as necessary to allow compatibility with satellite operation. As a result, services and user access are provided in a manner similar to that of terrestrial cellular systems. Additionally, this approach allows low cost dual mode terminal designs that can operate with either the GEM system or with local terrestrial cellular systems.

The GEM satellite utilizes state-of-the-art digital signal processing technologies to provide onboard digital beamforming and onboard switching. These technologies provide the capability for user-to-user connectivity through the satellite, eliminating the additional delay associated with a double hop connection through a ground gateway.

The GEM system is sized to provide 16,000 voice circuits for up to 2 million subscribers. The GEM waveform and link designs are consistent with providing high quality voice in the presence of shadowing up to 10 dB for some users. The GEM system provides these capabilities while making very efficient use of valuable system resources such as the mobile link spectrum and satellite EIRP. This efficiency translates to greater capacity and higher service quality for a given resource allocation. These efficiencies are the result of a very flexible and dynamically reconfigurable system design that is embodied in the common air interface, and each of the segment designs (i.e., ground, satellite, and user terminal segments).

2. SYSTEM ARCHITECTURE

The GEM system is a fully integrated mobile communications network which consists of the space, ground, and user terminal segments shown in Figure 1. The space segment features a Hughes geosynchronous satellite with a single L-band antenna and onboard digital signal processor. The ground segment consists of a primary gateway, one or more secondary gateways, and three uplink beacon stations. The primary gateway includes a network control center (NCC) and a satellite control facility (SCF) to provide overall control of the network and the satellite, as well as communications equipment to provide PSTN/PLMN connectivity. The user terminal segment consists of handheld, vehicular, and stationary types of terminals.

The satellite provides user terminal coverage of the MSS service area with over 200 beams, each approximately 0.7 degree or 450 km across. The maximum L-band coverage is an area approximately 12 degrees in diameter as viewed from geosynchronous altitude. A representative satellite coverage area over the continent of Africa is depicted in Figure 2. Feeder link coverage is provided at Ku-band (or C-band for customers or regions requiring that band) to gateways via an area coverage antenna. L-band-to-L-band links are connected by the satellite processor to support user-to-user calls in a single hop.

The satellite also provides Ku-band to Ku-band links accommodating an intranetwork communications subsystem (INCS) which transports control information among the NCC and secondary gateways. Bandwidth is also allocated in the feeder link to support payload monitoring and control. Uplink beacon stations transmit special L-band signals that the satellite tracks to maintain precise pointing of its mobile link beams.

Approximately 2000 carriers, each with 24 TDMA time slots nominally grouped to provide eight voice channels, are available to distribute traffic to mobile users in various beams. The network control and resource allocation operations within the primary gateway and secondary gateways dynamically distribute these TDMA signals among the beams to match instantaneous traffic demand. This traffic demand can be spread nonuniformly across all of the beams covering the GEM service region. The satellite processor implements this function by mapping each unique feeder link frequency to a specific L-band frequency and beam location. This mapping function allows L-band frequencies to be reused in multiple beams across the coverage area, constrained only by the available beam isolation. The narrow beamwidth and large number of beams achieve high L-band frequency reuse.

System subscribers have a unique telephone number that allows them to receive a call when they are registered anywhere in the satellite coverage area. Registration is automatic when the mobile terminal is turned on. The mobile switch in the gateway tracks the location of the mobile terminal and announces the incoming call. Optional position determination can bar calls to or from restricted regions, promoting the free circulation of handsets.

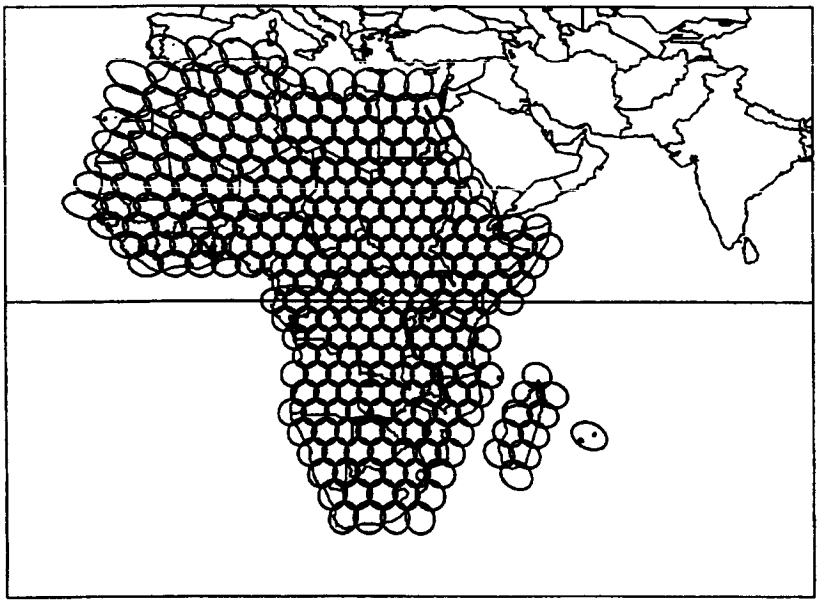


Figure 2. Representative Coverage of Africa

The system utilizes low rate encoded voice transmission and an efficient TDMA structure to pack 256 voice calls per megahertz of spectrum. The terminals may be either single or dual mode, dual mode terminals allowing communications either via the satellite or the local terrestrial cellular system.

The gateway stations and NCC elements are designed to support expandability and easy upgrades as technologies improve. To this end, the system supports additional gateways and modular expansion of capacity to individual gateways. Where appropriate, functionality is implemented in software to facilitate the upgrade process. In addition, the system can be reconfigured via the NCC to change features such as modulation, speech coding, and error coding.

2.1 Space Segment

The GEM space segment consists of a high power geosynchronous satellite based on the Hughes HS 601 bus structure, with onboard digital signal processing. To optimize the overall system design, the satellite utilizes the flexibility of digital beamforming to maintain beam positions with an inclined orbit. Large weight savings are achieved by avoiding the need to expend propellants for north-south stationkeeping while meeting performance requirements for at least a 12 year lifetime.

The communications payload, depicted in Figure 3, includes a single 12.25 meter L-band antenna for both transmit and receive and a digital signal processor. The L-band antenna, along with digital beamforming within

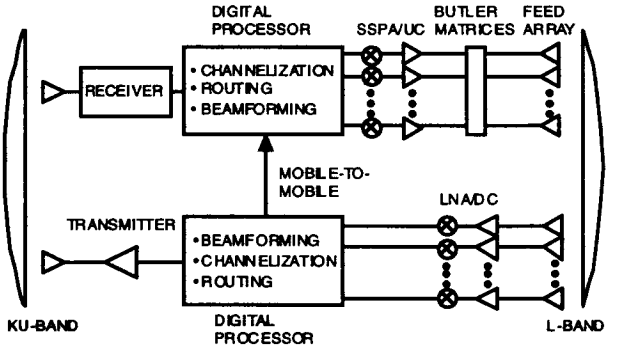


Figure 3. Communications Payload Block Diagram

the digital processor, provides several hundred narrow beams distributed over the L-band service region. Using a single reflector virtually eliminates congruency errors between transmit and receive beams. The L-band beams are populated by TDMA carriers under the control of the NCC to support the communications traffic in each beam. The large number of narrow beams with dynamically reconfigurable frequency allocation provides the system with the ability to support an efficient and high reuse of the L-band spectrum. The digital signal processor can utilize any or all portions of the 34 MHz of L-band designated for MSS service further enhancing spectrum efficiency. The processor also includes a mobile-to-mobile switch that allows connectivity between users without a connection through a ground gateway. The mobile-to-mobile switch supports up to 8000 full-duplex connections. Simultaneously, each half of these connections is provided to the feeder downlink for controlling the calls by the gateways.

2.2 Gateway Element

Gateway stations (GSs) provide connectivity between the PSTN/PLMN and user terminals via the space segment. Additionally, GSs manage and allocate system communications resources (e.g., L-band spectrum and downlink EIRP) necessary to support the communications traffic within their respective service areas, and also interface with partner cellular networks for interoperability functions. A single GS has very similar top-level functions and features as one complete land-based cellular station. GEM GSs use a modified GSM cellular switch to work efficiently in the satellite environment. The key functions of GSs are as follows:

- Manage call set-up and tear-down
- Generate call records for each call
- Allocate communications resources within its service area (frequency and power)
- Provide user authentication
- Support user roaming within the system coverage area

GS circuit capacity is modularly expandable up to 3000 circuits. Up to three gateways can be combined to form a gateway cluster of 9000 circuits. Figure 4 shows a simplified block diagram. Key elements of the gateway are as follows:

- Network switching system (mobile switch, HLR, VLR, voice mail, and short message systems)
- Gateway station subsystem (voice transcoding, satellite modems, radio resource management, and ancillary equipment)
- Gateway RF equipment (antenna, up/downconversion, amplification, and other RF equipment for transmission and reception at Ku-band)
- Traffic control subsystem (manages the power and bandwidth allocated to the gateway; sets up calls to and from the gateway and among user terminals in beams controlled by the gateway)
- Operation and maintenance terminals (operator terminals to support the network control subsystems in the gateway)
- Site power system (inverters/chargers, batteries, and other power sources for the gateway site)

The GS supports the voice encoding and channel encoding algorithms defined in the CAI. It implements data and facsimile by providing an interworking function (IWF) and protocol spoofing for the T.30 facsimile protocol.

2.3 User Terminals

User terminals provide the subscriber interface to the GEM system. Terminal types include handheld, vehicular, and stationary. Terminal designs support both single mode and dual mode operation. Dual mode terminals are capable of operation in both the GEM system and the local GSM cellular system. The distinguishing characteristics of the various terminals include the following: Handheld terminals resemble traditional ground cellular units. Vehicular terminals typically consist of handsets docked to an external, antenna, transmit power booster and dc power source. Stationary terminals feature directional antennas and can be configured to support from one to eight voice or data circuits.

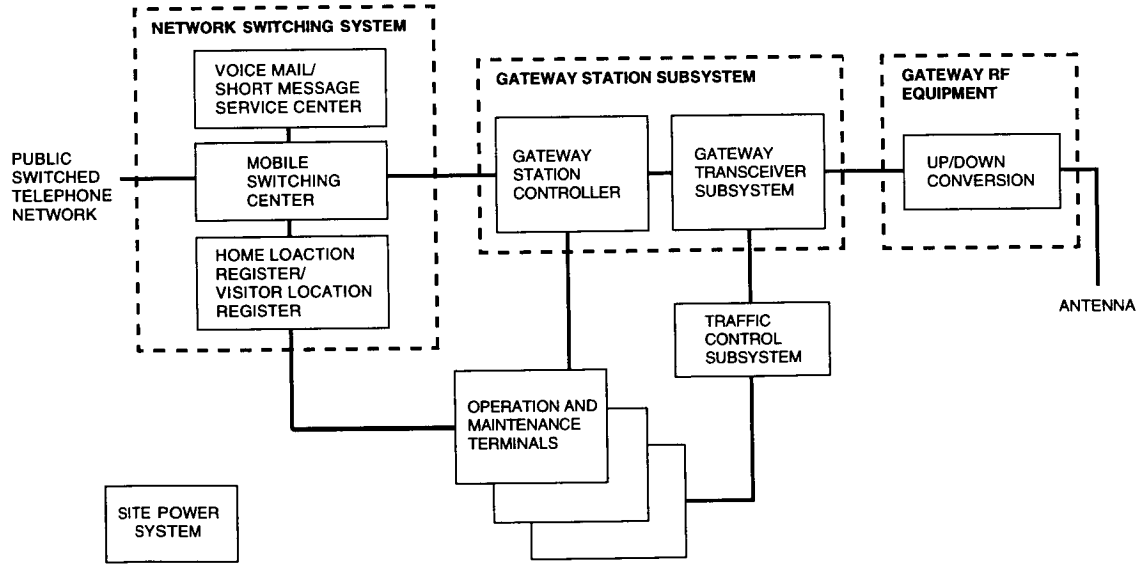


Figure 4. Gateway Block Diagram

Because the GEM system protocol stack is based on GSM, many of the functions and features of GSM are implemented in the terminal using the same hardware and software for both satellite and terrestrial operation. The differences are limited to the lower level protocol layers and some features, such as direct terminal-to-terminal calling, which have no counterpart in terrestrial GSM.

As for terrestrial GSM user terminals, GEM terminals employ a subscriber identity module (SIM) to uniquely identify the user of the terminal. The SIM contains the user's identity and nonvolatile memory to store the user's configuration data.

Dual mode terminals can automatically or manually choose either terrestrial or satellite access modes, depending on priorities and restrictions programmed into their SIMs. Terminals scan for available control channels and then choose the best cell (terrestrial mode) or beam (satellite mode). They then register in that cell or beam.

Terminals support authentication, link ciphering, and user identity protection as in GSM 03.20. A terminal adaptation function (TAF) provides support for data and facsimile in the user terminal. The TAF encapsulates the user's data or facsimile in a GSM-compliant packet which is then transported over the satellite or terrestrial network to the IWF in the gateway.

2.4 Common Air Interface

The GEM CAI defines the interface between the user terminals and the GEM satellite system and is an open interface. Figure 5 shows the layering and where each layer is implemented within the GEM system. The GEM CAI reuses the layer 3 mobility management (MM) and call management (CM) protocol of GSM including the short message service (SMS) and supplementary service (SS) protocol. Thus, the layer 3 procedures, state machines, primitives, information elements, and messages of GSM are reused. This strategy has several advantages:

- GS radio equipment can directly interface into a GSM switch at the GSM A-interface
- Bulk of the GSM terminal protocol software can be reused from GSM
- Services and features of GSM can be provided to the end users
- Testing, verification, and validation of these protocols is simplified
- Service provider may use the GSM authentication and privacy protocols
- Interoperability with terrestrial GSM systems is facilitated

The radio resource (RR) layer, the data link layer (DLL), and the physical layer protocols, however, are all optimized to the satellite link. This has several important practical advantages. The GSM protocol has been optimized for a Rayleigh fading channel with propagation delays of hundreds of microseconds and delay spreads on the order of a symbol duration. By contrast, the GEM channel is typically Gaussian or exhibits Rician fading with a K factor of 7 to 12 dB. Delay spreads are much less than a symbol duration and, for GEM, delay spread is not a problem.

The power control information bandwidth has been increased to 200 bps in GEM from the 33 bps of GSM for more effective power control.

The GEM CAI also provides for optimally routing both call deliveries and call terminations to the nearest gateway station to the PSTN point of entry. The latter two features, terminal-to-terminal connections and optimal routing, are not supported in the existing GSM RR layer protocol.

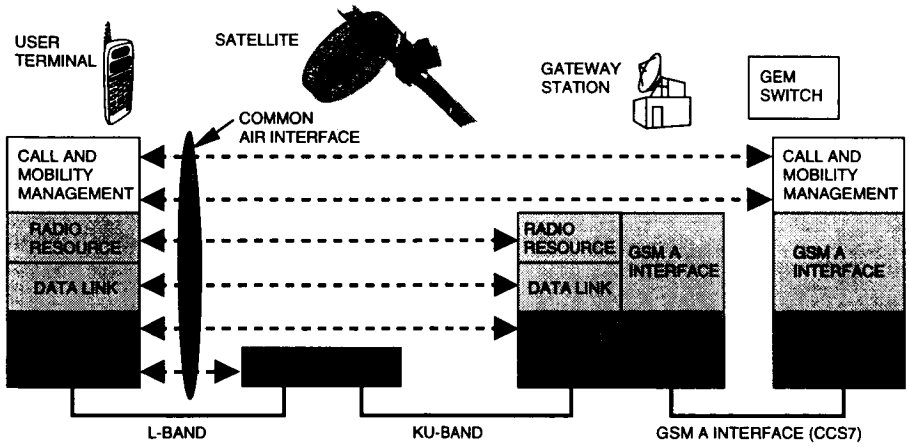


Figure 5. CAI Protocol Stack

The frame formats and modulation used in the GEM system take advantage of the unique features of the satellite digital signal processor. Because the frames and modulation are symmetrical between forward (Ku-L) and return (L-Ku), and because all transmissions are TDMA, it is easy to switch bursts in the satellite for user-to-user calls. The user terminals receive bursts from the gateway in the same format as those from another user terminal.

The modulation used is $\pi/4$ QPSK, which gives better channel spacing performance than, for example, GMSK. Because it is offset QPSK, however, it still retains much of the advantage of GMSK, in that it can drive the user terminal power amplifier nearly into saturation, saving user battery power.

The framing and modulation combine to provide the best efficiency in terms of users per megahertz of any mobile system today. In addition, the flexible slot structure allows for different rates of bearer channels to accommodate varying user bit rates and voice coding bit rates.

3. SUBSCRIBER SERVICES

The GEM system supports a multitude of advanced subscriber services; an example is shown in Figure 6. Subscribers with mobile terminals communicate with the PSTN/PLMN via the satellite and gateway stations or directly with each other via the satellite in a single hop. The GEM system is designed to provide a complementary service to existing or future GSM terrestrial networks, as well as form the basis of a number of services such as fixed rural, vehicular, and handheld telephony. The GEM system is modeled after GSM telecommunications services, permitting dual mode operation with GSM and provides similar user features. Because the system is modeled after GSM, many of the subscriber services available in GSM can be provided to the user either as a standard feature or as an option. These include the following:

- Voice, data, and facsimile telecommunications
- DTMF signaling
- Robust alerting in addition to conventional GSM paging
- Voice mail service
- Short message service
- Advanced calling features
  - Call forwarding
  - Call offering
  - Caller identification
  - Call waiting
  - Call barring
  - Multiparty calling
- Position determination

Digital voice, data, and facsimile services are provided between either a mobile user and a PSTN/PLMN or between mobile users. The system is designed for a voice data rate of 4.0 kbps, but the flexibility is provided to handle either higher or lower vocoder rates. Data and

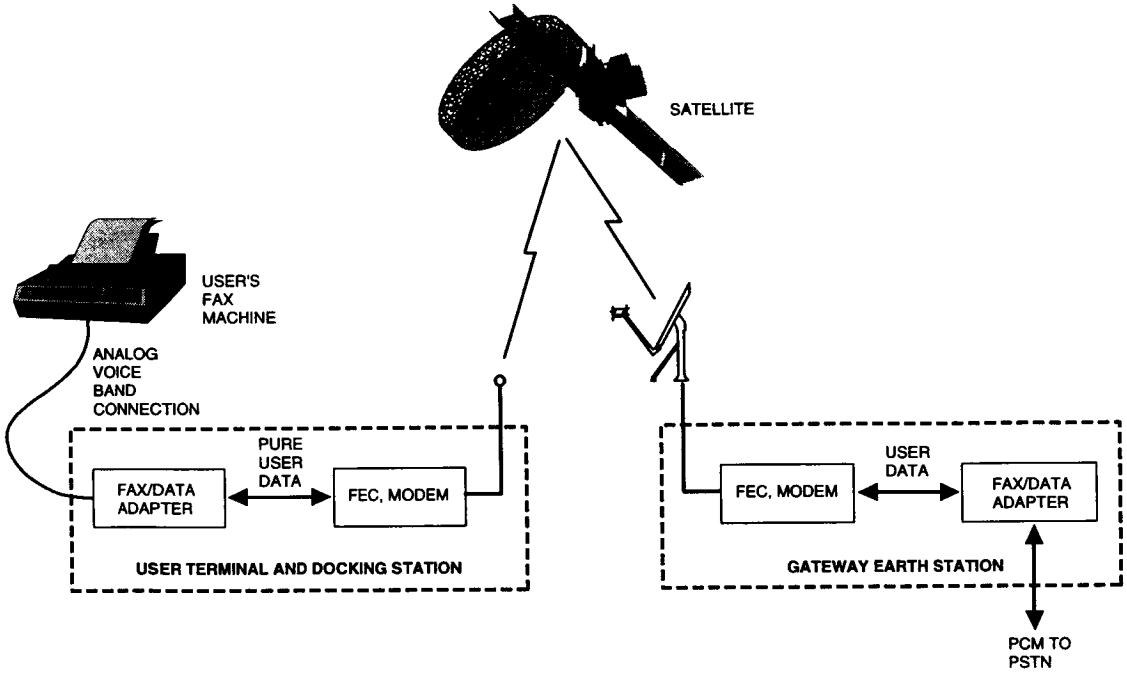


Figure 6. Mobile Fax Connection

facsimile services are accommodated with various rates including the standard rates of 2.4, 4.8, and 9.6 kbps. GSM NT mode is used for data transport, ensuring reliable transmission even in a harsh propagation environment Any data terminal equipment (DTE) data source, such as a laptop computer, may be used with a user terminal to access data services via V.24 and V.28 interfaces.

4. SYSTEM FLEXIBILITY

The GEM system has been designed to be flexible to support the MSS traffic requirements of varying markets and service providers. In addition, it is anticipated that these traffic requirements will change over time, and, therefore, future MSSs must be designed to be reconfigurable. The GEM system incorporates many features, as follows, to support these flexibility and reconfigurability needs.

- Add and expand GSs
- Modify coverage through digital beam forming on-orbit
- Dynamically reconfigure and remap L-band frequency allocations to beams
- Easily upgrade key functions such as modulation, speech coding, and error coding
- Partition satellite coverage area into independent service areas, each of which may be operated as a virtual MSS
- Exploit fragmented L-band allocations from any part or all of the 34 MHz associated with MSS service

SUMMARY

Hughes has developed a total system solution providing seamless mobile telephony coverage to users whether they are in a ground cellular or satellite spot beam. The solution preserves cellular/GSM services to the user in both satellite and GSM modes. State-of-the-art onboard digital processing provides flexibility to dynamically match system resources to traffic demand and to also efficiently accommodate changes in coverage region or spectrum allocation over the life of the satellite. A GSM-based open interface to user terminals supports a wide range of user services and packs an unmatched 256 voice channels per megahertz. This high bandwidth efficiency combined with high spectrum reuse (made possible by the small beam size and precision sidelobe control afforded by digital beamforming) results in maximum utilization of scarce mobile link spectrum.

ACKNOWLEDGMENT

The authors wish to thank the many GEM team members at Hughes Space and Communications and Hughes Network Systems for their hard work in designing the GEM system and their generous assistance in writing, reviewing, and publishing this paper. Special thanks goes to Dr. Eugene Kopp for his support in producing this paper.

Session 6
Propagation

Session Chairperson—*Faramaz Davarian*, Hughes Space and Communications, USA
Session Organizer—*David Rogers*, Communications Research Centre, Canada

Land-Mobile Satellite Channel Measurements in Athens City Center at 1800 MHz

Dr. A. Kanatas, N. Papadakis, P. Chatzopoulos, T. Sofos, Prof. P. Constantinou

National Technical University of Athens
Department of Electrical Engineering & Computer Science
Mobile Communications Laboratory
9 Iroon Polytechniou, GR-15773 Zographou, Athens, Greece
Tel: +301-7723849, Fax: +301-7723851
E-mail : kanatas@tinos.mobile.ece.ntua.gr

ABSTRACT

A Land Mobile Satellite System (LMSS) propagation experiment at 1.8 GHz was performed at Athens city center during June 1996 for high elevation angle channels. The transmitter platform used was a helicopter flying over predetermined paths to simulate high elevation angle (60°, 70°, and 80°) communication. A land vehicle specially prepared to accommodate the receiver system was performing the channel recordings. Four narrow streets with large building blocks were selected as representative, and the statistical analysis presented include cumulative distributions functions, level crossing rate, average duration of fades, fade and unfade duration distributions. The results show considerable increase of the fade depth compared to suburban and rural environments. In general the fade depth decreases with the increasing elevation angle, and the overall analysis indicates the strong dependence of the signal attenuation on the width of the streets, the average building height and the vegetation at the edges of the streets.

INTRODUCTION

The increasing demand for capacity (integrated voice and data services) and the high required link availability over wide areas has led to the concept of utilizing non-geostationary satellite systems as part of the future Universal Personal Telecommunications. Constellations of low earth orbit (LEO), medium earth orbit (MEO), and highly elliptical orbit (HEO) satellites have been proposed to serve the needs for lower path losses and higher elevation angles. A detailed characterization of the mobile satellite channel in all types of environments, over the whole range of elevation angle, and for all frequency bands allocated to Mobile Satellite Services (MSS) in WARC 92 and 95, is of great importance to system engineers.

Many experiments have been conducted in suburban, rural, open, wooded, and mountainous areas for elevation angles from 13° to 80°, [1-11], at UHF, L, and S-bands, and few of them have considered urban environments [6], [9-11]. Empirical models [12-13], giving attenuation scaling factors for the transition in different frequency bands, and calculating the required link margin for a specific outage probability, or the mean tree attenuation versus the path length through vegetation, have been extracted as regression fits to data gathered by these experiments, and facilitate the efficient design of system performance.

In June 1996, the Mobile Communications Laboratory of National Technical University of Athens conducted the first phase of a propagation measurements campaign to characterize the LMSS channel in urban environment. The measurements were performed in Athens city center in four typical narrow urban streets with heavy traffic and large building blocks. The transmitter platform used was a helicopter carrying the 1800 MHz source, and flying in predetermined paths to simulate high elevation angle (60°, 70°, and 80°) communication. A land vehicle specially prepared to accommodate the receiver system was moving with constant speed along the streets, performing the channel recordings, and collecting all the necessary information.

EXPERIMENTAL ASPECTS

Environmental Features

The propagation measurements were performed at Athens city center, characterized by heavily populated urban areas, with high buildings and narrow streets. Four streets were selected as representative, and may be described as follows :

1) Zaimi St. is a 10m wide, one way, one lane, 570m in length street, with two lanes of parked cars at both sides

of a moving car. There are sidewalks with small sour orange trees (2-3m high) placed along the street, while taller poplar trees (8-10m high) are rarely located. Ten perpendicular streets cross Zaimi St. enabling LOS communication even at low elevation angles. The building blocks between crossroads are approximately 40m in length, and 15-18m high (5-6 storeys).

2) Bouboulinas St. is a 10m wide, one way one lane, 430m in length street, with an extra lane of parked cars. The type of building blocks is similar to that described previously except a 50m wide block which is the back side of the National Museum. This is surrounded by a 7m wide yard, 3m below the street level, providing LOS at 20° elevation angle for approximately 70m.

3) Ippokratous St. is a 15m wide, one way two lanes, 1500m in length street, with one extra lane of parked cars. Twenty five perpendicular streets cross Ippokratous St. allowing an average width of 40m for the building blocks, for the most of the street length. The building's height ranges from 15m to 21m (5-7 storeys).

4) Asklipiou St. is the first parallel street to Ippokratous St. with opposite direction for the moving cars. It is a 11m wide, one way one lane, with two lanes of parked cars, 1400m in length street. The layout and characteristics of the building blocks are almost the same with those along Ippokratous St.

All the above described streets suffer from heavy traffic during the whole week, with many sort traffic jams per day. The large number of buses moving along these streets during working hours aggravate the traffic situation.

Experimental Configuration and Methodology

The primary target of this phase was to record the mobile satellite channel at high elevation angles in Athens city center, at L-band. The selection of a platform that would carry the transmitter system, was dominated by the need for constant elevation angle at practically achievable speeds, during each run. The ability of helicopters to keep almost constant flight paths for any desired elevation angle, with reasonable speeds and, the successful use in previous experiments were the best grounds to consider the helicopter as the most suitable platform. Hence a single engine, five-seater helicopter was used and was arranged in order to mount the transmitting antenna as shown in Figure 1.

Twelve flight paths were prepared (60°, 70°, and 80° for the 4 streets) and tested before the final measurements. The selection of these paths was based on the criterion of minimum deviation from the required elevation angle (±4°) for each run.

Figure 1 : The antenna mounted below helicopter

At the beginning of each run the helicopter was hovering, maintaining the desirable elevation angle for 10 seconds, thus allowing a LOS measurement with static mobile. The land vehicle, a 4WD vehicle adjusted and equipped with the receiver system, run all the specified routes with approximately constant velocity (40km/hr). Off-line verification of the achieved elevation angle, was possible through video recording and GPS data collected by the helicopter. Two video cameras were placed at both sides of the roof top of the 4WD car, recording the relative position of transmitter and receiver. A VHF radio link at 122.2 MHz was used for voice communication between the land vehicle and the helicopter to assist the synchronization during the experiment.

Transmitter & Receiver Systems

The transmitter hardware located on the helicopter consisted of an Anritsou MG3602A signal generator supplying a tone at 1800 MHz, a Trontech 10 Watt power amplifier at 1800 MHz, a Bird Model 42 watt-meter, a Crossed Drooping Dipoles antenna.

The 1800 MHz signal was frequency modulated with 1KHz audio, to assist the monitoring and to provide an indication of the received signal during measurements. The 13 dBm tone at the output of the signal generator was amplified by a 10 Watt power amplifier. The amplified signal was fed via coaxial cable prior to the antenna, to a watt meter in order to assist the monitoring of the system during the flights. The watt meter indication was continuously stable at 9.6 Watt. The antenna mounted at the helicopter was an inverted-V type crossed drooping dipoles antenna, and was developed in the Mobile Comms Lab. This antenna presents almost constant gain 4.5 dB, over elevation angles from 90° to 50°, at 1800 MHz. The whole arrangement was power supplied by a 12 Volt - 80 Ah dryfit type battery followed by a DC/AC converter. A GPS receiver, recorded all flight paths in a memory module, allowing off-line verification of the maintained elevation angles.

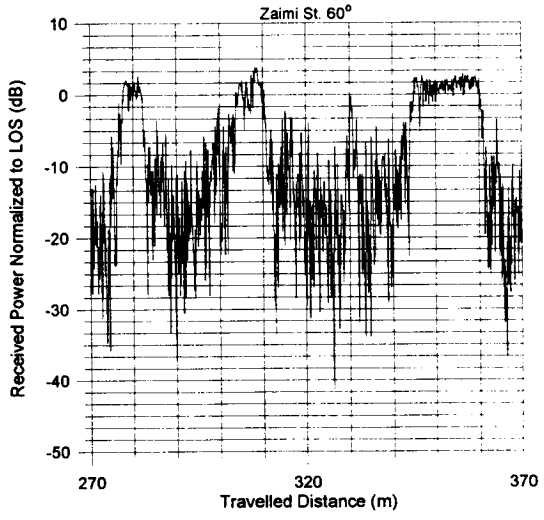
The receiver hardware located in the land vehicle consisted of an inverted-V Crossed Drooping Dipoles antenna, a band pass filter, a Low Noise Amplifier, an ICOM R7000 receiver, and the Data Acquisition System. The characteristics of the antenna are the same with the one used with the transmitter, and was placed on the vehicle's roof top. Located in the car were the BPF, an Apollo Microwaves Ltd. Model 13939, centered at 1800 MHz with 23 MHz 3dB bandwidth, and the LNA, a Mini Circuits ZHL-1724 HLN, with 30 dB minimum gain. The signal was fed via coaxial cable to the ICOM R7000 calibrated receiver, which DC voltage output was connected to the data acquisition system. This was developed at Mobile Comms Lab for mobile radio propagation measurements in urban and suburban environment. It is a 486/66 PC AT notebook based system using a DAQ-PAD MIO 16 XE-50 (16 bits) A/D card (20 Ksample/sec), an odometer, and National Instruments LabVIEW software. The odometer is connected to one of the rear wheels of the car, and is equipped with an optical encoding rotary device providing to the A/D card 100 triggering pulses in one full rotation. Hence, one TTL pulse is sent to the A/D card every 0.022 m. This is better than $\lambda / 7$, getting a sufficient distance resolution to measure the periodic fades that occur at half wavelength intervals. The software aided data collection and monitoring system was developed in such a way that enables the computer user to observe the signal fading on the computer monitor during the measurement process.

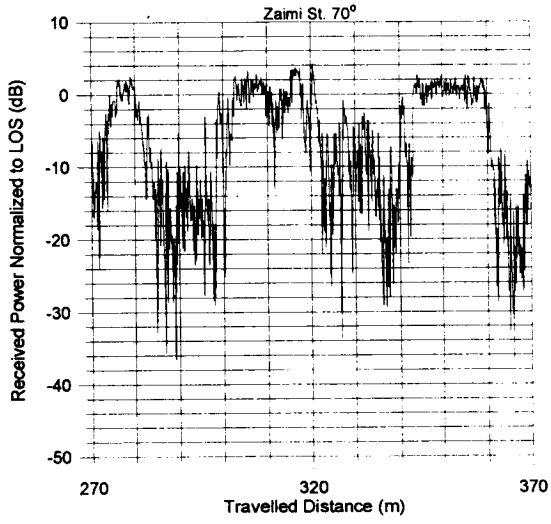
CHANNEL STATISTICS

A critical parameter from the very beginning of the statistical evaluation was the value of line of sight (LOS) for each run. The estimation of this value was based on the portions of run where unshadowed communication was possible, and on the initial recordings for each run where the mobile was static. Indicative channel recordings are shown in Figure 2 for 100m traveled distance along Zaimi St., and for 60°, 70°, and 80° elevation angle. The 0 dB value corresponds to the estimated LOS value. The first point one observes, is the a clear distinction between shadowed and unshadowed portions of the street. As the elevation angle increases, the portions where LOS communication is possible also increases. Moreover, new unshadowed portions added at 80°. Hence, while at 60° LOS communication was possible only at crossroads, at 70° and 80° the signal was unobstructed and well above buildings' roof tops for longer parts of the street. This was also evident from video recordings showing the helicopter flying above or below the roof tops correspondingly.

Cumulative Fade Distributions - (CFD)

The cumulative fade distributions derived for all streets, for 60°, 70°, and 80° are shown in Figure 3, where the ordinate represents the percentage of the traveled distance for which the fade depth is greater than the abscissa value. The ordinate may well be considered to represent the percentage of time since the land vehicle kept almost constant velocity throughout the runs. Negative values of fade depths correspond to signal enhancements caused by constructive multipath interference.
All distributions depict the effect of elevation angle on the fade depth levels. Hence as the elevation angle increases the fade depth decreases for the same percentage of traveled distance. For example at Bouboulinas St. and for 1% outage probability the fade depth reduces from 30,1 dB at 60° to 16,3 dB at 70°, and to 8,7 dB at 80°. Table I quotes the fade depth values for all streets and elevation angles for 1%, 10%, and 50%, of the traveled distance.





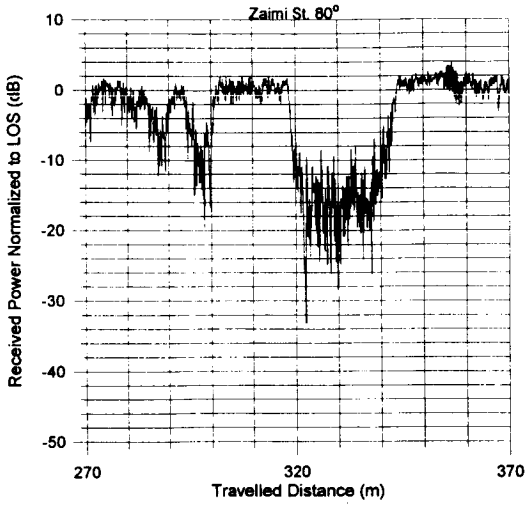
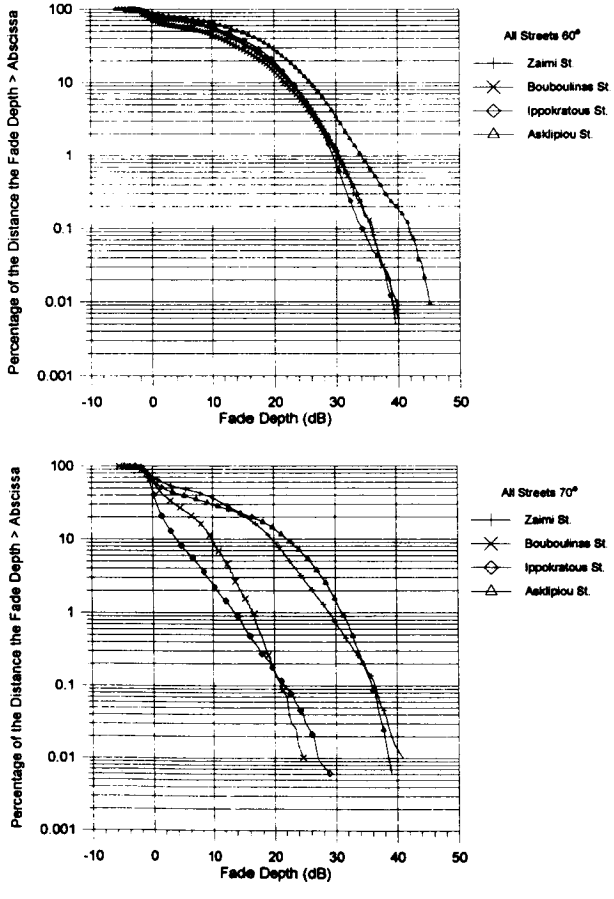


Figure 2 : Channel Recordings for Zaimi St. at 60°, 70°, and 80° elevation angles.



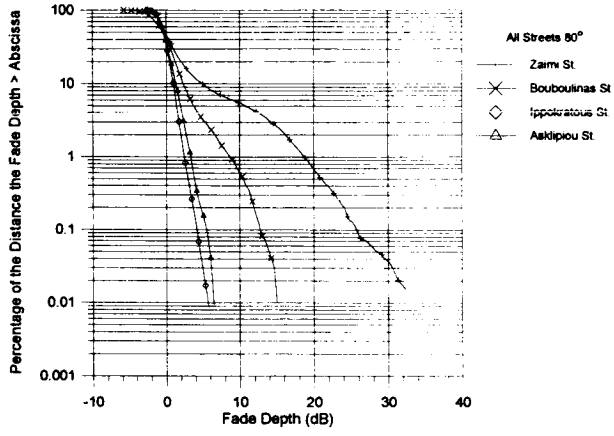


Figure 3 : Cumulative Fade Distributions classified by elevation angle.

Another interesting point extracted from Table I is the strong dependence of the fades on the environmental features. Ippokratous St. for example, a 15m wide street, is similar to Asklipiou St. as far as the building blocks are concerned, but presents lower fade depth at 70° since the last is an 11m wide street. Both these streets present similar values at 80° where the helicopter is visible for the whole run. Asklipiou St. seems to be the worst street for 60°, and 70° elevation, whereas for 80° Zaimi St. suffers from larger fades. This is due to the tall poplar trees in blossom that are responsible for shadowing of the transmitted signal even at high elevation angles, since Zaimi St. is a quite narrow street and the trees are close to the moving cars.

TABLE I : Values of fade depths for all streets.

Percentage		1%	10%	50%
Zaimi St.	**60°**	30,6	22,9	11,2
	70°	28,8	19,6	4,4
	80°	18,6	5	0
Bouboulinas St.	**60°**	30,1	21,8	8,2
	70°	16,3	9,8	0,8
	80°	8,7	2,4	-0,4
Ippokratous St.	**60°**	29,4	23,1	12,1
	70°	13,7	3,9	-0,1
	80°	2,4	0,9	-0,4
Asklipiou St.	**60°**	33,8	25,8	14,6
	70°	31	22,3	1,7
	80°	3,4	1,4	-0,3

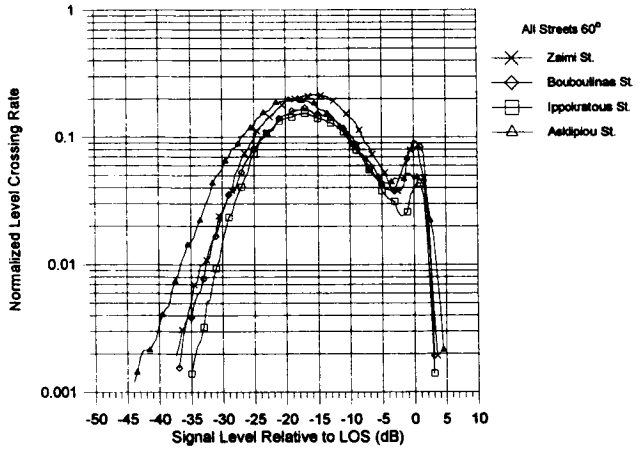
Level Crossing Rate & Average Fade Duration

This paragraph deals with second order statistics evaluated from channel recordings. Figure 4 depicts the Level Crossing Rate for all streets at 60°, and 80°

elevation angles. The ordinate represents the crossings per wavelength for the threshold indicated by the abscissa value. The latter corresponds to the received signal level relative to LOS. The first point to be observed is that at 80° elevation, the peak of LCR curves is at 0 dB level indicating the presence of LOS communication for the most of the traveled distance. At 60° where the LOS is obstructed, the peaks of the curves are concentrated around -18 dB and -15 dB. Figure 5 quotes the average duration of fades (AFD) for all streets at 60°, and 80° elevation angles. The ordinate represents the duration of fades in wavelengths for the abscissa value, which in turn corresponds to the received signal level in dB normalized to LOS. It is clear from all curves that at 0 dB level, the fade duration decreases since unobstructed view to the helicopter is maintained. As the elevation angle decreases the curves extend to lower signal levels.

Fade and Unfade Duration Distributions

The analysis of the fade durations is one of the most important parameters in the design of digital systems. The average duration of fades was cited in the preceding paragraph. The distributions of the fade and unfade durations are presented in this paragraph. These two statistics correspond to durations for which the signal attenuation is greater or less than a threshold value respectively. Figure 6 presents the fade durations distributions for Zaimi St., for 60°, and 80° elevation angles, and for several values of threshold.



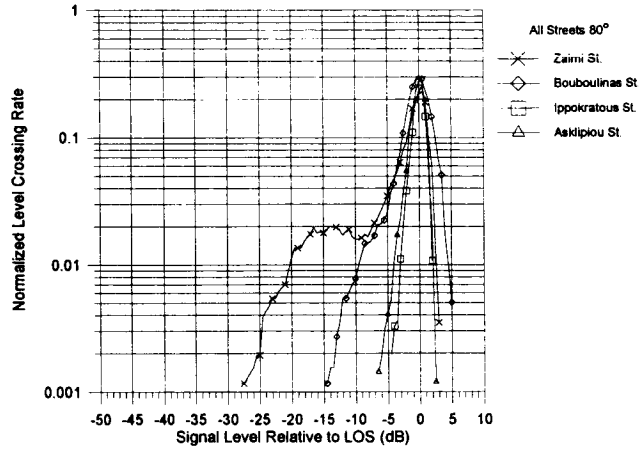
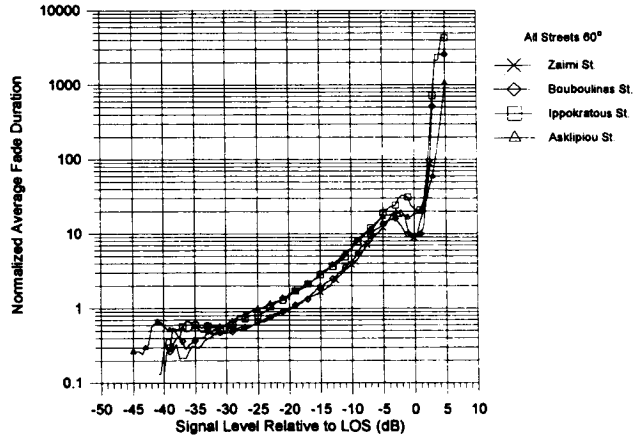


Figure 4 : Normalized Level Crossing Rate for all elevation angles and streets

Figure 7 depicts the unfade durations distributions for the same street. The ordinate represents the percentage of durations (fade or unfade) greater than the abscissa value. The latter is normalized to number of wavelengths. The dependence on the elevation angle is obvious since for a given percentage and threshold, the fade duration increases as the elevation angle decreases. It is worth mentioning, that the mean of the fade durations distributions is almost the same with the value of the average duration of fades for a given threshold.



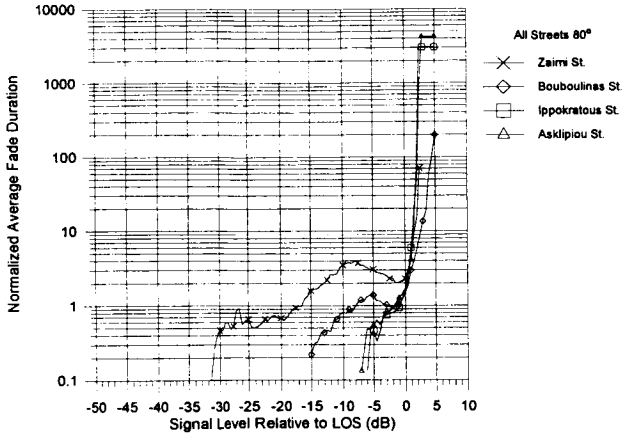


Figure 5 : Average Duration of Fades for all streets and elevation angles

CONCLUSIONS

The results from a high elevation angle (60°, 70°, 80°) LMSS channel propagation experiment have been presented. These measurements were performed during June 1996 at Athens city center and employed a helicopter and a land vehicle as the transmitting and receiving platforms respectively. The main objective of the campaign was to characterize the LMSS channel in typical urban narrow streets with large building blocks and heavy traffic. From the statistical analysis performed the following conclusions may be pointed out :

1) Athens city center presents greater fade depths for all elevation angles examined compared to suburban and rural environments investigated by other researchers, as expected. A more extensive comparison with other experiments requires the employment of frequency scaling factors.
2) The signal attenuation increases as the elevation angle decreases.
3) The signal fading is strongly environment dependent particularly in urban streets. The width of the streets, the average height of the building blocks and the vegetation at the edges of the streets, dominate the channel behavior. Indeed, the presence of tall trees on the sidewalks along narrow roads may shadow the transmitted signal and therefore is responsible for increased fade depths even at 80°. The latter has been confirmed by the cumulative distribution functions and the estimated values of time share of shadowing.

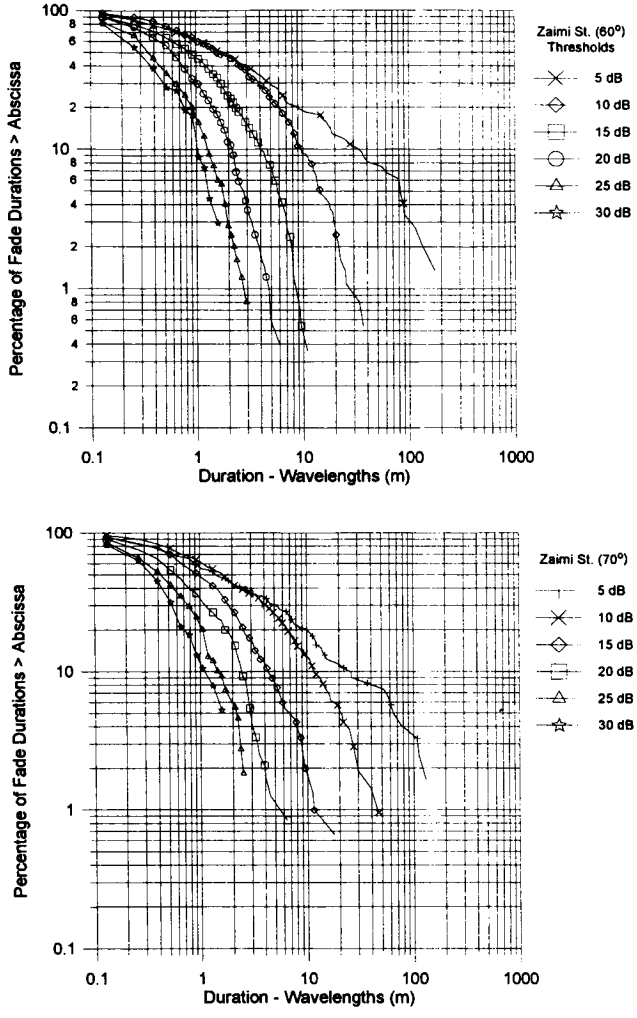
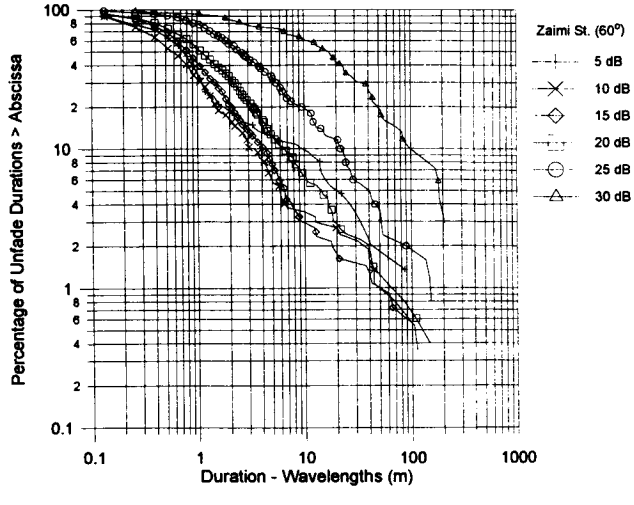


Figure 6 : Fade durations distributions for Zaimi St.



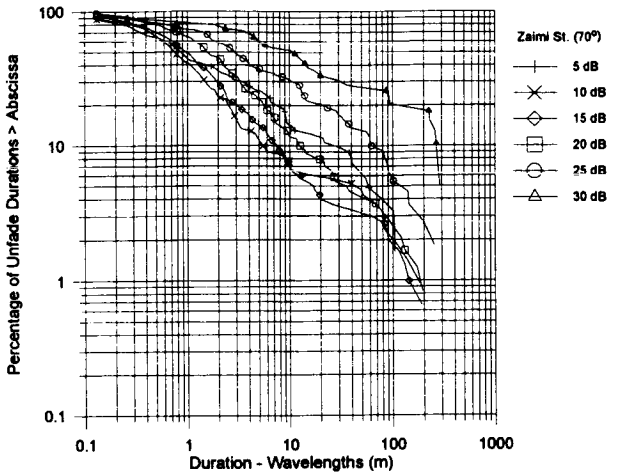


Figure 7 : Unfade durations distributions for Zaimi St.

REFERENCES

[1]. **J. Goldhirsh and W. J. Vogel**, " *Roadside Tree Attenuation Measurements at UHF for Land-Mobile Satellite Systems* ", IEEE Trans. on Antennas and Propagation, Vol. 35, No. 5, May 1987.

[2]. **W.J. Vogel and J. Goldhirsh**, " *Fade Measurements at L-Band and UHF in Mountainous Terrain for Land Mobile Satellite Systems*", IEEE Trans. On Antennas and Propagation, Vol. 36, No. 1, January 1988.

[3]. **J. Goldhirsh and W. J. Vogel**, " *Mobile Satellite System Fade Statistics for Shadowing and Multipath from Roadside Trees at UHF and L-Band* ", IEEE Trans. on Antennas and Propagation, Vol. 37, No. 4 April 1989.

[4]. **W. J. Vogel and J. Goldhirsh**, " *Mobile Satellite System Propagation Measurements at L-Band Using MARECS-B2* ", IEEE Trans. on Antennas and Propagation, Vol. 38, No. 2 February 1990.

[5]. **W. J. Vogel and J. Goldhirsh**, " *Multipath Fading at L-Band for Low Elevation Angle, Land Mobile Satellite Scenarios*", IEEE Journal on Selected Areas in Communications, Vol. 13, No. 2, February 1995.

[6]. **G.C. Hess**, " *Land-Mobile Satellite Excess Path Loss Measurements* ", IEEE Trans. on Vehicular Technology, Vol. 29, No 2, May 1980.

[7]. **R.W. Huck, J.S. Butterworth, and E.E. Matt**, "*Propagation measurements for land-mobile satellite services* ", Proc. IEEE 33rd Ann. Veh. Technol. Conf., 1983.

[8]. **G. Butt, B.G. Evans, M. Richharia**, " *Narrowband Channel Statistics from Multiband Propagation Measurements Applicable to High Elevation Angle Land-Mobile Satellite Systems* ", IEEE Journal on Selected Areas in Communications, Vol. 10, No. 8, October 1992.

[9]. **E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke**, " *The Land Mobile Satellite Communication Channel-Recording, Statistics, and Channel Model* ", IEEE on Vehicular Technology, Vol. 40, No. 2, May 1991.

[10]. **V.S.M. Renduchintala, H. Smith, J.G. Gardiner, I. Stromberg**, "*Communications Service Provision to Land Mobiles in Northern Europe by Satellites in High Elevation Orbits - Propagation Aspects* ", IEEE, 40th Intern. Conf. on Vehicular Technology May 6-9 Orlando, Florida.

[11]. **H. Smith, M. Sforza, B. Arbesser-Rastburg, J.P.V. Baptista, S.K. Barton**, "*Propagation Measurements for S-Band Land Mobile Satellite Systems Using Highly Elliptical Orbits* ", Proc. of 2nd European Conf. on Satellite Communications, Liege Belgium, 22-24 October 1991.

[12]. **J. Goldhirsh and W.J. Vogel**, "*Propagation Effects for Land Mobile Satellite Systems: Overview of Experimental and Modeling Results*", NASA Reference Publication 1274, Feb. 1992.

[13]. **A.G. Kanatas and P. Constantinou**, "*Narrowband Characterization of the Land Mobile Satellite Channel : A Comparison of Empirical Models* ", European Trans. on Telecommunications, Vol. 7, No. 4, July/August 1996.

Propagation Measurements for Land Mobile Satellite Communications at S-band with non-GEO ETS-VI satellite

Shin-ichi Yamamoto+, Eiji Okamoto+, Abdulrahaman Ali Aboudebra++ and Tetsushi Ikegami+

+ Communications Research Laboratory,
Ministry of Posts and Telecommunications
893-1 Hirai, Kashima, Ibaraki 314 Japan
TEL: +81-299-84-7128, FAX: +81-299-84-7158
e-mail: yamamoto@crl.go.jp

++ Department Electronics,
School of Engineering, Tokai University
1117 Kitakaname, Hiratsuka, Kanagawa
259-11 Japan

Abstract

Since the main problem in land mobile satellite communications is the frequent blockage of the satellite link by shadowing, this paper analyzes the propagation measurement results of land mobile communication obtained from a satellite at a high angle of elevation in different environments. From these measurements of the ETS-VI satellite, which has a sub-recurrent orbit and 10-20 degrees higher angle of elevation than a geostationary satellite, the shadowing probability is proved to become lower with increasing angle of satellite elevation.

Moreover, the cumulative distributions of the fade duration (FD), which is the duration when communication is unavailable and of the non-fade duration (NFD), which is the available duration, are analyzed to verify that the use of high angle of elevation has significant advantage in land mobile satellite communications.

1. Introduction

In land mobile satellite communications, a major problem is the frequent blockage of the satellite link by shadowing. Shadowing is the key factor determining the availability of land mobile satellite communications, comparing with statical analysis of fading in the line-of-sight state. For this reason, shadowing should be treated in two different states, the fade state, and non-fade state[1]. Propagation measurements in the L-band (1.5 GHz) from the geostationary satellite, Engineering Test Satellite 5 (ETS-V), have been performed and the shadowing probability was analyzed in different environments[2][3][4]. The angle of elevation of ETS-V from a vehicle is, depending on the area, about 47 degrees in the Kanto zone in Japan.

Engineering Test Satellite 6 (ETS-VI) failed to reach geostationary orbit and has a sub recurrent orbit. Propagation measurements can be done with the ETS-VI at 10-20 degrees higher angle of elevation than with geostationary satellites.

Considering the causes of shadowing, availability becomes higher as the elevation increases. Propagation measurements at higher angles of elevation using a helicopter have already been reported[5]. However real satellite conditions are essential for detailed evaluation.

In this paper, the propagation measurement results of the land mobile satellite channel at a maximum angle of elevation of 64 degrees using ETS-VI in the S-band (2.1 GHz) are analyzed and discussed.

2. Outline of measurement

Figure 1 shows the experimental setup. Ka-band (30 GHz) non-modulated waves were transmitted to ETS-VI from Kashima Space Research Center (Kashima city, Ibaraki pref.), and the on-board transponder (ETS-VI) converted the Ka-band signal into the S-band (2.1 GHz) signal and transmitted it back to the ground. The test van received the S-band signal with directional and omni-directional antennas, and the measured signal strength was recorded. The polarization of the received signal was LHCP. To fix the receiving frequency of the signal and the receiving power at line-of-sight, they were controlled in Kashima Space Research Center. Propagation loss and Doppler shift, caused by satellite transfer, and change of range between satellite and earth station, were compensated for by controlling the output frequency and level from the signal generator. The Doppler shift was calculated beforehand from the orbit of the satellite and output frequency of the signal generator is controled to calculated Doppler shift by a personal computer.

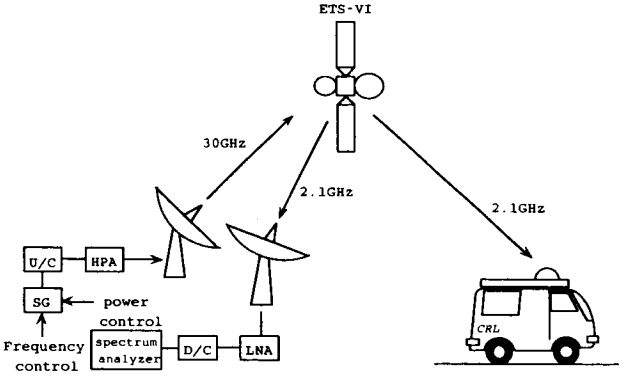


Fig.1 Outline of measurement

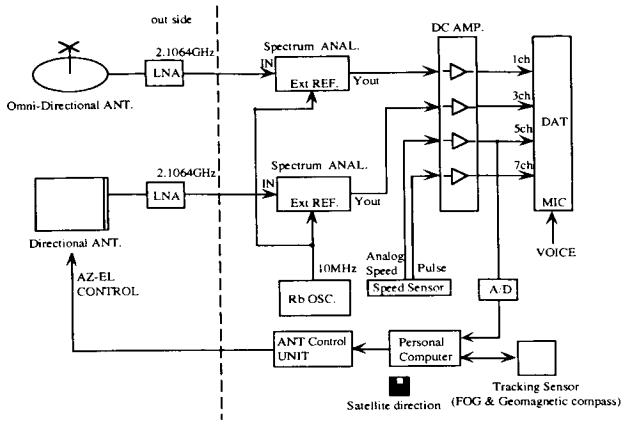


Fig.2 Block diagram of measurement system

Figure 2 shows the measurement system. The omni-directional antenna is a cross dipole with reflector. The directional antenna is a flat antenna composed of 3 × 4 lines of 12 elements patch antennas. Its gain is 15.5 dBi at the received frequency (2.1064GHz) and its beam width is 19.0 degrees in the horizontal plane and 27.0 degrees in the vertical plane. The directional antenna was set up with an antenna tracking system and controlled in both azimuth and elevation. An open loop antenna tracking method utilizing a fiber optic gyroscope (FOG) and a magnetic compass was used[6][7]. The accuracy of antenna tracking was about ± 5 degrees for the angle of direction. For the test van, only the change in direction was followed, because the environment was flat.

For the satellite, the change in direction was followed, using pre-calculated information, and the change in elevation was also followed according to satellite movement.

The signal received by the antenna was fed to the spectrum analyzer through a low noise amplifier. The spectrum analyzer worked in zero span as a receiver, and Yout output voltage was used that was proportional to the received power.

The received signal level, speed of the test van (analog voltage) and distance pulse (TTL) were recorded on a digital data recorder (DAT) through a DC amplifier. Necessary information, such as conditions of run and voice comments were also recorded. Distance pulses every 5 cm of the run were recorded, so the result could be analyzed without any influence from run conditions.

In this paper, we focus on the results obtained with the directional antenna. Results with the omni-directional antenna will be reported in the near future.

3. Propagation Measurement

Propagation was measured in three different environments: Kashima, Naha and Makuhari.

3-1 Cumulative distribution of received power

(1) Kashima

Kashima is a small city with few tall structures. Figure 3 shows the cumulative distribution of the received power. The abscissa shows the relative received power with respect to the line-of-sight signal power. The ordinate shows the probability that the received power is less than the abscissa value in Gauss scale. For each angle of elevation, the test van ran the same circular course. The difference in total

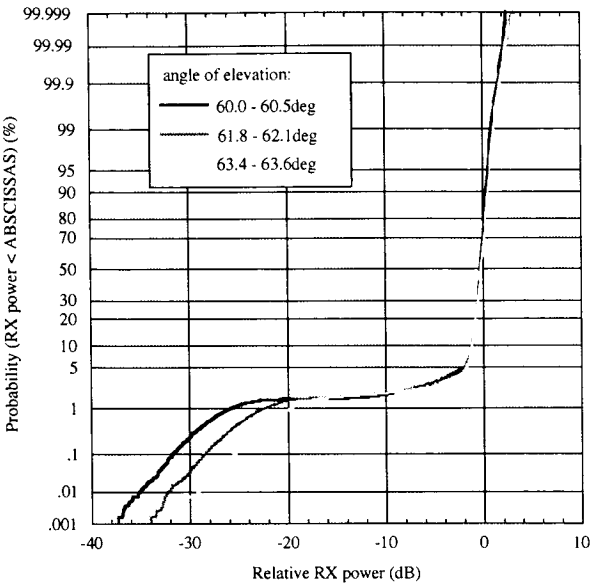


Fig.3 Cumulative distribution of received power (Kashima)

measurement duration was within 1% on average (about 9.4 km).

As the figure shows, the characteristic of each elevation changes at the -1 dB point, so a region above this point corresponds to line-of-sight and a region below to shadowing. Propagation in the line-of-sight state follows the Nakagami-Rice distribution, but it is close to a log-normal distribution considering random noise, scattered waves, antenna tracking error and the fact that antenna gain depends on the angle of incidence.

The measured data fit a straight line, which means a log-normal distribution.

Shadowing probability was about 5% of the total measurement duration, becoming lower as the angle of elevation increased. Within the limit that relative received power was from -1 to -20 dB, the transmitted signal wave was not blocked completely, that is, the received signal level was reduced by small structures, such as utility poles or trees, or by interference with scattered wave. The characteristic curve in this state inclined a little, independent of the angle of elevation. This means that the probability of this state was low. This tendency increased as the angle of elevation increased, particularly within the limits to -10 dB. This is caused by the reduced influence of, for example, interference waves or attenuation by utility poles or trees, as the angle of elevation increased. In the state where the signal wave is blocked completely, the characteristic curve follows the distribution of noise in the receiver system. Although the inclination is the same in this figure, received power falls further, compared with the line-of-sight state, as the angle of elevation increased. This is caused by the influence of diffracted waves added to system noise, because the antenna receives diffracted wave also being directed to the roof edge when the beam direction of the antenna increases as the angle of elevation increases.

(2) Naha (Okinawa)

Naha is a medium-size city, which has large streets and many buildings with more than ten floors. Figure 4 shows the cumulative distribution of received power. As Naha is situated to the south of Japan, the satellite angle of elevation

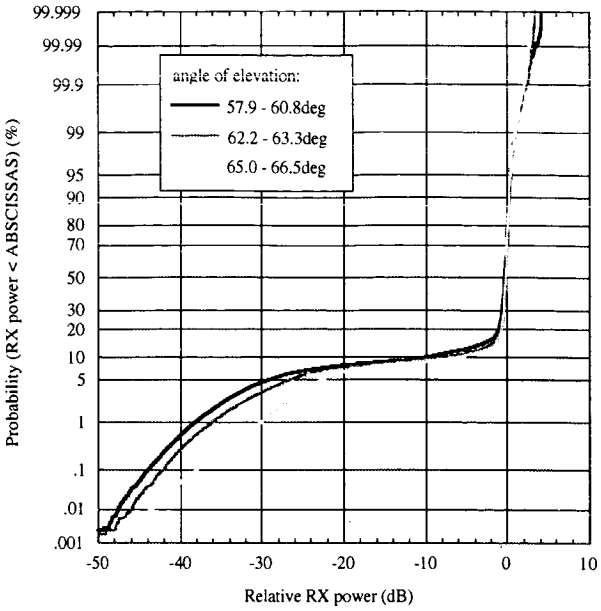


Fig.4 Cumulative distribution of received power (Naha)

is the highest found in Japan. For each angle of elevation, the test van ran the same circular course and the difference in total measurement duration was within ±2% on average (about 8.7 km). Shadowing probability in Naha was about 15% of the total measurement duration, and decreased as the angle of elevation increased, like in Kashima. For relative received power from -1 to -20 dB, the signal wave was not blocked completely, similar to Kashima. On the other hand, in the state in which the signal wave was blocked completely and relative receive power was less than -20 dB, shadowing probability did not vary with changes in the angle of elevation when the satellite angle of elevation was less than about 63 degrees, unlike Kashima. When the angle of elevation was more than 63 degrees, the curve changed but not significantly, compared with Kashima. We attribute the difference to the

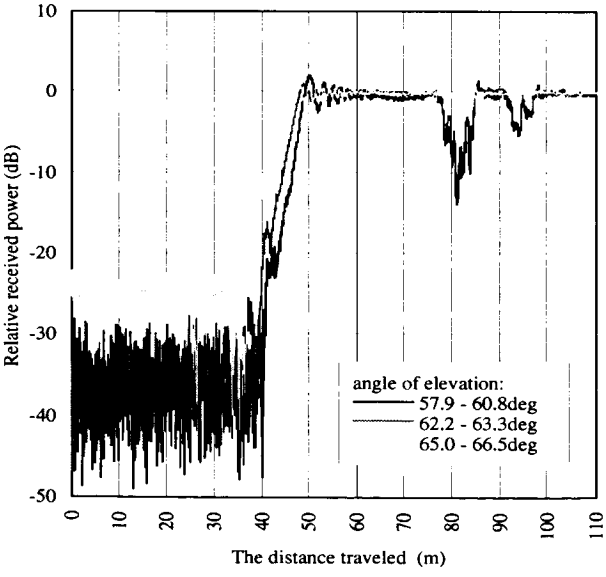


Fig.5 An Example of the signal received power

distribution of buildings.

Figure 5 shows the relative received power of each angle of elevation that acquired in same place of the measurement course in Naha. As the figure shows, the received signals level were reduced by shadowing. The cause of the shadowing that the distance traveled within about 50m was large structure such as building, of 80 and 95m points were small structure such as trees or utility poles. The reduction of received signal level were decreased as the angle of elevation increased. In the case of large structure such as building, there was regraded influence by diffracted waves from roof edge of buildings as the angle of elevation increased, and in the case of small structure, the reduction of received signal level were very small.

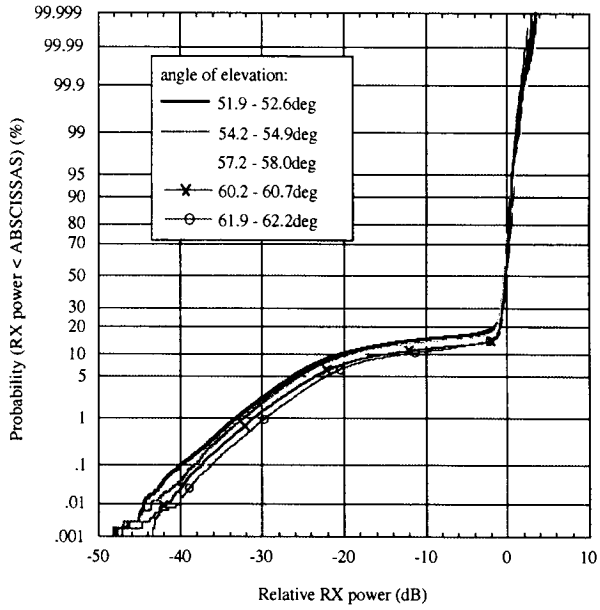


Fig.6 Cumulative distribution of received power (Makuhari)

(3) Makuhari (Chiba pref.)

Makuhari is an area with many tall buildings close together. The main causes of shadowing are artificial structures such as tall buildings or overpasses. Figure 6 shows the cumulative distribution of received power. For each angle of elevation, the test van ran same the circular course and the difference in total measurement duration was within ± 1% to average (about 5.4 km).

Shadowing probability in Makuhari was about 15-20% of the total measurement duration. This is different from the previous two cases; it fell stepwise as the angle of elevation increased. This phenomenon is peculiar to the environment in which shadowing is mainly caused by artificial structures and there is a clear distinction between line-of-sight and shadowing. On the other hand, in the state in which the signal wave was not blocked completely and relative received power was from -1 to -20 dB, the characteristic curve inclined a little. This means that data exists in the higher ranges. This must be independent of the angle of elevation and be constant, because the inclinations of different curves were not distinct. One reason is waves diffracted by the edge of buildings. In the state in which relative received power was less than -20 dB and the signal wave was blocked, there was little influence by diffracted waves. This means that the signal wave was blocked completely.

3.2Cumulative distribution of the fade/non-fade duration

The fade duration (FD) means the distance at which received power is continually below a certain threshold level. The non-fade duration (NFD) means the distance at which received power is continually above a certain threshold level. It is important for planning communication systems to analyze their statistical properties. In this paper, cumulative distributions of FD and NFD events at each satellite angle of elevation are used to analyze the dependence of FD and NFD distributions on angle of elevation. We can not compare the character simply, because the total number of FD events at each angle of elevation are different, so we consider their distributions. The resolutions of FD and NFD depend on the distance pulse, and the minimum was 0.05 m.

Figure 7 shows the cumulative distribution of FD events in Kashima. The abscissa shows the FD, and the ordinate shows the probability that FD was higher than the FD on the abscissa in total number of FD events. It is shown by a Gauss scale. The threshold level was -3 dB. Supposing that the threshold level was less than -3 dB, the characteristic curve did not show any big change as [2] shows clearly. As this figure shows, the distribution of FD events followed on almost log-normal distribution at each angle of elevation. And FD events, more than 0.1 m decreased as the angle of elevation increased. This means that FD events less than 0.1 m increased, in fact when the angle of elevation was 63 degrees, FD events less than 0.1

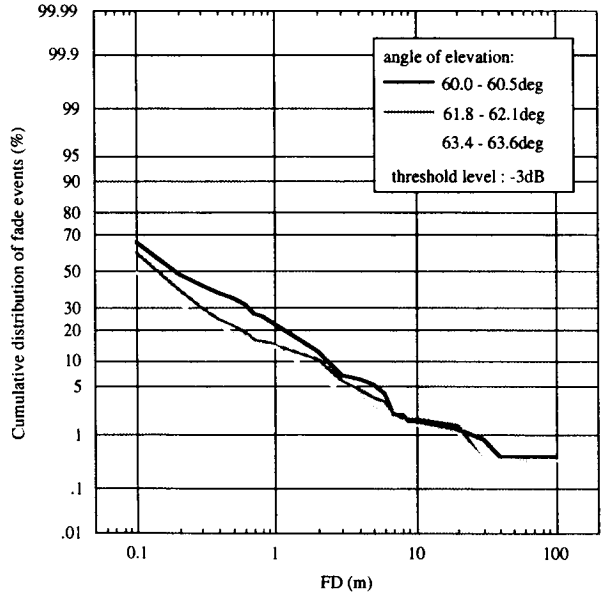


Fig.7 Cumulative distribution of fade events (Kashima)

m amounted to 50% of the whole distribution of FDs. Besides, as the angle of elevation increases even long FDs occur. In Kashima, the total number of FD events at each angle of elevation was almost similar, so FD events to 10 m decreased. Judging from the FD events for more than 40 m, which did not change, shadowing caused by tall buildings or overpasses will not be reduced even if the angle of elevation is higher than at present.

Figure 8 shows the cumulative distribution of NFD events in Kashima. The abscissa shows the NFD, and ordinate shows the probability that NFD was higher than NFD on

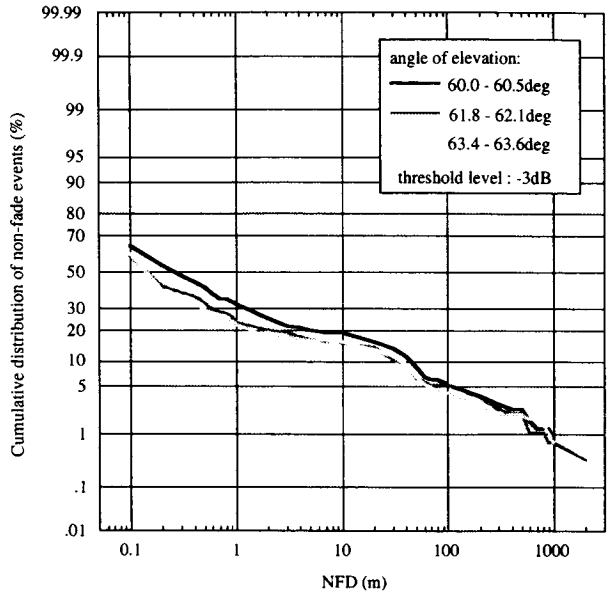


Fig.8 Cumulative distribution of non-fade events (Kashima)

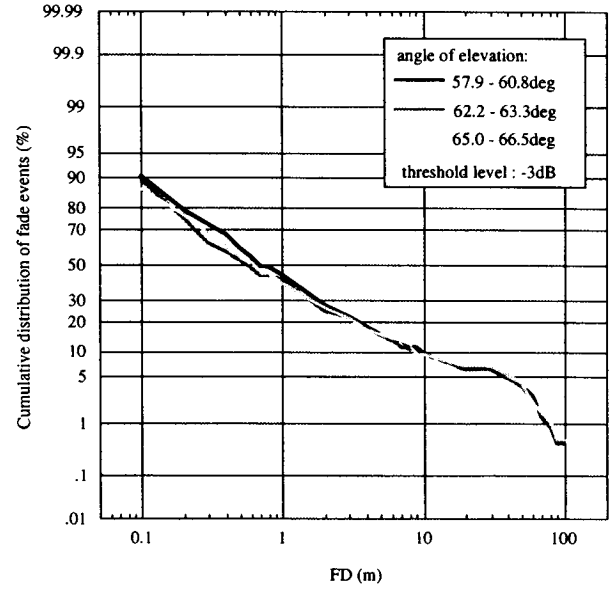


Fig.9 Cumulative distribution of fade events (Naha)

the abscissa in total number of NFD events. It is shown by Gaussian scale. The threshold level was -3 dB. The characteristic curve had an almost log-normal distribution, similar to the FD. The probability of NFD, from 0.2 to 4 m, decreased as the angle of elevation increased, but this probability for more than 500 m, increased. So it is easier to make a long NFD as the angle of elevation increases.

Figure 9 shows the cumulative distribution of FD events in Naha. In this case, the character curve was almost straight within the limit of 20 m and changed after this point. This is peculiar of an environment like a city center, in which the main cause of shadowing is artificial structures, because shadowing probability caused by structures accounted for a higher percentage of the total shadowing probability in this case. Regarding the dependence on angle of elevation, although the FD distribution changed a

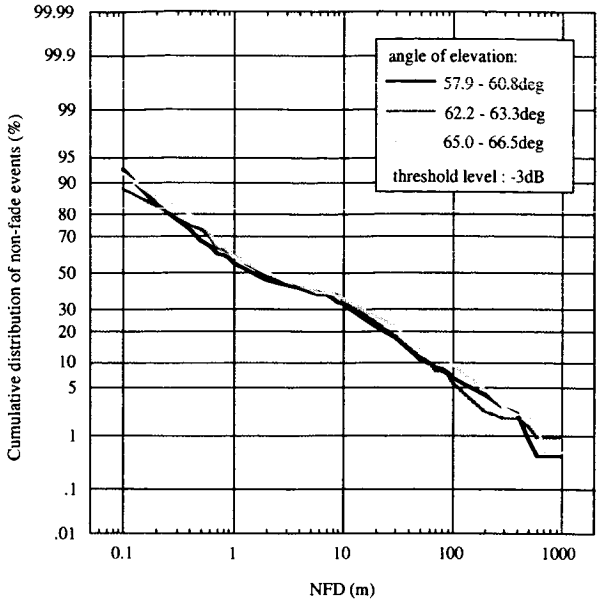


Fig.10 Cumulative distribution of non-fade events (Naha)

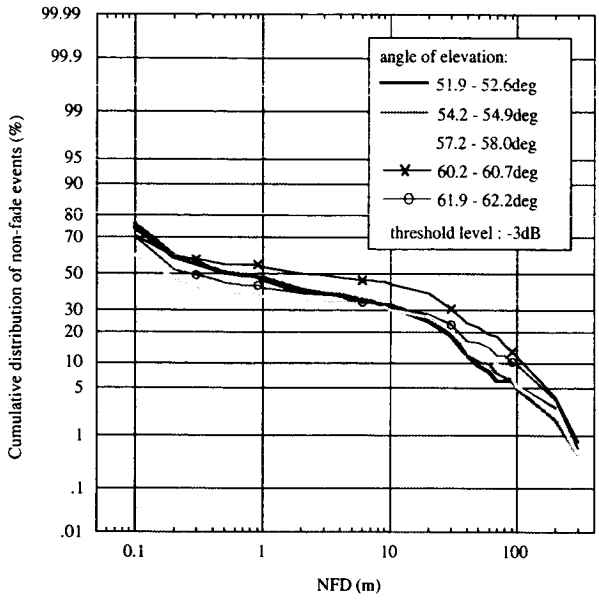


Fig.12 Cumulative distribution of non-fade events (Makuhari)

little within the limits to 2 m, it did not change so much in other regions. For 58-67 degrees of elevation, the FD distribution did not depend on the angle of elevation.

Figure 10 shows the cumulative distribution of NFD events in Naha. NFD has an almost log-normal distribution, unlike FD. And NFD more than 10 m, occurred more frequently as the angle of elevation increased. Although the NFD distribution decreased within 100-400 m, it increased in the region over 500 m. In this city, it is easier to make a long NFD, similar to Kashima.

Figure 11 shows the cumulative distribution of FD events in Makuhari, Chiba. This case has different characteristics from the other two cases. The characteristic curve of the limits to 8 m and the curve of over this limit differ clearly, because the shadowing caused by artificial structure is the main cause, like in Naha. It is seen even more clearly in Makuhari.

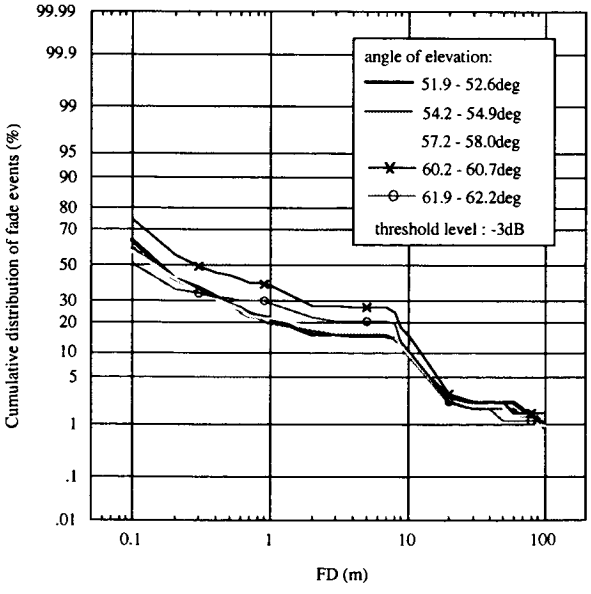


Fig.11 Cumulative distribution of fade events (Makuhari)

As regards the dependence on angle of elevation, this figure shows the tendency that FD events over 7 m increase while FD events less than 7 m decrease. Further it shows that FD events tends to decrease within 0.2-1 m and it gets longer in the limits to 0.1 m.

Figure 12 shows the cumulative distribution of NFD events in Makuhari. NFD does not form a log-normal distribution. It can be divided into two parts, a less than 10 m and over 10 m. As the angle of elevation increases, this tendency becomes clearer. Regarding the dependence on angle of elevation, NFD events within 0.2-10 m fall while they increase, in contrast, in the regions less than 0.1 m and more than 10 m as the angle of elevation increases. And as the angle of elevation increases still more, it falls in the region less than 0.1 m. This indicates clearly the distribution of NFD, it is particular in Makuhari. Another distinctive feature in Makuhari is that there is no NFD more than 400 m. Therefore, it will be difficult to improve long NFD in environments having many artificial structures such as tall buildings or overpasses, compared with other environments.

4. Conclusion

Through the measurements using ETS-VI, which has a sub recurrent orbit, we obtained the dependence on angle of elevation for the propagation in land mobile satellite links in three different environments. The results of the cumurative distribution of received power shows that the shadowing probability reduced as the angle of elevation increased. This is favourable for the land mobile satellite communications. From the results of the cumurative distribution of FD and NFD events, it is obvious that short FD events decreased and long NFD events increased, as the angle of elevation increased. These tendency are also favourables. However, from the results in Makuhari, in particular, we can conclude that the maximum NFDs for each angle of elevation within up to 62 degrees are similar, in an environment having many artificial structures. This clearly indicates that the distribution of NFD was under

the influence of size and place of the artificial structures. FDs are also under the influence of many artificial structures in the similar way.

These results can be used as basic data for personal satellite communication systems using low orbital satellites. Analysis of data taken with the omni-directional antenna is left for further study.

References

[1] Y. Hase, W.J. Vogel and J.Goldhirsh, "Fade-duration derived from land mobile satellite mesurements in Australia", IEEE Trns. Communications., vol.39, pp.664--668, May 1991.

[2] N. Obara, K. Tanaka, S. Yamamoto and H. Wakana. " Land mobile satellite propagation measurements in Japan using ETS-V satellite", IMSC'93, pp.313-318, June 1993.

[3] N. Obara and H. Wakana, " Fade/NON-Fade duration characteristics and a model for land mobile satellite communications channels" , IEEE. APS 1992.

[4] W.J.Vogel, J.Goldhirsh and Y.Hase, " Land-mobile-satellite propagation measurements in Australia using ETS-V and INMARSAT-Pacific", Jhons Hopkins University, Applied Physics Laboratory Tech. Rep. S1R89U-037, Aug. 1989.

[5] G.Butt et.al, " Narrowband channel statistic from multiband propagation measurements applicable to high elevation angle land mobile satellite systems", IEEE J-SAC, SAC-10(8), vol.10, no.8, Oct.1992, pp.1219-1226.

[6] S. Yamamoto, K. Tanaka, K. Suzuki and H. Wakana. "An antenna tracking method for land mobile satellite communications system. Estimation algorithm of antenna direction under magnetic disturbances", I.E.I.C.E. Tech. Report. SAT94-49~57 (1994).

[7] S. Yamamoto, K. Tanaka, H. Wakana and S. Ohmori, "An antenna tracking method for land mobile satellite communications system", Electronics and Communications in Japan, Part 1, vol.78, No.9, pp.91-102, 1995.

LMS Channel Measurements at EHF-Band

Axel Jahn, Erich Lutz
German Aerospace Research Establishment (DLR)
Institute for Communications Technology
P.O. Box 1116, D-82230 Wessling, Germany
Phone: +49 8153 28-2847 Fax: +49 8153 28-1442
{Axel.Jahn,Erich.Lutz}@dlr.de

I. Abstract

A measurement campaign is presented for high data rate satellite services at EHF-band. A land mobile terminal in a vehicle probes the received signal from an aircraft emulating the satellite. A high-gain antenna with a beamwidth of 25° is used for reception. The antenna is controlled by a GPS-based steering processor. Dual narrowband as well as wideband measurements can be performed simultaneously using independent antennas and receiver equipment. Results are presented showing the tracking performance. A comparison of narrowband signals using either an omnidirectional or a steered high-gain antenna is given, and the wideband characteristics of the LMS channel are illustrated.

II. Introduction

Next generation satellite systems will provide large-scale coverage for high performance services and seamless integration in UMTS. These services will offer high capacity links for multimedia applications in an extended service region. The spectrum allocated in the L-band and recently by WRC'95 in the FPLMTS-bands at 2 GHz will certainly not suffice to serve the increased demand for spectrum allocations. Higher frequency bands at Ka-band (20/30 GHz) and EHF-band (40/50 GHz) provide larger spectrum portions for LMS services. European institutions actively support research and technology development in these frequency bands. The ACTS project SECOMS (Satellite EHF Communications for Mobile Multimedia Services) investigates a 3rd generation GEO satellite system for European coverage that shall cover these topics [1]. In the framework of SECOMS, several field trials and service demonstrations are planned [2]. The paper describes a 40/50 GHz satellite propagation experiment. The experiments determine the narrowband and wideband characteristics of the satellite-to-mobile channel in various operational scenarios and environments.

A. Background

The effects of the land mobile satellite (LMS) channel play a big role in design and development of satellite-based communications networks. The mobile channel at EHF-band is still little known. To this end, tropospheric and ionospheric effects in feeder links have been investigated intensively [3]. At Ka-band, several measurement campaigns have also been reported for the LMS channel [4], [5], but not at EHF-band. The presented campaign tries to fill this gap.

The propagation path is predominantly subject to shadowing and multipath effects. Narrowband measurements describe mainly the shadowing of the channel. The attenuation caused by obstacles and the time-share of shadowing can be determined. Wideband measurements can determine the signal bandwidth at which the channel starts to become frequency-selective. This parameter is important when broadband services shall be transmitted through a channel. Wideband channel characterisation is especially important for the design of spread spectrum systems, e.g., to determine the frequency hopping intervals or the tap delays for the design of adaptive equalizers.

Since satellites are not yet available with sufficient link margins and transponder bandwidth to test the wideband and narrowband characteristics of the EHF-channel, a suitable aircraft carrier (CESSNA 207) was used to simulate the satellite. This approach allows also to test the channel properties for a wide range of environments in full azimuth and elevation ranges. The advantages and limitations of airborne test platforms are discussed in more detail in [6].

B. Objectives

The purpose of the measurement campaign is manifold:

- to perform narrowband measurements in order to collect a channel database with a wide range of environments, elevation angles and azimuth angles;
- to investigate the wideband characteristics of the land mobile satellite channel;
- to investigate antenna steering algorithms in presence of shadowing and fading, using high gain antennas.

The measurements have been performed at EHF-band between 39.5-40.5 GHz.

The user behavior for high bit rate services strongly depends on the terminal type. The following options

seem reasonable: (i) terminals in vehicles with roof mounted antennas (vehicle may drive or park), (ii) briefcase terminals with on-top antennas (user is likely to stay quasi-stationary), and (iii) handheld terminals with low data rates (the user may stay or move). The environments of interest are: open, rural with tree-shadowing (one or two sides), suburban, urban, or mountainous. The range of elevation angles for a user located in Europe (or North-America) and a geostationary satellite is restricted to 13°...50°. Higher elevation angles might also be interesting at EHF-band if one considers 2nd generation, Teledesic^{TM}-like low earth orbiting (LEO) systems.

III. Measurement Set-Up

The measurement set-up basically consists of a transmitter (TX) and a receiver (RX) part in a vehicle. The transmitter radiates a test signal, either wideband or narrowband, and the receiver is probing the channel for the transmitted test signal. The carrier frequency is 40.175 GHz. The receiver consists of two-stage demodulators transferring the 40 GHz EHF-signal to a 1.8 GHz intermediate frequency band. Commercially available channel sounders working in this band can now be used to measure the received signal. The wideband equipment uses a signal bandwidth up to 30 MHz and allows the measurement of 15 impulse responses per second. The dynamic range of the measurement system is better than 30 dB. Fig. 1 on the next page shows the block diagram of the receiver part.

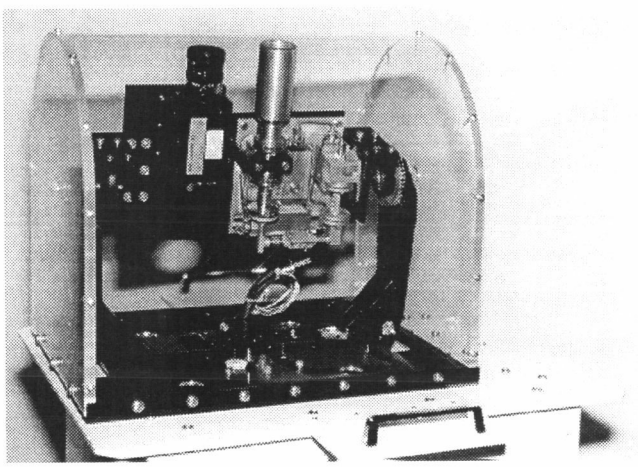


Fig. 2. Antenna steering platform with EHF-band receiver.

A SECOMS user terminal will likely need a high-gain self-steering antenna with a beamwidth of ±5°...±7°. For the channel measurements, the steering of the user antenna should be implemented in the most optimal way, in order to eliminate the effect of antenna steering. Therefore, a GPS-based implementation of the steering was selected. The aircraft sends its position through a telemetry link to the measurement vehicle. The receiver is equipped with a steering processor calculating the pointing angle to the aircraft emulating the satellite, cf. Fig. 1. A mechanical positioner is used to orient the user antenna, see Fig. 2. In order to achieve a high pointing accuracy of approximately 1°, the assessment of on-line differential GPS is required. A GPS reference station is therefore operated at DLR. This leads to a position accuracy of 1...2 m. A precise positioning reference system in the vehicle, providing the orientation with respect to a reference plane, is mandatory, too. An advantage of this antenna steering approach is that self-steering algorithms can be implemented in user terminals, e.g., using a signal strength criterion, which can be evaluated in a later stage of the project. The performance of phased-array antennas developed in the framework of the SECOMS project and different steering algorithms can then be compared with the ideal channel behavior measured with the GPS-based approach.

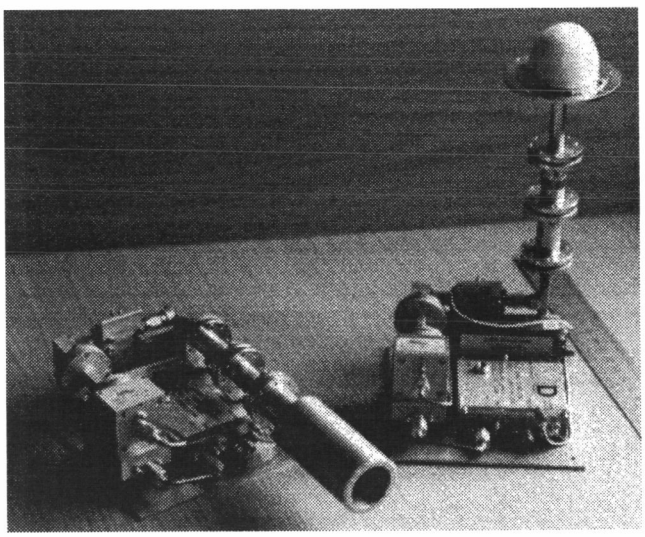


Fig. 3. 40 GHz receiver implementation.

Two independent receivers allow the simultaneous reception with different antennas or different receiver types. Thus, it is possible to measure narrowband and wideband signals, or narrowband signals with an omni-directional antenna and a high-gain antenna, for instance. High-gain antennas require steering which is only possible on one platform. The non-steerable platform is equipped with an omni-directional antenna. Fig. 3 shows the 40 GHz receiver implementation with mounted antennas. The different combinations of narrowband and wideband receivers with antennas and platforms allow

- the comparison of different antennas types,
- the comparison of narrowband and wideband measurements, and
- the investigation of the antenna steering algo-

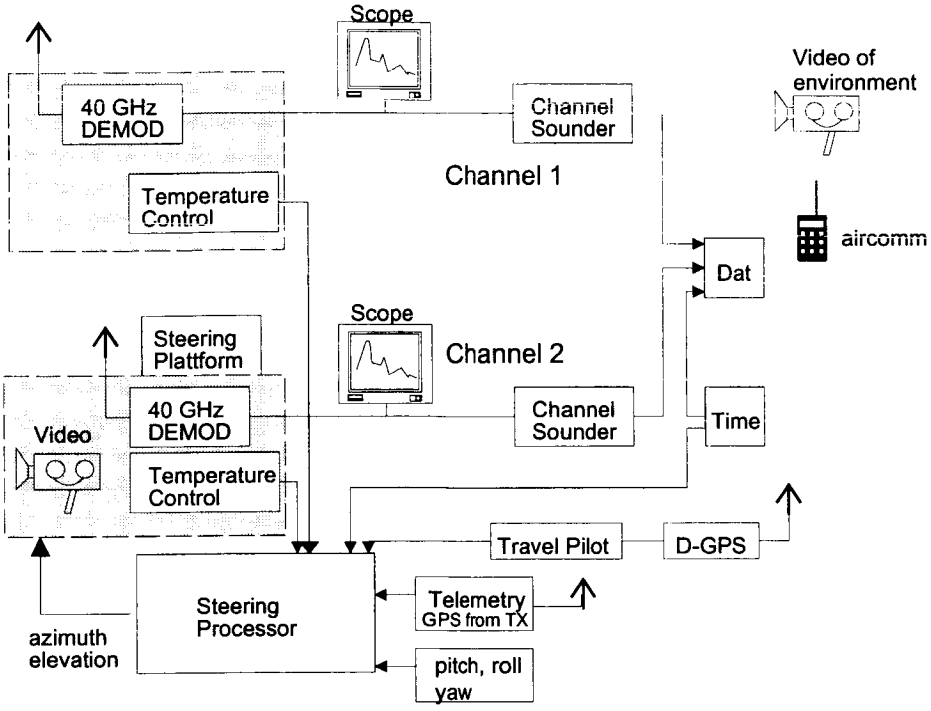


Fig. 1. Block diagram of the receiver.

rithms by comparison with an omni-directional reference antenna.

IV. Results

The results presented in what remains are taken from experiments in fall 1996 and spring 1997. The results have been used to set-up the data calibration and normalisation routines as well as to check the equipment functionality. All measurements were recorded in a driving vehicle. A more comprehensive set of measurements will be undertaken in summer 1997 to collect data for a channel database and to derive a channel model.

A. Antenna Tracking

Fig. 4 illustrates the performance of the GPS-based steering algorithm of the receive antenna. The pointing error between antenna boresight and the true direction towards transmitter is in general less than 1°. The mean deviation error is 0.45°. It should be noted that such a good performance can only be achieved through the assessment of the following preparatives:

- the use of differential GPS
- the use of an accurate position inertial system (3-axis controlled gyroscope)
- magneto-flux compensated compasses (we used also direction information from a travel pilot)
- prediction algorithms to compensate for transmission delays

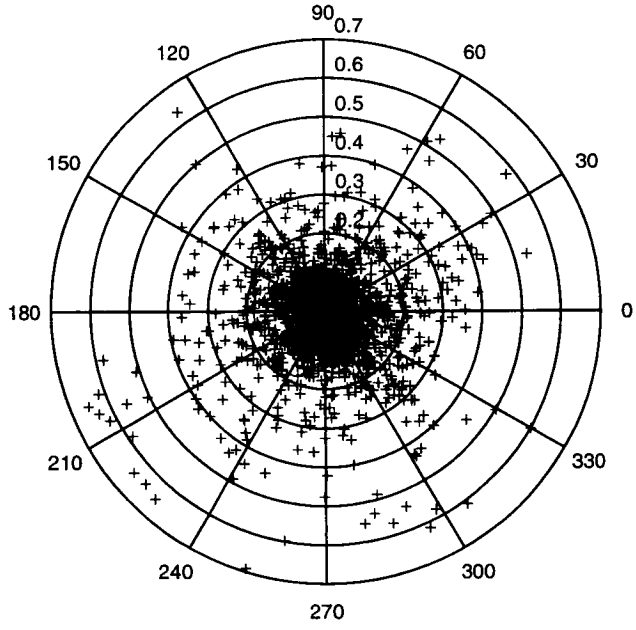


Fig. 4. Antenna pointing deviation from boresight in DEG

B. Narrowband Measurements

Narrowband power plots versus time in Fig. 5 compare the performance of a high-gain antenna (25°beamwidth) with an omni-directional one in a provincial road environment at an elevation angle of 25°. Little multipath fading can be observed beside short shadowing events. It is obvious that the omni-

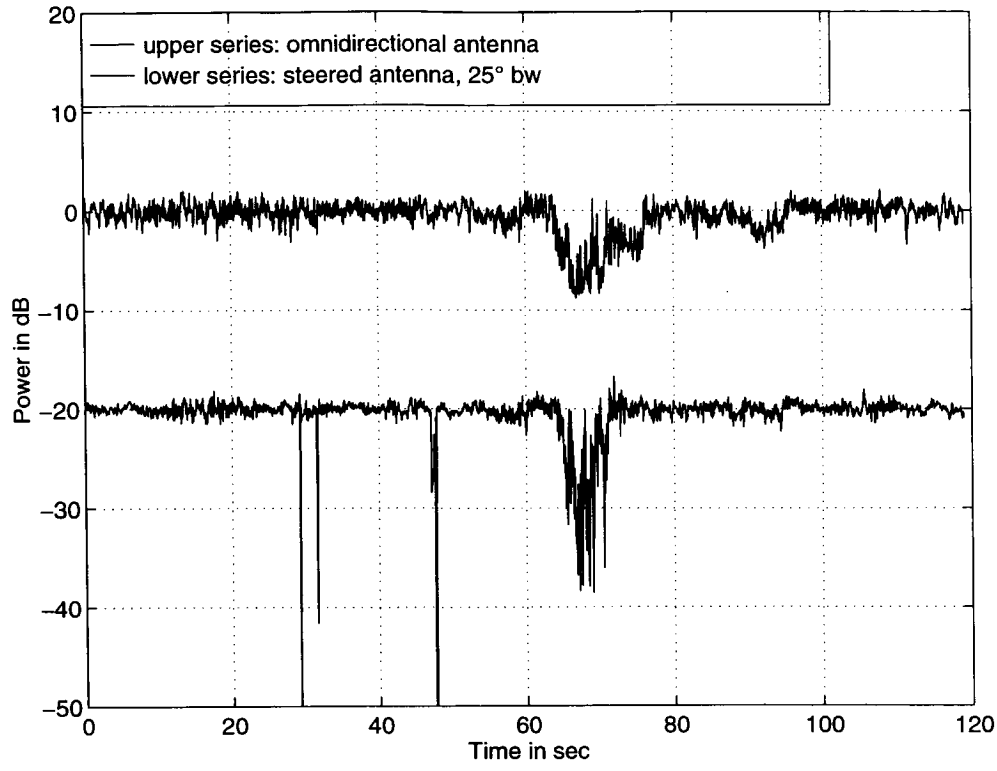


Fig. 5. Narrowband power series, provincial road, elevation: 25°

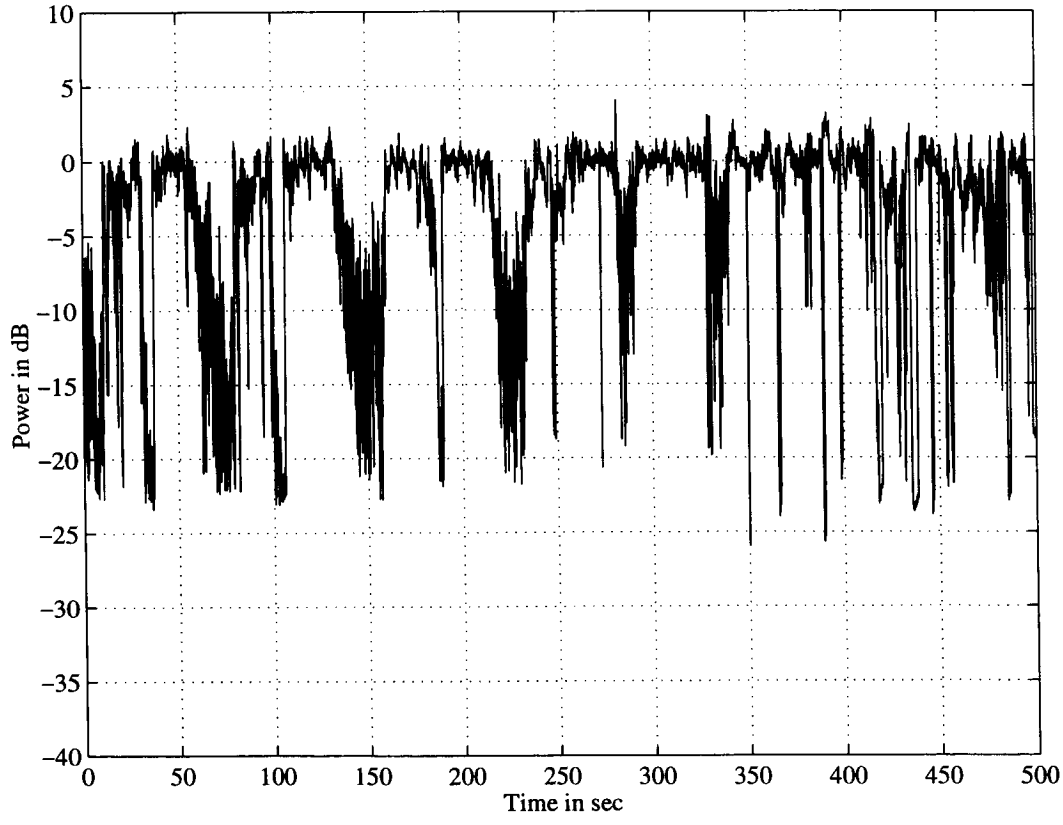


Fig. 6. Narrowband power series, urban, elevation: 25°

directional antenna picks up more multipath due to the lower discrimination of echoes, but for the same reason it is less sensible to short blockage. Note that for the purpose of better presentation, the series of the high-gain antenna was offset by -20 dB. The Rice-factor in non-shadowed conditions for the steered high-gain antenna is 21.5 dB, whereas the omni-directional antenna yields to an Rice-factor of 17 dB. In Fig. 6 a narrow-band measurement of the steered high-gain antenna in a urban environment is given. The power series is characterised by severe shadowing and blockage. In shadowed conditions, the channel attenuation is 10...20 dB. This relatively strong multipath power (also visible in the fading during LOS) indicates that the 25°-antenna can still pick up reflected and diffuse components.

C. *Wideband Measurements*

In Fig. 7 an example of the wideband results is shown. The measurement was taken in an tree-shadowed rural environment at 25°elevation using the omni-directional antenna. The 0-dB line corresponds to an undisturbed signal under line-of-sight condition. The signal on the direct path is subject to shadowing. Echoes appear very rarely, if any (note that signals below -29 dB are caused by the noise floor of the wideband receiver). Compared to results at L-band obtained at the same location, the number of echoes is much smaller and the echo attenuation is higher at EHF-band [7]. Fig. 8 shows the situation in the urban environment. The shadowing process for the direct component of the signal is now stronger. The attenuation reaches 15...25 dB if shadowed. Furthermore, the number of echoes is higher. There are few (2-4) echoes with short delays (30...100 nsec), and a couple of echoes with delays of 350, 600, and 850 nsec. The echo attenuation is in the range of 20...30 dB.

References

[1] G. Losquadro. Multimedia services using portable and mobile satellite terminals. In *ACTS Mobile Telecommunications Summit*, pages 411–418, Granada, Spain, November 27-29, 1996.

[2] G. Losquadro, B. Evans, A. Jahn, E. Lutz, A. Paraboni, P. Taaghol, and G. Tartara. SECOMS demonstrators for broadband mobile satellite links at Ka and EHF bands. In *ACTS Mobile Telecommunications Summit*, pages 165–172, Granada, Spain, November 27-29, 1996.

[3] U. Fiebig. Attenuation measurements of the 40 GHz beacon signal of Italsat. In *Proc. 3rd CEPIT Meeting*, Rome, Italy, October 9, 1995.

[4] M. Rice, J. Slack, B. Humpherys, and D. Pinck. K-band land-mobile satellite channel characterization using ACTS. *International Journal of Satellite Communications*, 14:283–296, 1996.

[5] L. Borghino, S. Buonomo, L. D'Amato, and C. Molinari. Measurements and analysis on Ka land mobile satellite channel. In *Mobile and Personal Satellite Communications 2 - Proceedings 2nd European Workshop on Mobile/Personal Satcoms (EMPS'96)*, pages 457–473. Springer-Verlag, London, 1996.

[6] P. Höher, T. Wörz, A. Schmidbauer, A. Jahn, and R. Schweikert. On satellite emulation by an airborne platform. In *Proc. IEEE GlobeComm'96*, pages 995–1000, 1996.

[7] A. Jahn, M. Sforza, S. Buonomo, and E. Lutz. Narrow- and wideband channel characterization for land mobile satellite systems: Experimental results at L-band. In *Proc. Fourth Int. Mobile Satellite Conf. IMSC'95*, pages 115–121, 1995.

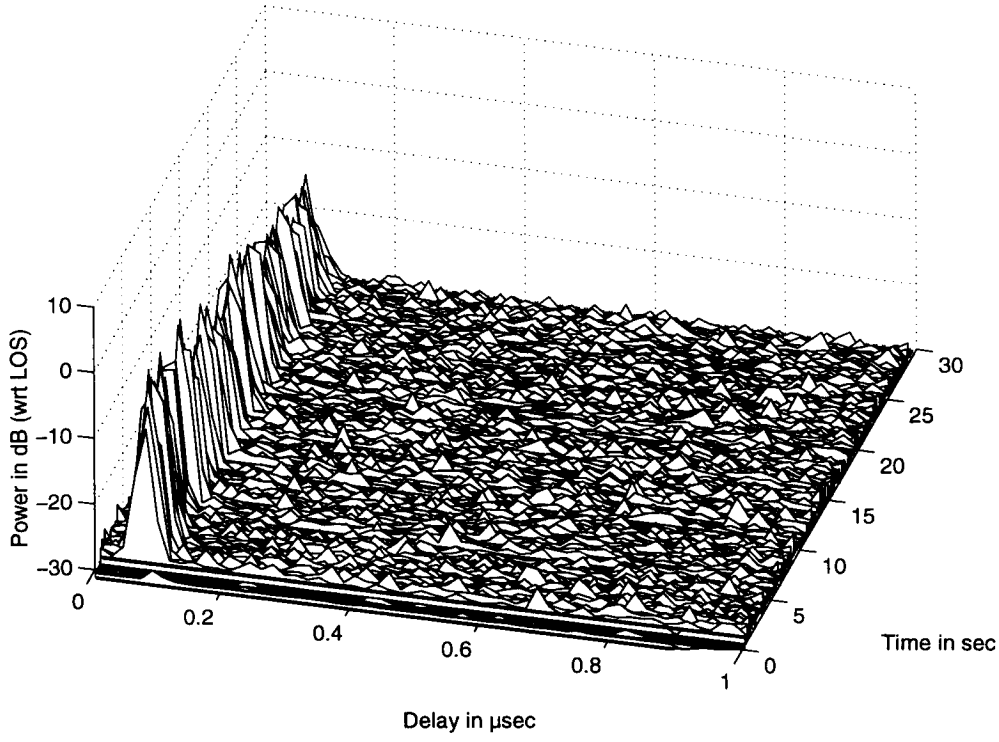


Fig. 7. Wideband power delay profile in an tree shadowed rural environment, elevation 25°.

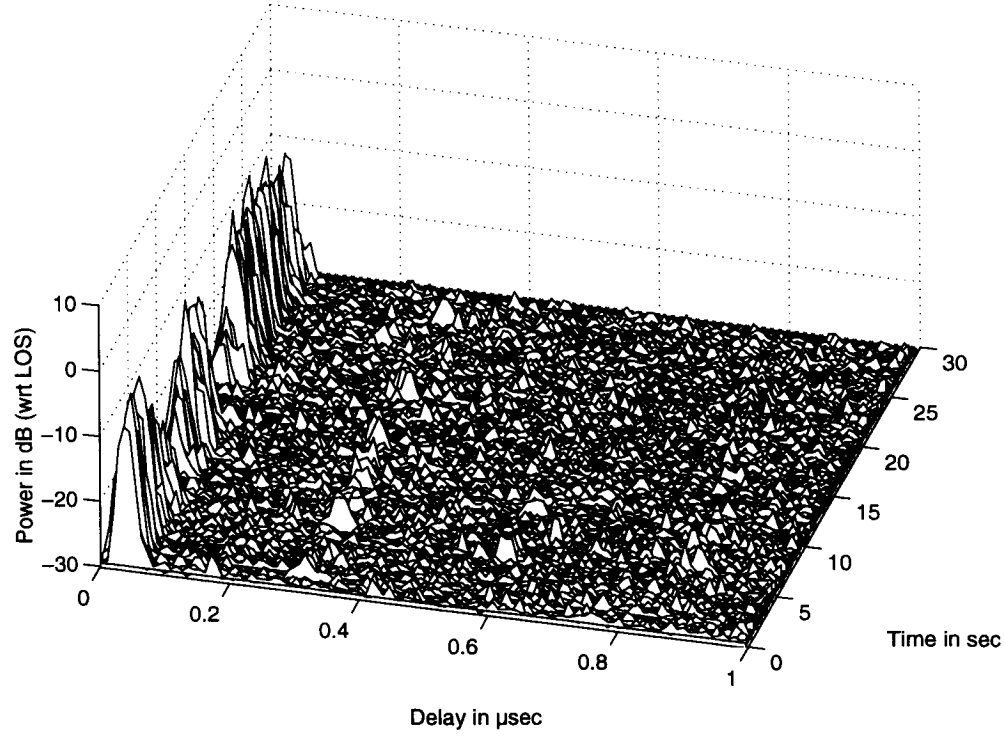


Fig. 8. Wideband power delay profile in an urban environment, elevation 25°.

Satellite Diversity as a Propagation Impairment Mitigation Technique for Non-GSO MSS Systems

Riza Akturan and Kent Penwarden
Globalstar, L.P.
3200 Zanker Road, P.O. Box 640670, San Jose, CA 95164-0670, USA
Tel (408) 473-6113, 473-4082. Fax (408) 473-6227
E-mail: Riza.Akturan, Kent.Penwarden@Globalstar.Com

ABSTRACT

This paper presents methodologies for predicting the improvement in performance of non-GSO MSS Systems which may be achieved by using satellite diversity. The status of the channel between satellite and user depends on the propagation state of the path. In this paper, propagation models which describe the propagation states of those links have been derived from optical and microwave measurements. The simulations of both of these models have shown that satellite diversity notably improves the availability of the link, and therefore, increases the performance of the satellite service to the users that are often shadowed or blocked by terrestrial objects.

1. INTRODUCTION

A path between satellite and user terminal is considered "available" when it may be used by the system for purposes of communication, e.g., when a call of a specified duration can be placed and completed, or a circuit can be established and maintained for a specified duration. Thus, the status of the channel between satellite and user depends on the state of the path. The path between a satellite and a user can simply be described in three path states; clear without any obstruction, shadowed by vegetation and blocked by solid objects where the shadowed and blocked paths cause link impairments.

Satellite diversity is a propagation impairment mitigation technique designed to improve path availability both for shadowing and blocking problems. The satellite diversity allows for the possibility of combining several paths which provide communication between more than one satellite and a single mobile earth station (MES). This method is illustrated in Figure 1.1 where the highest satellite is blocked, but another satellite is clear to the MES, thus, the link is available.

This paper presents two methodologies for predicting the performance improvement which may be achieved by using satellite diversity. The two methods herein described use a fading model derived from microwave measurements and a state analysis derived from the photographic images of environments [1] for the simulation of satellite diversity. The both methodologies incorporate a non-geostationary (non-GSO) satellite coverage model and predict a quantifiable improvement for the actual system implementation.

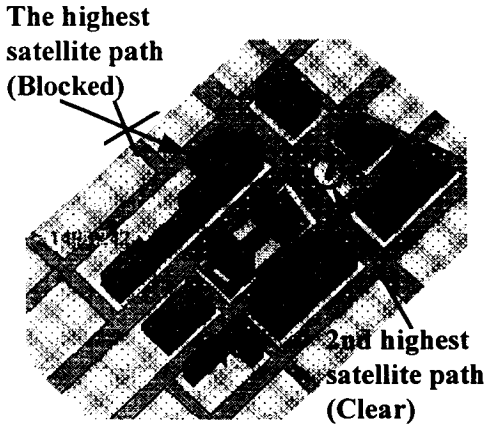


Figure 1.1 Basic description of satellite diversity.

The methods presented here are not dependent upon either a particular satellite constellation, or the propagation model employed, or any particular system parameters, therefore, the conclusions in this paper are the general descriptions of satellite diversity. They also do not represent structure-induced multipath and body blockage conditions. The structure-induced multipath in non-GSO mobile-satellite service (MSS) systems for line-of-sight (LOS) conditions will be small compared to the fading induced by foliage shadowing and/or user-body blockage. Also, the body blockage effect can be ignored assuming that the user is co-operative and antenna is above head.

Other propagation impairment mitigation techniques which may be applied in concert with satellite diversity are: 1) maintaining pre-established fade compensation margin in the link budgets (for receiver automatic gain control), 2) transmitter power control, 3) digital interleaving, and 4) forward error correction. These communication techniques are effectively employed in digital terrestrial cellular radio systems to improve the reliability, capacity, and quality of communications and to reduce power requirements

mobile earth stations, and that they are available to be applied to non-GSO MSS systems.

2. SATELLITE DIVERSITY IMPLEMENTED WITH CDMA ACCESS TECHNIQUE

Satellite diversity is a scheme to provide service-link communication paths for a MES (mobile earth station) by way of more than one satellite. To implement such a scheme requires that means be employed to distinguish the signals received and transmitted by way of the several satellites.

To implement satellite-diversity reception, separate parallel receivers within a MES and within a gateway must be tuned to signals from the several satellites. Means must be implemented to maintain signal independence for receiver processing to enhance quality and then to combine the outputs of the several receivers in an optimum manner to result in a signal superior to any single signal.

In the diversity scheme considered in this paper, all the multiple paths communicate with a single MES home onto a single gateway earth station; that is, the multiple satellites employed would be at once in view of both the MES and the gateway site. On the forward-link, a MES needs to be capable of separately receiving, resolving and optimally combining the multiple signals received from the multiple satellites. On the return-links, signals transmitted from a single MES and received by way of multiple satellites at a gateway earth station would be distinguished by path diversity. Separate tracking directional antennas would operate for each satellite in view.

Digitally implemented CDMA waveforms provide the means for the parallel-receiver processing of images of a signal which have been received by way of multiple satellites and also the multiple beams within a single satellite. In the CDMA MSS system, signals may be sent from a single gateway earth station by way of multiple probable useful paths to an MES, with unique code delays generated for each path. In this scheme, the satellite transponders are linear bent pipes without on-board processing, which act analogously to reflection surfaces in terrestrial multipath generation. The CDMA spread spectrum systems typically use rake receivers. Each rake tine is a separate demodulator. Images of the same signal received by way of multiple satellites are processed in separate parallel demodulators comprising the rake. Each rake receiver allows signals on individual path arrivals to be tracked independently and the received signals to be combined. While there may be fading on the separate paths through the satellites, the fades are anticipated to be independent, and demodulation based on the combined signals should then be much more reliable than the signal through any one satellite.

3. AN OVERVIEW OF A SATELLITE DIVERSITY SYSTEM: GLOBALSTAR

Globalstar is a satellite-based digital telecommunications system that will offer wireless telephone and other telecommunications services worldwide beginning in 1998 [4]. Globalstar will provide low-cost, high quality telephony and other digital telecommunications services such as data transmission, paging, facsimile and position location to areas currently underserved or not served by existing wireline and cellular telecommunications systems.

Globalstar service employing satellite diversity will be delivered through a 48-satellite LEO constellation. Users of Globalstar will make or receive calls using fixed, hand-held or vehicle-mounted MESs similar to today's cellular telephones. Because Globalstar will be fully integrated with existing fixed and cellular telephone networks, Globalstar's dual-mode MES units will be able to access to telephony services by both conventional cellular network and the Globalstar satellite network. Using the dual-mode capability, in urban environments the Globalstar MESs will most likely access a terrestrial cellular radio-telephone system, but in suburban, rural and wilderness areas they will have access to Globalstar. Each subscriber terminal will communicate through satellites to a local Globalstar gateway which will in turn connect into existing telecommunications networks.

Globalstar expects that its CDMA technology (by Qualcomm) combined with continuous multiple satellite coverage and path diversity will enable the Globalstar system to provide service to a wide variety of locations, with less potential for signal blockage from buildings, terrain or other natural features. MESs will be able to operate with one or more satellites in view. Typically two to four satellites will be overhead. As satellites are constantly moving in and out of view, they will be seamlessly added to and removed from the calls in progress, thereby reducing the risk of call interruption. Therefore, the loss of an individual satellite will not result in any gap in global coverage compared to systems which have overlapping coverage areas.

4. SIMULATIONS USING EERS MODEL

Long-duration average propagation statistics pertaining to land mobile satellite system design have been collected for MSS systems. The Extended Empirical Roadside Shadowing (EERS) Model described in [6], was used to determine path availability along a tree-shadowed road as a function of satellite elevation angle.

There is strong influence of elevation angle on fading, therefore, the instantaneous elevation angle of each satellite must be considered in the prediction of

performance of a non-GSO MSS satellite system employing satellite diversity.

A computer model of a simulated satellite constellation was used to calculate the geometry between a MES and the satellites at one minute intervals at selected latitudes. At each interval, the simulation calculated the elevation angle of each satellite. In the system modeled, the satellites where those which were co-visible to both the MES and a gateway. The constellation simulated is the 48-satellite Globalstar constellation with circular orbits of 1400 km altitude arranged in a Walker "delta-pattern" at 52° inclination for uniform distribution of overlapping coverage around the world.

Separate simulations were run for gateways located at a series of different latitudes; the satellite constellation simulated is approximately independent of longitude. For these simulations, the MES and the gateway were approximately in the same location, where the local propagation impairments surrounding the MES are predominant in determining availability. Separating the MES and gateway could reduce the diversity gain by lowering the number of satellites available in view.

The EERS model provides a cumulative distribution function of shadowing fade at 1.6 GHz. The cumulative fade distribution is of the form A(el,p), where A is the fade exceeded (dB) with respect to free-space propagation, p is the percentage over which the fade is exceeded, and el is the path elevation angle (degrees) to the satellite.

At each interval of the simulation for each satellite, el is provided by the simulation, p is generated as a random number, and a random fade A(el,p) is calculated. At each interval the value of A is compared to a pre-established fade compensation margin, and if A is less than that fixed value, the path is declared "available".

A separate simulation was performed for the following diversity schemes: 1) the highest satellite without diversity (handing off when the next-highest satellite rises to become the highest satellite, 2) at least one of the two highest satellites, and 3) at least one of the three highest satellites. The number of intervals where a link was available in the duration of the simulation was accumulated, divided by the total number of intervals and multiplied by 100% to determine the percentage of time that scheme was available. This was repeated for the other schemes.

Since there are several parameters which can be varied, several views of the modeling results are possible. The parameters which may be varied are: 1) depth of fade to the blocked condition (or fade compensation range required), 2) latitude, 3) probability (%) of being blocked by trees (being unavailable for communication), and 4) number of satellites in diversity scheme.

One useful example is to calculate "Percent availability" as a function of "Fade compensation" for a fixed latitude. Figure 4.1 shows the calculated path availability for a given fade compensation margin for each of three diversity schemes evaluated. Observation of Figure 4.1 shows that in this tree-shadowed environment at 1.6 GHz and 40° latitude, an availability goal of 98% may be achieved with only 6.9 dB of fade compensation given two-satellite diversity, where 12 dB may be required for a highest-satellite hand-off scheme without diversity. While three-satellite diversity is projected to further lower the required fade margin, it is evident that with this constellation the greatest single benefit would be achieved with two-satellite diversity.

Another view of the results of the same simulation are presented in Figure 4.2 where the fade compensation was held fixed at 6.9 dB and the latitude was varied.

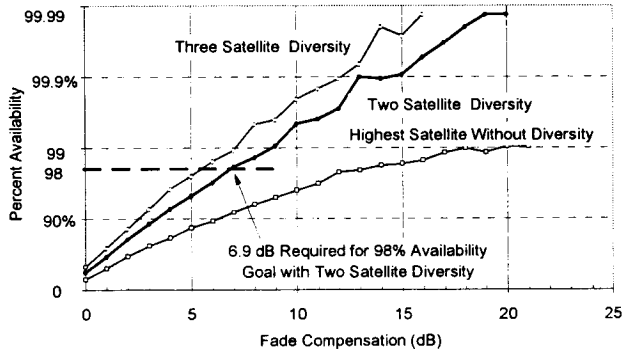


Figure 4.1 Propagation through trees at 1.6 GHz. Availability versus fade compensation at 40° latitude.

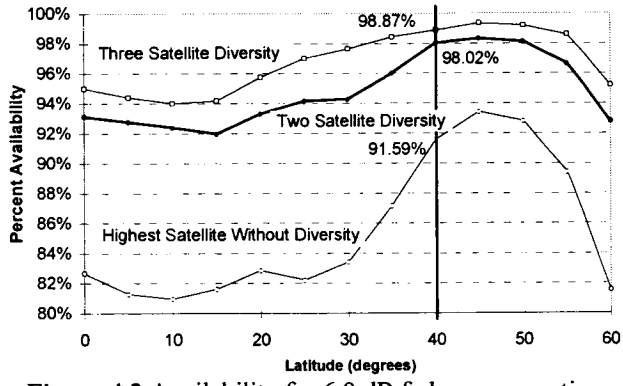


Figure 4.2 Availability for 6.9 dB fade compensation.

5. SIMULATIONS USING PHOTOGRAMMETRY

The state of the LOS path basically depends on local features of the environment. Based on these assumptions, satellite diversity has been analyzed with optical (photogrammetric) measurements. Briefly, this method is based on measuring the propagation characteristics by

processing fisheye lens photographs of the environments which results in a three-state description of environments as clear, shadowed, and blocked [1]. An example photogrammetric fisheye image is shown in Figure 5.1, on the left. On the right, it's processed version is shown as a three-state image. The clear, shadowed and blocked portions of the image were identified in different colors.

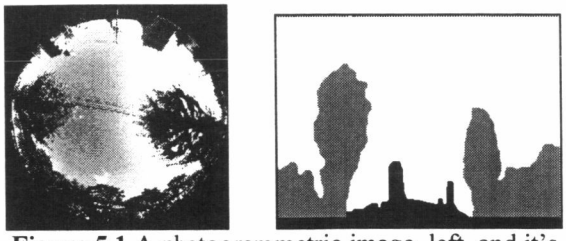


Figure 5.1 A photogrammetric image, left, and it's processed version; a three-state image, right [2].

With the low transmitter signal power levels available both to the non-GSO MSS satellite and to handheld MESs, building blockage cannot be overcome with practical increased fade margins; only a diversity path will provide mitigation to this impairment. Availability of a path will be in proportion to the openness of the sky above a MES. Any clear (neither blocked nor shadowed) signal path available to just one of the satellites in view must be sufficient to provide the requisite availability. In rural and suburban environments, blocking by man-made structures should be minimal for a non-GSO MSS system employing satellite diversity to avoid shadowing.

Using the photogrammetric method previously in [2], path diversity gain has been investigated for latitude dependence and different diversity schemes. It has been shown that the diversity gain levels out after using 2 or 3 satellites. Latitude dependence is specific to each constellation. The potential of satellite diversity to mitigate tree-shadowing or blockage depends on the variability and structure of the skyline with azimuth at any particular location and on the satellite constellation. The optimal separation of satellites is a function of the environment. In [1], it was also shown that best azimuth separation for the highest diversity gain between two satellites is 180° and the worst is 0°. The mean separation for a reasonable diversity gain is at least 60°. Globalstar's mean separation is 97°.

In this section, environment dependence of satellite diversity was determined using the photogrammetric method. This was done by measuring the diversity gain obtained by receiving signals from up to four best satellites from the Globalstar constellation in different user environments. Urban, suburban and rural Central Texas environments were considered.

5.1 Improvement of Link Margins

Visibility of the Globalstar constellation has been simulated for a MES in Central Texas. The results of this simulation should be valid at most temperate zone locations. The simulation resulted in the elevation and azimuth angles of the visible Globalstar satellites as a function of time. The satellite visibilities have been compared with the three-state photographic images of the Central Texas locations. Hence, each visible satellite's path state was identified as clear, shadowed or blocked as a function of time.

Then, the propagation characteristics of these paths were described by distinct signal distribution models according to their states. For the urban environment, the Urban 3-State Model [2] was used to acknowledge the presence of specular reflections in the blocked urban environments, where Rician distribution represents clear-state signal distribution, Loo's distribution represents both shadowed-state and blocked-state signal distributions. In suburban and rural environments, Rural 3-State Model [2] has been used, where Rician distribution represents clear-state signal distribution, Loo's distribution represents shadowed-state signal distribution, and Rayleigh distribution represents blocked-state signal distribution. All models represent L-Band measurement results. Using these procedures, path attenuation statistics have been estimated for diversity and no-diversity operations for the three Central Texas environments using the procedures in [1].

"Satellite diversity gain is the decrease in fading at a given cumulative probability when more than one satellite can be accessed by a mobile user terminal, compared to the fading experienced when only one satellite, usually the highest one, is available for communications" [2]. Using this definition, combining diversity gain has been estimated from the path attenuation statistics for employing up to four satellites in the link where the usable signals received from visible satellites are added in order to form an acceptable signal level at the user terminal. Results showed that diversity gains obtained in the three different environments have similar characteristics, a visible decrease in fading which in turn results in a better quality of service with less usage of the system resources. In urban Austin, 10 dB gain is obtained at 95% with 2-best-satellite diversity as shown in Figure 5.2. The satellite diversity gains were 7 dB for the suburban and 4 dB for the rural Austin environments as depicted in Figures 5.3 and 5.4, respectively, at 95% availability.

Also, it has been noted that increasing diversity beyond 3-best satellites achieved a small additional gain in all environments, where the best satellites provided the highest power among all visible satellites. The

diversity gain was the least in the rural environment although there was still a significant gain compared to the non-diversity operation simulation. Most of the visible satellites were in the clear-state in the rural environment because of the low average skyline elevations, 5.3° [3]. This condition also applied to the suburban environment where the average skyline elevation was 7.4°. The highest gain was obtained in the urban location where the average skyline elevation angle was 27° and its standard deviation was high [3].

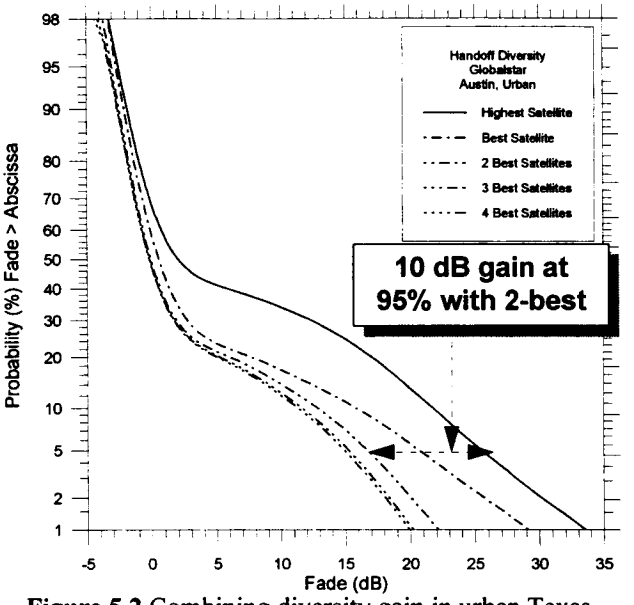


Figure 5.2 Combining diversity gain in urban Texas.

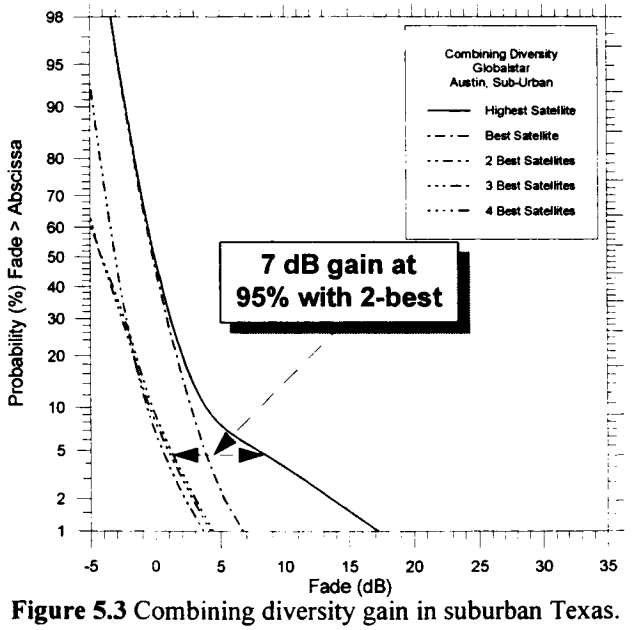


Figure 5.3 Combining diversity gain in suburban Texas.

Figure 5.5 shows the results of a comparison between four environments for the service availability of 90% by receiving signal from only the highest satellite or two best satellites. It is seen from this figure for 90% service availability that the improvement obtained by employing diversity with two best satellites provides 9 dB, 3 dB and 2 dB diversity gains in urban, suburban and rural environments, respectively, compared to employing only the highest satellite, non-diversity operation.

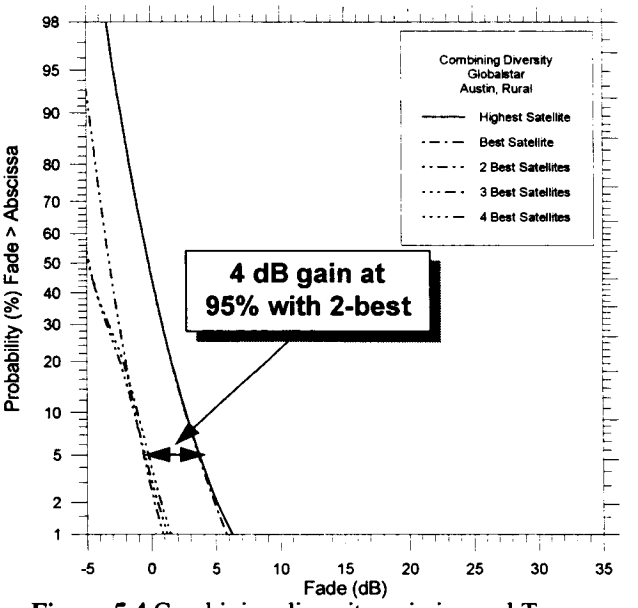


Figure 5.4 Combining diversity gain in rural Texas.

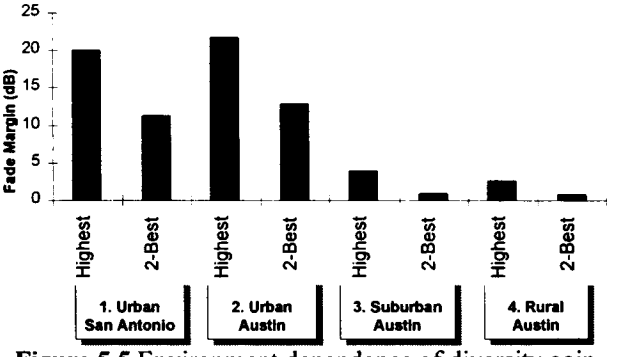


Figure 5.5 Environment dependence of diversity gain for 90% service availability.

5.2 Improvement of Call Durations

A call interruption and availability analyses has also been implemented for different environments for diversity and non-diversity operations. For the non-diversity operation, it has been assumed that a good call is established if the highest satellite is in the clear-state. For the diversity operation, it has been assumed that a good call is established if, at least, one visible satellite is in the clear-state.

In this analyses, the highest and the two best satellite states have been recorded as a function of time for the Globalstar constellation. Then, random calls with

various lengths have been placed and the call success and interruption statistics have been calculated. Calls with more than 18 seconds interruption length were dropped. The results are as follows.

Figure 5.6 shows that employing diversity instead of non-diversity in urban environments increases 3-minute good-call durations by about 25%. The same operation in the suburban environment increases good-call durations by 12% (from 87% to 99%). The increase in the rural environment is only 1% (from 99% to 100%). This concludes that satellite diversity increases the service quality of non-GSO MSS systems in the hard-to-serve urban and suburban environments.

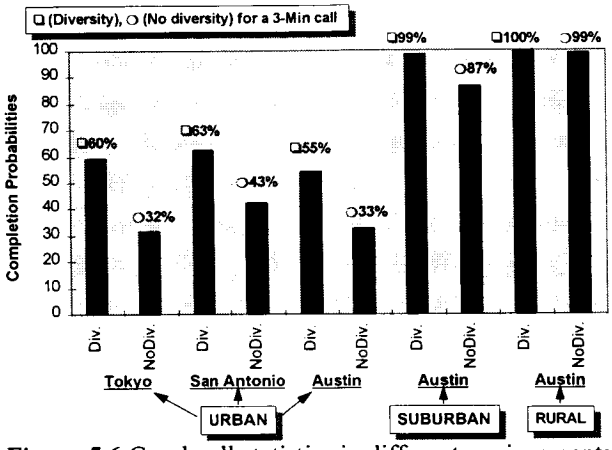


Figure 5.6 Good-call statistics in different environments.

6. CONCLUSIONS

The status of the channel between satellite and user depends on the propagation state of the communication path. Multiple satellites used in a diversity arrangement could be an effective means of overcoming propagation impairments. In this paper, two independent propagation models have been incorporated into the simulation of a non-GSO satellite constellation to predict link availability for diversity and non-diversity operations.

Simulations using the EERS method showed that in a tree-shadowed environment at 1.6 GHz and 40° latitude, an availability goal of 98% may be achieved with only 6.9 dB of fade compensation with satellite diversity operation, where 12 dB may be required with non-diversity operation; a 5.1 dB diversity gain. Simulations using the photogrammetric method supported the results of the EERS method. In the simulations of the suburban and rural areas, diversity gains of 7 dB and 4 dB were obtained. In the urban area, the diversity gain was 10 dB with 2-satellite diversity operation, compared to a non-diversity operation.

Also using the photogrammetric method, it has been estimated that employing diversity instead of non-diversity in urban environments increases 3-minute uninterrupted (good) call durations by about 25%. The same operation in the suburban environment increases good-call durations by 12%. Therefore, the service quality of non-GSO MSS systems employing satellite diversity notably increases in the hard-to-serve urban and suburban environments.

The above simulations have shown that satellite diversity notably improves the link availability, and therefore, increases the performance of the communication system and the quality of service to MESs that are shadowed or blocked by terrestrial objects. As a result, the satellite diversity technique lowers the need for high link margins and other complex fade mitigation methods to provide a good quality of service.

High link margins require bigger and heavier satellites to supply higher power levels to MESs on the Earth. The satellite diversity technique is implemented within the CDMA technology of the Globalstar system. Thus, satellite diversity allows the Globalstar to built smaller, lighter and less expensive satellites providing better call completion probability and superior performance with smaller link margins.

REFERENCES

[1] **R. Akturan**, "Photogrammetric Mobile Satellite Service Prediction," Doctoral Dissertation, The University of Texas at Austin, Austin, TX, Aug., 1996.

[2] **R. Akturan, and W. J. Vogel**, "Path Diversity for LEO Satellite-PCS in the Urban Environments" Accepted for publication by IEEE Transactions on Antennas and Propagation.

[3] **R. Akturan, and W. J. Vogel**, "Photogrammetric Mobile Satellite Service Prediction," Proceedings of the Eighteenth NASA Propagation Experiments Meeting, NAPEX XVIII, Vancouver, BC Canada, June 16, 1994.

[4] **J. Schindall**, "Concept and Implementation of the Globalstar Mobile Satellite System", Proceedings of the fourth International Mobile Satellite Conference, Ottawa, Canada, June 6, 1995.

[5] **J. Goldhirsh and W. J. Vogel**, "Propagation Effects for Land Mobile Satellite Systems: Overview of Experimental and Modeling Results," NASA Reference Publication 1274, Feb. 1992.

[6] Documents 3M/39 and 3/45, Revisions to Recommendation ITU-R P.681-2.

Design of Mobile Satellite Communication Systems Affected by Rain Attenuation: Results at 19.77 GHz derived from Meteorological Radar Measurements

Emilio Matricciani
DEI and CSTS-CNR, Politecnico di Milano
Piazza L. da Vinci, 32 - 20133 Milano (Italy)
Phone: 39-2-2399-3639 Fax: 39-2-2399-3413
email: matricci@elet.polimi.it

ABSTRACT

The simulations of vehicles driven in zig-zag routes with speed probability distribution modelled as lognormal for urban (Model I) and suburban (Model II, higher values of speed) traffic, in a statistically significant set of rain rate maps derived from a meteorological radar located in Spino d'Adda, have shown: (*a*) in the mobile system a given value of *A* (dB) is exceeded, on the average, for less time than in the fixed system; (*b*) the higher the probability of going straight, the less is the time *A* is exceeded for the vehicle. When the vehicles are driven in loops, then there is no significant difference between mobile and fixed systems. Mobile systems performance improves when Model II is used. The gain *G* (dB) obtainable may be of the same order of magnitude of that provided by FEC codes, or even larger.

INTRODUCTION

After [1] we have investigated how rain attenuation (*A*, dB) cumulative probability distributions (cdfs) of fixed satellite systems are transformed to those applicable to mobile satellite systems working in the same frequency bands and weather conditions, in the special case of vehicles driven in zig-zag patterns, to simulate city streets. The vehicles speed was modelled as a lognormal random variable, a mathematical model derived [2] from measurements performed in Italy. As for rain, we have used a significant number of rain rate maps sampled every 1.5 minutes by a 2.8 GHz meteorological radar located at Spino d'Adda (45.5° N, 9.5° E) in 1989, and operated by CSTS-CNR (Centro di Studi per le Telecomunicazioni Spaziali of Italy's National Research Center, CNR). On these maps we superposed a square grid to represent city streets and simulated moving terminals at 19.77 GHz, towards a geostationary satellite located at 19° W, i.e. the former Olympus satellite position, the 19.77 GHz beacon of which was received at Spino d'Adda for CSTS propagation experiments, with slant path elevation angle of 30.6°. In this paper we compare the cdfs of the mobile system with those of the fixed system in the same geographical area.

RADAR DATA BASE

In 1989 the radar was operated by CSTS to collect reflectivity (Z, mm^6/m^3) maps in polar coordinates, for 33 rainy days randomly sampled from February to December, in a way that allowed to reconstruct *Z* maps belonging to a plane parallel to the Earth surface at an altitude of about 1 km, to minimize clutter and still retain good estimates of *Z* at ground. The total number of maps processed for 1989 is 10,037, each of which lasts 1.5 minutes. The reflectivity maps were then converted into rain rate (*R*, mm/h) maps by inverting the well-known relationship $Z=200R^{1.6}$. The minimum observable rain rate is 0.5 mm/h. The geographical area observed at the altitude of 1 km around the radar site is circular with 40 km radius and slant range resolution of 150 m. The circular area of 0.9 km radius could not be observed because of the finite time required to switch the radar between the transmit and receive mode.

The original data were then routinely inspected by CSTS, the clutter was identified and labelled accordingly. From these data further processing gave 25,600 square (Cartesian) cells of side 500 m (linear spatial resolution for our simulations) relative to a square geographical area of side 80 km in which the originally observed circular area is inscribed. To determine the rain rate value *R* of each square cell of side 500 m, its 0.25 km^2 area was divided into 64 (8x8) "pixels" and then the 64 original rain rate values found in polar coordinates were averaged to yield the value of *R*. Cells outside the circular area were removed and the border cells with less than half valid pixels were also excluded. The same method of validation was used for the inner cells affected partially by clutter. However, for the present simulations, we gave these inner cells a value of *R* extrapolated from the values of *R* found in the surrounding cells with valid data: this procedure is necessary because

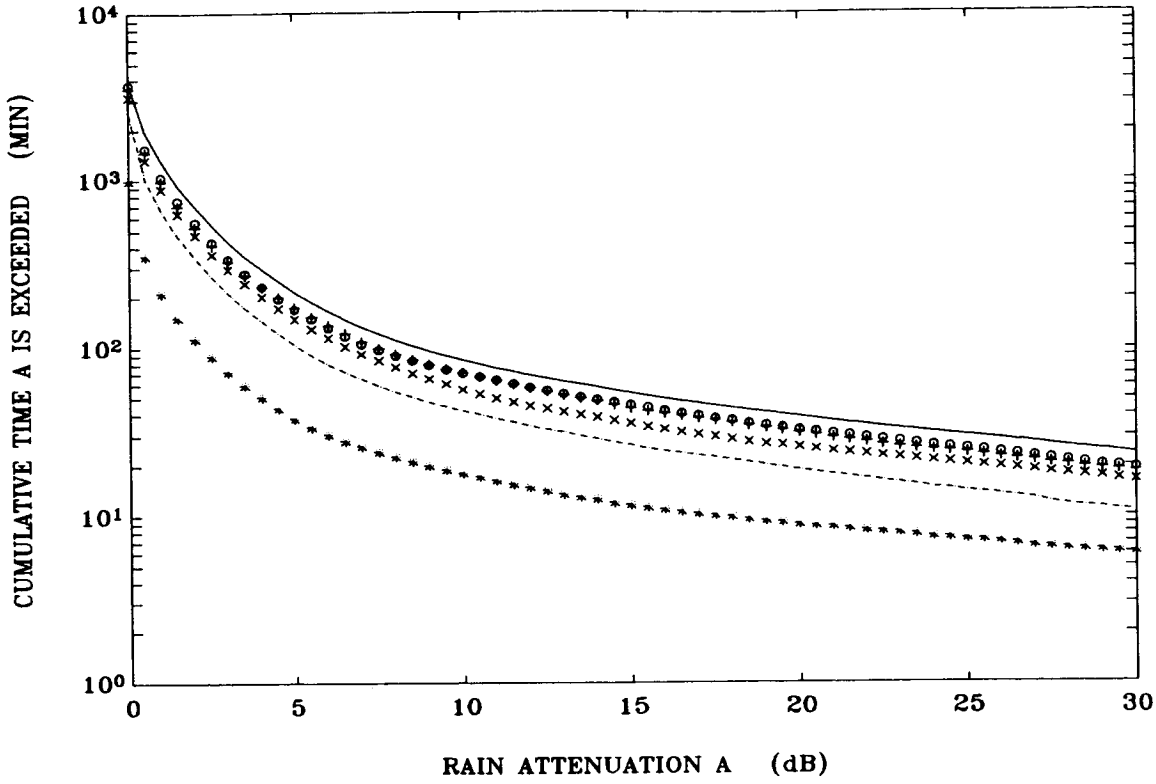


Figure 1: Cumulative time exceeded (min) vs. A (dB) in the fixed system (continuous line) and in the mobile system (zig-zag), p_S =0.0: "o", p_S =0.2: "+", p_S =0.5: "x", p_S =0.8: "-.", p_S =1.0: "*". Vehicles started inside the geographical area, speed Model I (urban).

we need retain as much as possible a more correct continuity in the spatial extent of rain. The procedure was also adopted for the area of 0.9 km radius around the radar site.

Since the linear spatial resolution is 500 m, and never more than 20 surrounding cells are necessary for this averaging, we think that, in the long term, the error in R is not significant and even less significant is the error in A, since A is computed as a line integral of R over several cells. The total number of useful cells amounts to 20,300, that is a geographical area of 20,300x0.25km^2=5075 km^2, to be compared with the circular area really observed of 5026,5 km^2 (the difference is due, of course, to the artificial square cells obtained above).

SIMULATIONS

Rain attenuation calculation: To calculate A in the 30.6° slant path we modelled the vertical structure of precipitation with two layers, as in [3], with precipitation height given by the 0°C isotherm height suggested by ITU-R (reported in eq. (1) of [4]). The lower layer models rain, the upper layer models the melting layer. A is calculated as a line integral as in eq. (20) of [4] and according to the values of R found in the cells sampled by the projection at ground of the slant path (see [3] for details on the constants k and α that appear in the relationship $Y=kR^{\alpha}$, being Y (dB/km) the specific attenuation). Since in the real world both fixed and moving terminals would measure the same value of A (but for different intervals of time), A is calculated only once for each rain rate map, by assuming that the fixed terminal is located at the center of the cell. To force the moving terminal to also measure the same value of A, not only it is driven through the center of the cell, but its attenuation does not change along the distance through that cell. The quantity that does change from vehicle to vehicle is the fade duration. The number of the useful cells defined above is less than 20,300 as we excluded other cells because of the following argument: since the rainy path length L towards Olympus is 6.36 km and its projection at ground is thus $D=L\cos 30.6°=5.47$ km, we excluded those cells for which it is not possible to integrate Y for the entire path towards the satellite. By doing so, we were sure that both fixed and moving terminals could always "measure" a value of A relative to the complete slant path, if rain were somewhere along the path. At last, we obtained 18,205 cells for each map, i.e. we covered 4551 km^2. Notice that the number of fixed terminals is thus 18,205, a very large number, more than sufficient to yield a reliable estimate of the fixed system cdf

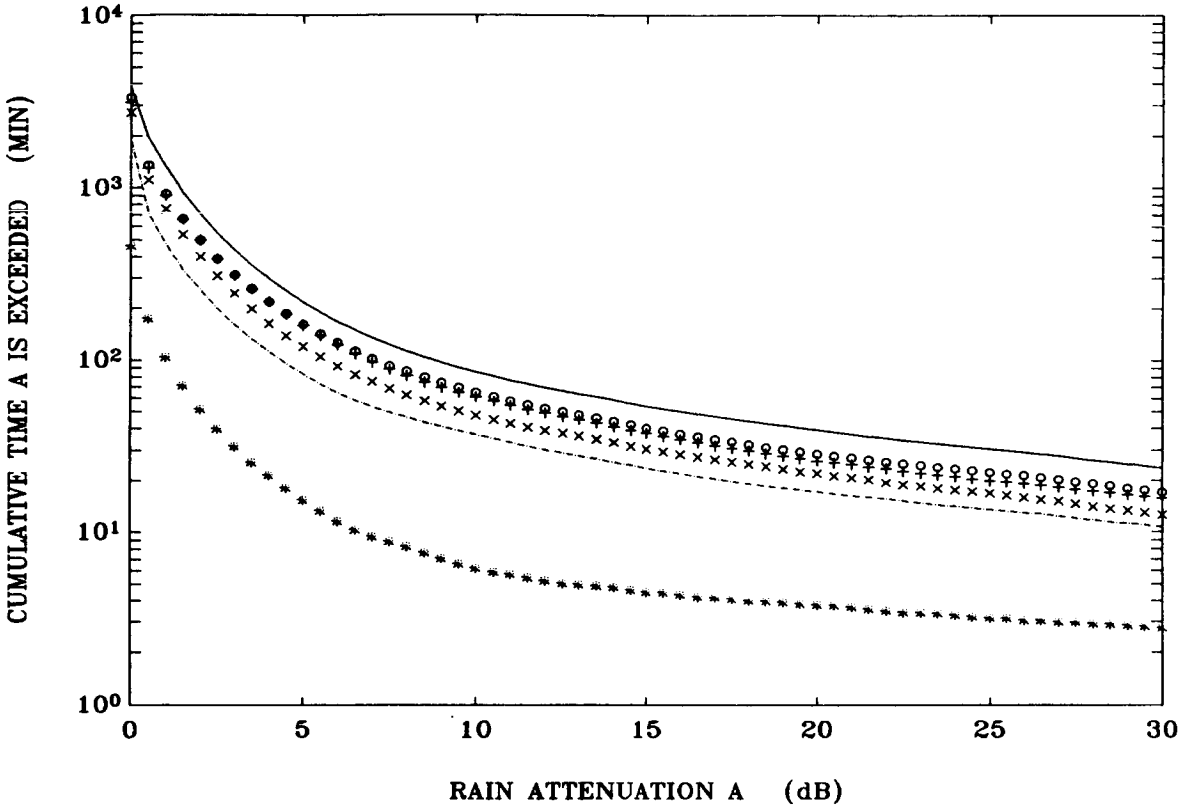


Figure 2: Cumulative time exceeded (min) vs. A (dB) in the fixed system (continuous line) and in the mobile system (zig-zag), p_S =0.0: "o", p_S =0.2: "+", p_S =0.5: "x", p_S =0.8: "-.", p_S =1.0: "*". Vehicles started inside the geographical area, speed Model II (suburban).

of A, even if A is calculated every 1.5 minutes (duration of a map).

Moving terminals in zig-zag routes: To make simulations simpler, but nevertheless significant, we considered a grid of squares of side d=500 m, i.e. the same dimensions of the rain rate cell, with intersections coincident with the centers of the rain rate cells. As mentioned above the fixed terminal is located at these intersections. A moving terminal is driven with a constant speed v along each cell distance d, independently from the previous or the following value, and randomly extracted from a lognormal probability distribution with median value M=30.13 km/h, or M=56.01 km/h, and standard deviation equal to 0.221 for both (the average value of the Gaussian distribution is $\log_e M$). These probability distributions may model, respectively, urban traffic in slow routes (Model I) and suburban traffic in faster routes (Model II) [2].

The vehicle starting point was chosen randomly and uniformly within the useful geographical area (18,205 cells) probability p_S, or turning to the right or to the left with equal probability $p_R=p_L=(1-p_S)/2$; no U-turn was allowed. However square loop routes within the area were also investigated. When one map is over (after 1.5 minutes), space continuity is maintained by driving the vehicle in the same segment but in the new rain rate map. The attenuation measured in a cell by a vehicle lasts a time interval $T_M = d/v$. Since v is random, T_M is also random and thus A is not sampled uniformly. This fact is taken into account in the statistics. Once the vehicle leaves the geographical area it is not further considered (the vehicle enters of course a contiguous area). To obtain reliable statistics we found that 50 independent vehicles must be driven in the simulations (not shown for brevity). In summary, the results below are averaged over 18,205 terminals for the fixed system, and over 50 terminals for the mobile system.

RESULTS

Figure 1 shows the cumulative number of minutes exceeded for the fixed (averaged over 18,205 terminals), $T_{T,F}$, and the mobile systems (averaged over 50 terminals), $T_{T,M}$, with Model I (urban traffic) for v and vehicles started being driven from inside the useful area, for the indicated values of A. Once normalized to the total time (i.e. the time corresponding to A>0 dB, 3800 minutes) these curves yield the conditional probability distributions. To assess the reliability of these results, the curve of the fixed system was compared to that measured with Olympus or predicted at

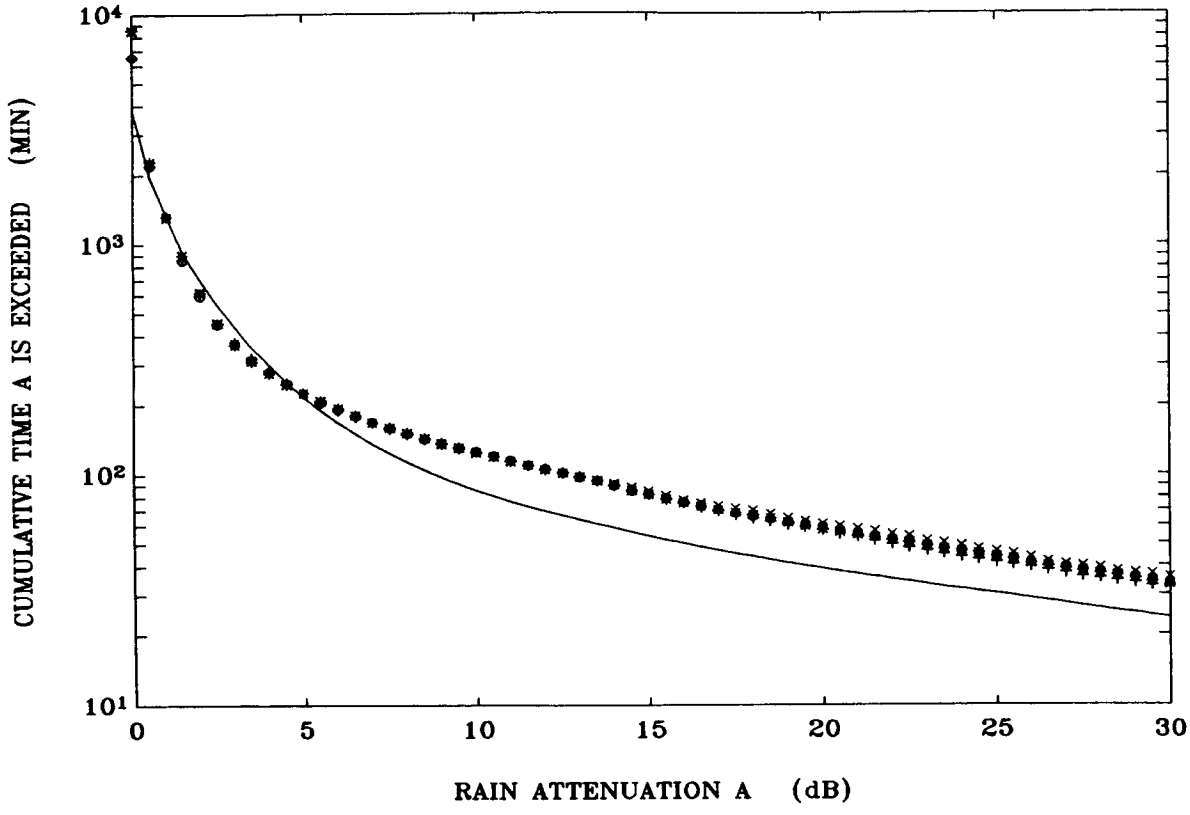


Figure 3: Cumulative time exceeded (min) vs. A (dB) in the fixed system (continuous line) and in the mobile system (square loops), side 8 km: clockwise "o", counterclockwise: "+"; side 4 km: clockwise "x", counterclockwise: "*". Speed Model I (urban).

19.77 GHz from SIRIO long term 11.6 GHz statistics. The agreement (not shown for brevity) is excellent and thus, as a consequence, we not only can compare the present results for the fixed and mobile systems, but we can also use them for reliable long term predictions.

In Figure 1 we notice that: (*a*) a given value of A in the mobile system is exceeded, on the average, for less time than in the fixed system; (*b*) the higher the probability of going straight, p_S, the less is the time exceeded for the vehicle. This behavior is reasonable: when the vehicle is driven in zig-zag routes with shorter straight lines it tends to stay more within the area of observation than that driven in longer straight lines. Similarly, Figure 2 shows the results for Model II, again for vehicles started being driven from inside the useful area. Now the results are even more favorable to the mobile system.

Figure 3 shows the results obtained in square loops centered on the radar site, with sides 4 or 8 km, i.e. perimeters of 16 or 64 km. The curves are again compared to the overall fixed system curve. Now, there is no improvement for the mobile system, but rather its performance may become even worse. If the latter results are compared to the fixed system curve obtained by considering only the fixed terminals of the square loop (i.e. 32 and 128 cells, respectively), the curves almost coincide with those of the corresponding fixed systems.

Figure 4 shows $T_{T,M}$ for Model I and vehicles started being driven from a point, randomly and uniformly selected, at the perimeter of the useful area. For these terminals the performance is much better, even neglecting the obvious more favorable cases of p_S =0.0 and p_S =0.2.
Notice that in the real world, a fraction q_{in} of the total number of vehicles driven in the area would be started *inside* the area (and eventually leave it, as in our simulations, by entering a contiguous area), and a fraction $q_{out}=1-q_{in}$ would be driven into the area from the perimeter. We can consider this overall situation by weighting the results, e.g. those shown in Figures 1 and 4, according to the values of q_{out} and q_{in}, once the latter values have been estimated. Calculations of $T_{T,M}$ for different values of q_{out} and q_{in} are not shown for brevity but, obviously, the value of $T_{T,M}$ obtained is bounded by the upper limits of Figure 1 (q_{in} =1.0) and the lower limits of Figure 4 (q_{in} =0.0), for Model I.
For system design a useful figure of merit is the "gain" of the mobile system [1]:

$$G = A_F - A_M \qquad (1)$$

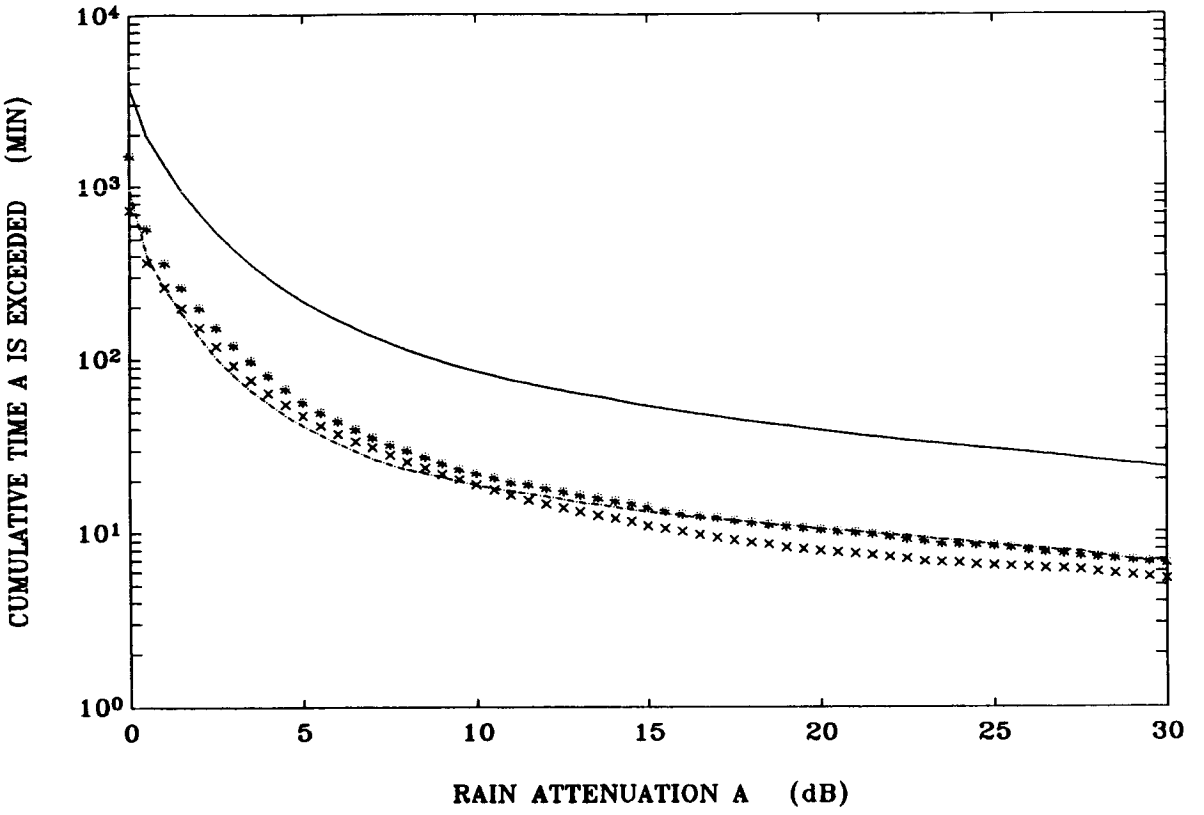


Figure 4: Cumulative time exceeded (min) vs. A (dB) in the fixed system (continuous line) and in the mobile system (zig-zag), p_S =0.5: "x", p_S =0.8: "-.", p_S =1.0: "*". Vehicles started at the perimeter of the geographical area, speed Model I (urban).

i.e., the difference, for a given probability (or cumulative time) exceeded, between the attenuation of the fixed system, A_F, and that of the mobile system, A_M. Figure 5 shows some results of the normalized gain

$$g = \frac{G}{A_F} \qquad (2)$$

(which is more independent of frequency than G), for Model I (urban traffic) and vehicles started being driven from inside the area: we observe that, for a given cumulative time, i.e. $T_{T,M} = T_{T,F}$, g increases as p_S increases, and that g is a significant quantity. A larger gain is surely obtained when the vehicles entering the area (from the perimeter) are included.

CONCLUSIONS

The simulations of vehicles driven in zig-zag routes have shown that: (*a*) on the average, in the mobile system a given value of rain attenuation A (dB) is exceeded for less time than in the fixed system; (*b*) the higher the probability of going straight p_S, the less is the time exceeded for the vehicle. When the vehicles are driven in loops, then there is no significant difference between mobile and fixed systems. Mobile systems performance improves when Model II (suburban traffic) is used or vehicles enter the area from its perimeter. The gain (dB) obtainable can be large and of the same order of magnitude of that provided by FEC codes, or even larger. More results on these routes and on freeways (i.e. straight routes) will be reported in another work.

Acknowledgment: Marco Scotti and Diego Terreni are gratefully thanked for having performed the simulations for their graduation degrees in Electronics Engineering at Politecnico di Milano.

REFERENCES

[1] **E. Matricciani**, *Transformation of Rain Attenuation Statistics from Fixed to Mobile Satellite Communication Systems*, IEEE Trans. on Vehicular Technology, vol.44, 565-569, 1995.

[2] **E. Matricciani and S. Moretti**, *Rain Attenuation Statistics Useful for the Design of Mobile Satellite*

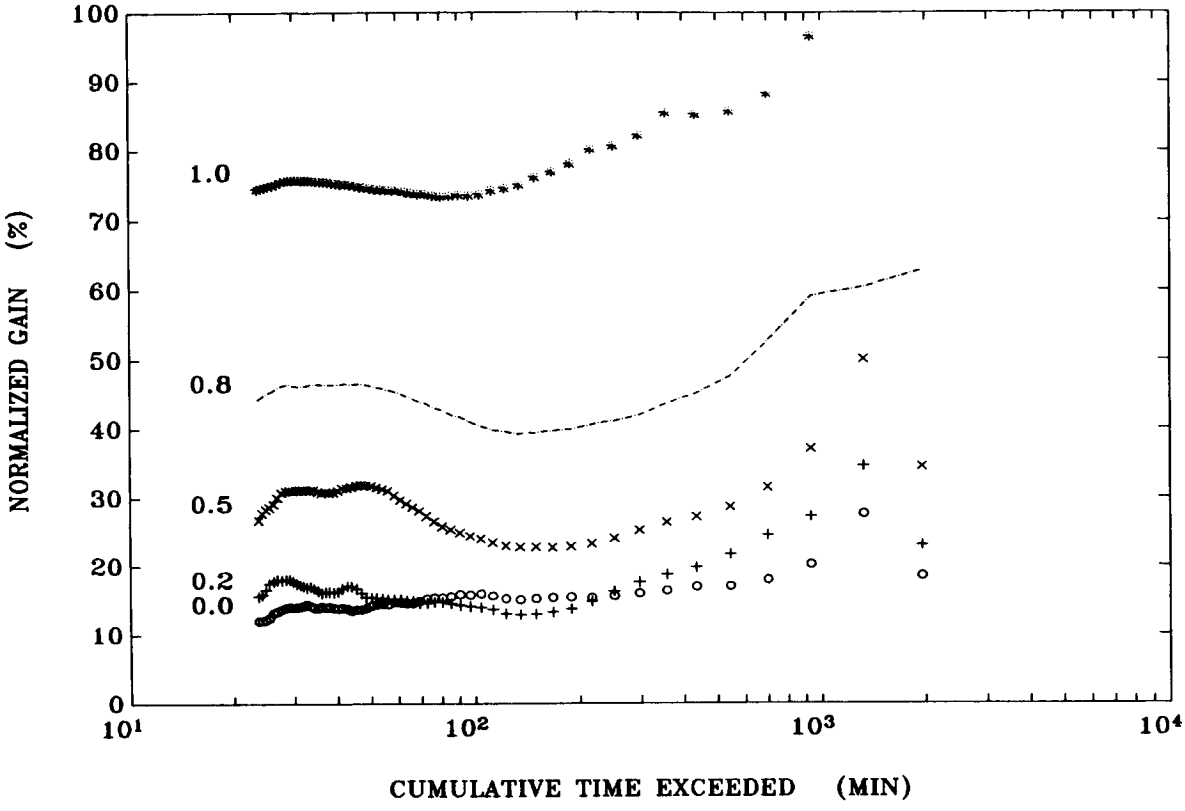


Figure 5: Normalized gain (%) versus cumulative time exceeed (min) for vehicles started inside the geographical area for different values of p_S . Speed Model I (urban).

Communication Systems, under revision for IEEE Trans. on Vehicular Technology.

[3] **E. Matricciani**, *Rain Attenuation Predicted with a Two-Layer Rain Model*, European Transactions on Telecommunications and Related Technologies, vol.2, 715-727, 1991.

[4] **E. Matricciani**, *Physical-mathematical model of the dynamics of rain attenuation based on rain rate time series and a two-layer vertical structure of precipitation*, Radio Science, vol.31, 281-295, 1996.

Propagation Analysis of the ACTS Maritime Satellite Channel

Erik Perrins and Michael Rice
Department of Electrical & Computer Engineering
Brigham Young University
Provo, UT 84602
Phone: 801 378 4469 FAX: 801 378 6586
email: perrinse@ift.ee.byu.edu and mdr@ee.byu.edu

Abstract

The ACTS Mobile Terminal (AMT) has been deployed aboard the USS Princeton (CG 59) in an experiment to test the viability of K and Ka band technology in maritime satellite applications. Pilot tone data recorded on-board the USS Princeton (CG 59) show that multipath interference is negligible and that most variations in the received power level result from antenna tracking errors and possibly obstructions by the ship's superstructure. This paper presents the results of a study of ship maneuvers conducted during October 1996 which provide data illustrating the relationship between pitch/yaw/roll and received power levels. Plots showing received power (both time series and per-cent availability) are presented to aid in link budget calculations.

Introduction

The ACTS mobile terminal (AMT) [1] has been deployed aboard the USS Princeton (CG 59) in an experiment to test the viability of K and Ka band technology in maritime satellite applications. The satellite link consists of a ground terminal at the Jet Propulsion Laboratory in Pasadena, California and the mobile terminal on board the USS Princeton in the eastern Pacific Ocean. The link uses one of the ACTS steerable spot beams.

Part of the AMT is a small slotted dipole antenna which is mechanically steerable in both azimuth and elevation [2]. The antenna controller tracks the satellite location by locking on to a pilot tone transmitted from the ground terminal through ACTS. The received pilot tone levels are tracked by a phase-lock-loop and sampled by an on-board data acquisition system at a rate of 4000 samples/second [3]. In addition, the data acquisition system records the pitch, roll, yaw, and location of the mobile terminal. These data are used to evaluate the effectiveness of ACTS K and Ka band technology to provide reliable two-way ship-to-shore communication via satellite.

This paper focuses on a data set recorded during maneuvers in the eastern Pacific during October 1996. The data are used to determine the amount of time received signal fades occurred. Fades due to multipath were not observed in the data sets. The observed fades were due primarily to antenna tracking errors. It is possible that shadowing by the ship superstructure contributed to the fades, but the data was inconclusive. Due to the narrow beam width of the ship-board antenna, tracking errors can cause significant attenuation.

Analysis

A plot of the received pilot tone power during the maneuvers is shown in Figure 1. In this plot, the received level is normalized to the line-of-sight power level observed prior to the maneuvers. We note six fades of significance, four of them at about the 10 dB level and 2 severe fades in excess of 30 dB. Plots of the ship yaw, roll, and pitch during the maneuvers are shown in Figures 2, 3, and 4, respectively. The ship executed a series of west to south to west turns during the experiment.

For each data point, the elevation angle and ship oriented azimuth (forward is zero degrees, with positive angles measured clockwise) were computed using the method outlined in [4]. This data was used to determine if the line-of-sight path was obstructed by the ship superstructure. As shown in Figure 5, the elevation angle never dropped below 40 degrees. For the ship oriented azimuth angles associated with the maneuvers (see Figure 6), the line-of-sight path is clear for elevation angles as low as 30 degrees. Thus it does not appear that shadowing by the ship superstructure accounts for the observed fades. (Note: there are structures on the ship which can block the line-of-sight signal with an elevation angle as high as 45 degrees, but the AMT-to-ACTS path was never oriented in these directions.)

An interesting correlation exists between the yaw and the received pilot power level as illustrated in Figure 7. It appears that the deepest fades occur just after the ship changes direction after a long turn. This suggests that the antenna tracking may be source of the fades. Since the beam width of the antenna is narrow,

relatively small tracking errors (on the order of 10 degrees) can cause noticeable fades. The apparent correlation between the deep fades and the end of long turns suggests there may be mismatch between possible inertial calculations performed by the antenna controller and that required by hard ship maneuvers to maintain a sufficiently small tracking error.

Results and Concluding Remarks

The cumulative effects of the signal variations during the ship maneuvers are illustrated by the histogram in Figure 8. This plot quantifies the percentage of time that fades exceed the levels on the abscissa. For example, the 2%, 5%, and 10% fade levels are 20 dB, 15 dB, and 8 dB, respectively. Thus, a communication system must include an 8 dB fade margin to realize a 90% availability during hard maneuvers like those performed in this experiment.

References

[1] B. Abbe, M. Agan, and T. Jedrey. ACTS mobile terminals. *International Journal of Satellite Communications*, 14(3):175–190, January 1996.

[2] M. Agan, D. Nakamura, A. Campbell, R. Sternowski, W. Whting, and L. Shameson. ACTS aeronautical experiements. *International Journal of Satellite Communications*, 14(3):233–247, January 1996.

[3] R. Frye. ACTS mobile terminal data acquisition system (DAS). Technical Report AMT-002, Jet Propulsion Laboratory, March 1993.

[4] M. Rice, B. Mott, and K. Wise. A pointing error analysis of the ACTS mobile terminal. In *Proceedings of the International Mobile Satellite Conference*, Pasadena, CA, May 1997.

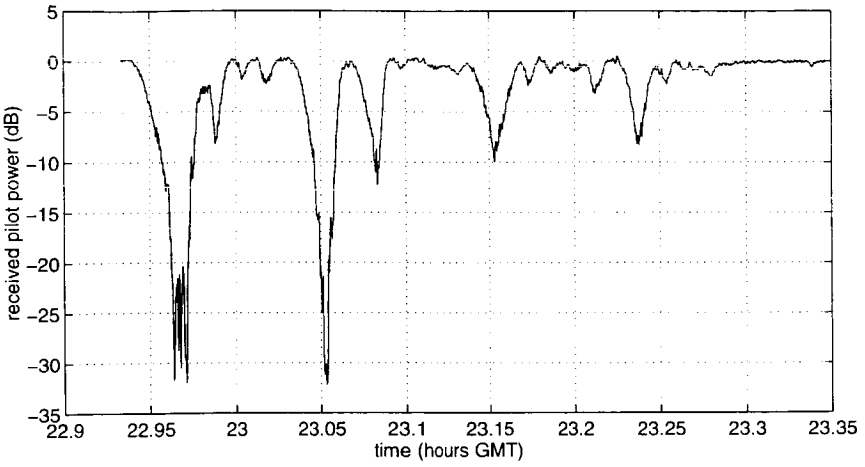


Figure 1: Received Pilot Power During Ship Maneuvers

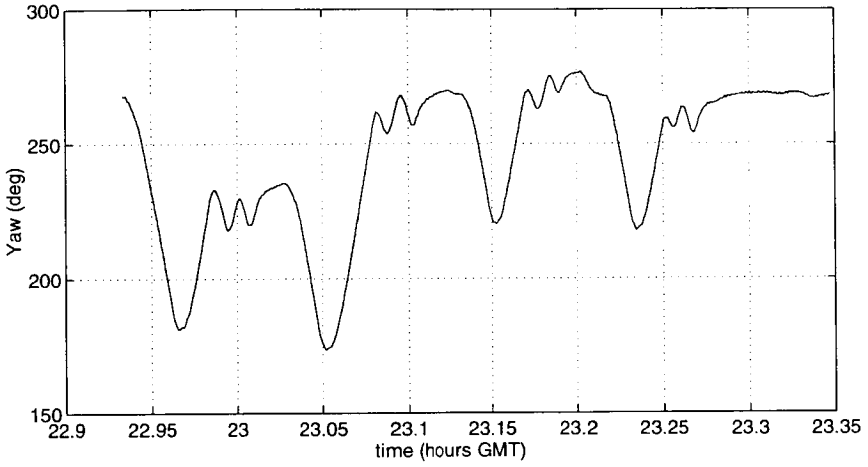


Figure 2: Yaw During Ship Maneuvers

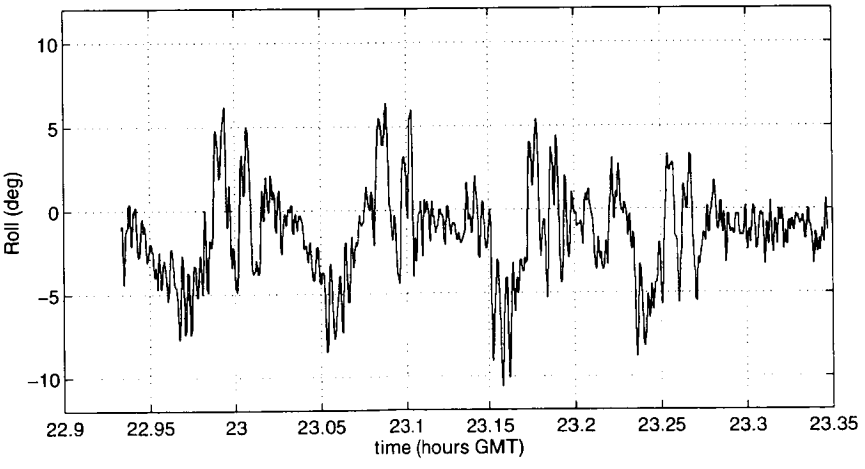


Figure 3: Roll During Ship Maneuvers

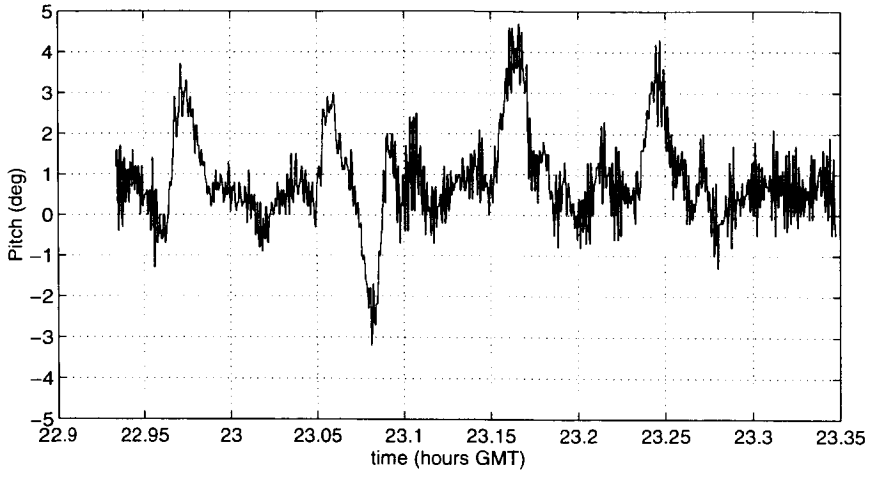


Figure 4: Pitch During Ship Maneuvers

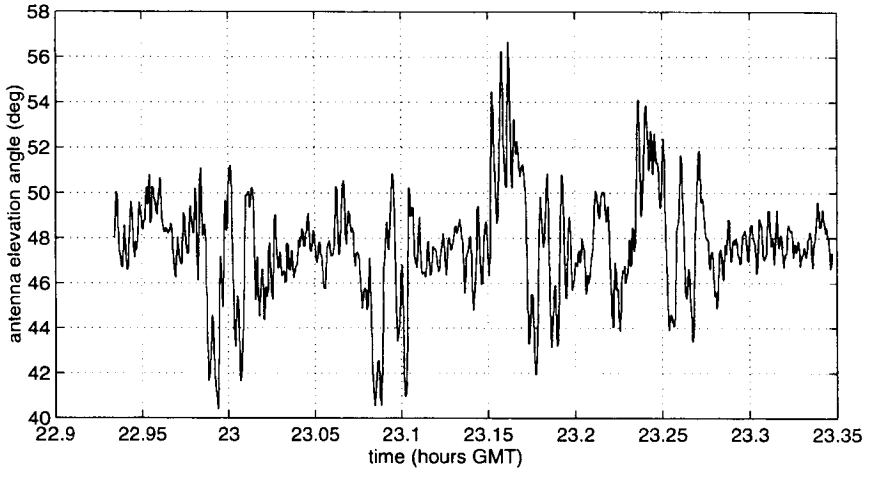


Figure 5: Ship to ACTS Elevation Angle During Ship Maneuvers

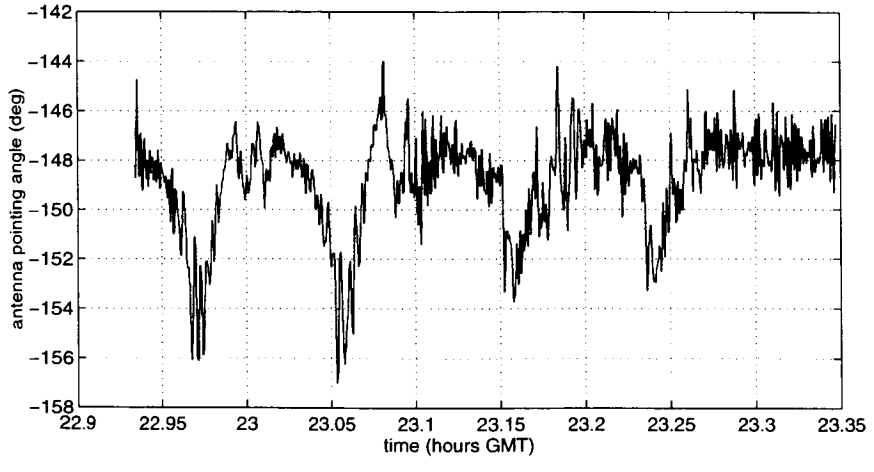


Figure 6: Ship Oriented Azimuth During Ship Maneuvers

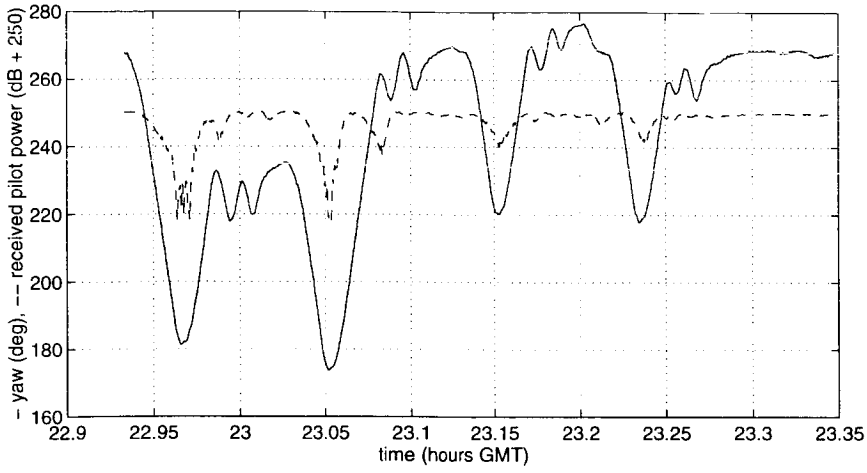


Figure 7: Yaw and Received Pilot Power During Maneuvers

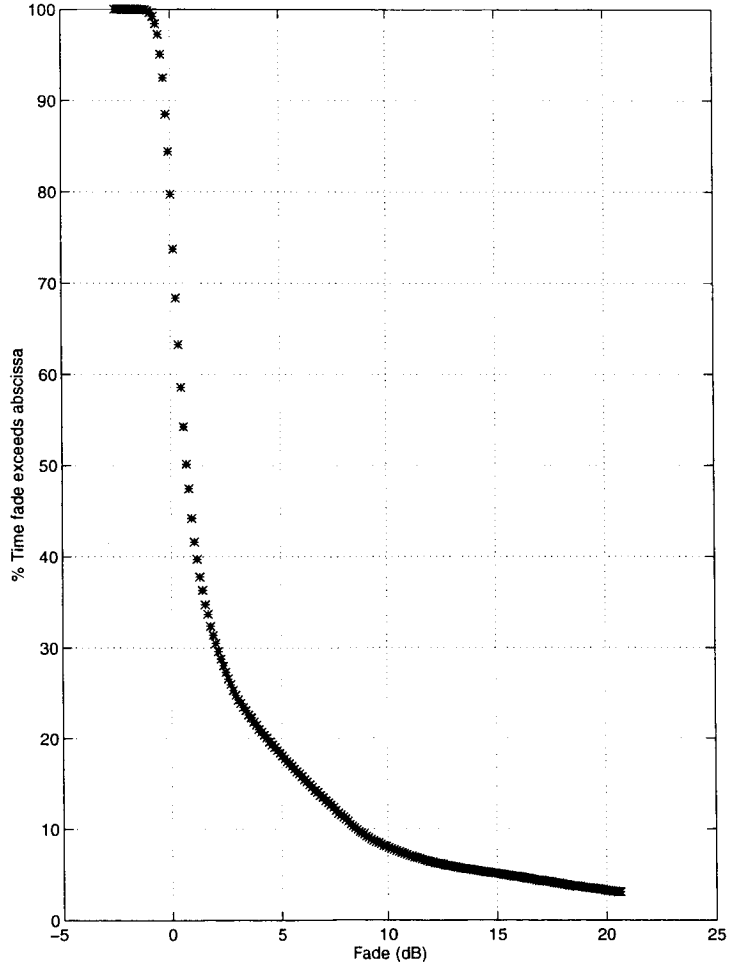


Figure 8: Histogram Showing Fade Durations During Ship Maneuvers

A New Model for the ACTS Land Mobile Satellite Channel

Michael Rice and Brian Humpherys
Department of Electrical & Computer Engineering
Brigham Young University
Provo, UT 84602
Phone: 801 378 4469 FAX: 801 378 6586
email: mdr@ee.byu.edu

Abstract

The ability of the statistical models outlined in NASA publication 1274 to model the effects of shadowing and multipath interference on land mobile satellite channels are relatively good for those models which use a time share between shadowed and unshadowed propagation. The ability of these same models to describe the effect of antenna pointing errors is limited. A new model, the Generalized Total Shadowing Model, is introduced which accounts for the ointing errors resulting from the narrow beam antenna of the ACTS Mobile Terminal (AMT) used in land mobile satellite experiments. A summary of the results show that the new model accurately represents the fade statistics of the experiment runs and that the parameters of the model reveal insight into the mechanisms affecting the received power on the land mobile satellite channel.

Introduction

The ACTS Mobile Terminal (AMT) [1] developed by the Jet Propulsion Laboratory was used in field tests conducted in 1994 to collect propagation data [2]. One of the interesting aspects of the AMT study is the effect of the antenna used these experiments. The AMT antenna, described fully in [3], is a small ($8'' \times 3''$) high-gain reflector with a 3 dB beamwidth of $\pm 9°$ in elevation and $\pm 6°$ in azimuth which tracks the satellite signal in azimuth for a fixed elevation angle. The narrow beamwidth rejects the off-axis reflections of the transmitted wave front arriving at the mobile. The result is a system which, relative to systems using hemispherically omnidirectional antennas typical in land mobile satellite applications, displays significantly lower levels of multipath interference but more severe signal attenuation during shadowed propagation (i.e. when the ine-of-sight propagation path is either partially or com-)letely obscured).

Statistical models, summarized in NASA Publication 1274 [4], have been developed to characterize the fade statistics observed on the land mobile satellite channel at L-band. This paper examines the applicability of three of these statistical models in representing the fade statistics at K-band using the AMT. The shortcommings of these models (as applied to the AMT data) are summarized and a new model is proposed.

Experiment Description

Propagation experiments were conducted in a variety of environments in the Pasadena, CA area and are fully described in [2]. Following Jakes [5], the environment associated with each run was classified as either *open* or *suburban*. Due to the wide range of characteristics observed with the runs in the suburban category, two sub-classifications were defined: suburban I and suburban II. The classifications are summarized as follows:

Open: area where there were very few obstacles such as tall trees or buildings in the path [5]. (In the AMT experiments, this was a limited access multilane freeway.)

Suburban I: a broad suburban thoroughfare lined with trees and buildings. The tree canopies cause intermittent blockage and the buildings are either too far removed from the road side or not tall enough to cause significant blockage.

Suburban II: a small, two-lane roadway lined with trees and buildings. The tree canopies often cover the entire roadway and buildings are close enough to contribute to the fading process.

For all runs, the elevation angle was 46°. Since the Los Angeles, California represents a seasonally invariant suburban environment, the weather conditions were constant: semi-arid, partly cloudy days, with no rain. The transmitted pilot tone (with downlink frequency 19.914 GHz) was received with a composite $C/N_0 = 55.63$ dB [1] and tracked by the AMT antenna tracking subsystem [3]. Pilot tone levels were recorded by the AMT data acquisition system using 4000 samples/second in a bandwidth of 1.5 kHz.

Received Signal Characterization

For pilot tone tests, the received signal consists of an attenuated version of the transmitted pilot tone and a

number of reflected copies of the pilot tone so that the received signal may be expressed as

$$r(t) = D\cos\omega_0 t + N_c(t)\cos\omega_0 t - N_s(t)\sin\omega_0 t \tag{1}$$

where D is the amplitude of the received line-of-sight signal and where $N_c(t)$ and $N_s(t)$ are independent Gaussian random processes each with zero mean and common variance σ_d^2. The received envelope voltage $\mathcal{V}(t)$ is given by

$$\mathcal{V}(t) = \sqrt{[D + N_c(t)]^2 + [N_s(t)]^2} \tag{2}$$

which is described by the Rice probability density function

$$p_{\mathcal{V}}(v; D) = \frac{v}{\sigma_d^2}\exp\left\{-\frac{D^2+v^2}{2\sigma_d^2}\right\} I_0\left(\frac{vD}{\sigma_d^2}\right). \tag{3}$$

The voltage models described in the technical literature attempt to describe the variations in the line-of-sight signal due to partial or complete shadowing observed on real mobile satellite channels. These effects are incorporated in the statistical model by treating parameter D in (3) as a random variable with an assumed density function.

The power in the received signal (1) is given by

$$\mathcal{S}(t) = \frac{\langle \mathcal{V}(t)^2 \rangle}{2} \tag{4}$$

where $\mathcal{V}(t)$ is given by (2). The statistics of $\mathcal{S}(t)$ are described by the non-central chi-square density function

$$p_{\mathcal{S}}(S; D) = \frac{1}{\sigma_d^2}\exp\left\{-\frac{D^2+2S}{2\sigma_d^2}\right\} I_0\left(\frac{D}{\sigma_d^2}\sqrt{2S}\right). \tag{5}$$

Statistical models describing the statistics of power fluctuations on real land mobile satellite channels follow the same approach as those for the envelope voltage.

Loo Model

The Loo model [6] assumes that the line-of-sight signal is never totally blocked, just attenuated and that under foliage attenuation, D (in Equation 3) follows a log-normal distribution[1]

$$p_D(D) = \frac{1}{\sigma D\sqrt{2\pi}}\exp\left\{-\frac{[\ln(D)-\mu]^2}{2\sigma^2}\right\}. \tag{6}$$

so that the density function of the envelope voltage is

$$p_{\mathcal{V}}(v) = \int_0^\infty p_{\mathcal{V}}(v|D)p_D(D)dD \tag{7}$$

While this model represents the statistics of shadowed propagation well, it does accurately model fade statistics with a combination of shadowed and unshadowed propagation as illustrated in Figures 1, 2, and 3. This poor performance of the Loo Model is due to the fact that the AMT system experiences shadowed propagation where the line-of-sight signal is completely obscured.

The Log-normal Shadowing Model

The Log-normal Shadowing Model proposed by Smith and Stutzman [7] is an extension of the Loo model which attempts to improve the modeling of shadowed propagation by associating different statistics with shadowed and unshadowed propagation and using a time share parameter to combine the two. During unshadowed propagation, the density of the received signal is given by (3) while during shadowed propagation, the density of the received signal is given by the Loo model (7). The time share parameter A represents the proportion of time the line-of-sight propagation path is shadowed so that the density function under these assumptions is

$$p_{\mathcal{V}}(v) = (1-A)p_{\mathcal{V}}(v; D_{\text{LOS}}) + A\int_0^\infty p_{\mathcal{V}}(v|D)p_D(D)dD. \tag{8}$$

As illustrated in Figures 1, 2, and 3, the log-normal model is an improvement over the Loo model, but is still limited in its accuracy since this model inherits the assumption from the Loo model that the line-of-sight path is never completely obstructed.

The Total Shadowing Model

The Total Shadowing Model developed by Lutz, et. al. [8] is similar to the Log Normal Model for voltage in that a time share parameter is used to distinguished shadowed and unshadowed propagation. The key differences are in the conditions assumed for shadowing: the line-of-sight component is totally blocked (i.e. $D = 0$) and the average power σ_d^2 is a random variable with a log-normal density function[2]. Rewriting (5) with $D = 0$ to be conditional on $S_0 = \sigma_d^2$, the density function for the Total Shadowing Model is

$$p_{\mathcal{S}}(S) = (1-A)\frac{1}{\sigma_d^2}\exp\left\{-\frac{D^2+2S}{2\sigma_d^2}\right\} I_0\left(\frac{D}{\sigma_d^2}\sqrt{2S}\right) + A\int_0^\infty \frac{1}{S_0}\exp\left\{-\frac{S}{S_0}\right\} \times \frac{10}{\ln(10)\sigma S_0\sqrt{2\pi}} \times \exp\left\{-\frac{[10\log(S_0)-\mu]^2}{2\sigma^2}\right\} dS_0. \tag{9}$$

1. A log-normal random variable z results from performing the transformation $z = e^x$ where x is a normal random variable with mean μ and variance σ^2.

2. The log-normal density used in (9) results from the transformation $z = 10^{x/10}$.

Figures 5, 7, and 9 illustrate the accuracy of (9) in predicted the fade statistics for the same three runs considered previously. This model works better for the AMT data since it accounts for the case where the line-of-sight signal is blocked. These figures show a mismatch between the model and the data in the transition region between unshadowed and shadowed propagation. The power density functions, shown in Figures 4, 6, and 8 illustrate the inaccuracies nicely. The mismatch is due to pointing errors which were shown in [9] to contribute up to 2 dB fluctuations in the received pilot power in the absence of any obstructions.

The Generalized Total Shadowing Model

The Generalized Total Shadowing Model (introduced here) incorporates the pointing errors into the model by extending the Total Shadowing Model to include a random variation of the parameter D. The power density functions shown in Figures 4, 6, and 8 show clearly that the received pilot power is concentrated at two distinct levels under the condition of unshadowed propagation. This characteristic suggests that the density function of D is bimodal, the simplest of which is

$$p_D(D) = (1-B)\delta(D-D_1) + B\delta(D-D_2) \tag{10}$$

so that the density function for the received pilot power is

$$p_S(S) = \tag{11}$$

$$(1-A)\int_0^\infty p_S(S|D)p_D(D)dD$$

$$+A\int_0^\infty p_S(S|S_0;D=0)p_{S_0}(S_0)dS_0 \tag{12}$$

$$= (1-A)(1-B)\frac{1}{\sigma_d^2}\exp\left\{-\frac{D_1^2+2S}{2\sigma_d^2}\right\}I_0\left(\frac{D}{\sigma_d^2}\sqrt{2S}\right)$$

$$+(1-A)B\frac{1}{\sigma_d^2}\exp\left\{-\frac{D_2^2+2S}{2\sigma_d^2}\right\}I_0\left(\frac{D}{\sigma_d^2}\sqrt{2S}\right)$$

$$+A\int_0^\infty \frac{1}{S_0}\exp\left\{-\frac{S}{S_0}\right\}\times\frac{10}{\ln(10)\sigma S_0\sqrt{2\pi}}\times$$

$$\exp\left\{-\frac{[10\log(S_0)-\mu]^2}{2\sigma^2}\right\}. \tag{13}$$

The ability of (13) to fit the fade statistics is illustrated in Figures 4 through 9 which shows an improvement over the Total Shadowing Model (9).

Results and Concluding Remarks

Of the statistical models developed in the open literature, the Total Shadowing Model provides the most accurate description of the fading statistics for the K-band land mobile satellite channel using the ACTS Mobile Terminal. The accuracy of the Total Shadowing Model was improved by incorporating an allowance for pointing error in the equations to produce the Generalized Total Shadowing Model. Note that the Generalized Total Shadowing Model applies to systems with narrowbeam antennas which incur pointing errors. Without significant pointing error, the Total Shadowing Model will most likely be the most accurate model for land mobile satellite applications.

The narrowbeam antenna influenced the system in three ways:

1. The rejection of off-axis reflections mitigates to some degree the multipath interference. The degree of multipath interference is usually parameterized by the Rice parameter κ which is the power ratio of line-of-sight signal strength to diffuse signal strength:

$$\kappa = \frac{D^2}{2\sigma_d^2}. \tag{14}$$

The line-of-sight power fluctuations induced by antenna pointing errors produce different values of this parameter. These values are tabulated in Table in three columns:

$$\kappa_1 = \frac{D_1^2}{2\sigma_d^2} \tag{15}$$

$$\kappa_2 = \frac{D_2^2}{2\sigma_d^2} \tag{16}$$

$$\kappa_{\text{avg}} = (1-B)\kappa_1 + B\kappa_2 \tag{17}$$

The values of κ_{avg} range from 13 to 30 dB with 15 of the 18 greater than 20 dB. These high values of κ_{avg} show the high degree of multipath rejection realized by the antenna gain pattern.

2. The rejection of off-axis reflections results in severe signal outages when the line-of-sight component is either partially or completely shadowed. This is illustrated by the high values of the time share parameter A (which is the proportion of time shadowed propagation occurs) in the third column of Table . For the lone run in the open category, propagation is shadowed 0% of the time, as expected. For the runs in the suburban I category, A ranges from 2% to 57% with half of the runs in the 13–18% range. The suburban II runs display values of A ranging from 17% to 66% where three of the four are greater than 40%.

3. Pointing Errors. K-band frequencies allow the practical use of high-gain, narrowbeam antennas

in mobile applications. Such antenna systems invariably exhibit pointing errors and, consequently, pointing error losses.

The AMT antenna tracking system uses a dithering process to realized closed loop tracking which produces losses less than 0.3 dB [3]. Additional pointing errors result from vehicle pitch and roll [9]. The resulting power variations are tabulated in the seventh column of Table . The power variations ΔP less than 0.3 result from the dithering process only, while the others are a result of the pitch and roll already described.

REFERENCES

[1] B. Abbe, M. Agan, and T. Jedrey, "ACTS mobile terminals", *International Journal of Satellite Communications*, vol. 14, no. 3, pp. 175–190, January 1996.

[2] M. Rice, D. Pinck, J. Slack, and B. Humpherys, "K-band land-mobile satellite channel characterization using ACTS", *International Journal of Satellite Communications*, vol. 14, pp. 283–296, January 1996.

[3] A. Densmore and V. Jamnejad, "A satellite-tracking K- and K_a-band mobile vehicle antenna system", *IEEE Transactions on Vehicular Technology*, vol. VT-42, pp. 502–513, November 1993.

[4] J. Goldhirsh and W. Vogel, "Propagation effects for land mobile satellite systems: Overview of experimental and modeling results", Tech. Rep., NASA Reference Publication 1274, February 1992.

[5] W. C. Jakes, *Microwave Mobile Communications*, IEEE Press, New York, 1974.

[6] C. Loo, "A statistical model for a land mobile satellite link", *IEEE Transactions on Vehicular Technology*, vol. VT-34, pp. 122–127, August 1985.

[7] W. Smith and W. Stutzman, "Statistical modeling for land mobile satellite communications", Tech. Rep. Virginia Tech Report EE Satcom 86-3, Virginia Tech, Blacksburg VA, August 1986.

[8] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke, "The land mobile satellite communication channel — recording, statistics, and channel model", *IEEE Transactions on Vehicular Technology*, vol. COM-40, pp. 375–386, May 1991.

[9] M. Rice, B. Mott, and K. Wise, "A pointing error analysis of the ACTS mobile terminal", in *Proceedings of the International Mobile Satellite Conference*, Pasadena, CA, May 1997.

Table 1: Best Fit Parameters for Generalized Total Shadowing Model and AMT data

RUN	CAT.	A (%)	κ_1 (dB)	κ_2 (dB)	κ_{avg} (dB)	ΔP (dB)
020201	open	0.0	29.71	29.95	29.88	0.24
070901	sub I	4.3	29.36	29.59	29.52	0.23
070903	sub II	46.6	24.97	25.91	25.73	0.94
070905	sub I	57.2	23.55	21.77	23.28	1.78
070906	sub II	17.3	23.66	24.65	24.55	0.99
070907	sub I	15.0	22.65	24.11	23.96	1.46
070912	sub II	43.9	21.22	21.31	21.24	0.09
070914	sub II	66.1	16.95	13.08	15.54	3.87
071016	sub I	8.8	24.21	25.22	24.78	1.01
071017	sub I	8.5	23.84	24.94	24.39	1.10
072405	sub I	16.6	12.57	16.49	13.13	3.92
072406	sub I	15.6	24.72	23.89	24.47	0.84
072407	sub I	32.9	24.07	23.13	23.94	0.94
072408	sub I	13.6	26.79	26.74	26.77	0.04
072409	sub I	13.1	24.37	24.33	24.37	0.04
072410	sub I	44.4	17.46	19.72	19.39	2.26
072411	sub I	2.2	22.10	23.09	22.79	0.99
072412	sub I	18.4	18.89	20.80	20.60	1.92

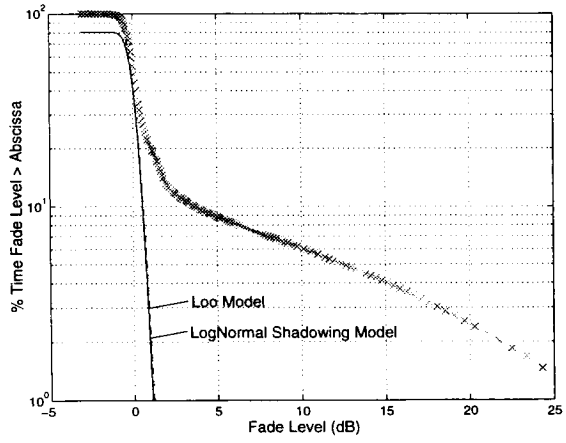


Figure 1: Cumulative Fade Depth Distribution: Histogram ('x') and Least Squares Curve Fit for the Loo Model ('–') and the Log Normal Shadowing Shadowing Model ('-') for Run 070907.

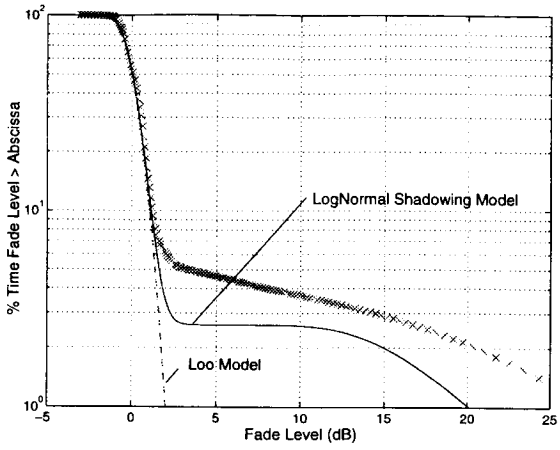


Figure 2: Cumulative Fade Depth Distribution: Histogram ('x') and Least Squares Curve Fit for the Loo Model ('–') and the Log Normal Shadowing Shadowing Model ('-') for Run 071016.

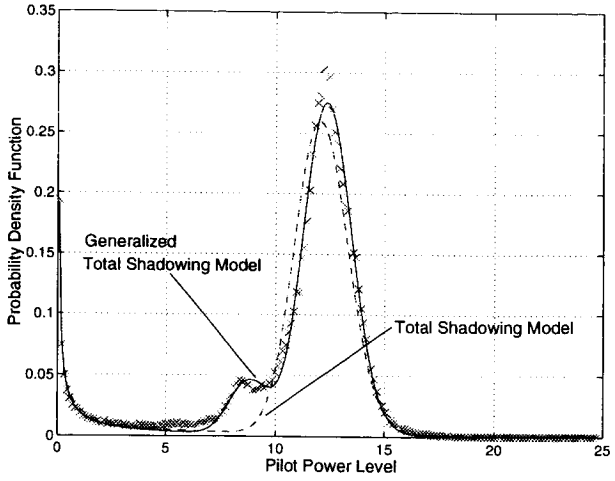


Figure 4: Probability Density Function of Received Pilot Power: Histogram ('x') and Least Squares Curve Fit for the Total Shadowing Model ('–') and the Generalized Total Shadowing Model ('-') for Run 070907.

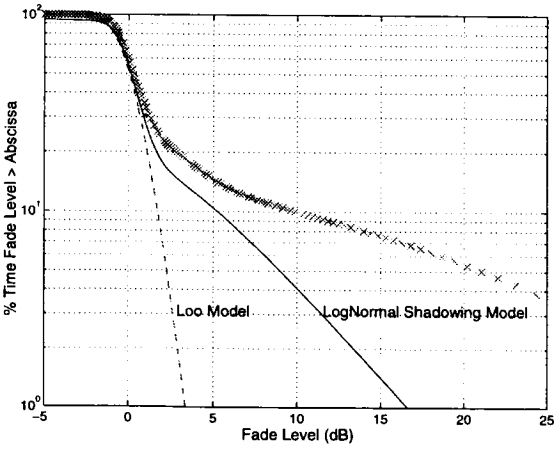


Figure 3: Cumulative Fade Depth Distribution: Histogram ('x') and Least Squares Curve Fit for the Loo Model ('–') and the Log Normal Shadowing Shadowing Model ('-') for Run 072405.

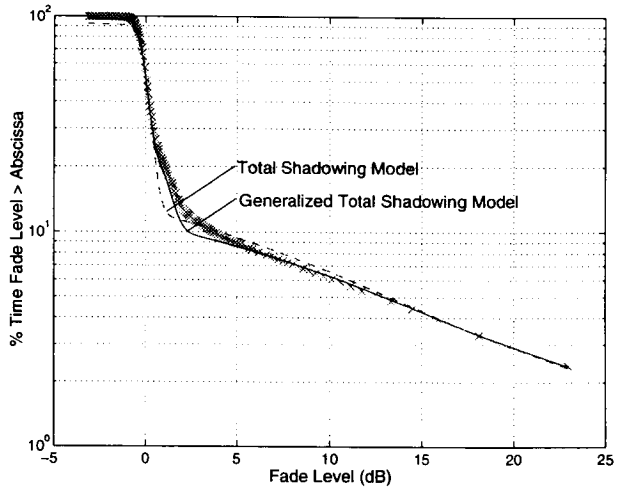


Figure 5: Cumulative Fade Depth Distribution: Histogram ('x') and Least Squares Curve Fit for the Total Shadowing Model ('–') and the Generalized Total Shadowing Model ('-') for Run 070907.

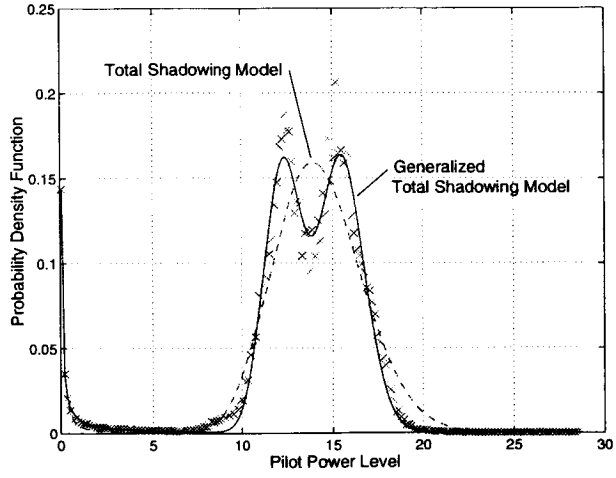


Figure 6: Probability Density Function of Received Pilot Power: Histogram ('x') and Least Squares Curve Fit for the Total Shadowing Model ('–') and the Generalized Total Shadowing Model ('-') for Run 071016.

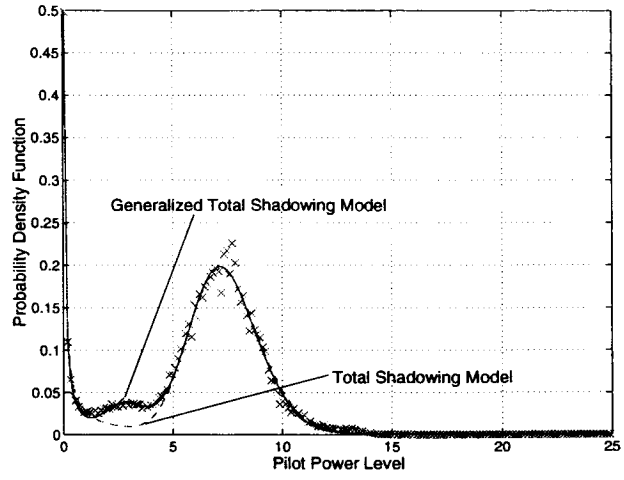


Figure 8: Probability Density Function of Received Pilot Power: Histogram ('x') and Least Squares Curve Fit for the Total Shadowing Model ('–') and the Generalized Total Shadowing Model ('-') for Run 072405.

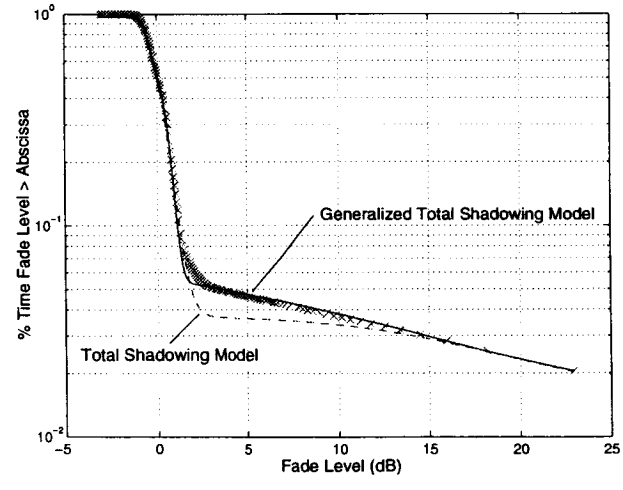


Figure 7: Cumulative Fade Depth Distribution: Histogram ('x') and Least Squares Curve Fit for the Total Shadowing Model ('–') and the Generalized Total Shadowing Model ('-') for Run 071016.

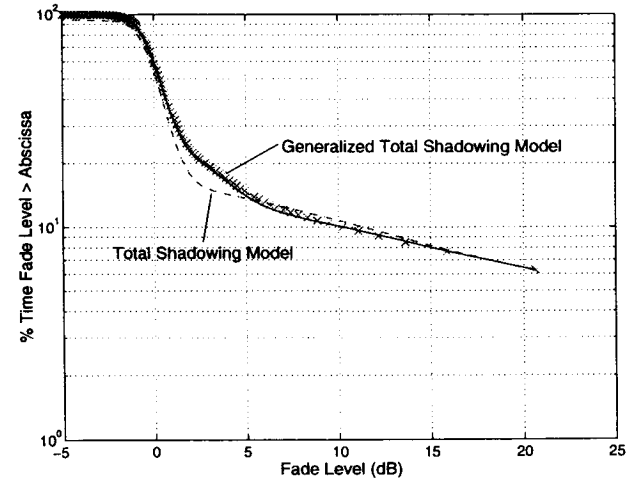


Figure 9: Cumulative Fade Depth Distribution: Histogram ('x') and Least Squares Curve Fit for the Total Shadowing Model ('–') and the Generalized Total Shadowing Model ('-') for Run 072405.

An Overview of the Revised Mobile Satellite Handbook: "Propagation Effects for Land Mobile Satellite Systems: Overview of Experimental and Modeling Results"

Julius Goldhirsh
The Johns Hopkins University, Applied Physics Laboratory
Johns Hopkins Road, Laurel MD 20723-6099, USA
Phone: 301-953-5042 FAX: 301-953-5458
email: julius_goldhirsh@jhuapl.edu

Wolfhard J. Vogel
The University of Texas at Austin, Electrical Engineering Research Laboratory
10100 Burnet Road, Austin, Texas 78758, USA
Phone: 512-471-8608 FAX: 512-471-8609
email: wolf_vogel@mail.utexas.edu

ABSTRACT

A review of the planned revised mobile satellite propagation handbook is presented. The revised manual will have a broadened scope such that it contains propagation information related to personal-mobile, air-mobile, and marine-mobile scenarios; subject areas; previously not covered in the former document. It will also review new experiments, revised and new empirical and theoretical models developed since the previous publication. It is expected that many of the chapters will have been made accessible on the World Wide Web (http://propagation.jpl.nasa.gov) by the time of the presentation of this paper. A formal document is expected to be published during the fall of 1997. As a sample of the contents of the revised handbook, the Extended Empirical Roadside Shadowing Model is described in detail.

INTRODUCTION

The previous version of the handbook was published in 1992 [1] and is presently available for downloading at the web address cited in the Abstract. This publication was an outgrowth of a series of joint mobile propagation experiments performed by the authors between 1983 and 1988. It contains a description of experiments, results, concepts, and models associated with propagation effects for land mobile satellite scenarios where investigations were described from within and outside the United States. Since its publication, a number of addition mobile satellite propagation experiments have been performed throughout the world. In addition, new areas of investigation such as personal access and mobile-aeronautical communications have approached a level of maturity and importance requiring descriptions of their propagation effects.

CONTENTS OF REVISED HANDBOOK

The following is a listing of the chapters in the revised handbook and a brief description of the contents.

Chapter 1: Introduction

(Rationale for text, background, and objectives.)

Chapter 2: Attenuation Due to Trees (Static Case)

(UHF, L band, and K Band static measurements, attenuation and attenuation coefficients, frequency scaling to K Band, seasonal effects.)

Chapter 3: Attenuation Due Trees (Mobile to Roadside Case)

(Background, time-series fade measurements from UHF to K-Band, Extended Empirical Roadside Shadowing Model, high elevation angle empirical model, validation of models, frequency scaling, seasonal effects on attenuation.)

Chapter 4: Signal Degradation for Line-of-Sight Communications

(Background, multipath for mountain and flat surface scenarios, multipath due to roadside trees, low

elevation angle measurements, dependence on antenna gain.)

Chapter 5: Fade and Non-Fade Durations and Phase Spreads

(Background, experimental aspects, cumulative distributions of fade durations, non-fade durations, and phase fluctuations.)

Chapter 6: Propagation Effects Due to Cross Polarization, Antenna Gain, and Space Diversity

(Background, frequency reuse, distributions from low and high gain antennas, space diversity simulation from measurements.)

Chapter 7: Investigations from Different Countries

(Measurement campaign results from Australia, Belgium, Canada, England, France, Germany, Japan, and the United States.)

Chapter 8: Personal Communication Propagation Effects

(Attenuation and multipath effects due to trees and building walls for personal communication scenarios, overview of experiments and results.)

Chapter 9: Aeronautical and Marine Propagation Effects

(Multipath effects caused by airplane structures and other ambient obstacles, multipath effects caused by ship structure and ocean surface, antenna diversity concepts.)

Chapter 10: Optical Methods for Assessing Shadowing and Blockage

(Background, methodology for hemispherical shadowing and blockage determination to extract fade distributions, efficacy of satellite diversity.)

Chapter 11: Theoretical Modeling Considerations

(Fundamental concepts, Loo's distribution models, Markov simulation, lognormal shadowing models, diffraction over buildings, multipath and attenuation models for personal communication.)

Chapter 12: Recommendations for Further Investigations

(Summary and review of recommendations contained in the text, gaps in experiments and theory, recommendations for future work.)

EXTENDED EMPIRICAL ROADSIDE SHADOWING MODEL

Background

As an example the information contained in the revised handbook, we review here the Extended Empirical Roadside Shadowing Model (EERS) which has been described previously in the literature [2], and has also been adopted (in 1997) by the ITU-R [3]. The Empirical Roadside Shadowing (ERS) model enabled the estimations of cumulative distributions of fades over the frequency range 870 MHz through 2.6 GHz, elevation angles from 20° to 60°, and probabilities ranging from 1% up to 20% [1, 4]. The Extended Empirical Roadside Shadowing Model (EERS) extends the ERS model as follows: (1) considers elevation angles as low as 7°, (2) considers frequencies up to 20 GHz; and (3) employs a percentage range of 1% to 80%. While this model has been validated at 870 MHz, 1.5 GHz, 3 GHz, and 20 GHz, it should be used with caution at other frequencies between 3 to 20 GHz until additional data become available for validation.

The ERS model was originally derived from the median of cumulative UHF and L-band fade distributions systematically obtained from helicopter-mobile and satellite-mobile measurements in the USA. The measurements were made over approximately 600 km of driving distance at path elevation angles of 21°, 30°, 45°, and 60°. The 21° case used the geostationary satellite MARECS-B2 [5]. Measurements for the other angles were obtained with a helicopter as the transmitter platform [6], where the configuration corresponded to a shadowing condition in which the helicopter flew parallel to the moving vehicle and the propagation path was approximately normal to the line of roadside trees (*e.g.*, azimuth of the path relative to the vehicle direction was 90°). Tree heights ranged from approximately 5 to 30 m. The simulated satellite paths were also predominantly along a 90° shadowing orientation, although deviations from this aspect did arise. The measurements were performed on two-lane highways (one lane each direction), and a four-lane highway (two lanes each direction), where the roadside trees were primarily of the deciduous variety.

In order to assess the extent by which trees populate the side of the road, a quantity called percentage of optical shadowing (POS) was defined, representing the percentage of optical shadowing caused by roadside trees at a path angle of 45° for driving on the right side of the road, where the path is to the right of the driver and the vehicle is in the right lane. The POS values for

the roads driven were predominantly between 55% and 75%, implying tree populations of at least these amounts.

In deriving the EERS model, use was made of the aforementioned body of data at UHF and L-band, as well as more recent data bases corresponding to mobile L-Band measurements of transmissions from MARECS B-2 in the western United States [7], static 20-GHz measurements [8], and mobile 20-GHz measurements using transmissions from the Advanced Communications Technology Satellite (ACTS) [2]. The latter measurements were performed within the first six months of 1994 during which a series of four 20 GHz mobile ACTS campaigns were performed in central Maryland (March, elevation 39°), Austin, Texas (February and May, elevation 55°) and Fairbanks, Alaska, and environs (June, elevation 8°).

EERS Formulation

In the ERS model, the (L Band) fading at $f_L = 1.5\,GHz$ for percentages, P, of distance traveled $20\% \geq P \geq 1\%$, and path elevation angles $60° \geq \theta \geq 20°$ is given by

$$A(P,\theta,f_L) = -M(\theta)\ln(P) + N(\theta) \quad (1)$$

$$M(\theta) = 3.44 + 0.0975\theta - 0.002\theta^2 \quad (2)$$

$$N(\theta) = -0.443\theta + 34.76 \quad (3)$$

In equation (1), $A(P,\theta,f_L)$ is the fading (dB) caused by foliage shadowing, previously validated for elevation angles between 20° and 60° and percentages of distance traveled between 1% and 20%. Substituting equations (2) and (3) into equation (1), $A(P,\theta,f_L)$ may be alternately be written as

$$A(P,\theta,f_L) = a(P) + b(P)\theta + c(P)\theta^2 \quad (4)$$

where $a(P), b(P), c(P)$ are tabulated in Table I for a series of fixed percentages of distance traveled in the interval between 1% and 80% as the first step in the extension of the model.

For $20\% \geq P \geq 1\%$ and $20° \leq \theta \leq 60°$, the relation between the fade distribution at L-band and another frequency f is found to be

$$A(P,\theta,f_L) = A(P,\theta,f_L)\exp\left\{1.5\left[\frac{1}{\sqrt{f_L}} - \frac{1}{\sqrt{f}}\right]\right\} \quad (5)$$

where $A(P,\theta,f_L)$ is the L-band fading calculated with equation (1) or (4).

For the same elevation angle range and percentages of $80\% \geq P \geq 20\%$

$$A(P,\theta,f) = A(20\%,\theta,f)\frac{1}{\ln 4}\ln\left(\frac{80}{P}\right) \quad (6)$$

where $A(P,\theta,f)$ is the fading (dB) at frequency f (GHz) exceeded for P (%) for an Earth-satellite path angle θ, and $A(P,\theta,f_L)$ is the corresponding attenuation (dB) at f_L = 1.5 GHz. The attenuation is defined relative to non-shadowed conditions with negligible multipath.

For the case in which $20° > \theta \geq 7°$, the distribution based on equation (5) is first calculated at $\theta = 20°$. The distribution for $\theta = 20°$ is subsequently assumed to be invariant at elevation angles down to 7°. That is, for $80\% \geq P \geq 1\%$ and $20° \geq \theta \geq 7°$

$$A(P,\theta,f) = A(P,20^o,f) \quad (4)$$

which implies that the probability distributions at elevation angles $20° > \theta \geq 7°$ are identical to those at 20°. Extension of the model to elevation angles smaller

Table 1: Listing of parameter values of $a(P), b(P), c(P)$ in Equation (4).

%, P	$a(P)$	$b(P)$	$c(P)$
1	34.7600	-0.4430	0.0
2	32.3756	-0.5106	1.3863 E-3
5	29.2235	-0.5999	3.2189 E-3
10	26.8391	-0.6675	4.6052 E-3
20	24.4547	-0.7351	5.9915 E-3
30	17.3022	-0.5201	4.2391 E-3
40	12.2273	-0.36754	2.9957 E-3
50	8.2910	-0.2492	2.0313 E-3
60	5.0748	-0.1525	1.2433 E-3
70	2.3556	-7.0805 E-2	5.7711 E-4
80	0.0	0.0	0.0

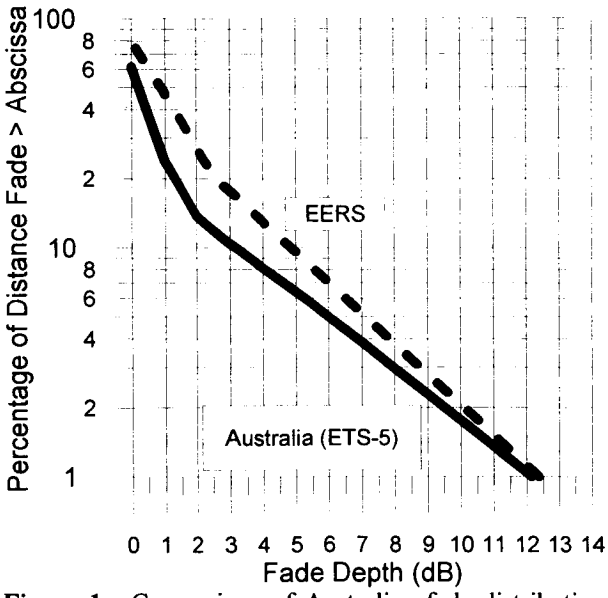


Figure 1: Comparison of Australian fade distribution with EERS model at 51° and frequency of 1.55 GHz.

than 20° is a complex task for the following reasons: (1) The EERS model tacitly assumes that the canopies of single trees shadow the Earth-satellite path. At lower angles, there may be a likelihood that the path cuts the canopies of multiple trees or multiple tree trunks. (2)†At low angles, there may also be a greater likelihood that the terrain itself may block the Earth-satellite path, creating high losses. (3) Ground multipath may also influence the distribution. Based upon empirical experience for cases where the above caveats do not arise, it has been found that, with good approximation, the EERS model at 20° elevation gives similar results to those observed at 7° or 8°. The rationale for this observation is the following: At 20° path elevation, the Earth-satellite path is already passing through the lower part of the tree canopies. Reducing the path elevation angle is likely to result in fading caused by tree trunks, which may tend to mitigate the signal degradation. On the other hand, attenuation effects may increase because of shadowing from tree canopies that are further offset from the road (as was the case in Alaska). The combination of these two effects generally results in median fade statistics that are relatively invariant at angles below 20°, although larger deviations relative to the median are expected because of occasional breakdown of the aforementioned underlying assumptions.

Example Validation of EERS Model

Australian Fade Distributions at L Band: A comparison of the EERS model with measurements made in Australia at 1.55 GHz [4] is shown in Figure 1. The cumulative distribution reflects measurements made over a distance of 403 km comprising 15 individual runs in the Sydney region using the Japanese ETS-V geostationary satellite. The elevation angle to the satellite was 51°. Two major vegetation zones were traversed: forests along the coastal roads and woodlands further inland. Forests ranged from dry sclerophyll, in which the crowns of contiguous trees did not touch each other, to tropical rain forests, in which the leafy crowns of the trees intermingled. The dominate tree genus in the forest was eucalyptus. Other than tree types, general similarities existed between the roads traveled in Australia and those in central Maryland, USA (*e.g.*, tree heights, POS, setbacks). Also shown is the EERS model for 51°. The maximum difference between the two distributions, which occurs at 14%, is less than 2 dB. The difference reduces monotonically above and below this percentage of distance traveled.

Example at K Band: In Figure 2 is shown a 20-GHz fade distribution (elevation angle 55°) for an approximate 10 km run along an evergreen (loblolly pine) tree-lined road in Bastrop, Texas. This distribution was obtained from measurements during the ACTS campaign [2]. The population of trees was in excess of 55%, and there were considerable segments of road where the trees formed a tunnel of foliage overhead. Also shown is the EERS model (dashed curve). The EERS model underestimates the fade by at most 5 dB for probabilities between 1% and

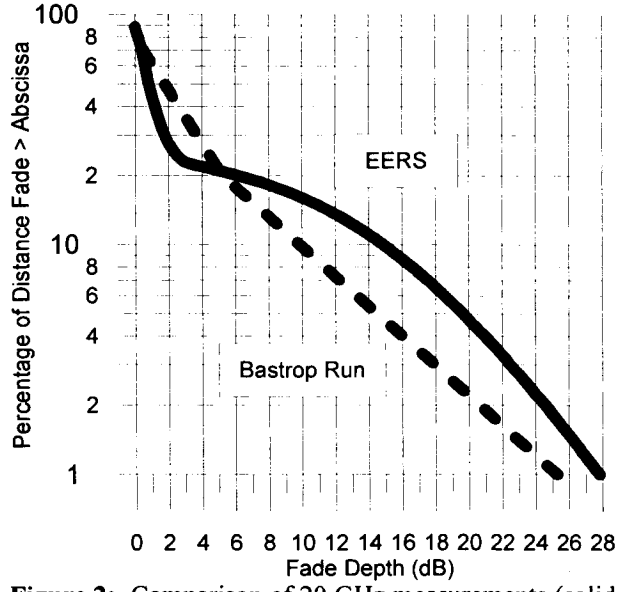


Figure 2: Comparison of 20 GHz measurements (solid line) made in Bastrop, Texas at elevation angle of 54.5° with EERS model.

Table 2: Fades exceeded at elevations of 60° and 80°.

Percentage, P	EERS Fade (dB) at Elevation = 60°		Fade (dB) at Elevation = 80°	
	f = 1.6 GHz	f = 2.6 GHz	f = 1.6 GHz	f = 2.6 GHz
1	8.2	11.0	4.1	9.0
5	4.8	6.5	2.0	5.2
10	3.4	4.5	1.5	3.8
15	2.6	3.4	1.4	3.2
20	1.9	2.6*	1.3	2.8*
30	1.4	1.8*	1.2	2.5*

20%. This deviation is within the variability expected in comparing the EERS model with measured distributions. The underestimation of the model at the smaller probabilities is most likely caused by the prevalence of foliage tunnels (extreme shadowing) giving a greater likelihood of fading at the higher elevation angles.

Step by Step Approach

The following represents a step by step procedure for application of the EERS model. The input parameters are f (frequency in GHz), θ (elevation angle to the satellite), P (percentage of distance traveled over which the fade $A(P,\theta,f)$ is exceeded.

Step 1: Calculate the fade distribution at f_L = 1.5 GHz valid for percentages of $20\% \geq P \geq 1\%$ at the desired path elevation angle, $60° \leq \theta \geq 20°$ using (1), where $M(\theta)$ and $N(\theta)$ are given by (2) and (3).

Step 2: Convert the fade distribution at . f_L = 1.5 GHz, valid for $20\% \geq P \geq 1\%$ to the desired frequency, f (GHz), where 0.8 GHz $\leq f \leq$ 20 GHz using (5).

Step 3: Scale the fade distributions to the desired percentages of distance traveled, where $80\% \geq P > 20\%$ using (6).

Step 4: For path elevation angles in the range $20° > \theta \geq 7°$, apply (7).

Extension to Elevation Angles Above 60°

The ITU-R [3] has recommended that for elevation angles greater then 60°, the EERS model be linearly interpolated with the fades listed in Table 2 at 80° (derived by Smith et al. [9]) for the frequencies of 1.6 and 2.6 GHz. Specifically, the EERS model at frequencies of 1.6 and 2.6 GHz can be extended to elevation angles above 60° by employing the following steps:

Step 1: Derive the EERS fade values at an elevation of 60° at the frequencies of 1.6 GHz and 2.6 GHz (Table 2).

Step 2: Linearly interpolate between the 1.6 GHz fade calculated for an angle of 60° and the fade for an elevation angle of 80° in Table 2 (second and fourth columns).

Step 3: Linearly interpolate between the 2.6 GHz fade calculated for an angle of 60° and the fade for an elevation angle of 80° in Table 2 (third and fifth columns) only for percentages of 15% and smaller. (At 20% and greater, the 60° values are not smaller than the 80° fade values.)

As an example, Figure 3 shows plotted the fade exceeded at 1.5 GHz versus elevation angle for a family of curves of equal percentage employing the EERS model with angle extension to 90°.

ACKNOWLEDGEMENTS

This effort was supported by the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California. Many thanks to Dr. David Rogers for his helpful suggestions.

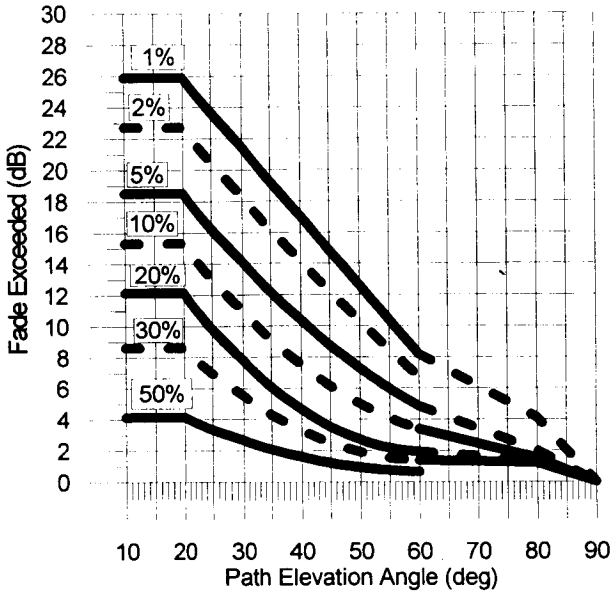


Figure 3: Fade exceeded at 1.5 GHz versus elevation angle for a family of equal percentage curves employing the EERS model with angle extension beyond 60° as described in text.

REFERENCES

[1] **J. Goldhirsh and W. J. Vogel**, *Propagation Effects for Land Mobile Satellite Systems: Overview of Experimental and Modeling Results,* NASA Reference Publication 1274, February, 1992.

[2] **J. Goldhirsh and W. J. Vogel**, *An extended empirical roadside shadowing model for estimating fade distributions from UHF to K-band for mobile satellite communications.* Space Communications., Vol. 13, No. 3, pp. 225-237, 1995.

[3] **International Telecommunications Union - Radio Communication Study Groups**, *Revision of Recommendation ITU-R P.681*, Document 3M/3 - February 1997.

[4] **W. J. Vogel, J. Goldhirsh, and Y. Hase**, *Land-mobile-satellite fade measurements in Australia.* AIAA J. Spacecraft and Rockets, Vol. 29, pp. 123-128, 1992.

[5] **W. J. Vogel and J. Goldhirsh**, *Mobile satellite system propagation measurements at L-band using MARECS-B2.* IEEE Transactions Antennas Propagation., Vol. AP-38, pp. 259-264, 1990.

[6] **J. Goldhirsh and W. J. Vogel**, *Mobile satellite system fade statistics for shadowing and multipath from roadside trees at UHF and L-band*, IEEE Transactions Antennas Propagation., Vol. AP-37, pp. 489-498, 1989.

[7] **W. J. Vogel and J. Goldhirsh**, *Multipath fading at L-band for low elevation angle, land mobile satellite scenario* IEEE Transactions on Selected. Areas of Communications., Vol. 13, pp. 197-204, 1995.

[8] **W. J. Vogel and J. Goldhirsh**, *Earth-satellite tree attenuation at 20 GHz: Foliage effects*, Electronic. Letters., Vol.†29, pp. 1640-1641, 1993.

[9] **H. Smith, J. G. Gardiner, and S. K. Barton**, *Measurements on the Satellite Mobile Channel at L & S Bands*, Proceedings of the Third International Mobile Satellite Conference (IMSC '93), June 16-18, Pasadena, California, pp. 319-324, 1993.

Channel Behavior and Power Control in Handheld Mobile Satellite Links

F. Davarian, S. Chen, and M. Shihabi

Hughes Space and Communications
Bldg S10, MS S352
2260 Imperial Highway
El Segundo, CA 90245
Tel: 310-662-5375 -- Fax: 310-662-7060
e-mail: fdavarian@ccgate.hac.com

ABSTRACT

Fading due to shadowing and blockage can impact the channel availability of a mobile satellite system. This paper examines channel behavior under open field and light shadowing conditions. Several prediction models are reviewed and key channel parameters are defined and determined using measured data. These parameters include signal autocorrelation function and bandwidth, as well as other statistical and dynamic channel attributes. Signal statistics are examined as a function of user environment. The objective is to develop tools to examine and predict the performance (availability) of the channel for a geostationary mobile satellite system. This paper also discusses different power control schemes and their advantages and disadvantages in combating shadowing-induced fades.

1. INTRODUCTION

While user acceptance is the key to the success of a new application, good channel availability is an important part of user acceptance of a communication network. Mobile satellite systems offer large regional and/or global coverage areas, making the service attractive to the customer. User environment, however, can impose limitations on channel availability because of induced signal fading. Channel characteristics are dependent on the environment that the user operates in. For example, the signal received by a handheld terminal operated in an open field is likely to experience mild variations, whereas signal reception in a heavily shadowed environment is likely to be plagued with frequent episodes of deep fades. To understand the effect of the environment on the channel, Figures 1-a and b are presented. Figure 1-a shows the signal level received by a handheld terminal in an open field. Figure 1-b shows the same signal received in a surrounding specified as light shadowing (In both cases the user is walking while operating the handset). In both figures, the abscissa shows time in seconds, and the ordinate shows signal power level in dB relative to the line-of-sight (path elevation angle 47^0, frequency 2 GHz).

To gain insight into channel behavior, signal histograms are presented in Figures 2-a and b. The signal from the open field is concentrated around the 0 dB level and has a near-symmetric histogram, indicating a Rician channel (strong line of sight and some multipath). The signal from the light shadowing environment has a skewed histogram indicating shadowing and blockage events. The cumulative distribution functions for these signals are given in Figure 3 b. For the open field (OF) signal, only 0.1% of all the observed samples experience a fade larger than 10 dB, whereas for the light shadowing (LS) signal, 10 dB fade occurs at 10% probability level.

Figures 1-3 demonstrate the effect of the user environment on channel behavior. Much work has been done to model the propagation channel. The next section summarizes some of the available models.

2. CHANNEL MODEL

Figure 4 shows the mobile satellite communication system. It consists of a forward link from the gateway station to the user terminal and a return link from the user terminal to the gateway. Without the loss of generality, we can assume that the satellite is a simple repeater of the signal. In this figure, user and gateway frequencies are assumed to be L- and Ku -band, respectively, however, one should note that other band allocations also exist. For example, the S-band is another popular region of frequency spectrum for mobile satellite communications.

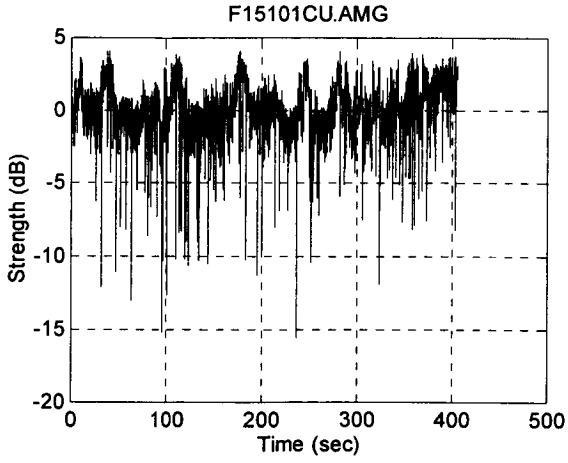


Fig. 1-a: Time Series of the Received Signal Power in an Open Field (f = 2 GHz)

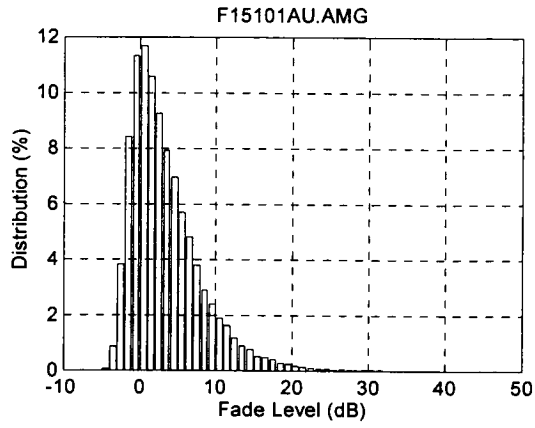


Fig. 2-b: Histogram of Fades (Light Shadow)

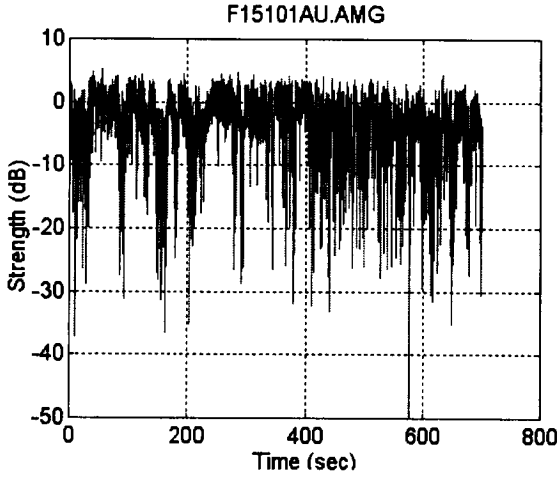


Fig. 1-b: Time Series of the Received Signal Power in Light Shadowing (f = 2 GHz)

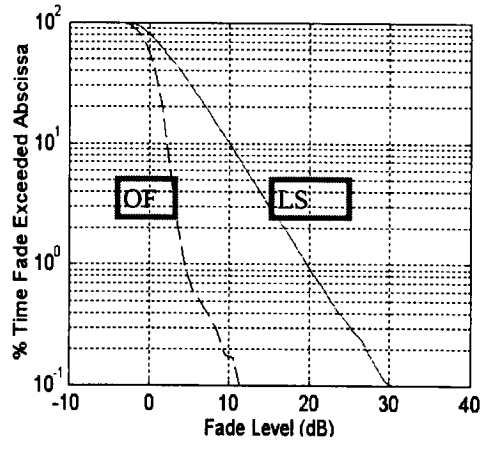


Fig. 3: Cumulative Distribution of Signal Power Levels

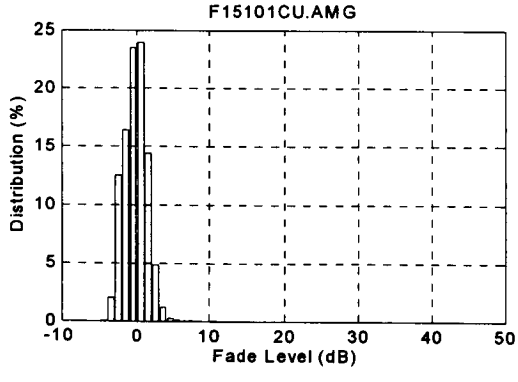


Fig. 2-a: Histogram of Fades (Open Field)

Although the two ends of the channel, satellite and user terminal, are separated by a vast distance, e.g., 35,000 km for the geostationary orbit, the environment affecting the channel is normally only a few meters from the user terminal. This environment introduces multipath and shadowing (or blockage) effects. Therefore, most analytical models, treat the channel as some combination of shadowing and multipath effects. In what follows, we review a few of these models.

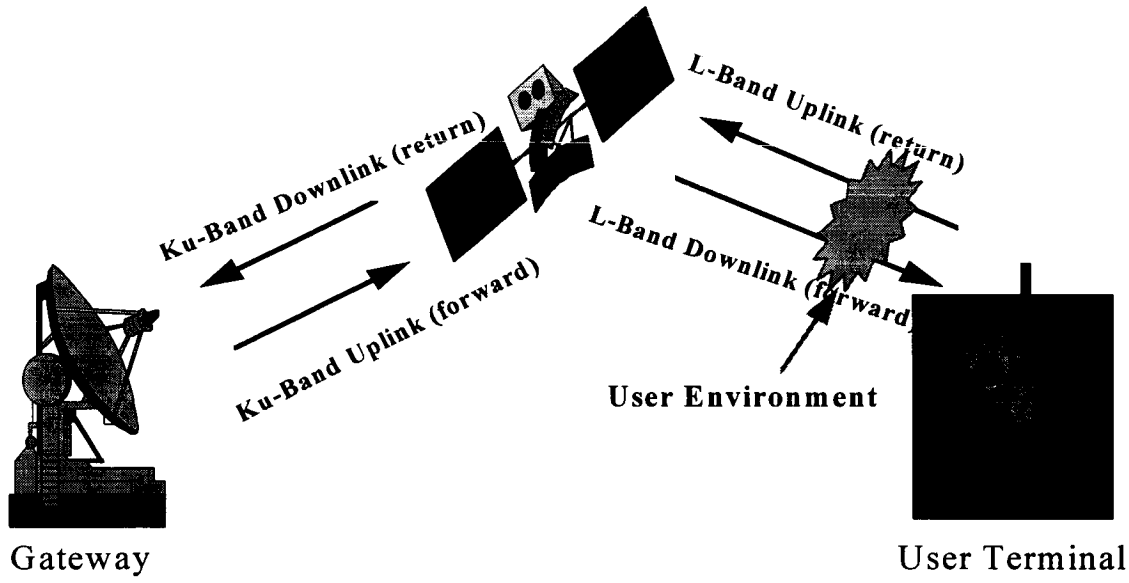


Fig. 4. The Mobile Satellite Channel

2.1 Loo's Model

This model, developed by Loo [1], assumes the received voltage due to the diffusely scattered power is Rayleigh distributed and the voltage variations due to attenuation of the line of sight signal are lognormally distributed. A constant intensity Rayleigh distributed scattering voltage is combined with a lognormally shadowed line-of-sight signal voltage, Figure 4. The mean scattered power in the model is set at a constant level depending on the environment. Model parameters are derived from field measured data.

2.2 Lutz, et al, Model

The model proposed by Lutz, et al, [2] consists of two propagation states: shadowing and no shadowing. In the unshadowed state, signal envelope statistics are Rician with a constant K factor. In the shadowed state (roadside trees) there is no line of sight; what remains is Rayleigh distributed multipath signal, Figure 5. The average strength of the multipath is lognormally distributed. Model parameters are derived from field measured data.

2.3 Smith and Stutzman Model

Two propagation states of shadowing and no shadowing are associated with this model [3]. In the unshadowed state, signal envelope statistics are Rician with constant K factor. In the shadowed state, the lognormally distributed amplitude of the direct signal is combined with a constant level diffuse multipath component, Figure 6.

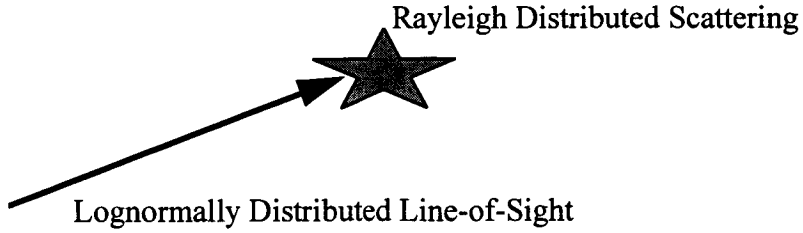


Figure 4. Model consisting of a lognormally distributed line-of-sight component and constant-power Rayleigh distributed multipath components.

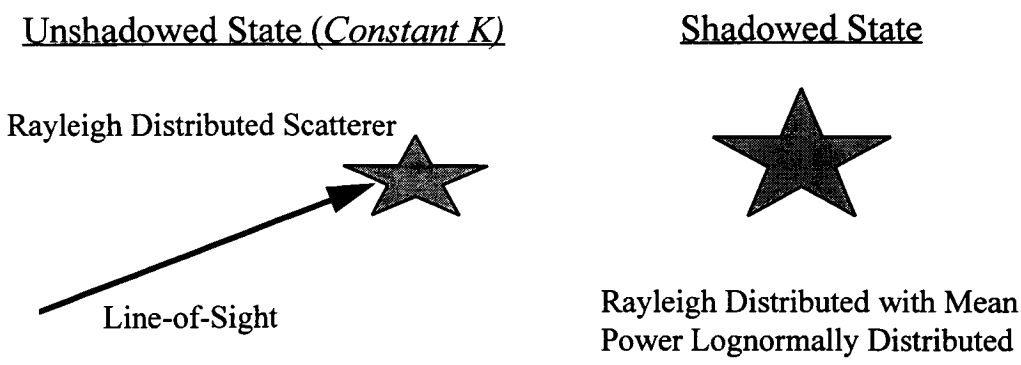


Figure 5. Model Consisting of Two States ($K = \frac{P_{LOS}}{P_{mul}}$).

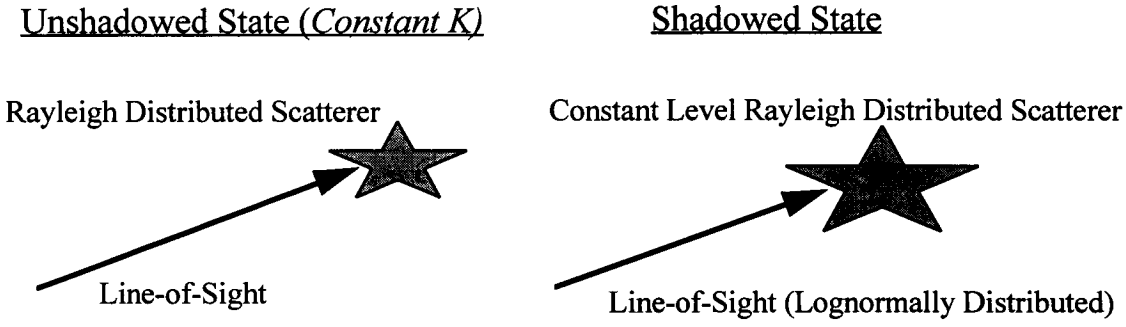


Figure 6. Smith and Stutzman Model

2.4 ITU-R Model

Unlike the previous models, the ITU-R model is an empirical scheme which is based on vehicular mesurements in the U.S., hence known as EERS for Extended Empirical Roadside Shadowing [4], It only applies for roadside tree shadowing conditions described by the percentage of optical shadowing (POS) between 55 and 75 % [5]. Figure 7 shows the fade statistics predicted by this model. The abscissa indicates fade depth and the ordinate shows percentage. This figure compares the model to measured data in Australia. Figure 8 shows another presentation of the ITU-R model to emphasize the effect of the elevation angle on signal statistics.

3. CHANNEL DYNAMICS

In this section the dynamics of the fading channel are examined by determining the autocovariance and power spectral density of the signal. The observed waveform is detected and recorded via its inphase and quadrature components. The input to the fading channel is a CW signal. The CW signal is modulated by the fading (shadowing and multipath) process. The output signal experiences Doppler spread and amplitude and phase fluctuations. The power levels of the output signal for two user environments are shown in Figures 1-a and b.

Let a(t) and b(t) denote the inphase and quadrature components of the observed signal. We form a two dimensional random vector v(t) = [a(t) b(t)]. The mean of v(t) is given by

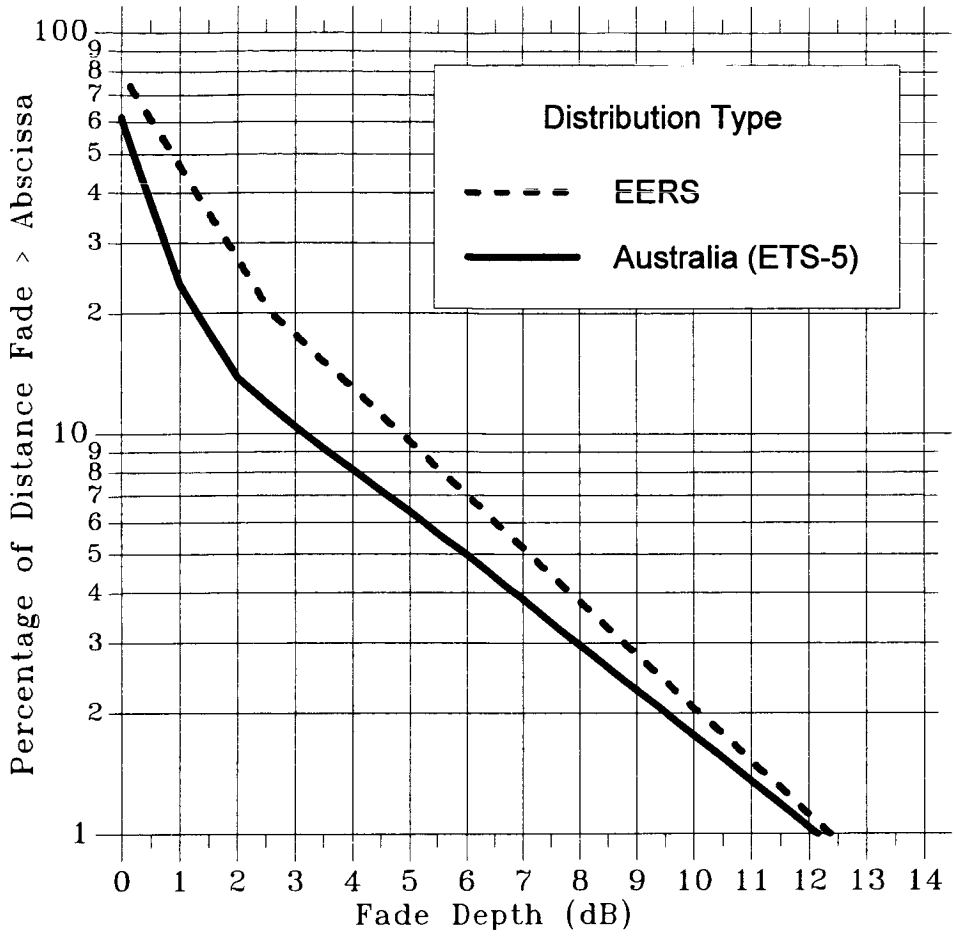


Fig. 7: Cumulative Fade Distribution Prediction by the ITU-R Model for L-Band

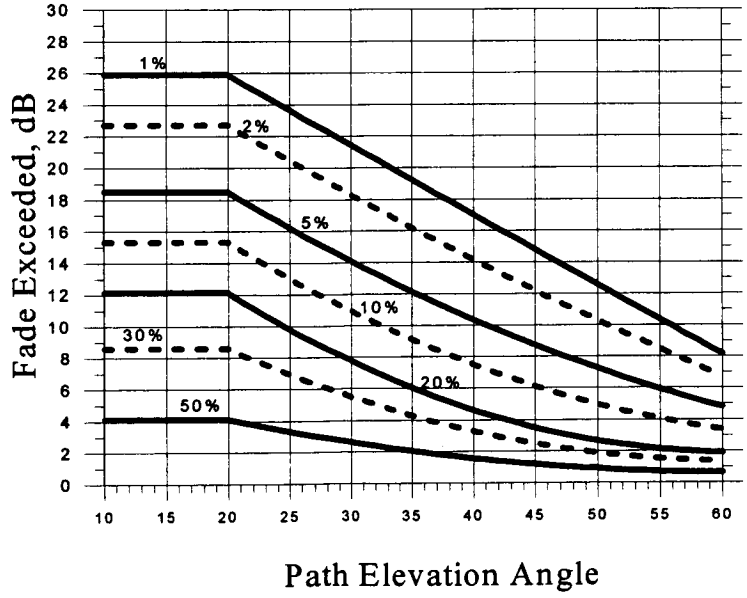


Fig. 8: Fade Exceeded versus Path Elevation Angle for a Family of Constant Percentages (L Band)

$$m = [\bar{a}(t) \quad \bar{b}(t)] \tag{1}$$

The autocovariance function of v(t) can be computed as

$$R(\tau) = E[v(t)v^{t}(t+\tau)] - mm^{t} \tag{2}$$

where superscript t denotes the transpose operation. The power spectral density of the signal can be obtained from the Fourier transformation of the autocovariance function. Signal bandwidth can be determined from the power spectral density. The autocovariance function can be normalized with respect to the signal power as

$$r(\tau) = \frac{R(\tau)}{R(0)} \tag{3}$$

Note that

$$0 \le r(\tau) \le 1 \tag{4}$$

Figure 9 shows the autocovariance functions for the open field and light shadowing data, respectively. Note that the mean values of the signals are subtracted from the observed waveforms to have zero mean processes for obtaining autocovariance functions. This implies that, to a great degree, the line of sight component has been taken out and that Figure 9 shows the autocorrelation function of the multipath component. The data used to obtain Figure 9 are complex-valued signals with amplitude and phase information preserved. It appears that the multipath components experience a 50% loss of correlation in less than a second. The light shadowing case (ground reflections and tree scattering) shows slower decay than the Open field case (ground reflections). The Fourier transformation of the autocovariance functions will result in the power spectra of the signals as shown in Figure 10. Channel bandwidth can be computed from Figure 11 . This figure provides out-of-band relative power levels for the two environments of open field and light shadowing, respectively. Note that the relative power levels in Figure 11 are with respect to signal total power (line-of-sight and multipath). The 90% bandwidth for the two environments, open field and light shadowing, are 0.4 Hz and 4 Hz, respectively. Clearly the presence of trees have caused more of the signal power to be spread in the frequency domain.

The amount of correlation between two signals can be measured from the cross-covariance function of two signals. In particular, we may be interested in knowing the correlation between the uplink and downlink frequencies of a link. Let $v_1(t)$ and $v_2(t)$ denote two simultaneous measurements at frequencies f_1 and f_2. These two-dimensional vectors are given as $v_1(t) = [a_1(t) \; b_1(t)]$, and $v_2(t) = [a_2(t) \; b_2(t)]$, where the elements inside the brackets represent the in-phase and quadrature components of the signals at frequencies f_1 and f_2, respectively.

The cross-covariance function of $v_1(t)$ and $v_2(t)$ is given as

$$R_{1,2}(\tau) = E[v_1(t)v_2^{t}(t+\tau)] - m_1 m_2^{t} \tag{5}$$

where vectors m_1 and m_2 denote the mean of $v_1(t)$ and $v_2(t)$. We may normalize the above function with respect to signal power:

$$r_{1,2}(\tau) = \frac{R_{1,2}(\tau)}{\sqrt{R_1(0)R_2(0)}} \tag{6}$$

Note that

$$-1 \le r_{1,2}(\tau) \le 1.0 \tag{7}$$

In a recent measurement campaign, Hughes Space and Communications Company collected dual frequency data similar to the waveforms shown in Figures 1-a and b. Unfortunately the processing of these data have not been completed at the time of the writing of this paper. Therefore, the dual-frequency data will be presented in a future publication.

An important factor in describing channel dynamics is the distribution of fade and nonfade events. Figure 12 shows the fade duration distribution for the light shadowing data with a fade threshold of 7 dB. Only 5% of all fade events have a duration of 2 s or longer. Figure 13 shows the distribution of nonfade events at a fade threshold of 7 dB. About 50% of all nonfade events have a duration of 6 s or longer. The reader should note that in the definition of fade and nonfade events, all fades and nonfades lasting less than 80 ms are discarded. In other words, the scenario consisting of two nonfade events separated by a fade event lasting less than 80 ms, constitutes a single nonfade event. This approach is adopted to prevent the degradation of event statistics due to a few noisy samples.

4. FADE MITIGATION

Several techniques can be used to lessen the effect of propagation on system performance. A nonexhaustive list of these options is given below:

1. Additional margin,
2. High path elevation angle,
3. Time diversity
4. Satellite diversity, and
5. Power control.

In what follows a brief description of Item 5, power control, will be given

4.1 Power Control

Power control refers to a fade mitigation technique that maintains a near constant received signal power. It changes transmitter power based on signal quality measurements made on the received signal. Power control can be performed in open loop or closed loop fashions. In the open loop scheme, the user terminal (or the gateway) measures the downlink signal quality and adjusts uplink signal power. In the closed loop scheme, measurements at the gateway (or user terminal) downlink channel are used to adjust the uplink signal power of the user terminal (or gateway). In a satellite based system, the propagation delay in the closed loop approach is nontrivial. Therefore the open loop approach responds quicker to fade events than the closed loop approach. However, the open loop method does not work well if the fades on the uplink and downlink frequencies are not correlated. A hybrid method which uses a combination of both open and closed loop schemes may be more suitable for some conditions. Table 1 summarized the advantages and disadvantages of the three options.

In designing a power control scheme, close attention should be given to channel dynamics. For example, Figure 12 indicates that a small percentage of all fade events last more than 1 second. Therefore in this case, the power control system should have a time constant less than 1 second to be of any use. Figure 12 is plotted for a fade threshold of 7 dB. A lower fade threshold will change the fade duration statistics, causing fade events, on the average, to have a longer duration.

4.2 Thoughts on Hybrid Power Control

In this section a method for combining open and closed loop power control is proposed. Let us define the following parameter.

$$F = r(\tau) - r_{1,2}(0) \qquad (8)$$

where $r(\tau)$ and $r_{1,2}(\tau)$ are given in Equations 3 and 6, respectively. Clearly, for positive values of F, there would be a preference to selecting the closed loop approach, whereas F < 0 indicates a preference to selecting the open loop approach. Note that when $r_{1,2}(0)$ is negative, the open loop approach is not a viable one. Let Q_c and Q_o in dB denote the closed loop and the open loop signal quality estimates in a power control system. The combining strategy can be presented by the following equations

$$Q = \frac{1}{2}Q_c(1+F) + \frac{1}{2}Q_o(1-F) \qquad (9)$$
$$\text{if } r_{1,2}(0) \geq 0$$
$$Q = Q_c \qquad \text{if } r_{1,2}(0) < 0$$

For a given environment, the value of F can be computed from field measurements.

Example

Given τ = 1 s, $r(\tau)$ = 0.7, and $r_{1,2}(0)$ = 0.7, we compute F as

F = 0.7 - 0.7 = 0

Therefore,

$$Q = \frac{1}{2}Q_c + \frac{1}{2}Q_o$$

In this example, signal quality measurements of the open loop and the closed loop approach are averaged to yield a signal quality indicator that can be used for power control.

REFERENCES

1. C. Loo, "A Statistical Model for Land Mobile Satellite Link," *IEEE Trans. on Vehicular Technology*, Vol VT-34, No. 3, August 1985, pp. 122-127.
2. E. Lutz, et al, "Land Mobile Satellite Communications Channel Model, Modulation and Error Control," *Seventh International Conference on Digital Satellite Communications*, 12-16 May, 1986, pp. 537-543.
3. W. Smith, and W. Stutzman, "Statistical Modeling for Land Mobile Satellite Communications," Virginia Tech Report EE Satcom 86-3, Virginia Tech, Blacksburg, VA, August 1986.
4. J.Goldhirsh and W. Vogel, "Extended Empirical Roadside Shadowing Model from ACTS Mobile Measurements," Proc. NAPEX XIX, JPL Publication 95-15, August 1995, pp. 91-102.

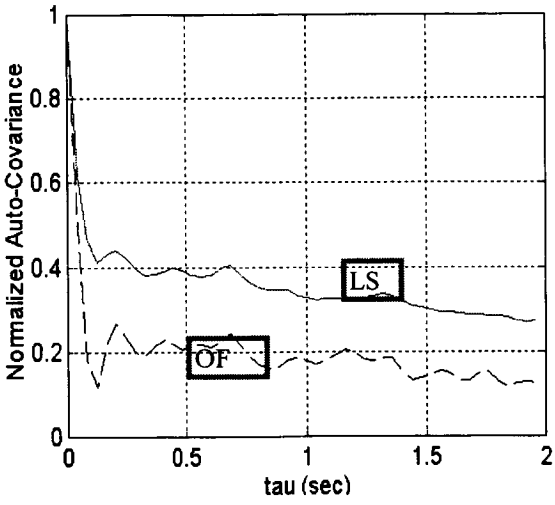


Fig. 9: Autocovariance Function of the Channel

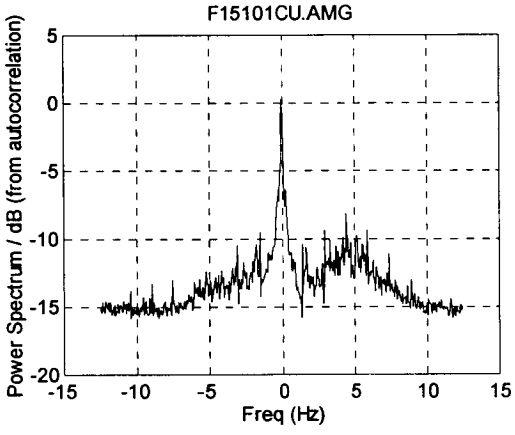


Fig. 10-a: Channel Power Spectrum (Open Field)

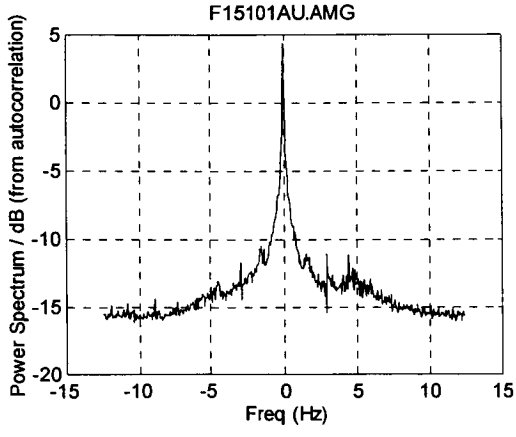


Fig. 10-b: Channel Power Spectrum (Light Shadow)

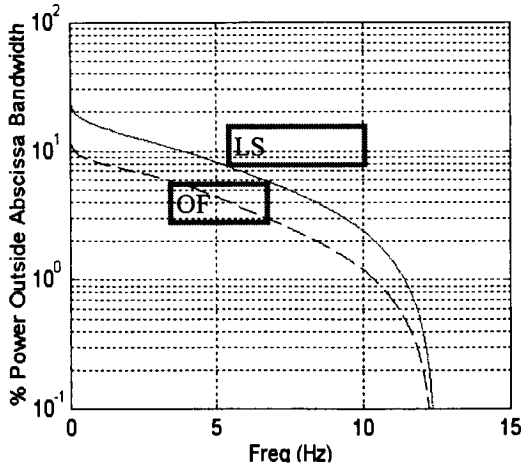


Fig. 11: Fractional Out-of-Band Power

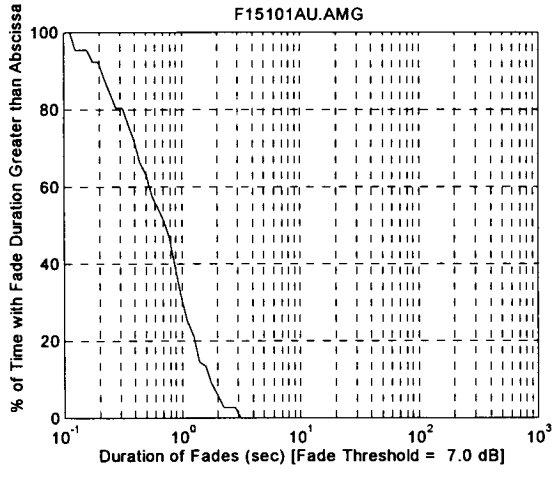


Fig. 12: Fade Duration Distribution (Light Shadow)

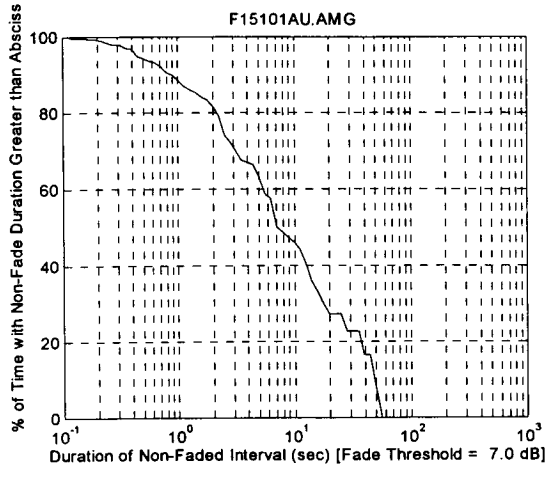


Fig. 13: Nonfade Duration Distribution (Light Shadow)

Scheme	Advantage	Shortcoming
Open Loop	Fast response time (good for fast changing channels)	Lack of correlation between uplink and downlink can cause large errors
Closed Loop	Good for slow changing channels	Slow response time due to propagation delay
Hybrid	A compromise between open and closed loop	Added complexity

Table 1. The Advantages and Shortcomings of Power Control Schemes

Session 7
Terminal Technology

Session Chairperson—*Hans Haugli*, Vistar, Canada
Session Organizer—*John Sydor*, Communications Research Centre, Canada

Dual Tunable Type Low Spurious PLL Synthesizer Driven by DDS for Hz-order Tuning used in North American MSAT Mobile Terminal

Kenji ITOH, Ken'ichi TAJIMA, Shuji NISHIMURA and Yoji ISOTA
Information Technology R&D Center Mitsubishi Electric Corp.
5-1-1 Ofuna, Kamakura-city, Kanagawa 247,JAPAN
Phone:+81-467-41-2549 FAX:+81-467-41-2519
email: itoh@micro4.isl.melco.co.jp

ABSTRACT

A dual tunable type low spurious L-band PLL synthesizer driven by a DDS is proposed for mobile terminal used in North American MSAT systems. The proposed synthesizer can achieve Hz-order tuning and low phase noise characteristics with small sized single loop PLL by employing the DDS. By such a fine frequency tuning function, the synthesizer can realize AFC tuning (Hz-order) by a demodulator and a narrow channel selection (6 kHz) at the same time. Furthermore, two techniques are indicated to suppress DDS's spurious components. First one is a dual tunable concept by the DDS and the PLL that can avoid high level spurious components from a pass-band of the PLL. Another is an employment of a frequency converter that can reduce transfer gain from the DDS to the PLL output. A developed L-band PLL synthesizer achieved the spurious level below -63 dBc, phase noise of -71 dBc/Hz@1 kHz offset and frequency setting speed of 2.7 ms.

INTRODUCTION

In digital mobile radio transmission systems [1], [2] like North American MSAT network, frequency synthesizers that are used in up- and down-converterts are required narrow frequency spacing, fast frequency setting speed, low phase noise and low spurious level at the same time [1], [2]. PLL synthesizers are often used in such systems because of its compact size. But PLL synthesizers cannot achieve the fast frequency setting speed and low phase noise in a case of narrow frequency step size, because PLL's reference frequency must become lower in such a case. A PLL synthesizer driven by a DDS [3] is an effective technique to achieve fine frequency step size, fast frequency setting speed and low phase noise [4]. But this PLL synthesizer has a disadvantage with high spurious level generated by quantization errors in digital calculation [5] and nonlinearity of digital to analog converter (DAC). The PLL cannot suppress such spurious components when they exist in a pass-band of the PLL.

In this paper, a dual tunable type PLL synthesizer with a frequency converted DDS is proposed for suppressing spurious components of the DDS. The proposed PLL synthesizer can change its output frequency in fine step size by a DDS and in coarse step size by PLL's variable frequency divider (VFD). This dual tunable concept by the DDS and PLL can avoid high level spurious components by DDS from in a pass-band of the PLL. And the frequency converter can suppress spurious components, because the transfer gain from the DDS to the PLL output can be reduced. This paper will discuss low spurious characteristics of the proposed synthesizer by analytical and experimental approaches. A developed low spurious PLL synthesizer with the frequency converted DDS is used in MSAT mobile terminals [6] for a channel selection (6 kHz spacing) and AFC tuning (Hz-order tuning) of a demodulator.

CONFIGURATION

Fig.1 shows a schematic diagram of a developed mobile terminal for North American MSAT network [6]. The proposed synthesizers used in up- and down converter are controlled by channel setting data from "Logic&signaling processor" and AFC data from "Baseband processor". Transmitting and receiving frequencies are corrected by the AFC data generated in a demodulator that calculates frequency error caused by TCXO. Fig.2 shows a schematic diagram of the proposed dual tunable type PLL synthesizer with a frequency converted DDS. This synthesizer is controlled by channel setting data and AFC data that is generated by a demodulator. So transmitting and receiving frequency errors by crystal oscillator's imperfection can be corrected by the synthesizer located in RF section.

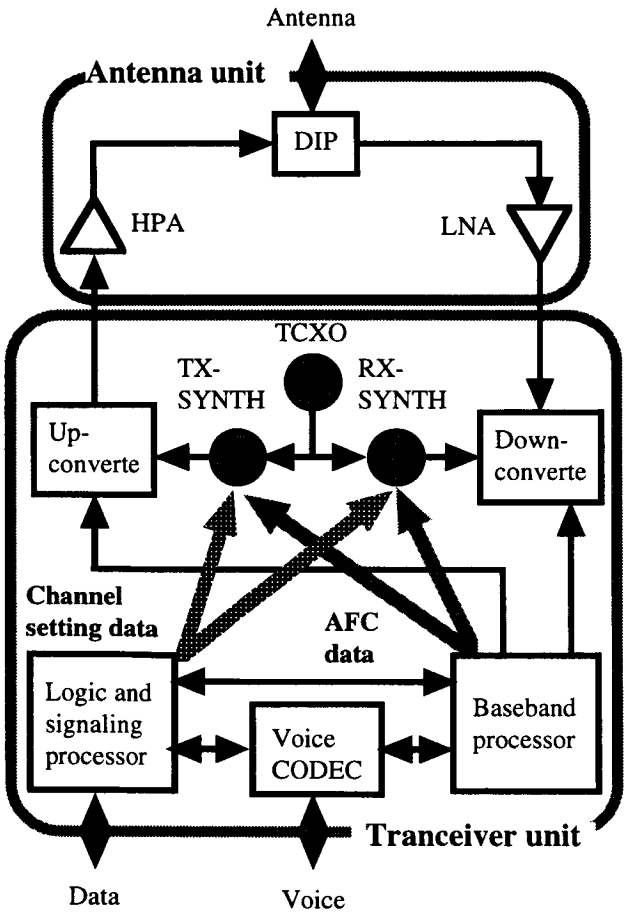
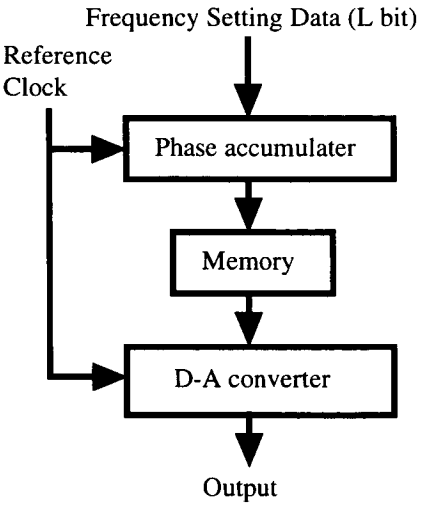
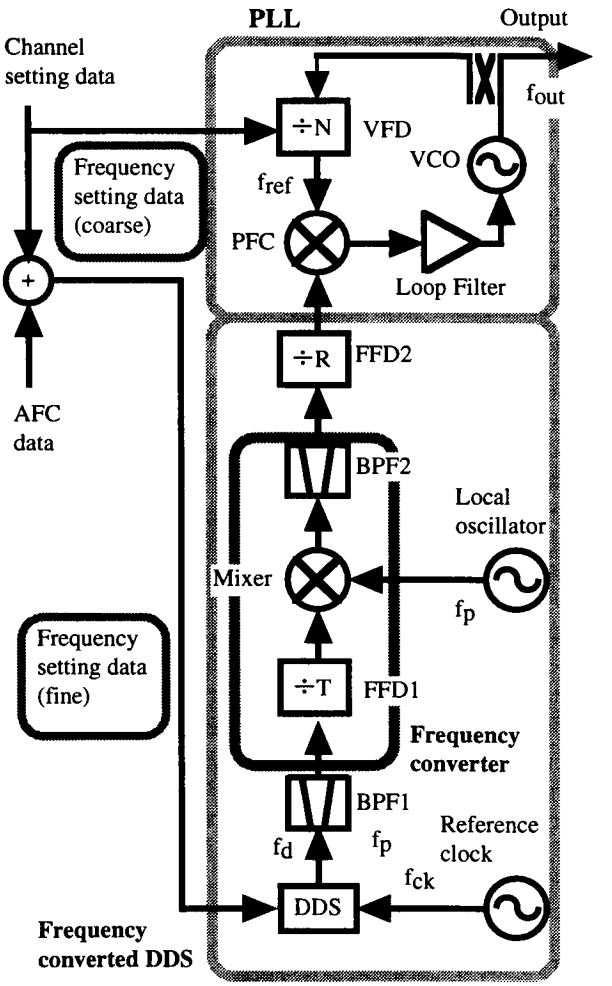


Fig.1 Schematic diagram of a developed mobile terminal for North American MSAT network.



(a) Direct digital synthesizer



(b) proposed synthesizer

Fig.2 Schematic diagram of the proposed dual tunable type PLL synthesizer with a frequency converted DDS.

The output frequency of the DDS f_d is converted to the reference frequency of the PLL by the frequency converter. The output frequency of the PLL f_{out} is

$$f_{out} = \left(\frac{N}{R}\right)\left(f_p \pm \frac{f_d}{T}\right),\ f_d = k \cdot f_{ck} / 2^L \qquad (1)$$

where T, R and N are the dividing ratio of the fixed frequency divider FFD1, the FFD2 and the VFD respectively, f_p is the local frequency of the mixer, f_{ck} is the clock frequency of the DDS, k is the frequency setting data of the DDS and L is the word length of k. (1) indicates that fout is changed by N with coarse steps size of $(1/R) \cdot (f_p \pm f_d/T)$, and by k with fine step size of $N \cdot f_{ck}/(T \cdot R \cdot 2L)$. The proposed synthesizer can change fout almost continuously by the DDS with larger L for practical use in microwave transceivers.

SPURIOUS ANALYSIS OF PROPOSED SYNTHESIZER

Spurious components of the DDS (frequency: f_{dsp}) can be considered as higher order mixing products by f_d and fck as follows:

$$f_{dsp} = |m \cdot f_d - n \cdot f_{ck}| \qquad (2)$$

Where m and n are harmonic numbers. Spurious level of the DDS S_{dsp} is not depend on n in a case of much lower f_{dsp} than f_{ck}. S_{dsp} is also considered to become higher as m becomes lower, because a spurious component with lower m is generated by DAC's nonlinearity and transient response. Fig.3 shows measured S_{dsp} versus m. The spurious level of the proposed synthesizer S_{out} which is generated by f_{dsp} is

$$S_{out} = |H(j2\pi \cdot \Delta f_{dsp})|^2 \cdot \frac{S_{dsp}}{T^2 \cdot R^2}$$
$$\approx \left|\frac{H(j2\pi \cdot \Delta f_{dsp})}{N}\right|^2 \cdot \frac{1}{R_{mix}} \cdot \frac{f_{out}}{f_d} \cdot S_{dsp} \qquad (3)$$
$$\Delta f_{dsp} = |f_d - f_{dsp}|$$

where $H(j2\pi\ \Delta f_{dsp})$ is the closed loop transfer function of the PLL, Δf_{dsp} is the offset frequency of f_{dsp} and R_{mix} is the output/input frequency ratio of the mixer. From (3), a frequency ratio (f_{out}/f_d) from the DDS to the PLL output is reduced by $20\log10(R_{mix})$ (dB) compared to one of the conventional synthesizer without a frequency converter. The closed loop transfer function $H(j2\pi\ \Delta f_{dsp})$ is

$$H(j2\pi \cdot \Delta f_{dsp})/N \rightarrow 1 \quad (\Delta f_{dsp} << f_n)$$
$$\rightarrow 0 \quad (\Delta f_{dsp} >> f_n) \qquad (4)$$

where f_n is the natural frequency of the PLL. (4) indicates that S_{dsp} can be suppressed by PLL in a case of higher Δf_{dsp} than f_n. But if f_d and f_{dsp} intersect each other during a frequency tuning, S_{dsp} cannot be suppressed because of no attenuation by the PLL. Therefore, the f_{out} should be designed that high level spurious components with lower m never intersect with f_d. Fig.4 shows the intersection between f_d and f_{dsp}. This intersecting condition of f_d and f_{dsp} is

$$f_d = f_{dsp} \quad (f_{d1} \le f_d \le f_{d2}) \qquad (5)$$

Using (2) and (5), the intersecting condition is

$$R_{os}(m \pm 1) = integer$$
$$R_{os} = f_d/f_{ck}\ (f_{d1} \le f_d \le f_{d2}) \qquad (6)$$

where R_{os} is f_d normalized by f_{ck}. Fig.5 shows m_{min} versus R_{os}. m_{min} is defined as the minimum m which satisfies (6). As shown in the figure, m_{min} becomes higher by narrowing the sweeping bandwidth Δf_d ($=f_{d2}-f_{d1}$) and selecting an adequate value of R_{os}. Moreover, Δf_d is

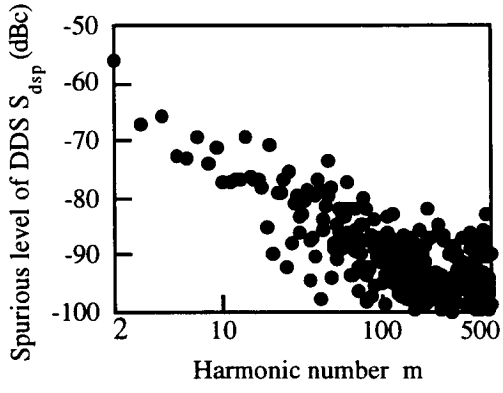
$$\Delta f_d = \frac{\Delta f_{out}}{M_{pll}} \cdot \frac{T \cdot R}{N}$$
$$\approx \Delta f_{out} \cdot \frac{R_{mix}}{M_{pll}} \cdot \frac{f_d}{f_{out}} \qquad (7)$$

where Δf_{out} is synthesizer's overall sweeping bandwidth and M_{pll} is the number of the coarse tuning channels. From (7), Δf_d is possible to be narrower by increasing M_{pll}. From above equations and discussions, design conditions can be derived as follows:

(a) Spurious level is reduced $20\log10(R_{mix})$ (dB) by the proposed frequency converter. On the other hand, Δf_d becomes wider as R_{mix} becomes higher for suppressing spurious components. This means that the high level spurious components will more likely intersect with fd. R_{mix} should be decided by trading off between a reduction

effect and intersecting spurious level.

(b) Δf_d can become narrower for higher m_{min} as M_{pll} will be higher. Then high level spurious components and the carrier signal will not intersect. But in such a condition, the phase noise and the frequency setting speed of PLL will be degraded, because the reference frequency of the PLL becomes lower. M_{pll} should be decided by trading off between Δf_d and PLL characteristics.



f_d	: 0.22 f_{ck}
Phase resolution	: 12 bit
Amplitude resolution	: 10 bit

Fig.3 Measured S_{dsp} versus harmonic number m.

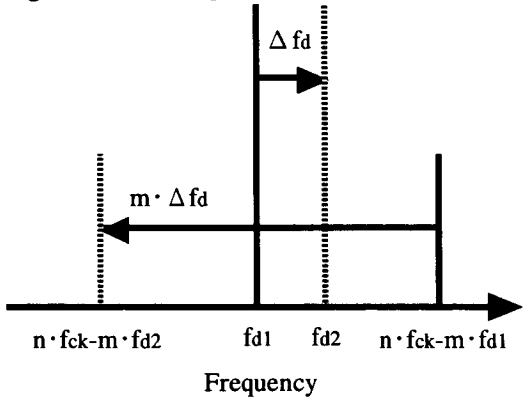
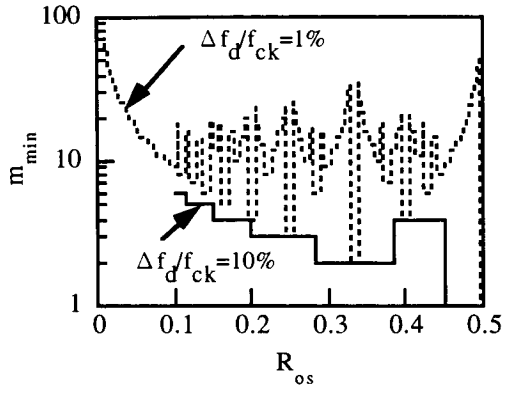
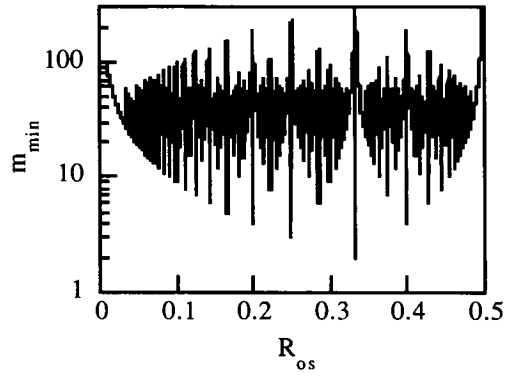


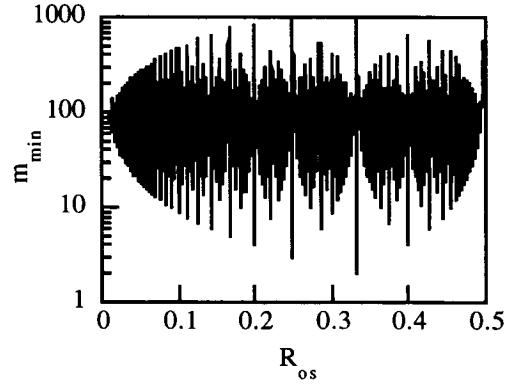
Fig.4 Intersection between f_d and f_{dsp}.



(a) $\Delta f_d/f_{ck}$=10%,1%



(b) $\Delta f_d/f_{ck}$=0.1%



(c) $\Delta f_d/f_{ck}$=0.01%

Fig.5 Minimum harmonic number m_{min} versus R_{os}.

MEASURED CHARACTERISTICS OF DEVELOPED PLL SYNTHESIZER

A L-band PLL synthesizer is developed for a mobile terminal [6] used in MSAT network [1]. Table 1 indicates the parameter summary. Coarse tuning step size

is 100 kHz and fine tuning step size is less than 10 Hz. This synthesizer has high enough frequency resolution for the digital controlled oscillator for the AFC of the demodulator.

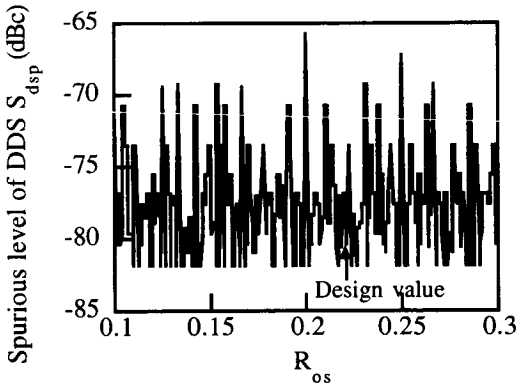
Fig.6 shows calculated S_{dsp} versus R_{os} which is obtained from the measured spurious level as shown in Fig.2. predicted spurious level varies between -66 dBc and -82 dBc. R_{os} is set to 0.221 by considering the realization of BPF2, and S_{dsp} is -80 dBc at this point.

Fig.7 shows the output spectra of the developed synthesizer. Spurious components are below the phase noise. Fig.8 shows the phase noise of the synthesizer. In this figure, the solid line indicates measured values and the dotted line indicates predicted values. The measured noise floor level is -71 dBc/Hz.

Fig.9 shows the transient characteristic of the synthesizer. In this figure, the solid line indicates measured values and the dotted line indicates calculated values by the numerically inverse Laplace transform. The frequency setting speed is 2.7 ms with frequency error of 1 kHz during the coarse frequency tuning.

Table 1 Parameter Summary of L-band synthesizer.

Output frequency f_{out}	L-band
Bandwidth Δf_{out}	30 MHz
Coarse step size	100 kHz
Fine step size	< 10 Hz
R_{os}	0.22
R_{mix}	23
M_{pll}	300
Natural frequency f_n	1.3 kHz
Word length L	32 bit



f_d	: 0.0015 f_{ck}
Phase resolution	: 12 bit
Amplitude resolution	: 10 bit

Fig.6 Calculated DDS spurious level S_{dsp} versus R_{os}.

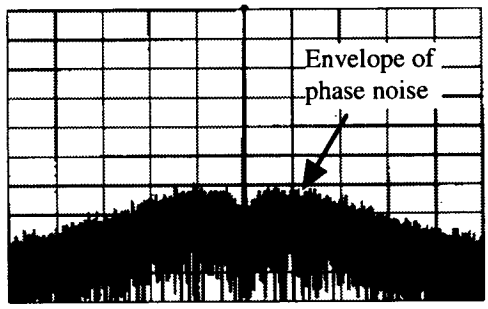


Fig.7 Output spectra of the L-band synthesizer.

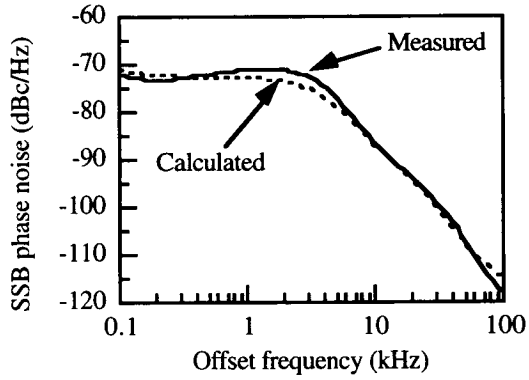
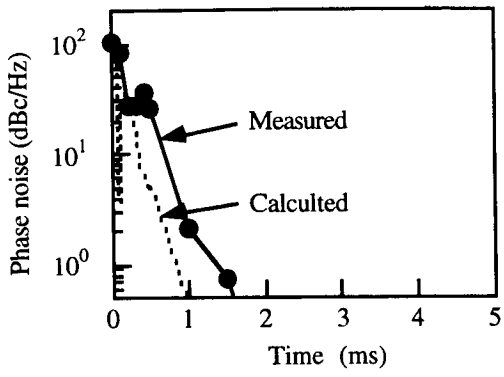
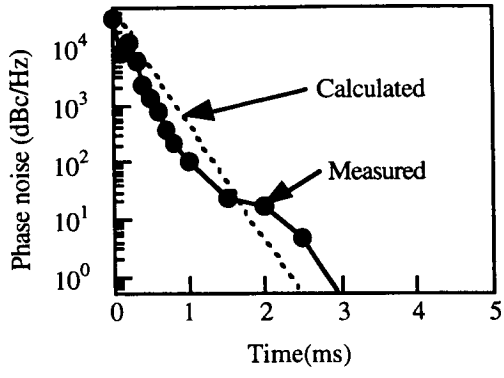


Fig.8 Phase noise of L-band synthesizer.



(a) fine tuning (Δ f=100 kHz)



(b)coarse tuning (Δ f=30 MHz)

Fig.9 Transient characteristic of L-band synthesizer.

CONCLUSION

The dual tunable type PLL synthesizer with the frequency converted DDS has been proposed and developed for MSAT systems. A low spurious characteristics was also described by analytical and experimental approaches. A developed L-band PLL synthesizer with channel steps less than 10 Hz achieved spurious level below -63 dBc, phase noise of -71 dBc/Hz@1 kHz offset and frequency setting speed of 2.7 ms.

REFERENCES

[1] **K. Fong, and G. Churan**, *Mobile earth terminals in the AMSC mobile satellite service system*, Proc. AIAA 16th International Communication Satellite System Conference, pp.245-255, 1994.

[2] **N. Nakajima and K. Kinoshita**, *A System design for TDMA mobile radios*, Proc. IEEE Veh. Technol. Conf., Orland, FL, 1990.

[3] **J. Tierney, C. M. Rader and B. Gold**, A digital frequency synthesizer," IEEE Trans. Audio and Electroacoustics, AU-19,1, pp.48-57, Mar.1971.

[4] **A. L. Bramble**, *Direct Digital Frequency Synthesis*, Proc.35th Ann. Freq. Control Symp., PP.406-414, 1981.

[5] **H. T. Nicholas and H. Samueli**, *An analysis of the output spectrum of direct digital frequency synthesizer in the presence of phase−accumulator truncation*, Proc.41st Ann. Freq.Control Symp., pp.495-502, 1987.

[6] **T.Fuji, M.Tsuchiya, Y.Isota and K.Aoki**, *Design and performance of mobile terminal for North American MSAT network*, Proc. International Mobile Satellite Conf. '95, pp.365-369, 1995.

Design and Development of a Baseband Processing Scheme Using a DSP for Mobile Satellite Communications

Yasuyuki NAGASHIMA[1], Takanori SHOJI[2], Masayuki DOI[2], Takashi KAWABATA[1], Yoichi MORITANI[1], and Tadashi FUJINO[1]

[1]Information Technology R&D Center, Mitsubishi Electric Corporation
5-1-1 Ofuna, Kamakura-shi, 247 JAPAN
Phone : 81-467-41-2453 FAX : 81-467-41-2419

[2]Communications Equipment Works, Mitsubishi Electric Corporation
8-1-1 Tsukaguchi-Honmachi, Amagasaki-shi, 661 JAPAN
Phone : 81-6-497-6378 FAX : 81-6-497-6381

E-mail : yasuyuki@radio.isl.melco.co.jp

ABSTRACT

The subject addressed in this paper is concerned with a baseband processing scheme applied to the North American mobile satellite communication system, MSAT. The baseband processing scheme can be applied to the multiple access schemes in MSAT, which are SCPC, TDMA, Slotted ALOHA and demand assignment TDMA, in order to support various service features such as voice, asynchronous data and packet data. The baseband processor is developed with considerations for reduction of processing load. The authors have achieved to reduce it less than 7 MIPS, which is sufficiently small to integrate it together with a demodulator and a voice codec in one DSP.

INTRODUCTION

In mobile satellite communications, especially in mobile terminals, integration techniques are essential to achieve miniaturization and cost reduction. The DSP has a possibility to provide a solution of the integration problem. The authors have developed a firmware which comprises a demodulator, a baseband processor, and a voice codec using only one DSP. It is employed by a mobile terminal for the MSAT[1][2][3].

In this paper, the authors describe the baseband processing scheme. The baseband processor forms transmission digital frames (it is called "framing"), extracts digital information from received digital frames (it is called "deframing"), and frame synchronization processes.

The authors' considerations to reduce a load of the baseband processing scheme are shown here. They are based on the block processing for a frame. The authors distribute the baseband processes into the process in a frame, in a symbol, and so on, to achieve reduction of load

Table 1 : Principal features of MSAT

Signaling channel	Outbound: 6.75kbit/s DQPSK Frame length : 120ms Superframe length : 3 frames
	Inbound: 3.375kbit/s π/2-shift BPSK Slot length : 120ms/90ms
Service features	Voice (6400bit/s IMBE Codec[4]) Asynchronous data (RS-232C) Packet data (X.25, X.29)
Data channel	6.75kbit/s DQPSK Frame length : 117ms (792 bits) Superframe length : 480ms (4 frames and 72-bit redundancy)
Voice channel	6.75kbit/s DQPSK Frame length : 117ms (792 bits) (consisting of 6 19-ms subframes) Superframe length : 480ms (4 frames and 72-bit redundancy)
Packet data channel	Outbound: 6.75kbit/s DQPSK Frame length : 238ms
	Inbound: 6.75kbit/s DQPSK Frame length : max 1118.8ms (variable)
Forward error correction	Convolutional Encoding/Viterbi decoding (K=7, R=3/4 or K=7, R=1/2) Interleaving
Access schemes	TDM/TDMA/Slotted ALOHA (Signaling channel) TDM/DA-TDMA/Slotted ALOHA (Packet data channel) SCPC (Communication channels)
Voice activation technique	Controlled per a frame (117ms) When voice transmission is paused, EOD is sent at the end of the frame Channel activity burst is sent every superframe (480ms)

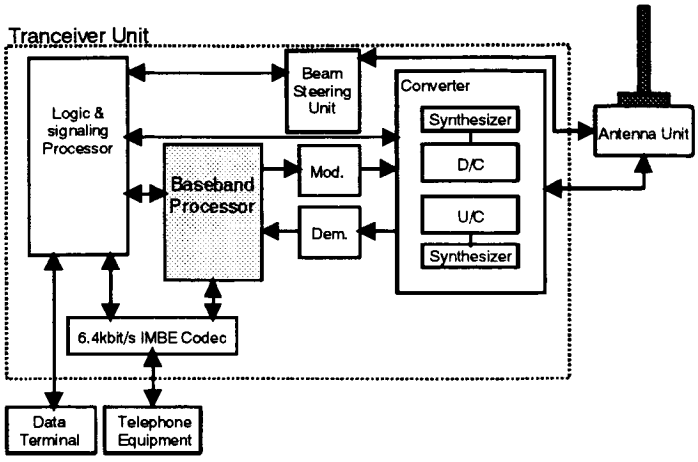


Figure 1 Configuration of the transceiver unit

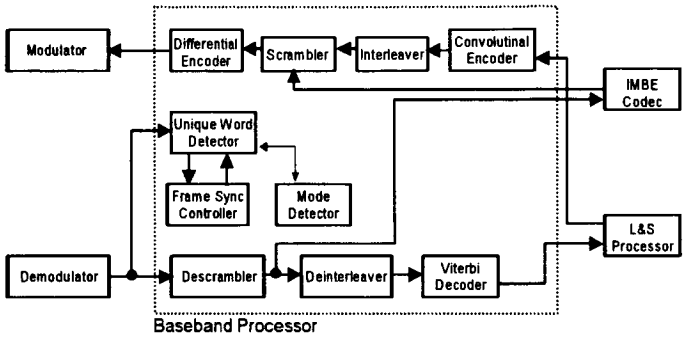


Figure 2 Functional diagram of the baseband processor

and timing preciseness of the transmission and reception simultaneously. The voice channel and the packet data channel have respectively more requirements such as small delay for the voice and applicability to long frame for the packet data channel. The baseband processor which was developed can also meet these requirements.

Under these considerations, the authors have achieved to reduce the processing load less than 7 MIPS, which is sufficiently small to integrate it with a demodulator and a voice codec in one DSP.

OVERVIEW OF MSAT

Table 1 shows principal system features of MSAT. In MSAT, voice and asynchronous data services are provided on the circuit switched channel, using SCPC. On the other hand, packet data service is on the packet switched channel, using TDMA and slotted ALOHA, and frame formats are different for each access scheme. Moreover, the frame formats of SCPC channels are also different for each communication mode. The baseband processor has to handle these various frame formats and access schemes.

OUTLINE OF BASEBAND PROCESSOR

Figure 1 shows an outline of the transceiver unit of the mobile terminal for MSAT[3]. The mobile terminal transmits inbound frames and receives outbound frames. As shown in Figure 1, the baseband processor makes interface to the logic and signaling processor, the IMBE voice codec, the modulator, and the demodulator.

Figure 2 shows a functional diagram of the baseband processor. The baseband processor has functions described in the following three sections.

(1) *Framing processes*

The baseband processor covolutionally encodes, interleaves, and scrambles digital data, and form them into the transmission digital frame. Actually, the baseband processor also operates differential encoding.

For the voice frames, error correction is operated by the IMBE voice codec. The baseband processor does not operate convolution encoding and bit interleaving. Furthermore, MSAT employs the voice activation technique. When voice becomes inactive, transmission is terminated with transmission of the EOD flag. While voice is inactive, a channel activity burst is sent at every superframe (480ms).

(2) *Deframing processes*

For the signaling frames and the data/packet frames, the baseband processor operates descrambling, deinterleaving, and Viterbi decoding.

(3) *Frame synchronization*

Frame synchronization is established by detecting a unique word. Except the packet data channel, any channels have a superframe as shown in Table 1. The baseband processor also processes superframe synchro-nization.

Information of unique word is also used for distinguishing communication mode in the SCPC channel.

BASEBAND PROCESSING SCHEME FOR DSP

(1) *Voice activity distinction*

In the voice mode, voice activation technique is employed as shown in Table 1. At the receiver's side, an interval of unique words becomes longer as 480ms. In this case, the baseband processor keeps frame synchronization by detecting unique words which are sent once per 480ms. In order to keep stiff frame synchronization, it must be as exact as possible to distinguish voice activity. As the authors' countermeasure, the "transitional" state is provided in the voice activity states as shown in Figure 3.

The voice-inactive state is usually established by detecting the EOD. However, due to a poor channel condition, EOD may sometimes be missed. In order to avoid a false state transition, the authors have set an additional

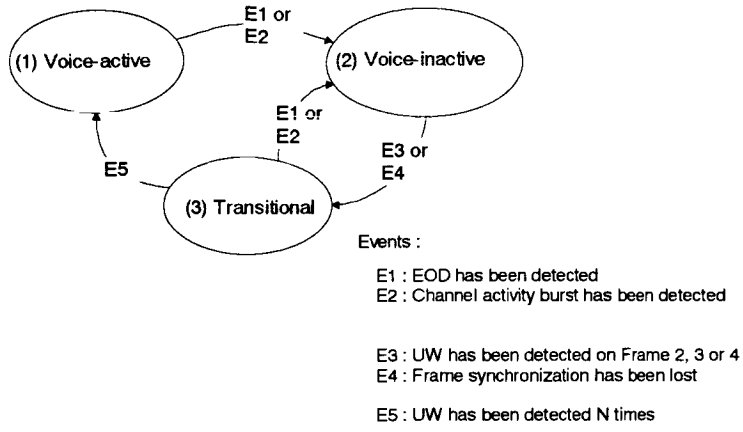


Figure 3 State transition diagram of voice activity detection

condition of transition from the state (1) to (2), of which the channel activity burst has been detected, that is, unique words have been detected at an interval of 480ms.

Even in the voice-inactive state, the baseband processor searches the unique word at every frame to detect the voice-active state. When the unique word has been detected on the frame timing except the superframe timing, a frame counter is reset and incremented. Afterward, the frame counter is incremented every unique word detection until its value reaches N. While this process is active, the state is the transitional state. In the transitional state, though recovery of voice is active as in the voice-active state, bit timing recovery and carrier detection are held still as in the voice-inactive state.

(2) *Considerations of process distribution*

Baseband processing schemes can be sorted into two kinds, which are the processing scheme which is processed collectively in a frame, and that which is subdivided in a symbol. Advantages and disadvantages of both schemes are shown in Table 2.

The authors consider that reduction of processing load is most important to achieve integration with the demodulator and the voice codec in one DSP. Since a DSP has an efficient function of repetition, it is advantageous to operate the same and/or similar processes collectively in one routine. Hence, it is desirable that any framing and de-framing processes are operated collectively. However, unique word detection and frame synchronization processes require preciseness of timing. Thus, these processes are operated with subdividing in each symbol. Figure 4 shows distribution of processes into process sections, such as symbol interruption, process per frame, background, and interface routine.

Baseband processing schemes are described below for three cases of the data/signaling channels, the voice channel, and the packet data channel.

(2-1) *Data/signaling channels*

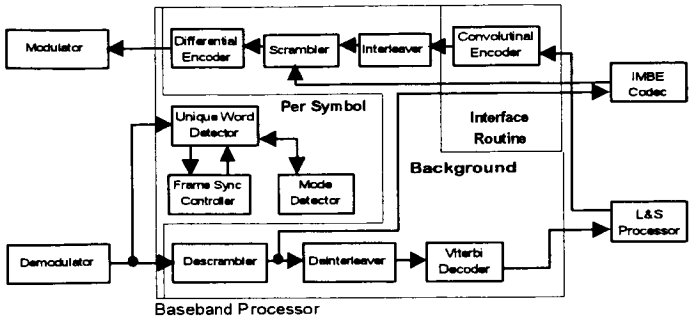
Figure 4 shows the functional diagram and the timing chart of the baseband process for the data/signaling channels.

Process per symbol, shown on the top of Figure 4(b), consists of receiving the bits of the outbound frames, unique word detection, frame synchronization, and communication mode detection. Since they all require the exact timing, the authors distribute them into the process in a symbol according to the considerations shown in Table 2.

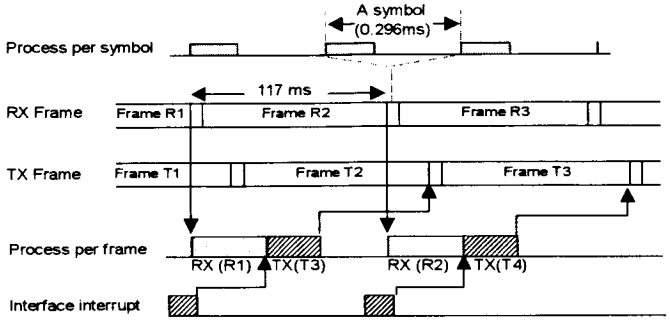
The rest of baseband processes are distributed into the process in a frame. The baseband processor initiates the process in a frame when a frame receiving has been completed. At this time, received bits of a frame are stored. In the process in a frame, first, deframing is initiated. Since the timing of the transmission frames is independent of that of the receiving frames, it is difficult to relate the timing of the framing to that of the transmission frames because adjustment of timing of the framing and the deframing in one DSP becomes complicated. Thus, the authors dispose the framing at the time when the deframing has been completed. They are processed collec-

Table 2 Comparison of processing schemes

	Advantages	Disadvantages
Block Processing Scheme (on the unit of a frame)	Smaller processing load Low peak of load	Hard to apply to exact timing Larger delay of processing
Subdivided Processing Scheme (on the unit of a symbol)	Easily applicable to exact timing Smaller delay of processing	Larger processing load Sharp peak of load



(a) Process distribution



(b) Timing chart

Figure 4 Baseband processing for the data/signaling channels

tively on the unit of a frame. Only convolutional encoding, which is the first process of framing processes, is processed coupled with the interface process in the interface routine for convenience of the development.

The process in a symbol has a higher priority than the process in a frame. Therefore, even while the process in a frame is active, the process in a symbol acts the same as described above. When the process in a symbol becomes active, the process in a frame pauses.

(2-2) *Voice channel*

Figure 5 shows the functional diagram and the timing chart of the process for the voice channel.

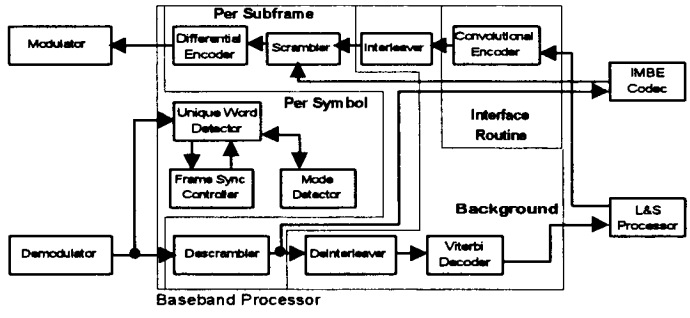
In the voice channel, requirement of a delay limit is more stringent than that of the data channels. Therefore, the framing and deframing processes are carried out in a subframe (20ms). Similarly to the data channels, the timing of the transmission frames and that of the receiving frames are independent of each other. Hence, the order of the baseband processing is nearly the same as that of the data frames, that is, the deframing is initiated first at the time when receiving of a subframe has been completed. The time unit of the baseband processing is a subframe (20ms), delay due to the processing is much smaller than that of the data frames. Unique word detection and the related processes are in the symbol interruption similarly to the data frame's processing.

In MSAT, the frame formats (i.e. the communication mode) of the transmission channel and that of the receiving channel do not change exactly simultaneously. That is to say, there is a case where the framing and the deframing are for the different communication modes. Considering this case and the delay requirement, the authors assign the subframe processing the higher priority than the frame processing.

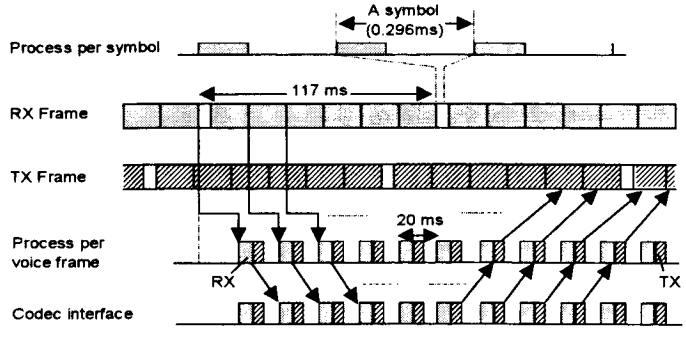
(2-3) *Packet data channel*

For the packet data channel, the baseband processing scheme is basically similar to that for the signaling channel. However, in the packet data channel, there is a case where length of an inbound TDMA burst becomes much longer than that of an outbound TDM frame. That is to say, the framing process may have much more load than the deframing process. In this case, it is advantageous for avoiding temporary overflow of process in the DSP to subdivide the framing process into a moderate number of sub-processes.

Figure 6 shows the functional diagram and the timing chart of the process for the packet data channel, especially in the case where length of the transmission frame is much longer than that of the receiving frame. Number of processed bits of a sub-process is approximately identical to that of an outbound frame, sometimes slightly different from it for convenience of processing.

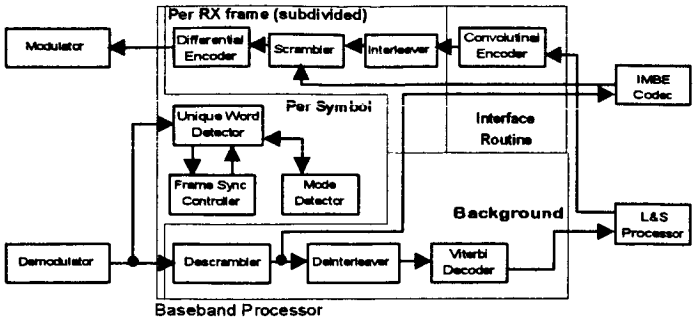


(a) Process distribution

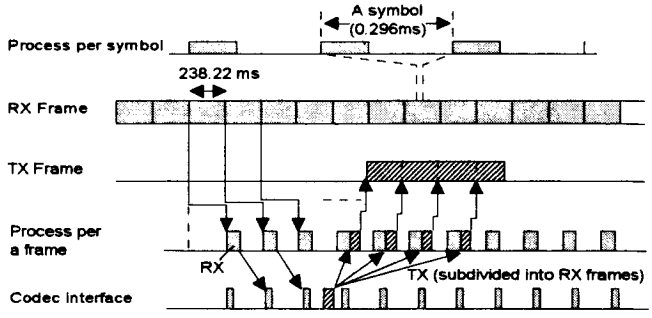


(b) Timing chart

Figure 5 Baseband processing for the voice channel



(a) Process distribution



(b) Timing chart

Figure 6 Baseband processing for the packet data channel

AMOUNT OF PROCESSING LOAD OF BASEBAND PROCESS IN DSP

Table 3 shows an amount of processing load of the baseband process in a DSP. In Table 3, processing loads for the data, voice, and packet data channels are given. As shown in Table 3, although some differences among the different operations exist, processing loads amount less than 7 MIPS. It is sufficiently small to integrate the baseband processor with the demodulator and the voice codec in one DSP.

Table 3 Amount of processing load

Kind of channel	Load (MIPS)
Voice	6.06
Data/Signaling	6.96
Packet data	6.60

CONCLUSION

This paper has presented a baseband processing scheme using a DSP. The described baseband processor can be applied to the MSAT service, which provides various service features, such as voice, asynchronous data, and packet data. It also has a sufficiently small processing load that it can be integrated with a demodulator based on the multiple differential phase detection[6] and an IMBE voice codec. Actually, the authors have developed the mobile terminal with integrating the baseband processor, the demodulator, and the voice codec in one DSP.

REFERENCES

[1] G.A.Johanson and W.R.H.Tisdale, "Summary of the AMSC mobile telephone system", AIAA'94, AIAA-94-0940-CP, pp.234-244 (March 1994).

[2] K.Karam, T.Hearn, and D.Rohr, "Design and implementation considerations of a MSAT packet network", Proc. IMSC'93, pp.245-250 (June 1993).

[3] T.Fuji, M.Tsuchiya, Y.Isota, and K.Aoki, "Design and performance of mobile terminals for North American MSAT Network", Proc. IMSC'95, pp.365-369 (June 1995).

[4] M.S.Brandstein, P.A.Monta, J.C.Hardwick, and J.S.Lim, "A real-time implementation of the improved MBE speech coder", Proc. ICASSP'90, S1.2, pp.5-8 (April 1990).

[5] Y.Nagashima, T.Shoji, M.Doi, T.Kawabata, Y.Moritani, and T.Fujino, "A baseband processing scheme using a DSP for mobile satellite communications", Tech. Rep. IEICE, SAT96-70, CS96-103, *in Japanese* (Oct. 1996).

[6] K.Kojima, M.Miyake, and T.Fujino, "Differential detection scheme for DPSK using phase sequence estimation", Trans. IEICE, 76-B-II, 10, pp.783-792, *in Japanese* (Oct. 1993).

Development of an Inmarsat Standard D Terminal for Global Messaging and Data Broadcast Applications

Jennifer McKenzie, Sean Simmons, Les Tibbo, Peter Rossiter
Calian Communications Systems Ltd.

330 Legget Drive
Kanata, Ontario
Canada K1Y 1Y5

phone: (613) 592-8600 FAX: (613) 592-3378

e-mail: mckenzie@calian.ca

ABSTRACT

Calian Communications has successfully designed the first terminal to be Type Approved and used with Inmarsat D, the first truly global personal messaging and data broadcast service.

With a link margin of over 16 dB, and a novel modulation scheme, message reception is possible in many urban environments and from some indoor locations with Calian's DMR-100 terminal.

BACKGROUND

Inmarsat D is the first truly global combined personal messaging and data broadcast service. Launched in December 1996 at a London press conference, initial service is being provided by Station 12, Inmarsat's Dutch signatory, via the Inmarsat satellites. One-way service is currently available with two-way service planned for 1997.

Calian has designed the DMR-100 to operate with this service. The DMR-100 supports both data broadcast applications and personal messaging. With a link margin of over 16 dB, it can receive the signal in urban environments and operates from some indoor locations. The DMR-100 can be used to receive general information, such as stolen credit card numbers, weather bulletins, news or financial data, or it can be used as the basis for short messaging applications.

Presently, one-way coverage is offered in three ocean regions: Atlantic East, Atlantic West, and the Indian Ocean (see coverage map in Figure 4). The first application of the DMR-100 is the broadcasting of financial information, including stock prices and currency rates. This service is being launched by IFX (Denmark) and will ensure that foreign exchange dealers across the globe receive the latest rates for currencies and other rates as required. Information generated by IFX is sent over a land-line to the LES where it is then broadcast over the Inmarsat satellites to the DMR-100 for display on a personal computer screen (see Figure 3).

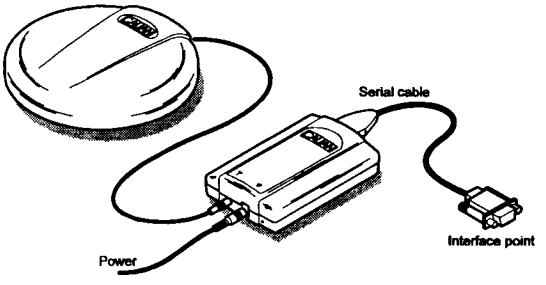


Figure 1 Calian DMR-100 Inmarsat D Terminal

SYSTEM DESCRIPTION

The Inmarsat D service was designed by Inmarsat and has two main applications: 1) personal messaging and 2) low bit rate data broadcasting. It can also be used for some SCADA applications for remote contact closure The information rate in the forward direction is less than ten bits per second, allowing for a very large link margin for penetration into urban environments and for some indoor applications with reasonable line-of-sight (see Link Budget in Table 1).

		INM-2	INM-3
LES EIRP (at edge of coverage)	[dBW]	55.4	54.4
Path Loss (5° elevation)	[dB]	200.9	200.9
Absorption Loss	[dB]	0.5	0.5
Satellite G/T	[dB/K]	-10.5	-10.5
Mean Uplink C/No	[dBHz]	72.1	71.1
Mean Satellite C/Io	[dBHz]	66.6	67.1
Transponder Gain	[dB]	166.0	167.0
Satellite Mean EIRP	[dBW]	20.0	
Downlink Losses	[dB]	189.9	
Receiver G/T	[dB/K]	-25.1	
Overall C/No	[dBHz]	33.6	
Required Theor. C/No ($BER<10^{-4}$)	[dBHz]	15.5	
Link Margin	[dB]	16.6	

Table 1 Worst Case Line-of-Sight Link Budget

Service Provider Interface

For data broadcasting applications, the information supplier is connected to the LES via a land-line, such as X.25, or in future, via the Internet. The data is broadcast to any number of terminals, each assigned to the same Group Identification Number. Each terminal can subscribe to up to three different Groups at a time.

In the case of global personal messaging, the service is modeled on terrestrial paging systems. The Caller submits a message to a Service Provider addressed to a unique Personal Identification Number (similar to a CapCode). The message may be submitted in one of several ways, depending on the Service Provider and may be Tone Only, Numeric, or Alpha-Numeric. The Caller may select one or more ocean regions or areas of spot beam coverage or, alternatively, the subscriber may inform the Service Provider of his whereabouts in advance. All aspects of the customer interface are handled by the Service Provider. Figure 2 shows the key players in the Service and their interaction.

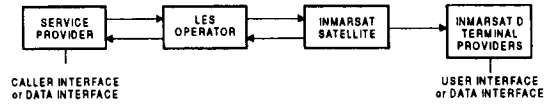


Figure 2 Key Players

Inmarsat Satellites

The message or the data is sent from the Service provider to the LES. The LES transmits the uplink signal to the Inmarsat satellite at C-Band which is then transmitted to the terminals at L-Band. The third generation satellites have up to five spot beams in addition to the global beam; however, with the second generation satellites, only global beam operation from each of the four satellites is possible. Inmarsat-D traffic can be transmitted in one spot beam only or in any combination of spot and global beams from up to four different satellites.

Inmarsat D Terminals

The DMR-100 is a small, portable unit which receives Inmarsat-D signals directly from the satellite at L-Band. It is assigned one Personal ID and up to three Group IDs enabling it to recognize and store messages and/or data addressed to it. These are subsequently

forwarded to a host over an RS-232 interface. A signal strength indicator allows the user to position the terminal in a location with adequate signal strength when inside a building. The terminal can be transportable for either indoor or outdoor use or vehicle based.

After power up, the DMR-100 will first tune to a fixed known frequency carrying the bulletin board information which directs the terminal to its specific traffic frequency. The traffic channel carries all the messages directed to the terminal. This dynamic allocation of traffic channels allows optimum configuration of global and spot beam coverage to be maintained. The terminal will remain locked on to the traffic channel frequency to receive messages. Synthesized in 2.5 kHz steps, the terminal is able to receive the full range of frequencies from 1525 MHz to 1559 MHz.

The modulation format is 32-ary FSK with 20 Hz tone spacing. The terminal is able to acquire with frequency offsets of up to +/- 1.6 kHz using high speed digital techniques for coarse and then fine acquisition. Frequency offsets can be stored and remembered for subsequent acquisitions.

The DMR-100 terminal for the Inmarsat D baseline service is a receive-only device. The terminal design is based on the Inmarsat D Satellite Definition Manual (SDM). It was Type Approved by Inmarsat in September of 1996.

In the future, the terminal will also include a transmitter section in order to provide a two-way acknowledgement and short message facilities. A two-way terminal is presently under development at Calian Communications.

PRODUCT DESCRIPTION

The Calian DMR-100 is the smallest and lightest Inmarsat terminal available today and weighs less than 300g, including the omni-directional antenna, 12.8cm in diameter and 2.5cm high and the receiver unit, 10cm x 6.5cm x 2cm. The unit will operate both outdoors in a mobile application or indoors under reasonable line-of-sight conditions.

The antenna has a G/T of -22.5 dB/K over 20 to 90° elevation angles. A block diagram for the DMR-100 is shown in Figure 5.

CONCLUSIONS

Calian Communications has successfully designed the first terminal to be Type Approved and used with Inmarsat D, a global messaging and data broadcast service.

With a link margin of over 16 dB, and a novel modulation scheme, message reception has been successfully demonstrated in many urban environments and from some indoor locations with Calian's DMR-100 terminal.

ACKNOWLEDGEMENTS

Calian would like to acknowledge the entire Inmarsat D Development Team at Inmarsat for providing the baseline design and continuing support for the Inmarsat D program. We would also like to thank the teams at Station 12 in the Netherlands and at IFX in Denmark.

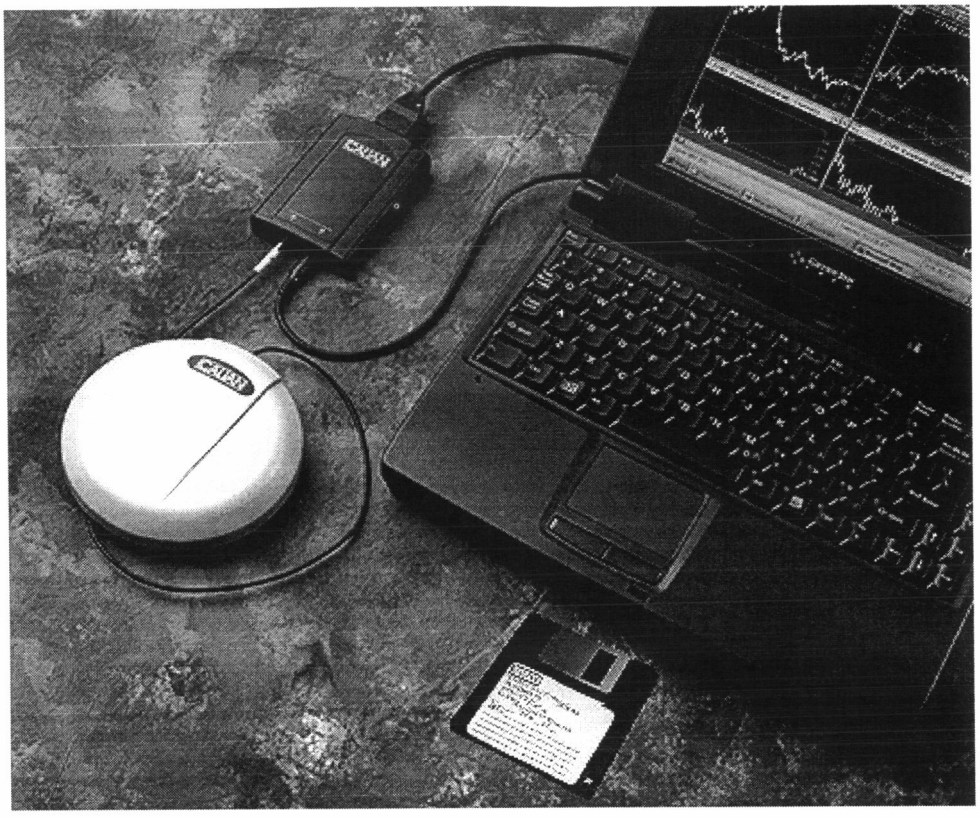


Figure 3 Terminal Configuration for Data Broadcast Applications

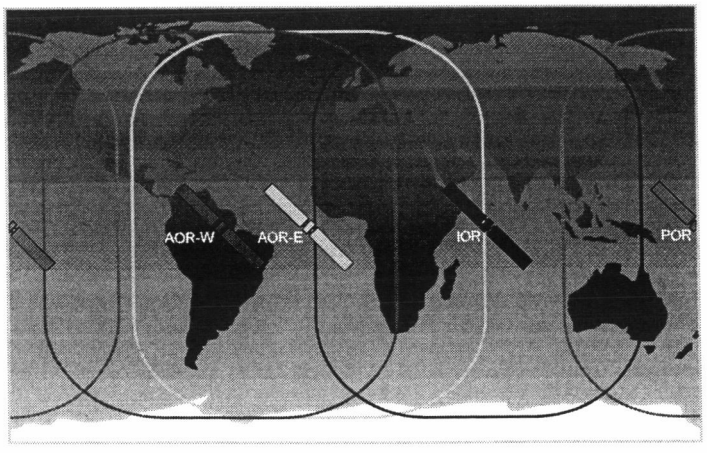


Figure 4 Coverage Map

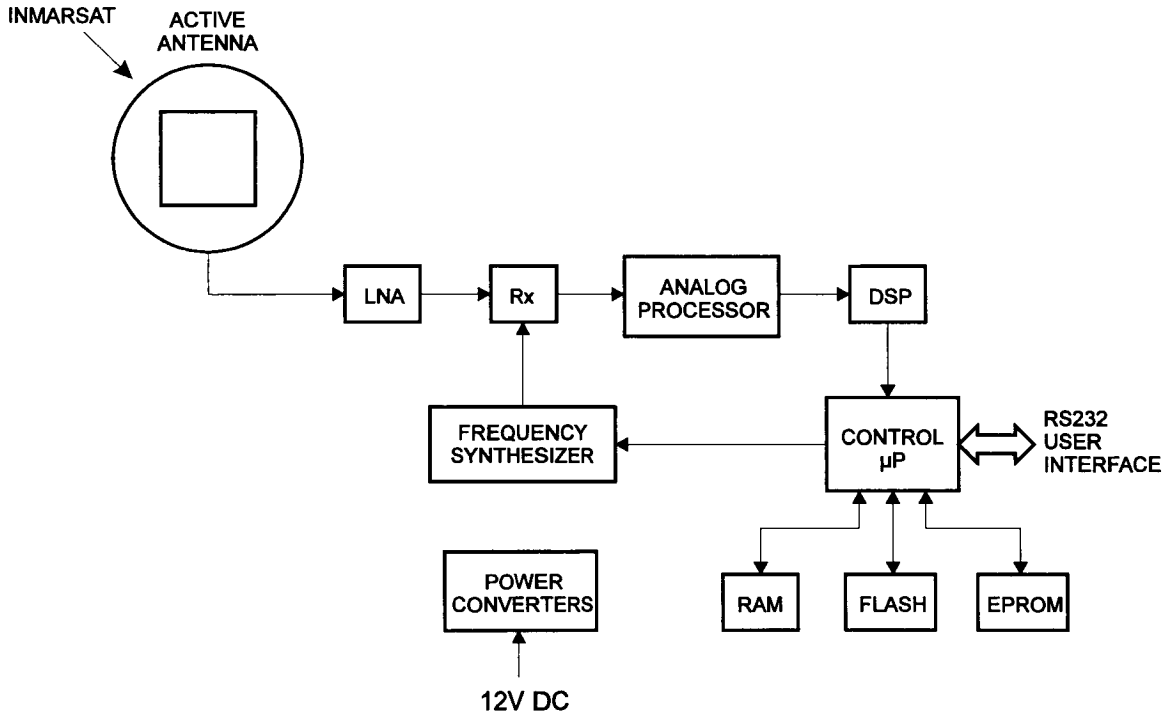


Figure 5 DMR-100 Block Diagram

A Ku/C-Band Portable CDMA Satcom Terminal for Personal Telephony

Harvey Chalmers, Ajit Shenoy, David Haschart, Zhangyi Wu, Ernest Ekelman, and Eric Kohls
COMSAT Laboratories, 22300 COMSAT Drive, Clarksburg, MD 20871, USA
Phone: +1-301-428-4000, Fax: +1-301-428-4534, e-mail: harvey.chalmers@ctd.comsat.com

Alain Zarembowitch
Mobile Satellite Services, 9055 Comprint Court, Suite 230, Gaithersburg, MD 20877, USA
Phone: +1-301-869-5900, Fax: +1-301-869-1478, e-mail: az@mobile-sat.com

Abstract

The system design and performance of a Ku/C-band spread spectrum terminal that employs code-division multiple access (CDMA) star or mesh network topologies for personal portable or maritime-mobile telephony is described. The use of spread spectrum technology dramatically reduces the size of the antenna and terminal compared to conventional VSAT Ku/C-band terminals. The CDMA terminal antenna ranges from 0.3 m (Ku-band) to 0.8 to 1.2 m (C-band). A unique forward and return link system architecture achieves several important overall system objectives, including network architecture capacity scalable in 1-MHz bandwidth steps, ease of frequency coordination with carriers in the same or adjacent satellites, and compliance with ITU-R regulations for off-axis emissions. The system employs widely available commercial digital signal processors and a GSM analog chipset, resulting in a low-cost user terminal.

Introduction

COMSAT Laboratories' new code-division multiple access (CDMA) portable satellite terminal uses spread spectrum technology to significantly reduce antenna and terminal size compared to conventional flyaway Ku/C-band satellite terminals. The entire terminal, including the antenna and power supply, is packaged in a case similar to a standard-size office briefcase, as shown in Figure 1.

In essence, the CDMA terminal combines the portability of some L-band briefcase satellite terminals with the advantages enjoyed by fixed VSATs in terms of lower cost Ku- or C-band space segment. When paired with a small, 0.36-m Ku-band dish or a 0.3-m-square flat plate antenna, the system is particularly well suited to complementing and extending the national telephone networks in underdeveloped or developing countries, using international regional or domestic Ku- or C-band satellites. Due to its low space segment cost, the CDMA terminal can serve both short-term (portable) and long-term (rural village) users.

The portability and range of communications services provided by the CDMA terminals can support a broad range of civilian and government applications, including:

- Switched telephony to national networks [public switched telephone network (PSTN) extension]
- Fax transmission
- News dissemination/collection
- Distress and emergency messages
- Lifeline telecommunications services.

The transmit power required for basic voice and data ranges from 50 mW to 2 W, depending on the space segment selected. Since the design is modular, the baseband electronics can be mated to a separable and remotely located Ku/C-band front end if desired.

System Overview

Currently, the most prevalent term nals available for portable satcom users are those designed for operation in the Inmarsat system. This is an L-band service with very large intersatellite spacing (≥30°), which easily allows small-aperture L-band antennas with 3-dB beamwidths of ±6° to be used without excessive interference to the adjacent L-band satellites.

For Ku- and C-band applications, however, intersatellite spacing is generally 3° and sometimes even 2°. Since 30-cm

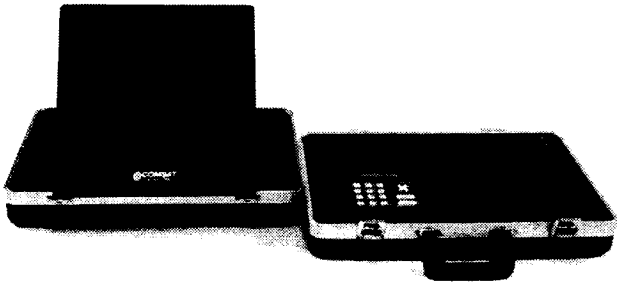


Figure 1. COMSAT's CDMA Terminal

Ku-band antennas have a 3-dB beamwidth of ±2°, significant interference to and from adjacent satellites will occur which will violate ITU-R and U.S. FCC limits on off-axis energy density. This problem can be solved by using spread spectrum technology, which spreads the transmitted energy over a sufficiently wide band to permit the use of small Ku- and C-band antennas while remaining within the ITU-R off-axis energy density limits.

Network Topologies

Depending on the network size, user privacy, and antenna size desired, network topologies for the CDMA portable system include:

- A star configuration (see Figure 2) with a dedicated hub station that provides interfaces to the PSTN.
- A full mesh configuration that allows each terminal to communicate directly with another terminal in the network, via the hub.

It is envisioned that a star configuration would be the most desirable initially, giving remote users immediate connectivity to the PSTN. However, as the terminal population grows, users will want to communicate directly with other remote terminals, creating the need for a full mesh configuration. Ultimately, the network must provide connectivity in either configuration. In both cases, the hub serves as the central point for call establishment, takedown, and the archiving of call statistics for billing.

The inclusion of mesh connectivity significantly affects the satellite link budget, since transmission between two antennas with apertures less than 1 m necessitates the use of higher transmit power in both terminals. Again, spread spectrum, combined with CDMA, makes such a configuration relatively straightforward because it mitigates the even higher off-axis EIRP from the small-aperture antennas.

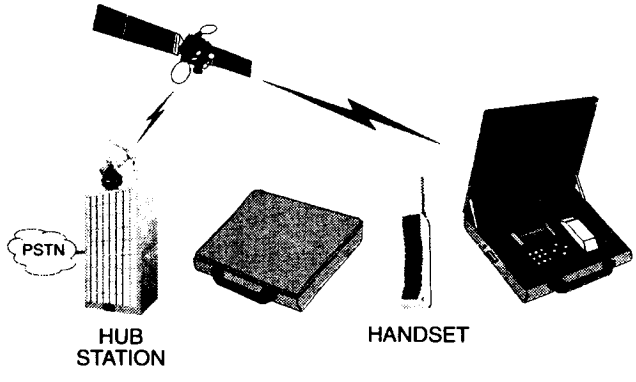


Figure 2. CDMA Terminal System Concept

Transmission Scheme

For thin-route applications it is also necessary to provide a scalable system that can readily adapt to a changing user population while permitting efficient use of the space segment as the system grows. Wideband CDMA (≥10 MHz) is therefore a poor solution, since large satellite bandwidth is required even for a minimal system. Consequently, a narrowband (1-MHz) CDMA/frequency-division multiple access (FDMA) concept is proposed for inbound transmission from the terminal. For transmissions from the hub, in both star configuration traffic channels and star/mesh signaling channels, CDMA is unnecessary due to the existence of a large-aperture antenna that provides emissions below the off-axis limits. Thus, for outbound transmissions, a time-division multiplexing (TDM)/FDMA approach will greatly simplify the receiver design.

With the TDM/CDMA/FDMA transmission scheme, between 6 and 24 full-duplex calls can be supported per megahertz, depending on the satellite type, terminal EIRP, and level of acceptable self-interference. Frequency planning is straightforward, since capacity can be allocated in (not necessarily contiguous) 1-MHz segments. A low-bandwidth outbound signal was deliberately chosen to permit the use of very low-cost GSM handset analog front-end chips and filters in the terminal receiver.

Figure 3 is a block diagram of the hub, and Figure 4 is a block diagram of the remote terminal. Major features of the terminal include:

- Small Ku/C-band dish or flat plate antenna
- High-quality digitized voice at 4,800 b/s
- Multirate data port up to 9,600 b/s
- Group-3 facsimile service at 2,400 b/s
- Auto-ringback for channel availability
- Sleep/power-down mode
- Auditory/visual call progress feedback, call timer, and menu-based operation
- Operation with INTELSAT or domestic satellites
- Easily transported and installed

CDMA vs SCPC Tradeoffs

Figure 5 shows the tradeoff between the number of supportable Ku-band carriers and antenna size for a typical 12-m hub and an 18-MHz INTELSAT VII transponder. For truly *portable* (not "transportable" or "luggable") satellite terminals, only diameters of less than 1 m are considered. Using traditional single channel per carrier (SCPC) access methods with such small-diameter antennas would allow only *one* carrier to be supported due to the inbound

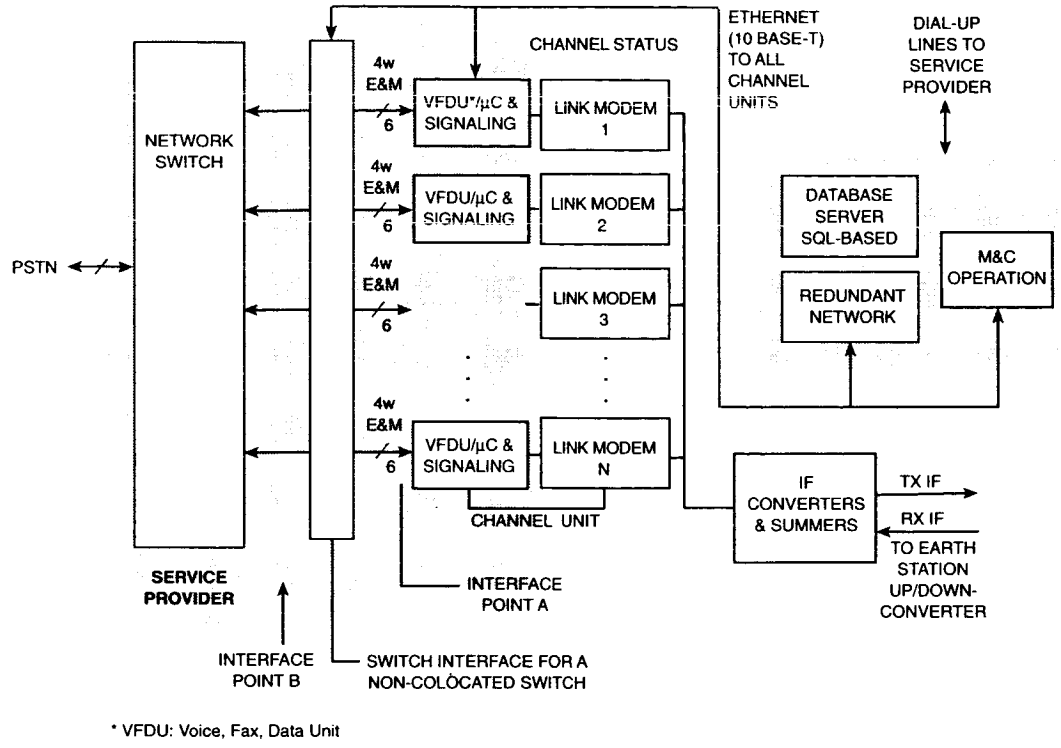


Figure 3. CDMA Network Hub Block Diagram

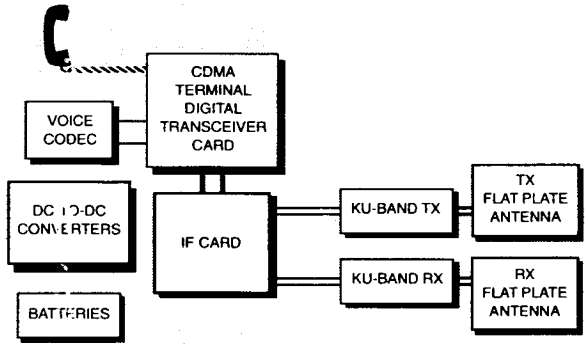


Figure 4. CDMA Terminal Block Diagram

off-axis interference resulting from the high carrier power concentrated in the narrow band, and the sidelobe gain of the small dish. Traditional SCPC methods require a 1.0-m dish in order to support 50 carriers per 18-MHz transponder, while the use of TDM/CDMA allows at least 100 carriers to be supported with 0.5-m dishes. Link budgets for a typical 1-MHz INTELSAT VII transponder lease and a 0.3-m-square Ku-band antenna require a satellite EIRP of 20.1 dBW to the terminal and -6.5 dBW to a 5.5-m hub, with a terminal-radiated SSPA power of only 400 mW. The hub-radiated power is also low (500 mW). The overall link is downlink-power-limited from the satellite to the terminal.

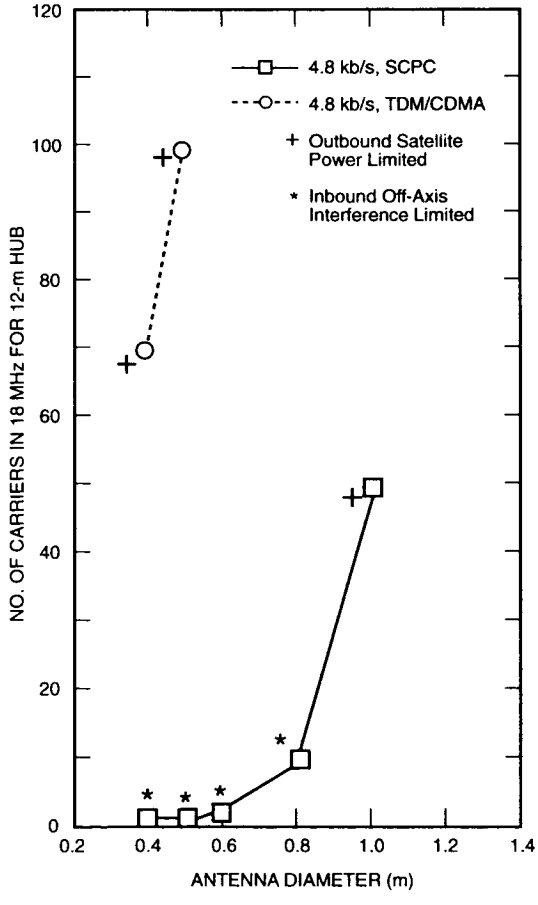


Figure 5. CDMA vs SCPC Tradeoffs

The proposed TDM/CDMA access method is optimized to maximize the number of carriers when the antenna is very small, and thus is perfectly suited for small portable satellite terminals.

TERMINAL ARCHITECTURE

Conceptually, the CDMA terminal can be broken down into two sections: one containing digital components, and the other comprising analog devices.

The Digital Section

The digital transceiver includes the microprocessor controller, which performs all control and signaling functions, including providing a multilevel, user-friendly menu on a 4 × 20 LCD display; downloading configuration parameters to all subsystems; implementing call setup and takedown protocols; performing RF power/frequency control; performing low-level data processing functions; monitoring the status of all modem functional blocks; controlling modem self-tests; scanning the keypad and hook-switch; and managing over-the-air software updates.

The modem is implemented digitally using digital signal processing (DSP) devices operating in tandem. To simplify transmit signal pulse-shaping, the modulator is based largely on a lookup table. The demodulator accepts a digitized baseband signal from the IF board and performs the functions of antenna signal acquisition, carrier acquisition and tracking, continuous-mode binary phase shift keying (BPSK) demodulation and data detection, automatic gain control (AGC) and automatic frequency control (AFC), and forward error correction (FEC) decoding.

The voice codec selected for this application is a high-quality 4,800 b/s digital vocoder. Its subjective quality has been found to be equivalent to the IS-54 8-kb/s digital cellular voice codec standard. The codec employs advanced voice coding and built-in error correction algorithms that can sustain high-quality operation under bit error ratios (BERs) as high as 1×10^{-2} (1 percent).

For facsimile transmission the system would operate at 2,400 b/s, allowing another 3-dB margin and yielding BERs less than 10^{-6}, which is ideal for fax operation. Note that higher data and facsimile rates (9,600 b/s) are also possible, depending on the solid-state power amplifier (SSPA) used.

The Analog Section

The analog section of the CDMA terminal contains the IF, RF, and Ku/C-band antenna modules.

IF Board. The IF board performs channel selection and signal conversion from baseband to L-band. It also provides the frequency reference to the RF modules.

RF Modules. The RF modules perform signal conversion to/from L-band and Ku/C-band. Phase-locked oscillators, locked to the IF reference, provide a convenient means of frequency control. A small correction in the reference is all that is necessary to accomplish a large correction at Ku-band frequencies.

Ku-Band Flat Plate Antenna. A low-cost, high-efficiency flat panel array antenna has been developed for consumer applications such as high-power direct broadcast satellite (DBS) and personal satellite communications (PSC) operations. For fixed installations, the flat plate antenna is easy to install on either the side or roof of a building and, in regions where restrictive covenants exist, it can be made less obtrusive than a reflector. For portable units, the flat plate antenna offers a lightweight assembly with a form factor amenable to integration within the lid of a briefcase.

The antenna is fabricated as a multilayer structure consisting of radiating slot elements and incorporating feed networks printed on low-cost metallized films. The configuration and number of layers vary, depending on the system requirements for the number and sense of polarization of electromagnetic radiation. The conducting, metallized film layers are suspended between nonconducting, low-cost foam substrates, with no direct contact between the radiating elements and the feed network. Excitation of the radiating elements is accomplished via electromagnetic coupling. This is one of the factors that contributes to the low manufacturing cost of the antenna compared to other printed-circuit antenna technologies.

C-Band Optimized Feed. A key feature of the CDMA terminal is its compactness, which facilitates portability, minimizes the installation space required, and keeps the terminal cost-effective. An axis-symmetric, prime-fed parabolic reflector antenna is the most desirable configuration at C-band. Due to reduced gain-to-noise temperature ratios *(G/T)* for 80-cm dishes, the feed element of the antenna is a critical factor in antenna performance. The use of C-band waveguide devices for the feed system is attractive due to the resulting low loss; however, such devices are large and present a significant blockage problem for a small dish. Printed-circuit devices provide reduced size, but with increased loss.

A spiral helix printed on a cone was designed and fabricated which judiciously combines the needed mechanical traits with desirable RF electrical parameters. The spiral is supported by a coaxial boom, which provides a 6-GHz coaxial cable to feed the transmit energy to the back-

radiating spiral, as well as a return path for the 4-GHz receive energy coupled into the forward-radiating spiral. The spiral is designed to provide one sense of circular polarization for the transmit, and the opposite sense for receive, both with an illumination function that is quite good for the reflector over both frequency bands. The spiral also facilitates mounting the receive low-noise block converter (LNB) behind the base, thus minimizing loss before the LNB and resulting in a good *G/T*.

HUB ARCHITECTURE

The hub earth station equipment for the CDMA terminal system is assumed to interface to an existing earth station facility at a 70-MHz IF. Major elements of the equipment include the channel units, network management system (NMS) hardware, and NMS software. Figure 6 is a photograph of the hub equipment.

The Channel Units

Each channel unit card performs the TDM transmit and time-shared multicarrier CDMA receive functions. A channel unit is designed to accommodate up to six simultaneous CDMA channels. Thus, for every increment of 1 MHz, the service provider will add one more channel unit card (if there are six users per MHz) or four more cards (if there are 24 users per MHz). The PN sequences assigned to the channel unit(s) despreaders serving the same 1-MHz frequency slice are the same for all channel units. Their differentiation is inherent in that the return links are frequency-division multiplexed. Thus, each channel unit is pre-tuned to a different 1-MHz slice of the transponder bandwidth.

An innovative channel unit digital design and local area network (LAN) connectivity, coupled with Internet Protocol (IP) addressing, allows a fully digital voice path with direct E1/T1 outputs, echo cancellation (performed on the channel unit without the need for external devices), and a direct interface to the Internet, if desired.

NMS Hardware

For configurations of up to 90 channels, the entire NMS/hub is confined to a single standard 19-in. rack. The rack contains general-purpose computers, a fully digital telephone switch, and a universal chassis for housing the channel units, frequency generation unit, channel bank, and other specialty modules. All modules are hot-swappable, so units can be removed and inserted without any service interruption.

The telephone switch provides full interconnectivity between the PSTN and the satellite networks. Multiple telephone interfaces and protocols are supported, including E1, T1, DTMF, and R2D signaling. The telephone switch also provides recorded voice announcements and tone generation and detection. The associated channel bank extends the telephone interfaces to two-wire (for the local telephone operator) and four-wire E&M for direct PBX interfaces.

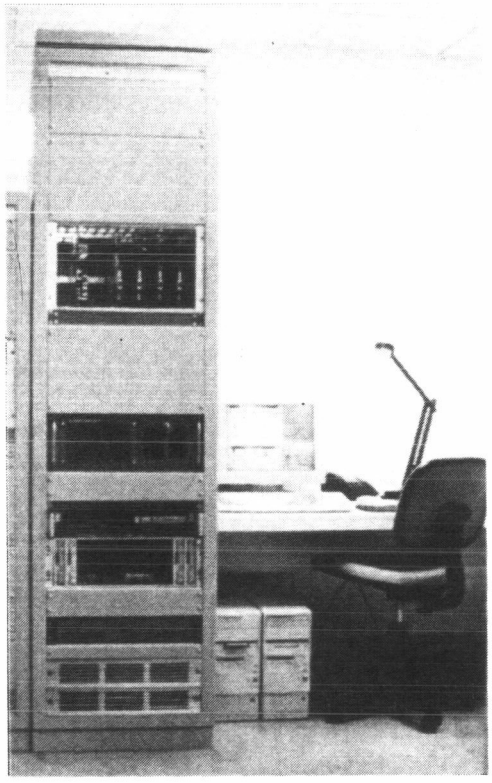


Figure 6. The CDMA Hub Rack and Operator Console

The frequency generation unit provides a highly stable, low-phase-noise frequency reference to the entire satellite network. The unit is locked onto Global Positioning System (GPS) time to eliminate the need for frequency calibration.

NMS Software

A distributed computational architecture is used for the NMS/hub. The overall NMS/hub functionality is split between several processes running on multiple computers connected by an Ethernet LAN. All computers are Pentiums running Windows NT 4.0 in a secure environment.

At the core of the NMS/hub is a Structured Query Language (SQL) database server which stores all pertinent information: the terminal authorization table, satellite configuration, equipment configuration, call record (for billing information), and event record (for alarms). This information can be consulted locally or remotely by using a user-friendly

human-machine interface. As the exclusive container for NMS/hub information, the database is the only file that needs to be backed up regularly. The database server features mirrored disk drives and replication to protect valuable information against hardware failure. The system administrator can tighten information security by customizing Select/Insert/Update/Delete access individually for each database table.

Each module and process reports its status periodically over the LAN. Automatic failure reporting, failure detection, redundancy switchover, and alarming ensure the highest possible network availability, whether the NMS/hub is operated in an attended or unattended environment.

Field Trials

Proof-of-concept 80-cm C-band spread spectrum terminals have been successfully designed, constructed, and demonstrated in live satellite loopback at COMSAT Laboratories. Following INTELSAT certification under Standard-Gx, the 80-cm CDMA terminal was tested over an INTELSAT V inclined-orbit satellite for a period of 8 weeks, during which time extensive BER data tests and voice tests were conducted. The modem implementation loss was only 1 dB from theory at a BER of 10^{-3}. Voice tests were conducted in a variety of languages, with male and female speakers, over three sets of link conditions: high-margin (BER < 10^{-6}), normal (BER = 5×10^{-3}), and highly degraded (BER = 2×10^{-2}). Excellent voice performance was reported by the test subjects.

Other Terminal Variations

Several packaging options for the CDMA terminal allow at least three broad-based applications:

- *Portable.* A portable Ku-band personal satellite terminal for voice, data, and fax. This option provides a low-cost satellite phone for government, civilian, and commercial personnel.
- *Village.* A very low-cost Ku/C-band single- or multi-line telephone terminal for remote villages. This fixed option provides national telephone connectivity to rural and remote coastal, mountain, or desert villages.
- *Maritime/Railroad.* A low cost C-band voice, data, and fax mobile terminal with a motion-stabilized antenna for shipboard and railroad users. This version of the CDMA terminal provides a low-cost telephony, data, and facsimile service into merchant marine vessels, oil rigs, pleasure yachts, and passenger railroads.

Conclusions

COMSAT Laboratories' new Ku/C-band CDMA portable and maritime/railroad-mobile terminals offer a low-cost and convenient alternative to existing flyaway C- or Ku-band terminals. The implementation of spread spectrum technology allows the use of dramatically smaller antennas (0.8 m instead of 2.4 m at C-band, for example). The star/mesh system design and inherent flexibility allow network growth and compatibility with almost any space segment and existing gateway earth station.

With lower Ku- and C-band space segment costs, space segment and user per-minute charges are expected to be significantly lower than with other satellite systems. With a small, low-cost personal terminal, the proposed personal voice, data, and fax services are cost-competitive when compared with other two-way voice flyaway or fixed Ku/C-band VSAT systems, offering tremendous opportunities to serve a variety of customer needs and applications in many developing and Third World countries.

ACTS Mobile Terminals

Brian S. Abbe, Martin J. Agan, Thomas C. Jedrey
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive
Pasadena, California, USA 91109
Phone: (818) 354-3887, Fax: (818) 354-6825, email: Brian.S.Abbe@jpl.nasa.gov

ABSTRACT

The development of the Advanced Communications Technology Satellite (ACTS) Mobile Terminal (AMT) and its follow-on, the Broadband Aeronautical Terminal (BAT), have provided an excellent testbed for the evaluation of K- and Ka-band mobile satellite communications systems. An overview of both of these terminals is presented in this paper.

INTRODUCTION

Since the early 1980's, NASA, through the Jet Propulsion Laboratory (JPL), has been involved in the development and demonstration of system concepts and high risk technologies to enable the introduction of commercial mobile satellite services. Initial efforts focused primarily on L-band [1], with the focus turning to K- and Ka-band in the early 90's. K- and Ka-bands have outstanding potential for higher data rate communications.

The AMT is a fully tested, mobile satellite communications terminal that has been used in a variety of application specific experiments and demonstrations since December 1993 [2]. Development of a follow-on higher data rate BAT, has recently been completed and numerous experiments and demonstrations are already under way. Details relating to the design and development of the AMT and BAT are presented in this paper.

ACTS MOBILE TERMINAL

The AMT incorporates system and subsystem solutions that have been devised to overcome the challenges of K- and Ka-band mobile satellite communications. The operational configuration utilized by the AMT is shown in Figure 1. A fixed station communicates through the satellite with a mobile terminal. An unmodulated pilot signal and a data signal are transmitted in the forward direction from the fixed station to the mobile terminal through ACTS [3]. The pilot is used by the mobile terminal to aid antenna tracking, and as a frequency reference for Doppler offset correction and pre-compensation. In the return direction only the data signal is transmitted.

Transmissions from both terminals occur in the band 29.634 GHz +/-150 MHz. The satellite is operated in a bent pipe mode, and transmissions from the satellite are in the band 19.914 GHz +/- 150 MHz. The operational data rates that the terminal was initially designed for range from 2.4 to 64 kbps. During the experiments, bi-directional data rates of 768 kbps have been achieved.

The block diagram of the mobile terminal is provided in Figure 2. The architecture of the fixed station is identical to that of the mobile terminal with the primary exception of the antenna and the power amplifier. The mobile terminal utilizes a low-profile, tracking elliptical reflector antenna, while the fixed station has a stationary, high gain, 2.4 meter parabolic dish. For power amplification, the mobile terminal contains a 10 Watt traveling wave tube amplifier (TWTA)[1], and the fixed terminal a 65 Watt TWTA. Descriptions of the subsystems that comprise the mobile and fixed terminals are presented in the remainder of this section.

Three different data rates are supported by the speech codec: 9.6 kbps (Motorola MRELP), 4.8 kbps (U.S. Government standard CELP), and 2.4 kbps (U.S. Government standard LPC-10). The voice quality ranges from near toll to communications quality, and generally improves

[1] The baseline terminal design calls for a 1W power amplifier.

monotonically as the data rate increases. The operational Bit Error Rate (BER) is specified to be less than or equal to 10^{-3}, providing little or no perceptible degradation in voice quality. Re-synchronization from outages due to shadowing occur within 350 milliseconds.

The terminal controller has four main functions: (1) coordinate the AMT subsystems, (2) direct the point-to-point communications operation of the link, (3) provide status and terminal information to a data acquisition system for recording, and (4) provide a user interface. It also contains the algorithms that translate the rain compensation algorithm [4] and communications protocol [5] into the operational procedures and interfaces to the various terminal subsystems.

An all digital low rate modem suitable for use in the land-mobile satellite communications channel was developed for use with the terminal. The modem design was simplified by the shadowed Rician channel model with K=30 (due to the antenna directivities).

This modem is a multi-rate DPSK modem that transmits and receives at 2.4/4.8/9.6 and 64 kbps. The coding used is a rate 1/2, constraint length 7 convolutional code. A key feature of the demodulator is the algorithm used to estimate and correct for the received frequency offset. This algorithm has been designed to handle frequency uncertainties on the order of 10 kHz changing at a maximum rate of 350 Hz per second [6]. The performance specification of the modem is a bit error rate of 10^{-3} at an E_b/N_o of 7.0 dB in an AWGN environment.

The transmitter converts the modem output at a 70 MHz IF to the transmit frequency centered at 29.634 GHz. It also provides the necessary power amplification for the antenna. The receiver translates the received signal centered at 19.914 GHz to a 70 MHz IF that is input to the modem.

Key functions of the receiver are pilot tracking and Doppler precompensation. The downconverted pilot is tracked in a dual bandwidth phase lock loop and may be used as a frequency reference in the mobile terminal. The output of the phase lock loop is also used to generate a reference for the upconversion stage to precompensate the transmitted signal from the mobile terminal for the estimated Doppler on the return link.

A small reflector antenna [7] to be used in conjunction with a separate high power amplifier, low noise amplifier, and diplexer was the primary antenna developed and used in the land-mobile experiments. The antenna subsystem satisfies the following specifications: (1) initial satellite acquisition occurs within 10 seconds; (2) the minimum gain specifications are met for vehicle turn rates up to 45°/second and pitch and roll up to +/-6°; (3) the tracking system maintains point during signal outages of up to 10 seconds; (4) the transmit gain is a minimum of 20.1 dBi over a 12° elevation beamwidth; (5) the receive gain is a minimum of 18.8 dBi over a 12° elevation beamwidth; and (6), the minimum receive system G/T is better than -6 dB/K over the elevation beamwidth. The antenna has minimal cross-polarization and sidelobe levels no greater than -15 dB. The antenna resides inside an ellipsoidal water-repelling radome of outside diameter approximately 8 inches (at the base) and maximum height of approximately 4 inches. A picture showing the reflector antenna is presented in Figure 3.

The initial characterization of the AMT performance was accomplished through a series of BER tests. Data of this nature was collected for all three of the lower operational data rates. BER values were recorded in the range from 10^{-5} to 10^{-1}. The main specification that this terminal had to meet was a BER of 10^{-3} at an E_b/N_o of 9.0 dB or better. This BER specification was achieved at 6.5 dB, 6.5 dB, and 8.5 (E_b/N_o) dB for 9.6 kbps, 4.8 kbps, and 2.4 kbps, respectively. The significantly higher BER specification for the 2.4 kbps case is due to the performance of the Doppler estimation and compensation algorithm in the modem.

Following baseline terminal performance characterization, the mobile terminal has been installed in multiple land-mobile platforms. The terminal has traveled throughout the continental United States and been used to perform various experiments ranging from basic land-mobile to personal, to control of a robotic vehicle, where the communications link using the point-to-point protocol (PPP) provided an Ethernet extension between the robotic vehicle and a base station.

BROADBAND AERONAUTICAL TERMINAL

With several modifications to the AMT design, much higher rate communication capabilities could be achieved for operation in the aeronautical environment. The resultant terminal, the BAT, provides full-duplex communications up to a data rate of 1.544 Mbps with a fully tracking antenna.

The system configuration for aeronautical operations is shown in Figure 4. The operation of this terminal differs slightly from that of the AMT. While the terminal is capable of using one of the high gain spot beams of ACTS [4], the primary satellite antenna used is the lower gain steerable beam. To allow the satellite antenna to be used to track the aircraft, position information from the aircraft is multiplexed in with the data on the return link, and forwarded to the satellite control center. As the aircraft traverses the spot for the steerable beam, the satellite control center will position the steerable beam to keep the aircraft within the 1 dB contour of the beam. The operational up/down frequency bands and data/pilot transmissions are identical to those used for the AMT.

Operating at higher data rates compared to the baseline AMT allows for the use of a coherent detection scheme. This is due to the fact that the high close-in carrier phase noise on the satellite has a negligible effect on coherent demodulation techniques at higher data rates than the baseline rates for the AMT. BPSK was chosen as the modulation with a coherent demodulator, and a concatenated coding scheme, a convolutional inner code (rate 1/2, constraint length 7) and a Reed-Solomon outer code (rate 239/256).

The demodulator loop parameters for this modem were optimized to track Doppler frequency offsets of up to 30 kHz, varying at 900 Hz per second. The synchronization algorithms (carrier, bit, and decoder) were also optimized to allow re-synchronization within one second of signal presence after loss of signal. Commquest Technologies, Incorporated modified a commercially available satellite modem to meet the specifications for the aeronautical environment. The maximum data rate for this modem is 2 Mbps. It achieves a BER of 10^{-6} at an E_b/N_o of approximately 3.0 dB (AWGN, laboratory testing).

The high gain aeronautical antenna employs an azimuth and elevation pointing system to allow it to track the satellite while the aircraft is maneuvering. The aeronautical antenna and radome have been developed by Electromagnetic Sciences (EMS) Technologies, Incorporated. The antenna design utilizes a slotted waveguide array, is mechanically steered in both azimuth and elevation, and is designed to enable mounting on a variety of aircraft. The radome has been shaped with a peak height of 6.7 inches and a 27.6 inch diameter. The radome was designed for low loss at both frequency bands of operation, and to withstand the aerodynamic loads on a jumbo jet. Antenna installation requires a 3.5 inch diameter protrusion into the fuselage to allow the necessary signals to pass to and from the antenna. A picture of the antenna is shown in Figure 5 illustrating both the slotted waveguide and the polarizer.

The antenna is capable of tracking a full 360° in azimuth and -5° to zenith (90°) in elevation. The antenna has a transmit gain of 30 dBi and a receive G/T of 0 dB/K. Circular polarization has been implemented[2], and there exists the capability to transmit up to 120 Watts of RF output power through the antenna. The actual dimensions of the combined transmit and receive array apertures are less than 16 inches wide, and less than 4.5 inches in height. The receive array has 161 elements with an ideal directivity of 30.4 dB at the band center. The receive 3 dB beamwidths are 4° in azimuth and 7.6° in elevation. The transmit array has 366 elements with an ideal directivity of 34.1 dB at the band center. The transmit 3 dB beamwidths are 2.6° in azimuth and 5.0° in elevation. Both arrays were designed to have maximum inband VSWR of 1.3:1 and the first sidelobe level at least 13 dB down from the peak.

The antenna tracking mechanism is required to maintain pointing within 0.5 dB of beam peak throughout all phases of flight. The antenna's narrowest beamwidth of 2.6° thus requires fine accuracy as the aircraft turns. The antenna positioner utilizes an elevation over azimuth

[2] ACTS uses linear polarizations.

mechanism, with a precision of a few hundredths of a degree. This positioner is controlled by a tracking algorithm that utilizes three sources of information: (1) a 3-axis inertial rate sensor, (2) the aircraft Inertial Navigation System (INS), and (3) the pilot signal strength feedback from a conical scanning of the beam. The rate sensor with a 50 Hz bandwidth and mounted on the main antenna assembly, provides the majority of the pointing information for the tracking system. A very low bandwidth (0.5 Hz), small displacement (0.5°) conical scan feedback control system is used to cancel, in the steady state, and continually adjust to any changes in the three axis inertial sensor offset and drift rates. The INS is used for satellite acquisition and correction of the rate sensor drift. The overall system accommodates tracking rates up to 60°/sec and 30°/sec^2 in azimuth, and 30°/sec and 15°/sec^2 in elevation.

The terminal has been successfully integrated into two different aircraft and has flown many flights [8], demonstrating the robustness of the terminal design. All operational aspects of the system, including the steerable beam tracking the aircraft, have worked well. As part of one experiment, a live demonstration of the system capabilities was held on U.S. National Public TV. In this demonstration, an Internet link to/from the ground was also integrated into the terminal.

SUMMARY

The development of ACTS, the AMT, and the BAT has proven to be an excellent proof-of-concept technology development for K- and Ka-band mobile satellite communications. Through this work, many advancements have been made in the area of K- and Ka-band mobile satellite communications.

REFERENCES

[1] Jedrey, T.C., Dessouky, K.I., Lay, N.E., "An Aeronautical-Mobile Satellite Experiment," IEEE Transactions on Vehicular Technology, Vol. 40, No. 4, pp. 741-749, November 1991.

[2] Abbe, B. and Pinck, D., "AMT Experiment Results," International Mobile Satellite Conference 1995 (IMSC '95), pp. 147-157, June 6-8, 1995, Ottawa, Canada.

[3] Gargione, F., et. al., "ACTS: Design and On-Orbit Performance Measurements", International Journal of Satellite Communications, Vol. 14, No. 3, pp. 133-173, May-June, 1996.

[4] Satorius, E. S., Tong, L., "Analysis of a Rain Compensation Algorithm for K/Ka-Band Communications," International Journal of Satellite Communications, Vol. 14, No. 3, pp. 297-311, May-June, 1996.

[5] Lay, N., Dessouky, K., "A Communications Protocol for Mobile Satellite Systems Affected by Rain Attenuation," IEEE Journal on Selected Areas in Communications, Vol. 10, No. 6, pp. 1037-1047, August 1992.

[6] Jedrey, T.C., Satorius, E.H., Agan, M.J., "A Frequency Offset Estimation and Compensation Algorithm for K/Ka-band Communications," International Journal of Satellite Communications, Vol. 14, No. 3, pp. 191-200, May-June, 1996.

[7] Densmore, A. C., Jamnejad, V., "A Satellite Tracking K- and Ka-Band Mobile Vehicle Antenna System," IEEE Trans. on Vehicular Technology, Vol. 42, No. 4, pp. 502-513, November 1993.

[8] Agan, M., Nakamura, D., Sternowski, R., Shameson, L., Sohn, P., "ACTS Aeronautical Experiments," International Journal of Satellite Communications, Vol. 14, No. 3, pp. 233-248, May-June, 1996.

ACKNOWLEDGMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

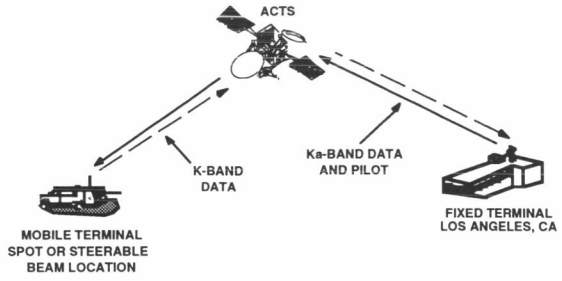


Figure 1 AMT Experiment Configuration

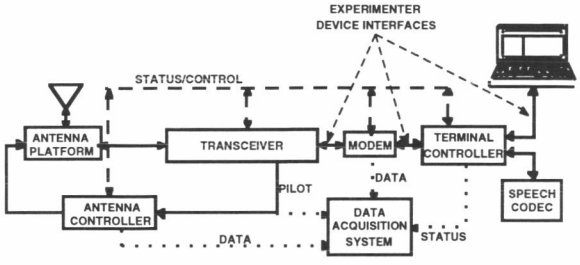


Figure 2 AMT Block Diagram

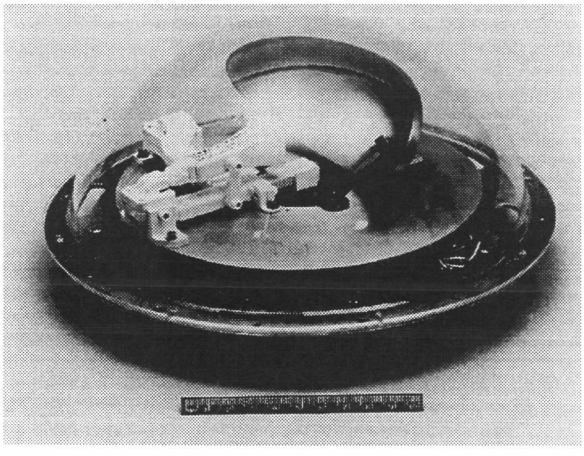


Figure 3 Small Reflector Antenna

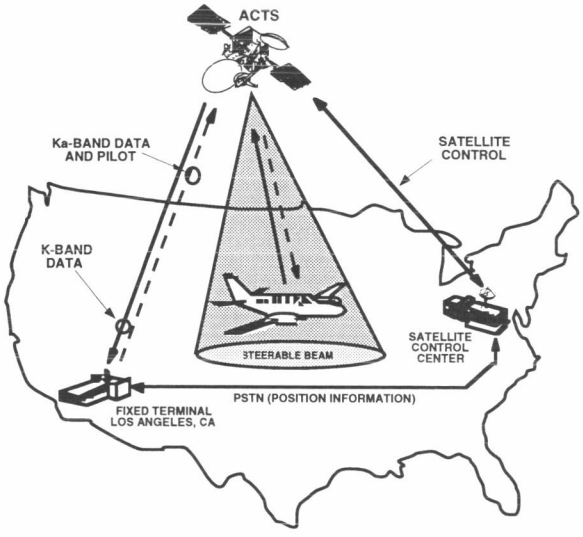


Figure 4 BAT Experiment Configuration

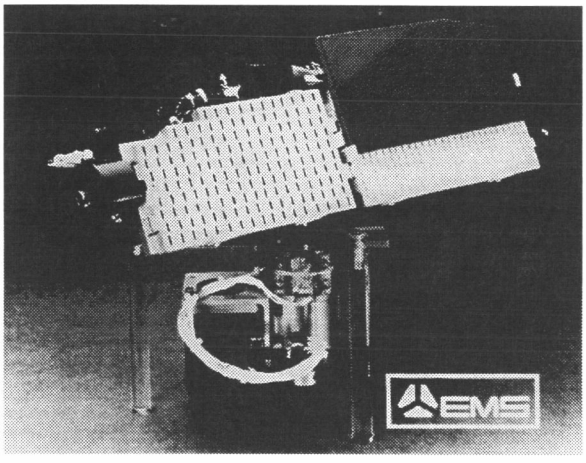


Figure 5 BAT Antenna

A Pointing Error Analysis of the ACTS Mobile Terminal

Michael Rice, Brian J. Mott, and Kevin D. Wise
Department of Electrical & Computer Engineering
Brigham Young University
Provo, UT 84602
Phone: 801 378 4469 FAX: 801 378 6586
email: mdr@ee.byu.edu

Abstract

The ACTS mobile terminal (AMT) developed by the Jet Propulsion Laboratory has been used to conduct experiments in land mobile satellite applications from 1991 to 1994. The AMT is equipped with a small (8" x 3") high gain reflector with a 3 dB beamwidth of 18 degrees in elevation and 12 degrees in azimuth. Satellite tracking was realized using mechanical steering in the azimuth plane only (the elevation angle was assumed fixed). While the narrow beam width rejected off-axis multipath reflections, received power fluctuations due to antenna pointing error were common. Theoretically, there should have been a less than 1 dB power fluctuation when the mobile terminal was level (due to the dithering associated with the tracking control). However, road surfaces, which are rarely level, induced pitch, roll, and yaw on the mobile terminal which lead to increased power fluctuations. A thorough analysis together with empirical road surface data showed that a 2 dB fluctuation is common.

Introduction

The ACTS Mobile Terminal (AMT) was developed by the Jet Propulsion Laboratory (JPL) as a test bed for proof-of-concept designs of K/Ka-band mobile satellite communication systems [1]. Pilot tone tests using the AMT were conducted to characterize the land mobile satellite channel at these frequencies. The system setup is illustrated in Figure 1 and a typical run is illustrated in Figure 2. The solid line in Figure 2 represents the one second average of the received pilot data while the dotted line identifies the maximum and minimum power levels received during each one second interval. Variations in the average power level are due to shadowing, multipath interference, thermal noise, and pointing error. The deep fades in the average signal power are due to shadowing while the short term variations indicated by the dotted lines are the result of thermal noise and multipath interference.

The interesting features here are the small changes in the average power on the order of 2 dB. We postulate that these variations are due to pointing errors. Usually, mobile satellite communications at L-band employ omni-directional antennas which do not require tracking systems and, as a result, incur no pointing errors. At K and Ka band, however, increased atmospheric and rain attenuation require higher antenna gains than provided by omni-directional antennas. These antennas necessarily have narrower beam patterns and thus require satellite signal tracking. The AMT is equipped with a small (8" x 3") high gain reflector with a 3 dB beamwidth of 18 degrees in elevation and 12 degrees in azimuth [2]. The antenna is mounted on a platform which tracks the satellite signal through 360° azimuth for a fixed elevation angle (46° for Southern California). Any pitch or roll imposed on the antenna platform through uneven or unlevel road surfaces induces a pointing error.

This paper describes an analytical procedure for predicting the azimuth and elevation pointing errors caused by changes in vehicle pitch, roll, and heading. These pointing errors are then coupled with the antenna gain pattern to generate an estimate of the loss in received power due to these effects. Comparisons with actual data recorded on the propagation runs confirms our assertion that pointing error is the cause of the power variations.

Pointing Error Calculation

The location of a ground station is usually specified by its latitude ϕ_g, longitude λ_g, and altitude h_g defining its location in spherical Earth Centered Fixed (ECF) coordinates. A geostationary satellite orbits the earth in the equatorial plane at an altitude of 35,786 km [3] and is specified by its longitude at the equator, or subsatellite point λ_s.

The ground station location in Cartesian ECF coordinates is

$$x_g = R_E \cos\phi_g \cos\lambda_g \quad (1)$$

$$y_g = R_E \cos\phi_g \sin\lambda_g \quad (2)$$

$$z_g = R_E \sin\phi_g, \quad (3)$$

while the satellite location in Cartesian ECF coordinates is

$$x_s = r\cos\lambda_s \quad (4)$$
$$y_s = -r\sin\lambda_s \quad (5)$$
$$z_s = 0 \quad (6)$$

The satellite location in Cartesian topocentric (or south-east-up (SEU)) coordinates is

$$\begin{bmatrix} x_{\text{seu}} \\ y_{\text{seu}} \\ z_{\text{seu}} \end{bmatrix} = \begin{bmatrix} \sin\phi_g\cos\lambda_g & \sin\phi_g\sin\lambda_g & -\cos\phi_g \\ -\sin\lambda_g & \cos\lambda_g & 0 \\ \cos\phi_g\cos\lambda_g & \cos\phi_g\sin\lambda_g & \sin\phi_g \end{bmatrix} \times \begin{bmatrix} x_s - x_g \\ y_s - y_g \\ z_s - z_g \end{bmatrix}. \quad (7)$$

Using the SEU coordinates the satellite location in spherical topocentric coordinates is

$$d = \sqrt{x_{\text{seu}}^2 + y_{\text{seu}}^2 + z_{\text{seu}}^2} \quad (8)$$
$$\tan AZ = \frac{y_{\text{seu}}}{x_{\text{seu}}} \quad (9)$$
$$\tan\theta = \frac{z_{\text{seu}}}{\sqrt{x_{\text{seu}}^2 + y_{\text{seu}}^2}} \quad (10)$$

where d is the slant range, AZ is the azimuth and θ is the elevation angle. The SEU coordinates are used to calculate the actual elevation angle and azimuth of the satellite with respect to the AMT.

Changes in the vehicle heading, pitch, and roll alter the orientation of the vehicle-mounted antenna. To track these changes, a coordinate system which follows the movements of the vehicle is required. For this purpose, a vehicular topocentric coordinate system is derived. A Cartesian vehicle-centered system (VEH) is defined where the positive X-axis points in the direction of the vehicle heading, the positive Y-axis points to port, and the positive Z-axis points "up". The VEH system can be derived from the SEU system by a series of coordinate transformations. First a transformation is needed to convert SEU to a north-east-down (NED) system (this is done for compatibility with transformations found in [4]). Following the convention found in [4], the yaw Ψ, pitch Θ, and roll Φ, defined in terms of the NED coordinate system, are

Yaw A rotation of the $X-Y$ plane about the Z-axis by Ψ degrees, measured positive to the right. This is the definition of heading.

Pitch A rotation of the $X-Z$ plane about the Y-axis by Θ degrees, measured positive with the nose up.

Roll A rotation of the $Y-Z$ plane about the X-axis by Φ degrees, measured positive with the left-side up.

The VEH system is therefore related to the SEU system by [4]

$$\begin{bmatrix} x_{\text{veh}} \\ y_{\text{veh}} \\ z_{\text{veh}} \end{bmatrix} = B \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_{\text{seu}} \\ y_{\text{seu}} \\ z_{\text{seu}} \end{bmatrix} \quad (11)$$

where B is given by Equation (16).

Equation (11) gives the coordinates of the satellite location in terms of the vehicle-centered Cartesian coordinates which incorporate pitch, roll, and heading. The actual azimuth and elevation angle between the vehicle and the satellite are given by

$$\tan(AZ_V + \Psi) = \frac{y_V}{x_V} \quad (12)$$
$$\tan(\theta_V) = \frac{z_V}{\sqrt{x_V^2 + y_V^2}}. \quad (13)$$

The azimuth and elevation angle pointing errors are given by

$$\Delta AZ = AZ - AZ_V \quad (14)$$
$$\Delta\theta = \theta - \theta_V \quad (15)$$

where AZ and θ are given by Equations (9) and (10), respectively.

$$B = \begin{bmatrix} \cos\Theta\cos\Psi & \cos\Theta\sin\Psi & -\sin\Theta \\ \cos\Phi\sin\Psi - \sin\Phi\sin\Theta\cos\Psi & -\cos\Phi\cos\Psi - \sin\Phi\sin\Theta\sin\Psi & -\sin\Phi\cos\Theta \\ -\sin\Phi\sin\Psi - \cos\Phi\sin\Theta\cos\Psi & \sin\Phi\cos\Psi - \cos\Phi\sin\Theta\sin\Psi & -\cos\Phi\cos\Theta \end{bmatrix} \quad (16)$$

Polarization Loss

A horizontally-polarized spot beam supported by ACTS was used to test the AMT antenna. The tilt of the antenna (caused by the pitch, roll, and heading of the vehicle) resulted in polarization mismatches between the antenna and the incoming ACTS signal. For any given vehicle heading, an upper bound for the tilt of the antenna is the angle formed by the roof of the van and the horizontal. This will be referred to as the maximum polarization error.

The SEU coordinates of three points p_1, p_2, and p_3, which define the SEU X-Y plane are individually transformed into VEH coordinates using Equation (11). The vehicular coordinates of p_1, p_2, and p_3 are then used to define a plane and its normal, $\vec{n}$ [5]:

$$\vec{a} = \vec{p_2} - \vec{p_1} \quad (17)$$

$$\vec{b} = \vec{p_3} - \vec{p_1} \quad (18)$$

$$\vec{n} = \vec{a} \times \vec{b} \quad (19)$$

The maximum polarization error ϕ_p is the angle between $\vec{n}$ and the normal to the vehicular X-Y plane $\vec{h} = [0, 0, 1]$. ϕ_p is given by [5]:

$$\cos(\phi_p) = \frac{\vec{h} \cdot \vec{n}}{\| \vec{h} \| \| \vec{n} \|}. \quad (20)$$

The loss in signal power received due to polarization mismatch is then upper bounded by [6]:

$$L_{p,\max} = \cos^2(\phi_p). \quad (21)$$

The angle γ between the azimuth direction and the axis of the vehicle tilt may be computed from the elevation angle θ_V and azimuth AZ_V. First, the pointing vector of the antenna $\vec{d}$ is calculated as follows:

$$d_x = \cos\theta_V \cos AZ_V \quad (22)$$

$$d_y = -\cos\theta_V \sin AZ_V \quad (23)$$

$$d_z = \sin\theta_V, \quad (24)$$

$$\vec{d} = [d_x, d_y, d_z] \quad (25)$$

A vector $\vec{t}$ in the direction of the axis of tilt of the van is found as the cross product of the normals $\vec{n}$ and $\vec{h}$:

$$\vec{t} = \vec{n} \times \vec{h}. \quad (26)$$

Therefore, γ is the angle between the vectors $\vec{t}$ and $\vec{d}$:

$$\cos(\gamma) = \frac{\vec{t} \cdot \vec{d}}{\| \vec{t} \| \| \vec{d} \|}. \quad (27)$$

We take the reflection of the maximum loss $L_{p,\max}$ in the direction the antenna is pointing in order to find the actual loss in received signal power due to polarization mismatch:

$$L_p = L_{p,\max} \cos(\gamma) = \cos^2(\phi_p) \cos(\gamma). \quad (28)$$

Using the conditions of the AMT test, the maximum polarization loss in received signal power $L_{p,\max}$ is 0.042 dB. This was found from Equation (21), using a fairly steep road tilt consisting of a 4° roll and a 4° pitch. For the data shown in Figure 2, the maximum calculated polarization loss L_p is 0.026 dB. When the polarization losses are compared to the losses due to pointing error (on the order of 1 dB), these levels are considered negligible.

Results and Concluding Remarks

To test the accuracy of Equations (14) and (15), measurements of the pitch, roll, and heading of the vehicle carrying the antenna system were taken at 0.1-mile intervals along the route circling the Rose Bowl in Pasadena, California. The azimuth and elevation angle errors were calculated using Equations (14) and (15), while the pointing error loss due to each was computed using logarithmic interpolation of the antenna gain pattern data. Plots of the pointing error losses using this data are illustrated in Figures 3 and 4. For the purposes of comparison, the actual pilot power measured during an earlier test run is included. For both runs, the maximum vehicle pitch angle was 3.3 degrees, while maximum roll angle was 4.5 degrees which resulted in maximum elevation angle and azimuth errors of 3.9° and 3.3°, respectively.

These results show that changes in vehicle heading, pitch, and roll can predictably account for an overall loss of 1.5 dB in the signal power received by a vehicle-mounted antenna. Differences between the measured data and the simulated data points are due to shadowing and multipath interference which are not modeled by the pointing error calculations developed in this paper. The losses generated by the simulation account for most of the long term average power variations observed in the measured data. We conclude that the small 1 to 2 dB variations in signal power observed on most of the AMT runs were due to antenna pointing error. Polarization mismatches were found to be fairly insignificant in the overall loss.

References

[1] K. Dessouky and T. Jedrey. The ACTS mobile terminal (AMT). In *Proceedings of the AIAA Conference*, Washington, DC, November 1992.

[2] A. Densmore and V. Jamnejad. A satellite-tracking K- and K_a-band mobile vehicle antenna system.

IEEE Transactions on Vehicular Technology, VT-42:502–513, November 1993.

[3] W. Pritchard, H. Suyderhoud, and R. Nelson. *Satellite Commnications Systems Engineering.* Prentice-Hall, Englewood Cliffs, NJ, second edition, 1993.

[4] B. Stevens and F. Lewis. *Aircraft Control and Simulation.* John Wiley, New York, 1992.

[5] E. Swokowski. *Calculus.* PWS-Kent, Boston, fifth edition, 1991.

[6] M. Jensen, 1995. Department of Electrical & Computer Engineering, Brigham Young University, private communication.

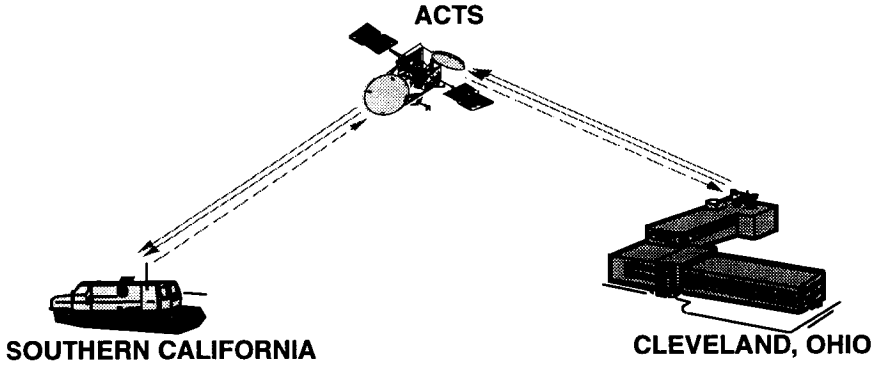


Figure 1: ACTS Mobile Link Diagram

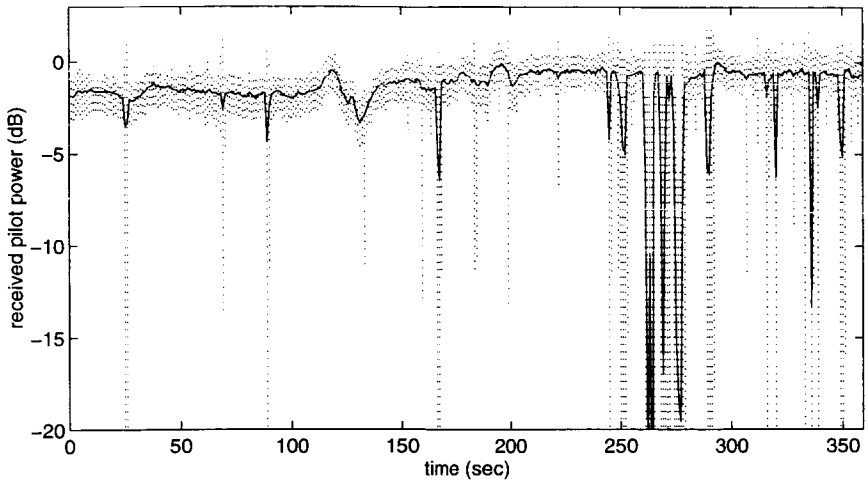


Figure 2: AMT received signal Power for a typical run in Pasadena, California. The solid line represents the received power averaged over one second.

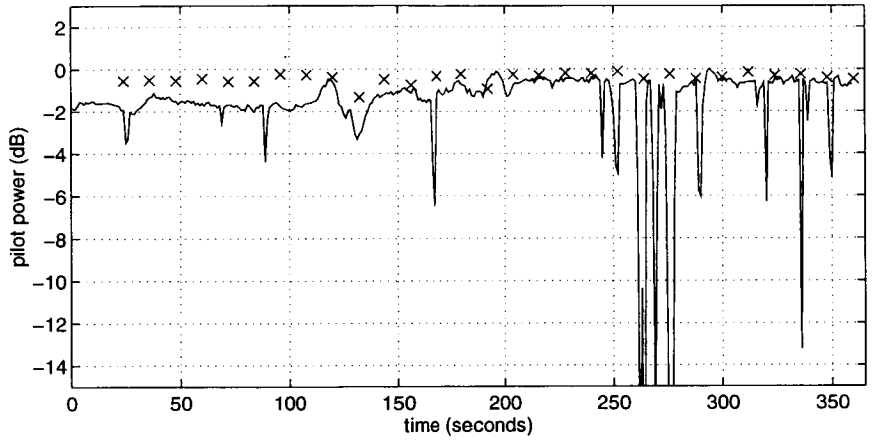


Figure 3: Pilot power (dB) vs. Time (s) for the Clockwise Rose Bowl Route (X = simulation, solid line = measured data).

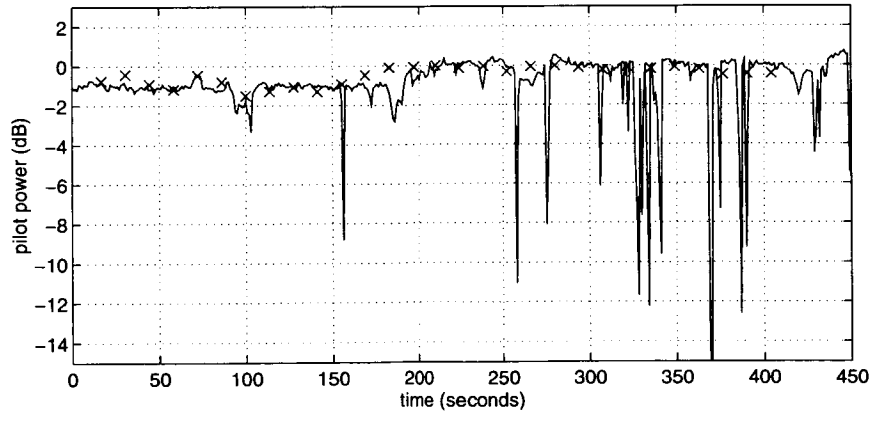


Figure 4: Pilot power (dB) vs. Time (s) for the Counter-clockwise Rose Bowl Route (X = simulation, solid line = measured data).

Session 8
Networking and Protocols II: Channel Assignment and Intersatellite Links

Session Chairperson—*Mala Laurin*, Motorola Satellite Communications, USA
Session Organizer—*Mike Moher*, Communications Research Centre, Canada

On Resource Management in LMS Systems

Axel Jahn
German Aerospace Research Establishment (DLR)
Institute for Communications Technology
P.O. Box 1116, D-82230 Wessling, Germany
Phone: +49 8153 28-2847 Fax: +49 8153 28-1442
Email: Axel.Jahn@dlr.de

ABSTRACT

Efficient resource management is mandatory to achieve maximum system capacity for next generation satellite systems. Resource management deals with the available spectral band, time, power, and space for a transmission signal. It includes (i) the frequency planning and (ii) the assignment of transmission resources to users. The paper presents a generalized notation as well as graph algorithms for resource management problems. Impairment graphs can be used for frequency planning, whereas flow graphs are suitable for resource assignment problems. To evaluate the performance of the resource management, service criteria (such as blocking or the carrier to interference ratio C/I) or efficiency critera (bandwidth requirements) can be derived from the graphs. Examples are provided for fixed and dynamic resource allocation schemes for a typical MEO satellite scenario. Satellite diversity can be also investigated.

I. INTRODUCTION

Next generation satellite systems will provide worldwide services using satellite constellations in non-geostationary (NGSO) orbits. The spectrum allocated in the L-band and recently by WRC'95 in the FPLMTS-bands at 2 GHz will certainly not suffice to serve the long-term demand for LMSS spectrum allocations. An efficient resource management (RM) is mandatory to achieve maximum system capacity. RM deals with the available spectrum, time, power, and space for a transmission signal. It includes:

- frequency planning before and during system operation; this is often called resource allocation (RA) that handles the mapping of available spectrum resources to network entities (satellite spotbeams, earth regions, or user terminals);
- assignment of transmission resources to users (channel assignment, (CA)) implying also the assignment of a transmission path together with a network access node;
- control of the transmitted power.

RM for NGSO satellite systems has not yet been addressed thoroughly in literature. It is different from the RM of terrestrial cellular due to the different network architecture and geometry. Fig. 1 illustrates the problems and ancillary conditions of NGSO satellite systems:

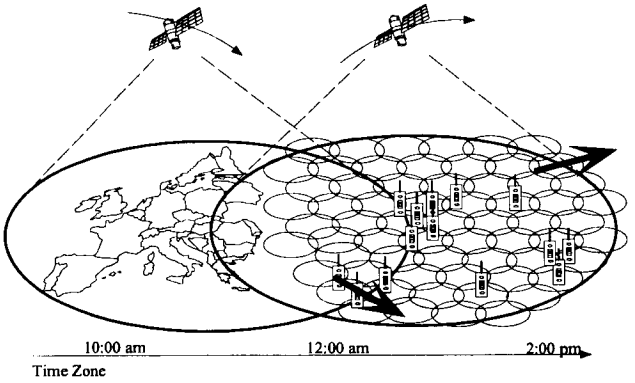


Fig. 1. Problems of resource management in NGSO satellite systems

- The satellites usually employ multibeam antennas to increase system capacity and spectrum efficiency. The dimensions of the covered areas may exceed the size of countries (for spotbeams) or continents (for footprints). Technological capacity limitations may apply to single spotbeams or the overall satellite.
- Spotbeam and footprint overlapping yields co-channel interference (or multiple access interference (MAI) in spread spectrum systems). Therefore, frequency planning is required. Beside the classical frequency planning using reuse patterns for the spotbeam antennas, satellite and/or orbit band segmentation is necessary. Another planning approach is the region-oriented resource allocation to earth-fixed cells [1].
- The constellation and thus the availability of network access nodes is time-variant.
- An inhomogeneous regional traffic distribution leads to time-variant unbalanced traffic loads for the different satellites and spotbeams. The regional traffic distribution depends also on local day time zones.
- RM must possibly cope with the use of multiple satellite visibility for diversity reception e.g., the

systems ICO [2] and Globalstar [3].
- Handover between spotbeams or satellites might be driven by the changing constellation geometry or propagation channel effects.
- Specific channels could be unavailable in selected regions or spotbeams due to other-system interference (e.g., from terrestrial fixed services) or due to administrative reasons (protection of other services).

Resource management must cope with the above mentioned problems. Powerful tools are needed for system engineering and performance test. Analytical approaches mostly fail for non-trivial systems. Simulations might be used for performance evaluation, but they suffer from modelling effort and extremely long processing times.

This paper presents a graph theoretical approach for the RM performance investigations. Starting from a graph representation of the problems, system criteria such as blocking, bandwidth demand and carrier-to-interference ratio can be derived for many applications. The advantages of a Graph model are manifold: (i) the results of evaluation are exact (no approximation due to an underlying stochastic process), (ii) theoretical bounds are known for many problems, and (iii) the algorithms itself can also serve for the final protocol implementation in the communications network.

The discussed methods can handle TDM/FDM as well as CDMA systems. Satellite diversity, other-system interference and channel shadowing can be considered in a straight-forward manner. Section III provides examples of different investigations for a MEO constellation system.

II. Resource Management

Resources in a communications network are manifold. They include (i) frequency spectrum, (ii) time, (iii) transmit power, (iv) space, (v) network nodes (e.g., channels in terms of required hardware), and (vi) human or technological resources (e.g., satellite battery power). An active (i.e. communicating) user needs a certain portion of the resources for information transmission. Some resources can not be dissipated (frequency band still exists while being used) but are withheld from usage by other users. The withhold can also be of administrative nature to locally protect other services from interference.

RM addresses several aspects:

(i) the spectral efficiency of signal transmission waveforms (modulation efficiency)

(ii) the segmentation of the spectrum resource into logical channels (multiplexing, TDM/FDM/-CDM);

(iii) the planning of frequency reuse (resource allocation (RA))

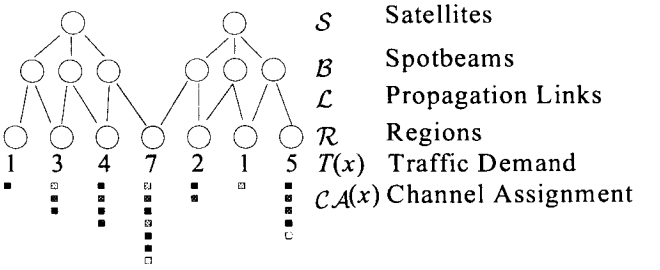


Fig. 2. Generalized network representation with channel assignment to users as a set assignment

(iv) assignment of channels to network nodes and users (channel (multiple) access (CA))

The items (i) and (ii) have been investigated extensively elsewhere (among many other references, see [4]) and will not be treated further. In [5] a first approach to the spectrum resource management was presented, restricted to terrestrial applications. However, there is still little known about RA and CA schemes considering the peculiarities of NGSO satellite systems. Thus, research for the techniques itself and evaluation methods are of prime importance.

A. Problem Formulation

In this section we develop a generalized description of RM, cf. Fig. 2 and 3. Let us denote the available frequency band and time by $\mathcal{F}, \mathcal{T}$. Through multiplexing, the spectrum resource is divided into a set of logical channels $\mathcal{C}$. The set $\mathcal{P}$ shall represent discrete values of the transmission power. Furthermore there are the network access nodes $\mathcal{B}$ (spotbeams) which can be hierarchically structured by switching entities $\mathcal{S}$ (satellites). Subscribers of the network are located in regions $\mathcal{R}$. The regions may represent spotbeam areas, but also administratively divided regions, or simply a grid on earth. Depending on the geometry several propagation links $\mathcal{L}$ are available between regions and network nodes. The set is time-variant due to the dynamic satellite constellation or the mobility of the users.

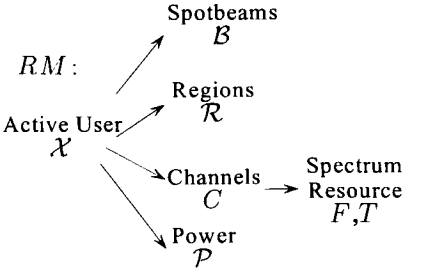


Fig. 3. Generalized assignment of resources to active users

In the regions, a traffic demand $T(x)$ is generated, modelled by stochastic processes. The parameters of

the processes can depend on regional traffic distributions and local busy hours. RM must assign a tupel $\{c_i \in \mathcal{C}, p_i \in \mathcal{P}, b_i \in \mathcal{B}, r_i \in \mathcal{R}\}$ for every active user $x_i \in \mathcal{X}, ||\mathcal{X}|| = T(x)$, cf. Fig. 3. The selection of the link $l_i \in \mathcal{L}$ is inherently included by the assignment of a network access node. In essence, RM is a multiple set assignment that determines the system performance. Fig. 2 exemplarily shows the channel assignment by colors to users in the regions. Equivalent to the graphical form in Fig. 3, the tupel assignment can also be written as

$$\{c_i, p_i, b_i, r_i\} = RM_{\mathcal{C},\mathcal{P},\mathcal{B},\mathcal{R}}(\mathcal{X}) \qquad (1)$$

It should be mentioned that the syntax and methods introduced here are also valid for other than satellite systems (for instance mixture between different terrestrial cellular and satellite).

B. RM Techniques and Terminology

To reduce complexity of generalized RM in Eq. (1), the assignment of different resources is usually splitted in several steps and RM is treated in a subsequent manner.

(i) Region assignment is mostly done first. In many cases, region assignment is trivial since only one service region exists. Example for region assignment: imagine a FPLMTS system with a terrestrial cellular and satellite component; the user has a dual-mode phone; the phone selects the cellular system with priority, but satellite is chosen if no cellular coverage is provided.

(ii) Power management includes the choice of a proper link margin ([6], [7]) and power control techniques. Power management is often independent from other RM schemes. In [5] a combined channel & power selection scheme is reported.

(iii) **RA before CA**; Frequency planning precedes the channel assignment to users. This is the classical approach of network dimensioning. Eq. (1) simplifies to

$$RM_{\mathcal{C},\mathcal{B}}(\mathcal{X}) = CA(RA(\mathcal{X})) \quad . \qquad (2)$$

The scheme is appropriate for bandwidth/interference-limited systems.

(iv) **CA before RA**; Radio resource allocation is done after the channel assignment. Eq. (1) simplifies to

$$RM_{\mathcal{C},\mathcal{B}}(\mathcal{X}) = RA(CA(\mathcal{X})) \quad . \qquad (3)$$

The scheme is appropriate for capacity-limited systems.

Satellite- or Region-Oriented Resource Allocation strategies (*SORA* [8], *RORA* [1]) belong to the group (iii), cf. Fig. 4. They are characterized by the network entity the radio resources are allocated to. Another classification of RM is according to the adaptiveness of the resource allocation: Fixed, Dynamic, or Hybrid Channel Allocation (*FCA, DCA, HCA*)[9]. While FCA counts to (iii), DCA and HCA belong to (iv). Here, the resource arrangement must be dynamically adopted to the time-varying traffic demand. Adaptive schemes are of great interest for satellite systems due to the inhomogeneous traffic distribution and the strong time dependence.

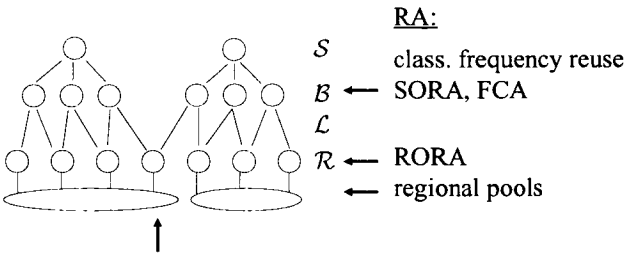


Fig. 4. RA and CA in a satellite network

In the Section III, Graph algorithms are developed to model RA and CA in satellite systems.

C. Performance Criteria

There are different performance criteria for RM: (i) system-oriented criteria such as spectrum efficiency, or (ii) service/user-oriented criteria such as grade-of-service (blocking, interruption) and quality of service (C/I, BER, outage probability). Other criteria could be satellite capacity requirements, technological effort (e.g., transponder, on-board switching), signalling and processing requirements, or satellite power requirements.

Depending on the form of RM, the solutions of Eq. (1) form a multi-dimensional finite state space $\mathcal{RM} = \{rm_1, rm_2, \ldots, rm_n\}$ with all possible tupel assignments. The goal is now to find optima for the desired performance criteria. This can be achieved by energy-functions E_x which assign a real value to each state of the state space $\mathcal{RM}$.

$$E: \quad \mathcal{RM} \rightarrow \mathbb{R} \qquad (4)$$

The search for optima of the energy function is a problem of combinatorics. Possible energy functions are for instance:

- $E_{C/I}$: minimal interference (relevant function for bandwidth/interference-limited systems),
- E_{p_b}: minimal blocking, (relevant function for capacity-limited systems) or
- E_{η_s}: maximal spectrum efficiency.

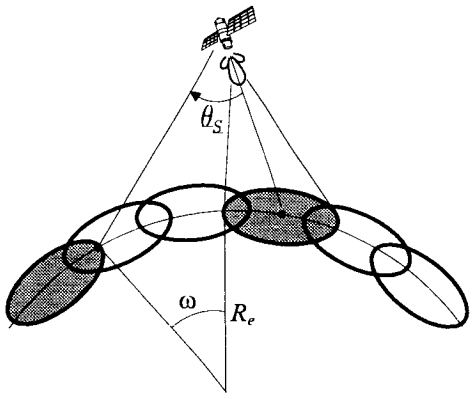


Fig. 5. Generation of an impairment graph

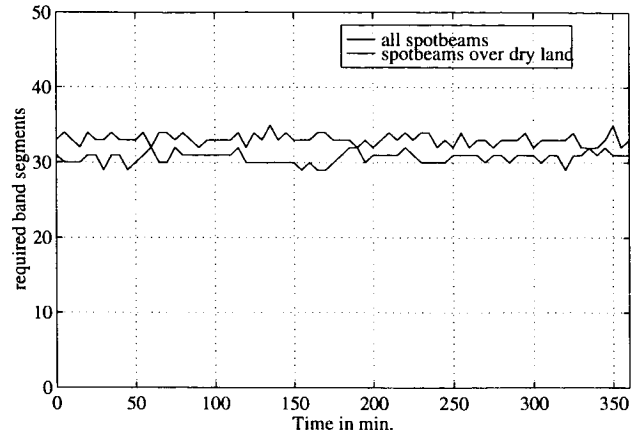


Fig. 6. Required band segments for SORA in a MEO constellation using 3-tier 19-spotbeam antenna

According to [10], [11] spectrum efficiency can be defined as $\eta_s = \frac{\text{carried information rate}}{M_s\text{: total required spectrum resource}}$ with

$$M_S = \iiint_{\frac{C}{I}(f,t,\vec{x}) \leq \text{ thres.}} \mathrm{d}f\,\mathrm{d}t\,\mathrm{d}\vec{x} \qquad (5)$$

being the used spectrum state space (a portion of spectrum space state is called *used* when the local interference exceeds a given threshold). The calculation of the C/I depends on the multiplexing, power control, frequency reuse, and satellite antenna, for details of TDM or CDM systems cf. [8], [12].

III. Graph Models

This section provides some examples of RA and CA functions in NGSO satellite systems that can be investigated and evaluated using Graph theory. Graph algorithms can derive exact numerical solutions, but also be implemented in the operating network. A comprehensive survey of Graph algorithms and their complexity is given in [13], [14]. D. E. Knuth has compiled a collection of easy-to-use and well-documented algorithms in C [15] for the public domain. Interesting links to graph pages on the WWW can be found in [16].

A. Satellite-Oriented RA using Graph Coloring

Satellite-oriented RA in the classical sense uses frequency reuse patterns for the spotbeams and band segmentation for satellites. Band segmentation for the satellites does not consider the varying distances between the satellites due to the constellation shift. Constructing an impairment (or interference) graph on spotbeam level and graph coloring can be adopted instead to allocate frequencies to the spotbeams. The construction of the impairment graph $G(v, e)$ is similar to the procedure in terrestrial network planning [17]. Each spotbeam corresponds to a vertex v_i of the Graph G. The construction of edges e_i is illustrated in Fig. 5. Two vertices (spotbeams) are connected by an edge if the antenna suppression between the edges of coverage is lower than a given threshold. Edge connections must be checked for every satellite in the system. Equivalently, the corresponding earth center angles can be used.

To determine the minimum number of frequency bands for the satellite constellation, the chromatic number $\chi(G)$ of the graph can be calculated. The total required system bandwidth corresponds to $\max(\chi(G)) \cdot \Delta f$ with Δf being the bandwidth of each band. Fig. 6 shows the number of required channels versus time for an ICO-type of constellation [2] with a 3-tier 19-spotbeam satellite antenna. The system can be colored by 35 colors. Since the total network has 190 spotbeams, this corresponds to a mean frequency reuse of 5.4. As an alternative, spotbeams covering ocean region can be switched off, resulting in a smaller number of 32 frequency bands (resulting frequency reuse: 5.9, switched-off spotbeams must be considered). Graph coloring algorithms (e.g., the simplex sequential algorithm) return additionally the proper coloring $c(v)$ (i.e. the resource assignment) for each vertex (spotbeam) $c(v) = COL(G)$, thus $RA(\mathcal{X}) = COL(G)$. The system bandwidth is then $B = \bigcup c(v)$. Fig. 7 displays the co-channel reuse (note that only the channel with maximum reuse is shown).

From the regional co-channel assignment, histograms of the C/I-value can be computed, cf. Fig. 8. The resulting C/I values reach very high values indicating that the chosen threshold for the generation of the impairment graph could be lowered. In general, graph coloring excels the classical frequency planning using reuse patterns since the actual satellite constellation geometry is taken into account.

For region-oriented RA, the impairment graph is generated in a similar way as described above, but considering the visibility of satellite spotbeams from earth cells.

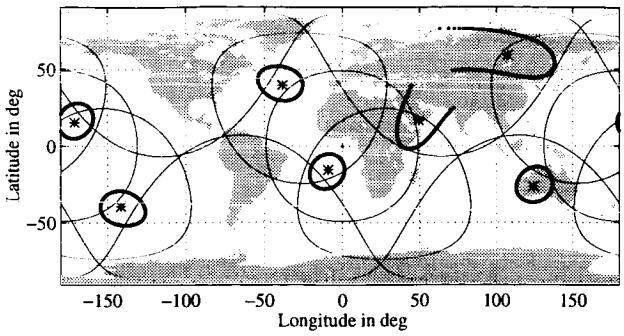


Fig. 7. Co-channel reuse in spotbeams on earth for a MEO constellation (19-spotbeam antenna)

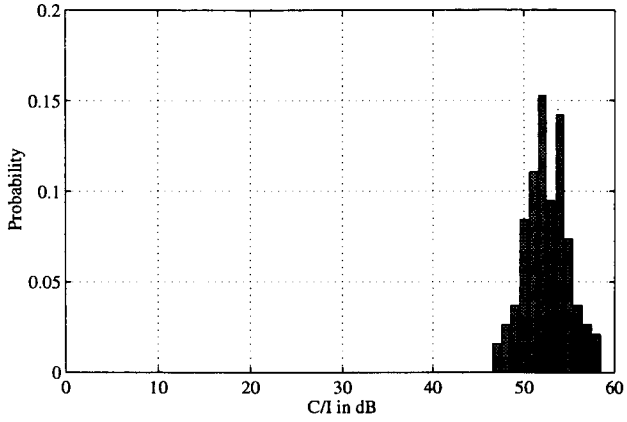


Fig. 8. C/I distribution of $RA = Col(G)$ for a MEO constellation (19-spotbeam antenna)

B. CA using Flow Graphs

CA can be investigated in an efficient way by flow graphs, cf. Fig. 9 for SORA and Fig. 10 for RORA, respectively. Let us first discuss the SORA case. The vertices of the graph are the network entities *satellites*, *spotbeams*, and *regions*. Additionally, there is a source and a sink vertex. The following connections belong to the edge set: (i) all regions are connected to the source vertex through the edges $e_{s,r}$, (ii) all propagation links $l_i \in \mathcal{L}$ between regions and spotbeams are edges $e_{r,b}$, (iii) all satellites are connected to their respective spotbeams $e_{b,sat}$, and (iv) all satellites are connected to the sink vertex $e_{sat,s}$. All edges are directed from the source to sink vertex (for simplicity the direction is not shown in Fig. 9 and 10). Furthermore, all edges in flow graphs are characterized by a capacity limit $cap(e)$. This property determines the maximum flow which can be transported through an edge. Maximum flow algorithms can then determine the network flow. The goal is now to dimension the capacity limits of the graph in such a way that the graph represents the CA problem. Thereto, the following capacities are assigned to the edges:

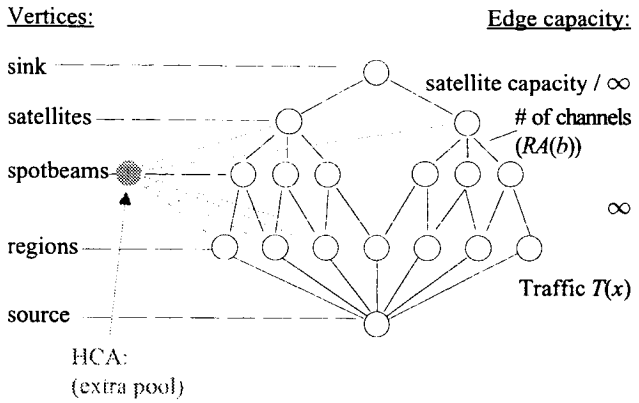


Fig. 9. Flow graph for satellite-oriented RA

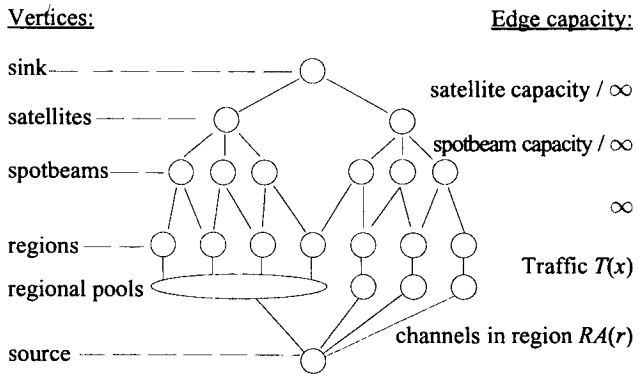


Fig. 10. Flow graph for region-oriented RA

(i) the capacity $cap(e_{s,r})$ for edges from the source to the regions is determined by the traffic demand $T(x)$ generated in the regions. This ensures that at most the generated traffic can flow through a region.

(ii) the capacity $cap(e_{r,b})$ for the links is set to infinity, so no restriction for the serving spotbeams is applied.

(iii) the capacity $cap(e_{b,sat})$ from the spotbeams to the satellite is determined by an appropriate total spotbeam capacity. The spotbeam capacity would correspond to the number of channels in a SORA scheme. Assigning a capacity to spotbeams ensures that at most the maximum number of channels per spotbeam can serve the traffic. Setting the limit to infinity means that all generated traffic is routed. This is useful to determine the required channels per spotbeam for DCA.

(iv) the capacity $cap(e_{sat,s})$ from the satellite to the sink vertex is determined by an appropriate total satellite capacity (or infinity).

The resulting function for CA is then $CA(\mathcal{X}) = FLOW(G)$. For HCA schemes (see Fig. 9), a slight modification can be applied: an extra pool with a certain amount of channels is connected to all regions and satellites. Thus, a satellite can borrow channels

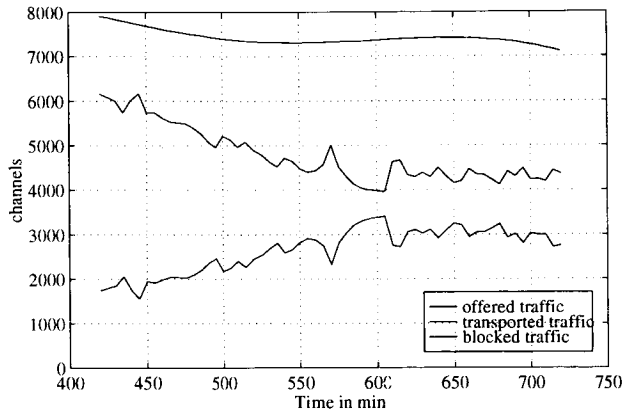


Fig. 11. Blocking of traffic for SORA, MEO constellation (19-spotbeam antenna), 250 channels per spotbeam, satellite capacity: ∞

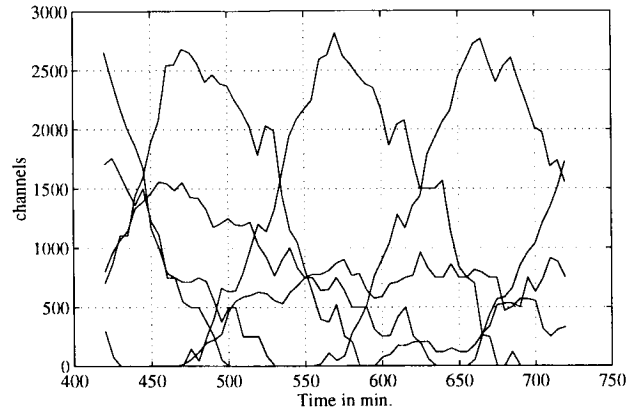


Fig. 12. Satellite load distribution for SORA, MEO constellation (19-spotbeam antenna), 250 channels per spotbeam, satellite capacity: ∞

from the common pool when spotbeam channels are exhausted and further traffic would be blocked. Now, $HCA(\mathcal{X}) = FLOW(G')$. In the RORA case (Fig. 10), radio resources are assigned to regions. In order to limit the channels per region in the flow graph , an additional layer is inserted in the graph G. The capacity limit $cap(e_{s,reg})$ for the edges between source and regions can be derived from network planning (cf. section III-A) or by genetic algorithms [1], thus $CA_{\mathrm{RORA}}(\mathcal{X}) = FLOW(G)$.

To determine the maximum network flow, the algorithms after Ford-Fulkerson or Edmonds-Karp [14] may be adopted depending on the graph's complexity. The algorithms return a traffic flow $f(e)$ over each network edge that enables the investigation of different aspects: (i) $cap(e_{s,r}) - f(e_{s,r})$ is the blocking in a region (note that $cap(e_{s,r}) = T(r)$ corresponds to the traffic demand),(ii) $f(e_{b,sat})$ is the traffic load sharing in the satellites' spotbeams, and (iii) $f(e_{sat,s})$ is the load distribution among the satellites. For example,

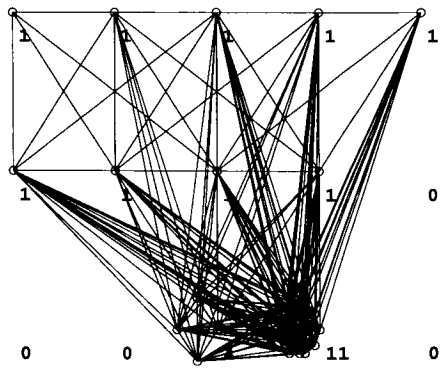


Fig. 13. Lexicographic expansion of an impairment graph for DCA; the numbers at the vertices correspond to the traffic demand.

Fig. 11 and 12 show results from a system simulation of the MEO constellation with 19 spotbeams. The traffic generation was based on an inhomogeneous subscriber distribution proportional to population densities and a busy hour profile. The capacity limit per spotbeam was set to 250 with unlimited capacity per satellite.

C. Satellite-Oriented DCA

In DCA, the problem of CA is preceding RA in order to achieve optimal resource management for unbalanced traffic demands. In section III-B we have seen that flow graphs are an appropriate tool to determine the traffic demand per spotbeam. Apart from that we still need a RA that matches the channel demand. Normal graph coloring is not applicable since it assigns just one color per spotbeam, not multiples. Here, lexicographic expansion of the graph leads to the wanted set assignment of colors [18], [19]. The procedure is illustrated in Fig. 13. Again an impairment graph as described in section III-A is used as baseline. Each vertex v_i represents a spotbeam (or region) with a certain traffic demand $T(i)$ (in the figure the numbers right of the vertex). The graph G is now expanded by lexicographic products G^*, i.e. each vertex is substituted with a clique $Cl(T(i))$ of size $T(i)$ (in the figure the cliques of size 4 and 11 have been already substituted). In this way each channel request in a region or spotbeam is now represented by a separate vertex. Graph coloring yields now a color set assignment with $||Col(v^*)|| = T(v)$, thus $RA(DCA(\mathcal{X})) = COL(G^*(CA(\mathcal{X})))$.

D. Modelling of Satellite Diversity in Flow Graphs

An important feature of NGSO satellite systems is the concept of satellite diversity to reduce link margin requirement and improve quality of service [6]. While diversity is inherently included in the construction of impairment graphs, flow graphs require special constructs for satellite diversity. We will assume that only

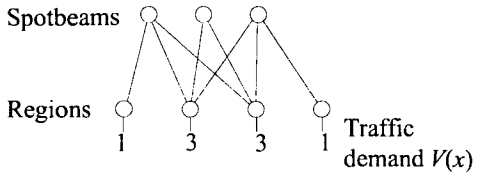


Fig. 14. Part of a flow graph without diversity.

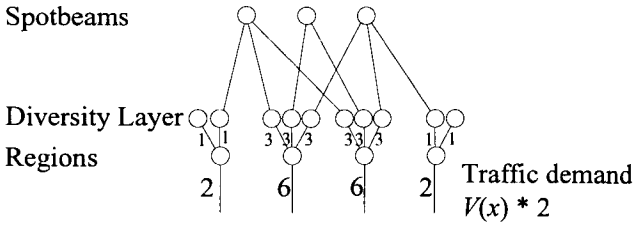


Fig. 15. Part of a flow graph with diversity

dual diversity is adopted. Fig. 14 and 15 explain the modelling of diversity in flow graphs. In Fig. 14, a clip of a flow graph is shown with an exemplary traffic demand in the regions. The traffic demand in fact is the capacity limit of the incoming edges, cf. section III-B. This situation is modified for diversity in Fig. 15. A diversity layer of vertices has been inserted in the graph. Since diversity doubles the demand for serving channels in the spotbeams, the capacity limit of the incoming edges must be doubled as well. In order to avoid that traffic of a single user is routed twice to the same satellite, all region vertices are splitted to $\max(2, \text{grad}^-(r))$ diversity vertices with $\text{grad}^-(r)$ the number of outgoing edges from the regions (i.e. propagation links). The capacity of these edges between the regions and the diversity vertices are limited to the elementary traffic demand. The diversity vertices are connected to one satellite only. Thus, each of the diversity vertices represents the part of the dual diversity traffic that is routed to the connected satellite. From the comparison of capacity limits and traffic flows, one can distinguish total blocking or diversity blocking of a traffic request (diversity blocking means that only one satellite is serving). Total blocking appears if $f(e_{s,r}) < cap(e_{s,r})$ while diversity blocking happens if $f(e_{s,r}) < 2 \cdot cap(e_{s,r})$.

IV. Conclusions

In the paper a concept for notation of resource management is presented. It comprises radio resource allocation techniques, power management, as well as channel assignment schemes. The concept can be widely adopted to cellular, satellite and future 3rd generation mobile telecommunications networks. Furthermore, satellite specific aspects of resource management using graph theoretical algorithms are presented. Impairment graphs can be used to determine proper resource allocations whereas flow graphs are suitable for channel assignment. Examples showed the performance measures that can be derived from the graph models.

References

[1] S. Blondeau, G. Maral, and J. P. Taisant. Region oriented resource allocation methodology for non-geostationary mobile communications satellite systems. *IOS Press Space Communications*, 13:171–191, 1995.

[2] P. Poskett. Satellite systems architectures. In *Mobile and Personal Satellite Communications 2 – Proceedings 2nd European Workshop on Mobile/Personal Satcoms (EMPS'96)*, pages 485–500. Springer-Verlag, London, 1996.

[3] J. Schindall. Concept and implementation of the Globalstar mobile satellite system. In *Proc. Fourth Int. Mobile Satellite Conf. IMSC'95*, pages A11–A16, 1995.

[4] J. G. Proakis. *Digital Communications*. McGraw-Hill, New York, 1989.

[5] J. Zander. Radio resource management. In *IEEE VTS 46th Vehicular Technology Conference (VTC'96)*, pages 16–20, Atlanta, April 28 - May 1, 1996.

[6] T. Goerke, N. Hart, and A. Jahn. A discussion on mobile satellite systems and the myths of CDMA and diversity revealed. In *Proc. Fourth Int. Mobile Satellite Conf. IMSC'95*, pages 469–475, 1995.

[7] A. Jahn, H. Bischl, and G. Heiß. Channel characterisation for spread spectrum satellite communications. In *Proc. IEEE Fourth International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'96)*, pages 1221–1226, 1996.

[8] F. Vatalaro, G. E. Corazza, C. Caini, and C. Ferrareli. Analysis of LEO, MEO, and GEO global mobile satellite systems in the presence of interference and fading. *IEEE J. Selected Areas in Communications*, 13:291–300, 1995.

[9] I. Katzela and M. Naghshineh. Channel assignment schemes for cellular mobile telecommunications systems: A comprehensive survey. *IEEE Personal Communications*, June:10–31, 1996.

[10] L. A. Berry. Spectrum metrics and spectrum efficiency: Proposed definition. *IEEE Trans. Electromagn. Compat.*, 19:254–259, 1977.

[11] S. Heeralall. Discussion of spectrum efficiency and the factors that affect it. In *IEEE Intern. Conf. on Selected Topics in Wireless Communications*, pages 412–416, Vancouver, B.C., Canada, 1992.

[12] E. Lutz. Other-cell interference in satellite power-controlled CDMA uplink. In *Proc. Intern. Mobile Satellite Conf. (IMSC'97)*, page this issue, 1997.

[13] M. C. Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, New York, 1980.

[14] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Englewood Cliffs, New Jersey, 1982.

[15] D. E. Knuth. *The Stanford GraphBase*. Addison-Wesley, Reading, 1993.

[16] M. Kleyn. *Graph Resources Links*. `http://www.cs.utexas.edu/users/kleyn/Kleyn-Graphs.-html`.

[17] A. Quellmalz. *Graphenorientierte Planung von Sendernetzen*. Nomos-Verlagsgesellschaft, Baden-Baden, 1993.

[18] M. Grevel and A. Sachs. A graph theoretical analysis of dynamic channel assignment algorithms for mobile radiocommunication systems. *Siemens Forsch.- u. Entwickl.-Ber.*, 12:298–305, 1983.

[19] R. Künzel. Channel assignment scheme for cellular radio networks. *Siemens Forschungs- u. Entwickl. Berichte*, 11:139–142, 1982.

Resource Allocation for CDMA Satellite Systems

Samuel Blondeau, Jean-Philippe Taisant, Jean-Benoit Besset
France Telecom - Centre National d'études des télécommunications (CNET)
38-40 rue du Général Leclerc 92794 Issy Moulineaux Cedex 9, France
Phone : +33 1 45 29 61 67 Fax : +33 1 45 29 45 34
email : samuel.blondeau@issy.cnet.fr, jean-philippe.taisant@issy.cnet.fr,
jean-benoit.besset@issy.cnet.fr

Gerard Maral
Ecole Nationale Supérieure des Télécommunications (TELECOM Paris), site de Toulouse
10 avenue Edouard Belin BP 4004 31028 Toulouse Cedex, France
Phone : +33 5 62 17 29 92 Fax : +33 5 62 17 29 89
email : maral@tlse.enst.fr

1. INTRODUCTION

On and after 1998, several non geostationary satellite mobile communication systems will compete with one another. With such orbits, the satellites move continuously relative to the Earth's surface, and permanent communications require the use of a constellation of satellites. The traffic generated by an Earth terminal is then supported by successive satellites, and the traffic channel must be handed over from one satellite to the next. Apart from the above introduced satellite handover, there exists a beam-to-beam handover internal to a satellite which results from the use of multibeam satellites.
For allocating system resource, a novel approach has been explored which reduces the signalling overhead associated with handover : this approach is sometimes quoted as "Region Oriented Resource Allocation" (RORA) [1]. It consists of allocating a given resource to the mobile terminal for the entire duration of its call, whatever the spot beams or the satellites it makes use of.
The preferred access scheme for this study is asynchronous CDMA (Code Division Multiple Access) for both forward and return directions. Asynchronous CDMA is not associated with the use of a synchronisation channel being present in the spot beams and satellite used.

This paper aims at presenting a methodology for optimising the system capacity in a RORA/CDMA satellite system. This method is based on the maximum interference noise allowable by the system. As long as the interference generated by interfering mobiles is below this value, the system can continue allocating channels to new mobiles. The maximum density of users is determined when the total interference generated by interfering mobiles on the Earth equals the maximum interference noise allowable by the system.

The paper is organised as follows. Section 2 presents the main assumptions. Section 3 presents the methodology adopted for the determination of the maximum density of users. Section 4 discusses some results.

All examples presented in this paper deal with a Walker constellation named CHLOE, consisting of 72 satellites at an altitude of 1420 km distributed in 9 planes inclined at 51°. This constellation was designed for its low replication period (158 minutes), i.e. the duration between similar constellation/Earth geometric configurations.

2. ASSUMPTIONS

2.1 Propagation model used

For this study, an in-house propagation model based on a two state Markov model is used. The two state model incorporates the effect of the mobile terminal environment on the link, by taking into account the presence or absence of obstacles and the subsequent non-visibility or visibility of the satellite.

This model was developed based on the results of a helicopter propagation campaign performed by FT/CNET. Statistical distributions of mean attenuation for each state and state transition probabilities are determined for different elevation angles and propagation environments.

In the simulation, a state is randomly attributed to a mobile when it starts a new communication. Considering the statistical distribution of the mean attenuation of the state previously chosen, a value for mean attenuation is determined. By assumption, this propagation state is considered constant during 3 seconds for a mobile terminal and 10 seconds for a hand-held terminal. So, every propagation step, considering the state transition probability, a new state and an associated mean attenuation are determined.

The Rice factor of the channel is determined from the mean attenuation as described in [2].

2.2 Power control

A closed loop power control technique has been implemented to optimise the value of the transmit power necessary to guarantee the required quality, therefore, to minimise the interference noise.

Considering the power control delay with regard to the propagation channel fluctuations and inaccuracies in the estimation of the instantaneous communication quality, the actual transmit power is different from the required transmit power.

We have, $$\Delta P_{tx}|_{error} = P_{tx}|_{actual} - P_{tx}|_{ideal} \quad (1)$$

If the actual transmit power is less than the required transmit power, the required quality is not guaranteed. Therefore, a power margin must be added at the transmit power to immediately compensate for inaccuracies in the estimation of the instantaneous communication quality. This power margin is noted as $P_{tx}|_{error}$.

Due to inaccuracies in the estimation of the instantaneous communication quality, we consider that the power control error distribution can be estimated to a first order of magnitude by a centred Gaussian normal distribution with standard deviation equal than the minimum power control step size. Considering the properties of a normal Gaussien distribution, it is sufficient to add to the transmit power a power margin equal to two standard deviations, resulting in only 2.5% of cases where the margin is not sufficient.

2.3 Availability criterion

2.3.1 Definition : The instantaneous communication quality is greater than or equal to the required quality when the Eb/Nt at the receiver is greater than or equal to the required Eb/No.

i.e. when, $$\frac{E_b}{N_t} \geq \frac{E_b}{N_0}\bigg|_{required} \quad (2)$$

With, $$\left(\frac{E_b}{N_t}\right)^{-1} = \left(\frac{E_b}{N_0}\right)^{-1} + \left(\frac{E_b}{I_0}\right)^{-1} \quad (3)$$

By definition, the communication is considered available if the instantaneous communication quality is greater than or equal to the required quality during more than 85 % of the communication duration.

By extension, if the instantaneous transmit power (P_{tx}) necessary to guarantee the required quality is greater than the maximum mobile transmit power ($Pe|_{max}$) then the instantaneous communication quality is less than the required quality. The equation (2) also requires,

$$P_{tx} \leq P_{tx}|_{\max} \quad (4)$$

2.3.2 Expression of the transmit power : To establish the expression of the transmit power necessary to guarantee the required quality, the different losses experienced during a communication have been identified.

The initial scenario used is that of a mobile, located at the sub satellite point, experiencing a Gaussien propagation channel without attenuation and interference noise. The transmit power necessary to ensure "instantaneous communication quality" for such a mobile is called the nominal transmit power ($P_{tx}|nominal$).

We have, $$P_{tx} = P_{tx}|_{nominal} \quad (5)$$

The further losses experienced during a communication will now be studied individually.

Firstly, when the mobile described above experiences a Ricean propagation channel. The mean transmit power must be increased to compensate the variation in the received signal power.

Hence, $$P_{tx} = P_{tx}|_{nominal} + \Delta P_{tx}|_{rice}^{h(K)} \quad (6)$$

Secondly, interference noise is considered. From equation (2), the apparition of interference noise decreases the value of Eb/Nt. To guarantee a required quality, the transmit power must compensate this reduction.

Hence, $$P_e = P_e|_{nominal} + \Delta P_e|_{interference} \quad (7)$$

Thirdly, LoS blockage is considered. As before, to guarantee a required quality, the transmit power must be increased to compensate this extra.

Hence, $$P_{tx} = P_{tx}|_{nominal} + \Delta P_{tx}|_{blockage} \quad (8)$$

Fourthly, since the mobile isn't necessarily located at the sub satellite point, the free space loss and the transmit and receive antennae gain are different. The transmitter must be take these variations into account.

We have, $$P_{tx} = P_{tx}|_{nominal} + \Delta P_{tx}|_{position} \quad (9)$$

Finally, the different losses experienced during a communication being independent and taking into account the inaccuracy in the power control loop, the expression of the transmit power can be decomposed as follows,

$$P_{tx} = P_{tx}|_{nominal} + P_{tx}|_{error} + \Delta P_{tx}|_{error} + \Delta P_{tx}|_{rice}^{h(K)} + \Delta P_{tx}|_{interference} + \Delta P_{tx}|_{blockage} + \Delta P_{tx}|_{position} \quad (10)$$

2.3.3 Transmit power required to guarantee the availability of a communication:

$$\text{Let, } \Delta P_{tx}|_{required} = P_{tx}|_{error} + \Delta P_{tx}|_{error} + \Delta P_{tx}|_{rice}^{h(K)} + \Delta P_{tx}|_{interference} + \Delta P_{tx}|_{blockage} + \Delta P_{tx}|_{position} \quad (11)$$

Principally due to satellite motion and propagation channel fluctuations, the power margin required to ensure the instantaneous required quality fluctuates during a communication. Therefore, the value of the power margin $\Delta P_{tx}|_{required}$ is calculated at each propagation step. An exemple of the power margin for a 2 minutes communication is shown in figure 1 below.

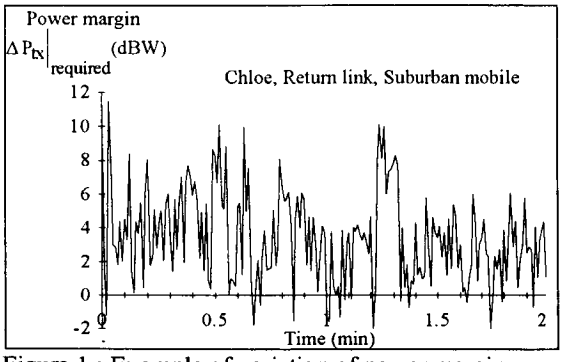


Figure 1 : Example of variation of power margin $\Delta P_{tx}|_{required}$ during a communication.

$$\text{Let, } Prob\left[\Delta P_{tx}|_{required} \le \Delta P_{tx}|_{necessary}\right] = 0.85 \quad (12)$$

Such that $\Delta P_{tx}|_{necessary}$ represent the value of the power margin $\Delta P_{tx}|_{required}$ necessary to ensure sufficient communication quality.

To determine this value, a distribution of the power margin required to ensure the instantaneous required quality during a communication is calculated. This distribution is indicated in figure 2.

$$\text{Let, } \Delta P_{tx}|_{available} = P_{tx}|_{max} - P_{tx}|_{nominal} \quad (13)$$

Based on these equations, the communication is considered available if the power margin $\Delta P_{tx}|_{necessary}$ is less than or equal to the power margin $\Delta P_{tx}|_{available}$.

$$\text{i.e. when, } \Delta P_{tx}|_{necessary} \le \Delta P_{tx}|_{available} \quad (14)$$

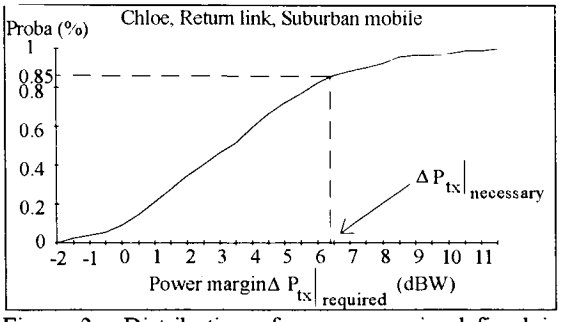


Figure 2 : Distribution of power margin defined in figure 1.

2.4 Availability tables

2.4.1 Introduction : Considering equation (11), only the power margin used to compensate interference noise depends on the number of mobiles communicating. The rest of the power margin required is to ensure sufficient instantaneous quality used to compensate mobile specific losses (position, blockage, etc.). Given that the power margin available is fixed and limited, a trade off between availability and capacity (number of mobiles communicating) must be made.

For these reasons, it is useful to establish, for a mobile without considering interference noise, a relation between availability and power margin necessary to ensure sufficient communication quality. The power margin is only used to compensate mobile specific losses. Whatever the capacity of the system, these mobile specific losses (noted MSL in equations) must be compensated to ensure sufficient communication quality.

$$\text{Let, } Prob\left[\Delta P_{tx}|_{required}^{MSL} \le \Delta P_{tx}|_{necessary}^{MSL}\right] = 0.85 \quad (15)$$

$$\text{with, } \Delta P_{tx}|_{required}^{MSL} = \Delta P_{tx}|_{rice}^{h(K)} + \Delta P_{tx}|_{blockage} + \Delta P_{tx}|_{position} + P_{tx}|_{error} + \Delta P_{tx}|_{error} \quad (16)$$

The relation between availability and power margin depends on latitude. Because satellite elevation angles distribution is different for each latitude.

2.4.2 Realisation : To establish these availability tables, mobiles are distributed on the Earth and interference noise is not considered. The system is simulated and for each mobile, the power margin

required to guarantee the instantaneous communication quality is calculated during the duration of a communication. Considering these values, the power margin necessary to ensure sufficient communication quality can be determined. This procedure is repeated on a replication period, resulting all configurations (satellite elevation angle, blockage) are performed for each latitude. From the different values obtained, a distribution of the power margin necessary to ensure sufficient communication quality is performed for each latitude. The availability is deduced directly based on the power margin obtained.

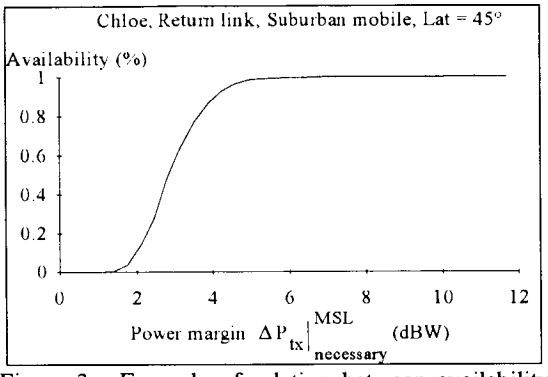


Figure 3 : Example of relation between availability and power margin.

Availability tables are constructed from distributions of the power margin necessary for each latitude. They are calculated for each propagation environment and type of terminal.

Considering availability tables, it is now possible to determine the power margin necessary to ensure sufficient communication quality for different availabilities for each latitude.

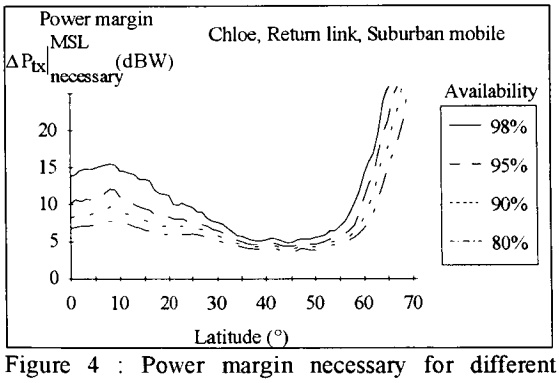


Figure 4 : Power margin necessary for different availability

2.4.3 Intrinsic maximum availability : The maximum availability is obtained when the power margin available is used totally to compensate for mobile specific losses. So, this value can be quoted intrinsic maximum availability.

If many mobiles use the system, interference noise cannot be neglected. Therefore, the real availability observed is necessarily less than the intrinsic maximum availability.

3. METHODOLOGY

3.1 Introduction

The objective of this methodology is to determine the maximum density of users. As long as the interference generated by interfering mobiles is below the maximum interference noise allowable by the system, the system can continue allocating channels to new mobiles.

As the interference noise is a function of interfering mobiles' transmit power, when a new mobile is added, interference noise for each mobile increases. So the transmit power required to guarantee the instantaneous communication quality must be increased. Therefore the previously calculated interference noise increases, and further iterations must be performed until a new stability point is reached. This iterative method is slow and cumbersome.

Our method is quite different. For each mobile, the transmit power required is calculated considering an interference noise equal to the maximum interference noise allowable by the system. Then the real interference noise is calculated for each mobile. When a new mobile is added, the real interference noise must be recalculated. As long as the new values of real interference noise is less than the maximum interference noise allowable by the system, the transmit power calculated is sufficient to ensure the instantaneous required quality. It is not necessary to determine again the required transmit power. This method, allows to determine easily the maximum density of users.

3.2 Induced interference noise

First of all, a value of target availability is fixed, representing the value from which the system is considered viable commercially. This value is necessarily less than the intrinsic maximum availability.

Considering the target availability, the power margin necessary to ensure sufficient communication quality for a mobile without considering interference noise can be deduced from availability tables.

The maximum power margin used to compensate the interference noise is equal to the difference between the power margin available and the power margin previously deduced.

i.e., $$\Delta P_{tx}\big|_{interference} = \Delta P_{tx}\big|_{available} - \Delta P_{tx}\big|_{necessary}^{MSL} \quad (17)$$

The maximum interference noise allowable by the system, termed the "induced interference noise", represent the value of interference which should be compensated by the power margin $\Delta P_{tx}\big|_{interference}$.

3.3 Maximum density of users

3.3.1 Aggregate interference noise per user : First of all, the land Earth is uniformly covered with a certain density of mobiles. For each mobile and each propagation step, the transmit power required to guarantee the instantaneous communication quality using methodology described in 3.1 has been calculated then the "aggregate interference noise" is determined for each user by summing the interferences from every mobile.

3.3.2 Algorithm for the distribution of users : The determination of the maximum density of users is realised every 2 minutes. This time duration represents the mean communication duration considered.

For each mobile, the maximum aggregate interference noise during the 2 minutes is stored. Considering this value and the induced interference noise, the initial user-density is continuously modified to ensure that the total aggregate interference noise generated by interfering mobiles on the Earth is equal to the induced interference noise.

This distribution of users, depending on their geographic position, represents the maximum density of users for the target availability previously fixed. While the density of users is less than the maximum density, actual availability is necessarily greater than target availability fixed.

4. RESULTS

This paper aims at presenting a methodology for optimising the system capacity in a RORA/CDMA satellite system. The results presented in this paper depends directly on the associated assumptions. It is just interesting to compare the results themselves. For this raison, simulations are performed only on a communication duration (2 minutes). When a propagation environment is used, all mobiles considered are in this propagation environment.

4.1 Inputs

All examples presented in this paper deal with :
- a Walker constellation named CHLOE
- a transmit power maximum equal to 2 dBW
- simulated antennae S and L band 1-6-9 spot beam configuration

4.2 Maps of maximum density of users

The results are displayed in the form of maps with density areas. The position of satellites are indicated by a cross. Figure 5 and 6 represents the maximum density of users per carrier for respectively 98% and 90% of availability.

The interference noise varies only slightly between different users and with time due to averaging over a large number of interfering users.

The density of users in 30°-60° latitude is greater than the density of users at the equator because the visibility of satellite is best for the temperate latitude.

The density of users in the coastal area is generally greater than the density of users in the inland region because the number of potentially interfering mobiles is lower.

4.3 Number of users

Based on the maps of maximum density of users, the maximum number of users can be determined.

Table 1 represents the maximum number of users for a suburban mobile terminal and rural mobile terminal.

			Switched Diversity	Combined Diversity
Suburban Mobile	Forward Link	98 %	1887	1847
		95 %	2302	2086
		90 %	2498	2208
	Return Link	98 %	2552	4244
		95 %	3206	4865
		90 %	3469	5128
Rural Mobile	Forward Link	98 %	2615	2410
		95 %	2804	2515
		90 %	2891	2575
	Return Link	98 %	3077	4785
		95 %	3277	5000
		90 %	3382	5152

Table 1 : Maximum simultaneous number of mobiles

The rural environment being less attenuate than the suburban environment, the power margin necessary to ensure the sufficient communication quality is lower, resulting an upper number of mobiles.

In return link, diversity is natural due to the isotropic terminal antenna. So it is beneficial to combine the different links. In forward link, diversity is not automatic to use combined diversity. The signal must be transmitted from different transmitters. It is an

advantage to combine different links but in parallel interference noise increases. Globally the number of mobiles obtained from the combined diversity and from the switched diversity is equal.

5. CONCLUSION

This paper has presented a methodology for optimising the system capacity in a RORA/ CDMA satellite system. Considering this method, the maximum number of users per country can be estimated.

The importance of the assumptions used to run this simulations must be emphasised. Especially, the actual results depends directly on the availability criterion chosen here.

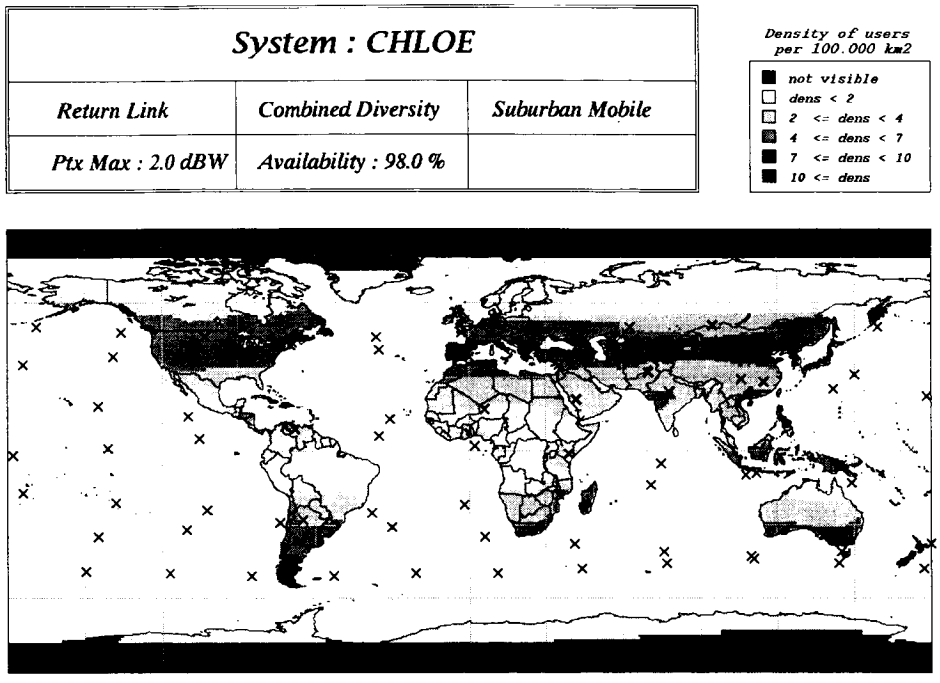


Figure 5 : Maximum density of users for 98% of availability in return link

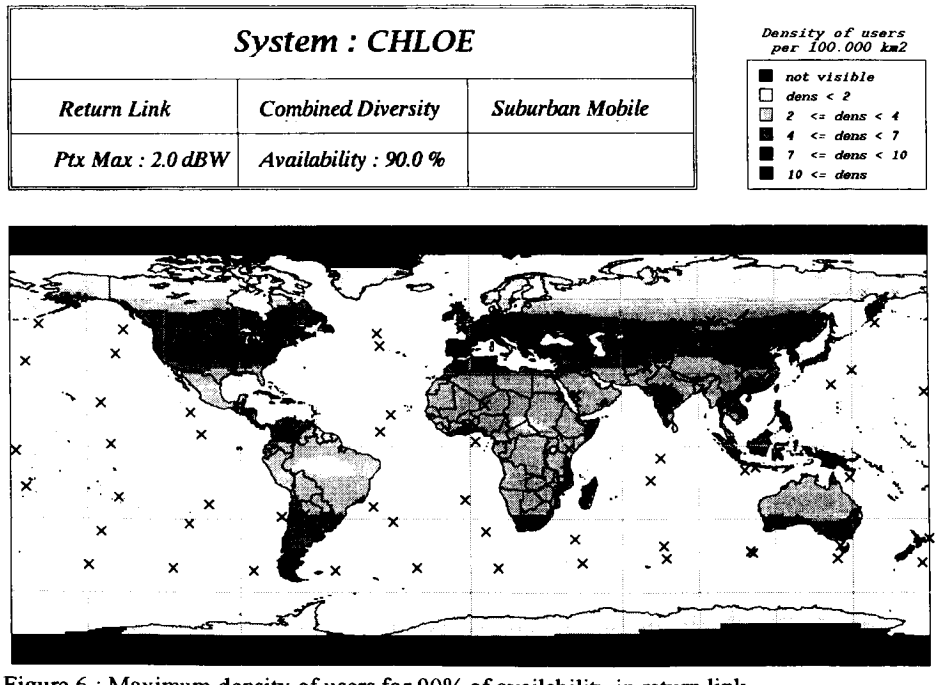


Figure 6 : Maximum density of users for 90% of availability in return link

[1] S.Blondeau & al., "Region oriented resource allocation methodology for non-geostationary mobile communications satellite systems", Space communications - special issue on satellite mobile and personal communications Volume 13, Number 3, 1995, pp 171-191.

[2] J.M. Conrat & P.Pajusco, "Fading and shadowing analysis of mobile satellite propagation channel at 1.5 and 2.2 GHz using a helicopter", ACTS Mobile Telecommunications Summit, Granada, Spain November 27-29 1996.

Traffic Flows and Dynamic Routing in LEO Intersatellite Link Networks

Markus Werner
German Aerospace
Research Establishment (DLR)
Markus.Werner@dlr.de

Gérard Maral
Ecole Nationale Supérieure
des Télécommunications (ENST)
maral@tlse.enst.fr

ABSTRACT

The focus of this paper is on dynamic routing schemes applied to low earth orbit (LEO) intersatellite link (ISL) trunk subnetworks in the context of both deterministic dynamic network topology and time-variable source traffic demand. A discrete-time modelling approach for both is presented in detail. The cause-and-effect relationship between routing scheme and network traffic flows is discussed. Considering a deterministic routing scheme as reference, special emphasis is laid on a traffic adaptive routing solution where this relationship becomes mutually interactive. The performance evaluation is carried out for Iridium as example LEO constellation, and covers distribution of global traffic flows, maximum and average ISL load statistics, and transfer delay statistics.

It is shown that the challenge of routing in the LEO ISL environment with considerable variation in both, topology and traffic, can be tackled particularly with distributed traffic adaptive approaches.

I. INTRODUCTION

With emerging low earth orbit (LEO) satellite systems for global personal and future broadband communications, the satellite communications community witnesses the advent of really new and severe networking challenges. This holds especially for advanced system concepts employing intersatellite links (ISLs), like Iridium [1], Teledesic [2], and M-Star [3]. Table I summarizes relevant parameters of the three candidates.

In particular, the routing of complex global traffic flows over dynamic network topologies requires sophisticated solutions. So far, this problem has stimulated only little (published) research work compared to other important topics in the LEO satellite field. Nevertheless, its importance for the design of future systems is expected to even grow in the light of broadband applications, probably accelerating the use of ATM networking concepts as well as of optical intersatellite links. Efficient routing is one of the key performance drivers in all broadband multiservice networks, especially if real-time connection-oriented services dominate.

Within this paper, we reduce the manyfold complexity of the topic by restricting ourselves to the following:

(1) All considerations and simulations are based on the Iridium satellite system and its ISL topology as an example. However they should principally be adaptable to other (potential) LEO systems employing ISLs.

(2) We assume a simple mono-service scenario considering plain old telephony (which is, by the way, the major market for the "first generation" LEO system Iridium). Fundamental results should nevertheless provide some indication on the performance to be expected in a connection-oriented multi-service environment.

(3) We slightly adapt two routing algorithms as known from terrestrial (computer) networks, with a clear indication that their use in virtual circuit networks does not impose fundamental problems.

The paper is organized as follows: First the time-discrete modelling of ISL topology (Section II) and of traffic demand/parameters (Section III) is presented in detail. Then two dynamic routing schemes – a deterministic and a traffic adaptive one – are introduced in Section IV. Simulation results for both schemes are finally reported and discussed in Section V.

TABLE I
PARAMETERS OF LEO SATELLITE SYSTEMS EMPLOYING ISLS.

	Iridium	Teledesic	M-Star
Orbit altitude	780 km	700 km	1350 km
Orbit period	100 min	99 min	113 min
Number of satellites	66	840	72
Number of orbits	6	21	12
Inclination	86.4°	98.2°	47°
Intraplane ISLs per sat.	2 perman.	up to 4	2 perman.
Interplane ISLs per sat.	up to 2	up to 4	2 perman.
Primary service	narrowband, voice	broadband, multimedia	broadband, real-time V/D

II. ISL NETWORK TOPOLOGY

One characteristic feature of LEO constellations is their inherent network *dynamics*. For the networking of such systems, and especially of the ISL space segment, it is therefore essential to exploit the fact that the constellations are also (periodically) *deterministic*. It is then straightforward to approach the problem by modelling it in time-discrete manner with sufficiently fine time resolution.

A. Dynamic Topology

The physical time-variable ISL topology of the system consists of all instantaneously existing direct links between pairs of satellites. Here a distinction has to be made between *intraplane* ISLs connecting subsequent satellites in the same

orbit and *interplane* ISLs connecting satellites in adjacent co-rotating orbits, as shown in the snapshot in Fig. 1. Whereas in the first case the distance between the participating satellites and the antenna pointing are fixed, interplane ISLs are subject to continuous variations of both, with obvious impact on the routing. In the extreme case of contra-rotating orbits, moreover, a permanent switching of ISLs would be necessary, a fact that has led to avoiding links across the resulting *seam* in general, cf. Fig. 1. Besides the *continuous* distance changes on co-rotating interplane ISLs there is also a *time-discrete* contribution to the ISL topology dynamics: The effectively implemented interplane ISLs (one per satellite and per neighbouring co-rotating orbital plane) are deactivated in polar regions; this means on/off switching of certain links in the topology. Lacking published information concerning the exact switching criterion, we have assumed always 4 interplane ISLs between a pair of adjacent planes (roughly centered around the equator), which results in a constant number of 106 ISLs in the topology. The two-fold dynamics of the ISL subnetwork significantly increases the complexity of connection-oriented network operation and has to be tackled by tailor-made routing strategies.

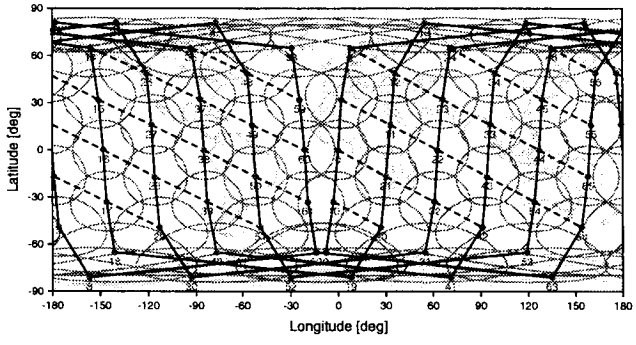


Fig. 1. Snapshot of the Iridium ISL topology. Solid lines: intraplane ISLs, dashed lines: interplane ISLs.

B. Deterministic Topology

The proposed LEO constellations are typically based on circular orbits and a regular phasing between the satellites (except Teledesic's random phasing). This implies that all satellites in the constellation – and with them of course all ISLs – periodically (with orbit period T_O) encounter exactly the same positions relative to each other.

However, since we are interested in the transport of traffic uplinked from Earth, we have additionally to take into account the rotation of the Earth beneath the satellite constellation. Firstly orbital mechanics guarantees that every satellite will orbit over the exact same points on Earth after a certain time, which is the smallest common integer multiple of the orbit period and one sidereal day (1436 min). On the other hand, one can expect that global traffic generation patterns typically vary with a period of an Earth day (1440 min), not accounting for e.g. weekly variations. Then – neglecting the small and systematic error introduced by assuming that sidereal day and Earth day are the same – we can define as *system period* T_s the smallest common integer multiple of the orbit period and one Earth day. For Iridium this yields T_s = 5 days = 7200 min corresponding to 72 times orbiting Earth. This system period is now an important parameter for all following investigations, because it specifies that the *complete* scenario, i.e. the constellation/traffic "pair", is also periodically deterministic – at least in a sense that allows the off-line investigation of traffic flows (based on a proper traffic model as presented in the next section) and also the off-line set-up of routing tables in a deterministic routing approach.

III. Traffic Modelling

The focus of this paper is on the ISL network *internal* traffic flows and routing schemes. The serving source/destination satellites are consequently regarded as collective traffic sources and drains, not considering single end user connections. Nevertheless we derive our figures for source traffic demand keeping in mind that single calls with their specific parameters form the "atomic" traffic entity for our study. The complete modelling consists of the following steps:

1. Service characterization (mono-service call model)

In order to limit the various parameters contributing to numerical results just from the beginning, we restrict ourselves to consideration of a mono-service telephone scenario. The most comprehensive market studies are also available for this case. It should of course be possible to learn general lessons from these investigations and to already project distinct results onto the situation in future integrated services networks. In the following, we assume the classical call model with Poisson arrival rate λ and negative exponentially distributed holding time, mean value $T = \frac{1}{\mu} = 3$ min.

2. Source traffic demand (addressable market model)

The long-term source traffic demand can be expected to be largely dependent on geographical subscriber distribution and other market factors like GDP, service penetration/acceptance, etc. "Long-term" means that daily or weekly demand variations are not considered yet. One of the most comprehensive studies that fits our needs in the context of voice over LEO satellite systems has been published as MSc thesis at MIT [4]. Essentially, it provides source traffic maps forecasting – with a resolution of 15° by 15° earth zones – the expected 100% market in millions of addressable minutes per year beyond 2000. Fig. 2 displays the earth zones and their traffic intensity levels, the latter encoding the traffic quantities according to Table III. The figures show the forecast for the year 2005. This traffic map has been adopted for our simulations.

3. Daily user activity (multihour model)

Trafic in a global communication system shows a multihour characteristic; thus a busy hour model is not valid since it introduces unacceptable inaccuracies in the basic source traffic assumptions. The daily evolution of traffic activity

per user is assumed to be the same for all users world-wide and according to Fig. 3. Further, the local time of each traffic zone is assumed to be equal to the solar time of the respective zone's centre longitude. Then it is possible to extract a daily activity coefficient from the histogram. All zones' (long-term) source traffic values are then weighted with their corresponding coefficient and scaled down properly so as to arrive at the effective demand (in number of call arrivals) in a given interval. The latter is equal to the interval used for time-discrete modelling of the constellation.

4. Call destinations (traffic flow model)

The second but last step takes care of a proper assessment of the destination zones for generated calls. For this purpose, the Earth is subdivided into six continental regions as displayed in Fig. 2, each one containing a unique subset of the 15°by 15°zones. The traffic flow between these regions is assumed according to Table III [5]. The figures reflect that most of the traffic is local. As far as the distribution of drain traffic *within* a given continent is concerned, we divide it as function of the instantaneous zone activity.

5. Traffic mapping (coverage mapping model)

The traffic generated on and delivered to Earth is allocated to the different satellites according to the fraction of the zones within their respective footprint.

TABLE II

TRAFFIC INTENSITY LEVELS AND CORRESPONDING EXPECTED MARKET (2005) IN MILL. OF ADDRESSABLE MINUTES/YEAR. [4]

intensity level	1	2	3	4	5	6	7	8
traffic(mill.min/year)	1.6	6.4	16	32	95	191	239	318

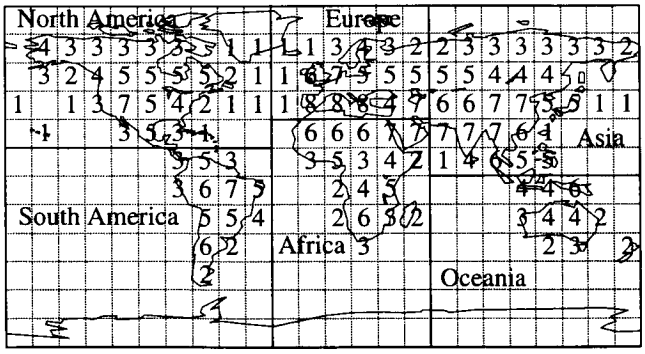


Fig. 2. Earth zone division and regional traffic intensity levels (2005). [4]

IV. Dynamic and Traffic Adaptive Routing

In the context of terrestrial (computer) networks, the concepts of *dynamic* and *adaptive* routing are often used synonymously since all those networks' topology is practically static (except in link or node failure situations) and the only real dynamics is introduced through traffic variations. A deterministic routing scheme is then usually also a static one based on fixed routing tables that are calculated off-line.

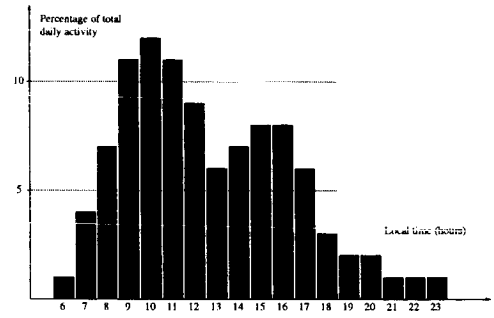


Fig. 3. Daily user activity.

TABLE III

REGIONAL TRAFFIC FLOW SHARES IN %.

	Destination					
Source	N.A.	Eur.	Asia	S.A.	Afr.	Ocea.
North America	85	4	4	3	2	2
Europe	4	85	4	3	3	1
Asia	5	5	83	1	2	4
South America	7	7	2	81	2	1
Africa	5	7	4	2	81	1
Oceania	5	2	7	1	1	84

In a dynamic topology scenario as encountered in LEO ISL networks, we have to redefine and clearly separate both concepts. Since the topology dynamics is (periodically) deterministic, is can be appropriately modeled by a time-discrete series of quasi-static topologies with a time step Δt_s in between. Routing in such a network becomes inevitably dynamic as it has to perform path search for every of these topologies and ends up with a corresponding series of updated routing tables. The update period is clearly linked to the selected time step, $\Delta t_u = \Delta t_s$. If a connection lasts longer than a step, it may be required to switch the path it follows according to the new routing table. This case of path handover with all involved problems like severe delay jitter has been extensively studied and reported in [6]. To sum up, the time-discrete approach inherently provides the adaptation of well-known static routing schemes to our dynamic scenario. For the purpose of our simulations, we have chosen to implement a deterministic routing scheme based on a version of the Dijkstra Shortest Path Algorithm [7] – the Moore-Dijkstra Algorithm (MDA) – using exclusively geometrical ISL lengths as link costs. Therefore the algorithm determines shortest routes respectively routes with minimum transfer delay. This scheme will be mainly used as reference for the performance evaluation of the considered traffic adaptive approaches.

Traffic adaptive routing is in the ISL scenario mainly motivated by the goal to use the on-board ISL capacity as efficiently as possible. This can be significantly improved by smoothening extreme peaks in the ISL loads as encountered without on-line traffic adaptive routing. Strong time variations in the ISL load can become particularly critical considering the expected high data rates in broadband systems, since

all satellites in the constellation have to be designed for the worst case – including ISL payload, power and antenna requirements. Using the expensive orbital resources sparingly will thus directly reduce the costs for every single satellite and the whole system.

The study of adaptive routing schemes has shown that efficient implementations work in *distributed* or *decentralized* manner. According to [8], a distributed traffic adaptive routing procedure generally consists of the following three components: (1) A measurement process that collects the required information on the actual traffic situation in the network, (2) an update protocol that distributes (e.g. by flooding) this information to all nodes, and (3) the shortest path computation to be performed in every single node, using the updated traffic data built into the respective link cost function. Clearly, such a scheme is also dynamic since all topological changes have to be reported to all nodes as well during the update period, or they are built-in because they are deterministic and therefore known in advance. An important parameter for adaptive routing is the interval Δt_u between successive updates of the routing tables. In particular, the relationship between the update period and the typical duration of connections is essential for a reasonable trade-off between performance gain and complexity. Typically, there is some lower bound for Δt_u below which only little performance improvement can be achieved at the expense of unacceptably high update protocol traffic.

For our simulations we have implemented a traffic adaptive scheme where the Distributed Bellman-Ford (DBF) Algorithm [7] is used for the path search. Three different path cost functions have been investigated in connection with this scheme: The first two are based on the sum of the costs of all participating *links*, where (1) the link cost is a function of link traffic alone, (2) the link cost is a metric combining link traffic and link length, whereas the third one (3) uses the sum of *node* costs (only traffic based).

V. Performance Evaluation

The numerical results reported in this section are extracted from an in-depth study of the topic [9].

The following parameter settings have been used: All simulations have been performed over the complete system period $T_s = 5$ days in order to guarantee that all possible pairs of source traffic patterns and orbital patterns are considered. The time step Δt_s has been 2 minutes throughout. For DBF's Δt_u, values ranging from 12 seconds up to 2 minutes have been evaluated in test runs; from that, an optimum value of 20 seconds has been fixed for all simulations.

Global traffic distribution. Fig. 4 displays the pronounced asymmetry of ISL loads at the beginning of the simulation (shortest length paths). This is mainly due to the combination of strong geographical source traffic variations and the bounded ISL topology. Fig. 5(a) clearly shows that this asymmetry is maintained with the deterministic MDA routing scheme. Note that with 106 maintained ISLs a uniform distribution across the network would imply roughly 1% of the total network traffic to be carried on each link. This is far from being the case: the ten most loaded links carry roughly half of it. In contrast to this, (b) proves that the traffic adaptive DBF scheme significantly improves a more even traffic distribution in the network: except one link, all carry less than 2.5% of total network traffic, the ten most loaded links carry roughly 20% together, and 40 links carry around 1% of the total traffic each.

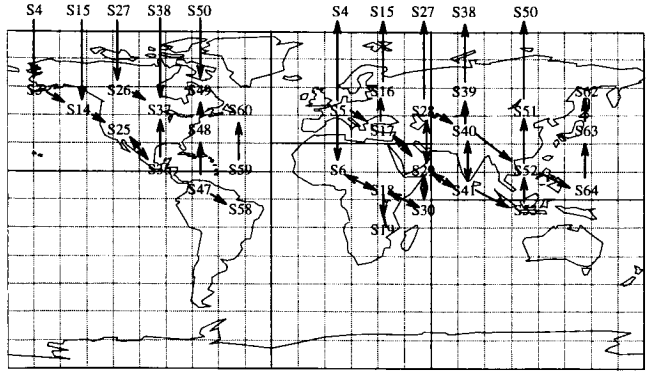


Fig. 4. Most loaded ISLs at $t = t_0$. Each displayed link carries more than 2% of the total network traffic.

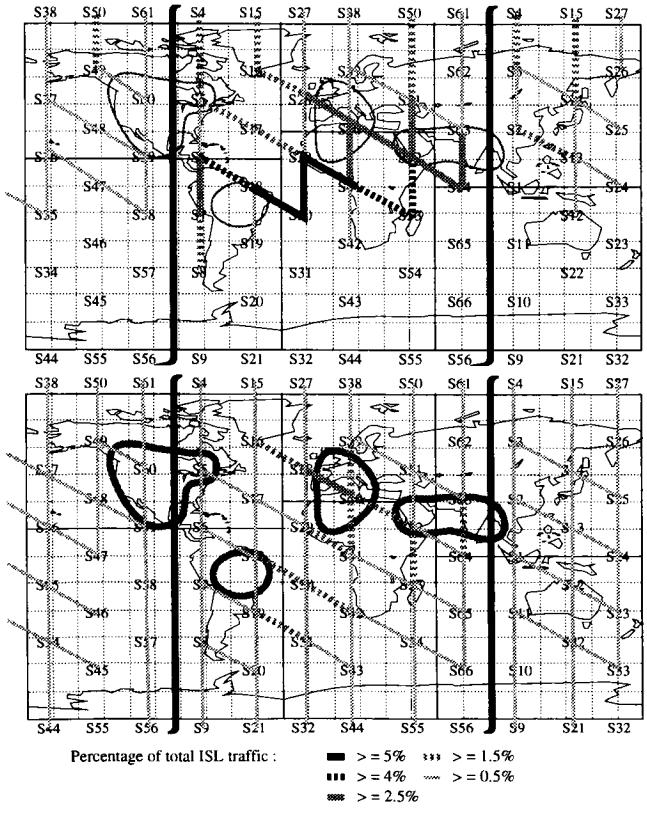


Fig. 5. ISL traffic distribution with (a) MDA and (b) DBF at $t = 300$min. The seam appears over Eastern U.S. and India/China. High source traffic areas are highlighted.

For the following statistical evaluations of ISL and satellite

load it should be noted that all figures are normalized to a certain constant total uplinked source traffic. By doing this, the strong influence of daily activity variations is excluded and the effects of the routing schemes can be exposed more clearly.

MDA/DBF comparison: ISL load statistics. The highest traffic load encountered on a single link at a given instant – which is called maximum ISL traffic in the following – is one of the most important performance parameters since it finally determines the ISL power requirements and thus the ISL link budget. It is therefore highly interesting to evaluate the reduction potential of DBF for this quantity. Fig. 6(a) illustrates that DBF can reduce the maximum ISL traffic by roughly 30% in the average, the instantaneous value being strongly dependent on the geographical source traffic pattern: The reduction tends to be smaller when most generated traffic is concentrated on a few adjacent zones on Earth. DBF's capability to distribute traffic more evenly in the network is also stressed taking another viewpoint, as illustrated in Fig. 6(b). The worst case link never carries less than 4.5% of total network traffic with MDA, but only 2-3% with DBF. The price to be paid for the reduction of the worst case ISL traffic is obviously quite a constant increase in the average ISL traffic load, Fig. 6(c). This is due to the fact that at a given time step DBF prefers previously less loaded links, whereas the load is decreased on the critical ones. Thus the group of ISLs carrying medium traffic loads grows compared to MDA, Fig. 6(d).

Influence of MDA cost function. Fig. 7 focuses on the comparison of a simple hop metric with the (reference) link distance metric for MDA. The clear difference in resulting average ISL traffic is due to the fact that the interorbital distance roughly doubles between the instant the ISLs are switched on and the instant they are crossing the equator. Then, given that we often have pronounced East-West traffic flows between regions near the equator (cf. Fig. 5), the distance-based algorithm would often tend to route this traffic over mid-latitude interplane ISLs due to their relative shortness, whereas the hop-based algorithm would select the "direct" (less hops) route along the equator.

Total satellite load and influence of DBF cost function. Besides the worst case ISL load it is also important to figure out the maximum total satellite load, which is simply the sum of all ISL loads for the worst case satellite. This value is a crucial input to the calculation of the total onboard power requirements for all maintained ISLs. Fig. 8 displays corresponding (normalized) figures for MDA and three versions of DBF. The link state DBF with a combined (logarithmic) traffic/distance cost metric ("log cost") shows an overall poor performance – like for ISL traffic, by the way. Both DBF schemes with pure traffic cost metric perform clearly better than MDA. As one expects, the node state version shows a slightly better performance than the link state one with respect to satellite (= node) traffic. However, the 4% gain in total satellite load has to be paid for by an increase of the same percentage in ISL load, compared to link state DBF.

Link hops and transfer time. Finally, Fig. 9 illustrates that transfer time need *not* necessarily be the price that has to be paid for any capacity saving achieved through traffic adaptive routing – which is of course a welcome result. The average number of hops is 4 for MDA and 4.5 for DBF. Based on Fig. 9(a), one additionally learns that DBF's efficiency in the ISL scenario does *not* result from changing paths from 1 to 2 hops but indeed from selecting typically 8-12 hop paths instead of 3-7 hop paths. This again implies that DBF's efficiency temporarily grows when there are two (or several) *remote* high source traffic areas. Despite the increase in the average number of hops, the transfer time keeps practically constant at roughly 55 ms, a result one would *not* expect at a first glance. Fig. 9(b) does of course show a slightly widened transfer time distribution for DBF, yet the average value is not affected. It is difficult to give a simple explanation for this phenomenon, since a lot of inluences come together. However one of the more obvious reasons is the different update period for both schemes, since the MDA scheme with (distance-based cost) yields on principle the lowest possible transfer time for a fixed update period.

VI. Conclusions and Outlook

The paper has essentially presented dynamic routing concepts for LEO intersatellite link networks. Based on a detailed time-discrete modelling of both, dynamic network topology and time-varying source traffic demand, a deterministic and a traffic adaptive scheme have been compared. The evaluation yields a superior performance of the more sophisticated traffic adaptive scheme in terms of significantly reduced maximum link and satellite loads and a more even distribution of network traffic. The performance gain is achieved mainly at the expense of a moderate increase in average link and network traffic. The average transfer time of a connection, on the other hand, is not seriously affected. Besides an optimization of several parameters in the routing procedures, future work should extend the investigations towards other ISL topologies and broadband integrated service scenarios.

References

[1] J. Hutcheson and M. Laurin. Network flexibility of the IRIDIUM global mobile satellite system. In *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, pages 503–507, Ottawa, Canada, June 1995.

[2] M. A. Sturza. Architecture of the TELEDESIC satellite system. In *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, pages 212–218, Ottawa, Canada, June 1995.

[3] J. C. Anselmo. Motorola unveils new satcom plan. *Aviation Week & Space Technology*, page 35, Sept.16, 1996.

[4] M. D. Violet. The development and application of a cost per minute metric for the evaluation of mobile satellite systems in a limited-growth voice communications market. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, Sept. 1995.

[5] M. Werner. Analysis of system connectivity and traffic capacity requirements for LEO/MEO S-PCNs. In E. Del Re, editor, *Mobile*

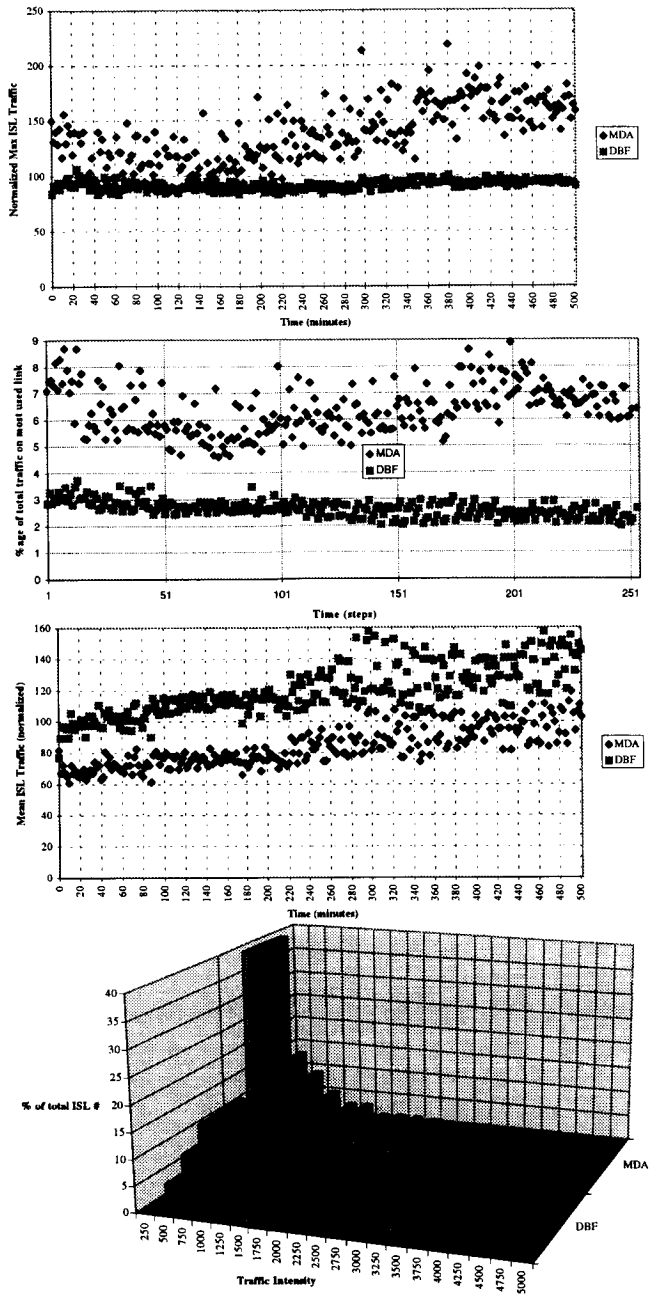


Fig. 6. MDA/DBF comparison: (a) maximum ISL load, (b) percentage of traffic on the most loaded link, (c) average ISL load, (d) ISL load distribution.

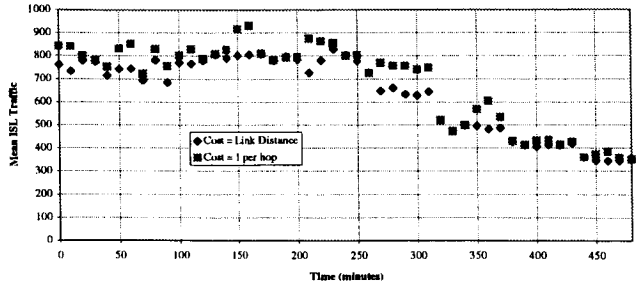


Fig. 7. MDA cost alternative.

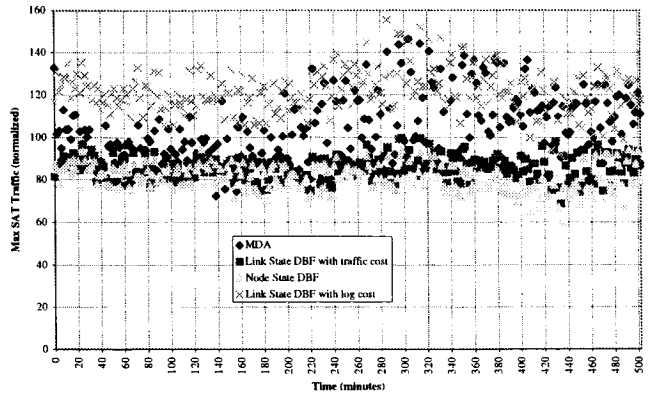


Fig. 8. Satellite load and DBF cost alternatives.

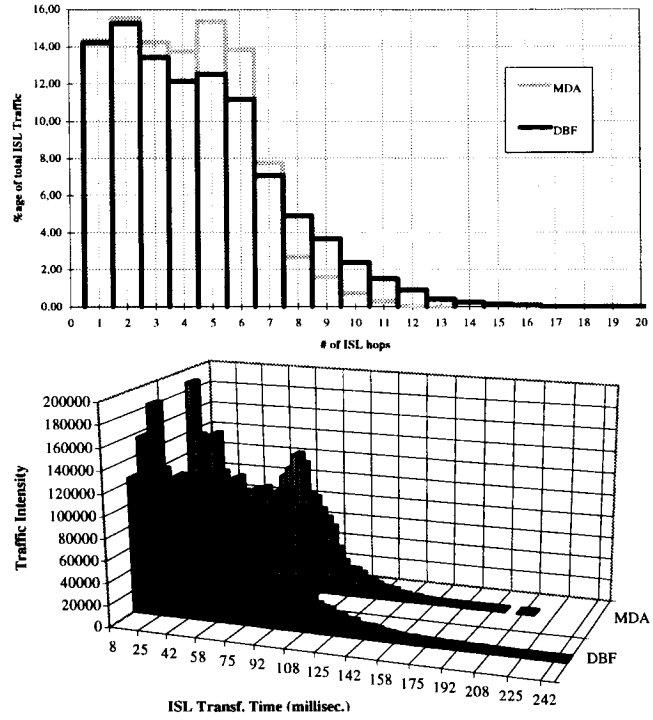


Fig. 9. MDA/DBF comparison: Distribution of (a) number of hops, (b) transfer time per connection.

and Personal Communications, Proceedings 2nd Joint COST 227/231 Workshop, pages 183–204, Florence, Italy, Apr. 1995. Elsevier.

[6] M. Werner, C. Delucchi, H.-J. Vögel, G. Maral, and J.-J. De Ridder. ATM-based routing in LEO/MEO satellite networks with intersatellite links. *IEEE Journal on Selected Areas in Communications*, 15(1):69–82, Jan. 1997.

[7] D. Bertsekas and R. G. Gallager. *Data Networks*, chapter 5.2. Prentice-Hall, Englewood Cliffs, N.J., 1987.

[8] M. Schwartz. *Telecommunication Networks: Protocols, Modeling and Analysis*. Series in Electrical and Computer Engineering. Addison Wesley, Reading, MA, 1987.

[9] B. Garnier. Network concepts for ATM traffic flows in low earth orbit satellite constellations with intersatellite links (ISL). Master's thesis, ENST Toulouse / DLR Oberpfaffenhofen, Toulouse, France / Weßling, Germany, Sept. 1996.

Analysis of intersatellite links load in a near polar LEO satellite constellation

Markus Werner
DLR Oberpfaffenhofen
Institut fuer
Nachrichtentechnik
Postfach 11 16
D-82230 Wessling, Germany
Markus.Werner@dlr.de

Oliver Kroner
NERA GmbH
Koenigsteiner Str. 76
D-65812 Bad Soden,
Germany

Gerard Marai
Ecole Nationale Superieure
des Telecommunications
(ENST)
Site of Toulouse
BP 4004
31028 Toulouse Cedex,
France
maral@tlse.enst.fr

ABSTRACT

The load of Intersatellite links (ISL) in a near polar low earth orbit (LEO) satellite constellation is analysed. A strong dissimilarity is demonstrated between ISLs connected to seam and non-seam satellites. Also, the load for intraplane ISLs looks quite different from that of interplane ISLs. Explanations are given for these differences, thus highlighting the mechanisms of ISL traffic routing, and the impact of the considered factors.

INTRODUCTION

Low Earth Orbit and Medium Earth Orbit (LEO/MEO) satellite constellations are alternatives to geostationary satellite systems for providing global mobile and fixed services. The use of intersatellite links has been identified as a means to provide global connectivity, and has been retained in the design of LEO systems like Iridium and Teledesic [1][2].

Polar or near polar LEO constellations comprise several orbital planes, with orbits intersecting at or near the poles. Satellites in adjacent planes are either co-rotating satellites, i.e. the satellites rotate in the same direction, or are, for two specific adjacent planes, contra-rotating. Each satellite is connected to its immediate neighbours by intersatellite links (ISLs). ISLs can be easily established within orbital planes (such links are named 'intraplane' ISLs), as a result of the nearly fixed geometry of satellites one with respect to another. ISLs connecting satellites in adjacent planes ('interplane' ISLs) are envisaged only between co-rotating satellites, as the relative geometry between contra-rotating satellites evolves too rapidly. The absence of ISLs between contra-rotating satellites introduces a discontinuity in the ISL network, named 'seam'. The length of intraplane ISLs is constant, while that of interplane ISLs varies with time and is the largest at the equator.

The load of intersatellite links depends on factors such as :
- the geographic pattern and time variation of the uplinked source traffic
- the procedure for routing the traffic,
- the inhomogeneity of the network topology as a result of the seam.

The paper aims at analysing the influence of the above factors on the load of individual intersatellite links. The investigations have been performed using a simulation package that generates traffic, establishes ISL routing, and displays the ISL loads. For the routing procedure, each ISL is attributed a 'link cost'. A route between any two satellites in the constellation is given a 'route cost', equal to the sum of the individual costs of the participating ISLs. The routing strategy then selects the route with minimum cost for information transport. Both the loads of interplane and intraplane ISL are investigated. In order to separate the influence of each factor, the source traffic is first considered to be homogeneous (no geographic nor time dependency). Then, an inhomogeneous traffic distribution is considered, incorporating a traffic matrix between regions of the world. To point out the influence of the seam, results are given in both cases for ISL links connected to a non-seam satellite, and ISL links connected to a seam satellite.

It should be noted that this study only aims at exposing the influence of the seam for two types of traffic generation, and does not claim at providing results for a realistic traffic generation case, as the variations of the uplinked source traffic with local time are ignored. This aspect is presently under investigation, and will be reported later on.

The paper is organised as follows : first the topology of the constellation and its associated ISL network are presented. Then the routing algorithm and its associated link cost function are introduced. Simulation results for the homogeneous traffic distribution are reported, followed by results for the inhomogeneous traffic distribution.

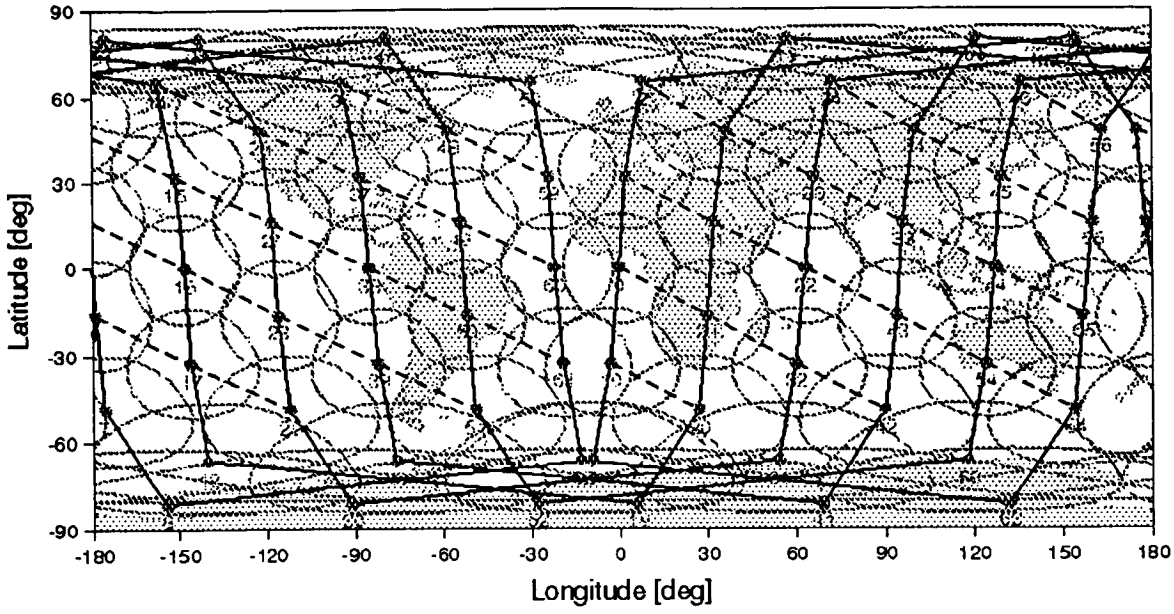


Figure 1 : The selected satellite constellation with ISLs (solid lines : intraplane ISLs, dashed lines : interplane ISLs)

SATELLITE CONSTELLATION TOPOLOGY

Figure 1 displays the constellation topology at a given time and its associated ISL network. The considered constellation is that of the Iridium system, which consists of 66 satellites distributed in 6 planes (11 satellites per plane) with inclination i = 86.4° with respect to the earth equatorial plane. Orbits are circular, and satellite altitude is 780 km, with orbit period equal to 100 minutes. Each satellite is represented with its associated coverage footprint. Solid lines between satellites indicate intraplane ISLs, while dashed lines show interplane ISLs. The seam appears clearly with no interplane ISLs over regions about meridian -10° at the time of the figure. As the earth rotates beneath the constellation, the seam moves with time towards the east, at a rate of 15 degrees per hour.

From this constellation, two satellites will be examined : satellite 42, one of the co-rotating satellites away from the seam, and satellite 11 on the seam, among the set of contra-rotating satellites. Figure 2 and 3 identify the ISLs originating from those two satellites respectively. Figure 4 to 7 indicate which ISLs are active while the satellites proceed northward and southward. These figures illustrate the effect of orbit crossing at the poles. Intraplane ISLs are active during the whole orbit, while the interplane ISLs are switched off when satellites approach the poles.

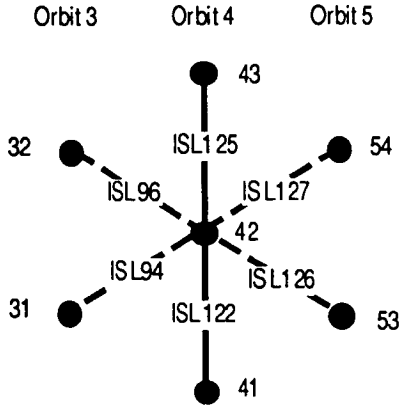


Figure 2 : Numbering of ISLs linking satellite 42 to its neighbours (solid line : intraplane ISLs, dashed line : interplane ISLs)

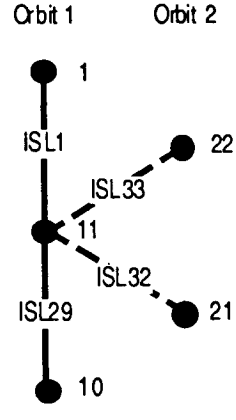


Figure 3 : Numbering of ISLs linking satellite 11 to its neighbours (solid line : intraplane ISLs, dashed line : interplane ISLs)

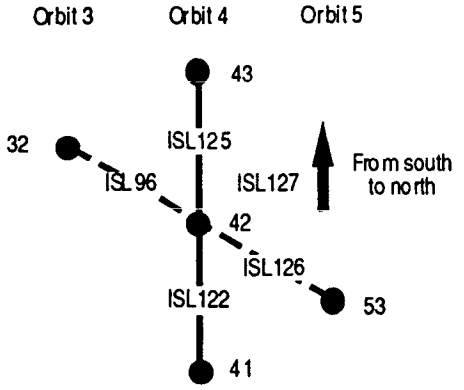


Figure 4 : Active ISLs of satellite 42 while the satellite proceeds northward

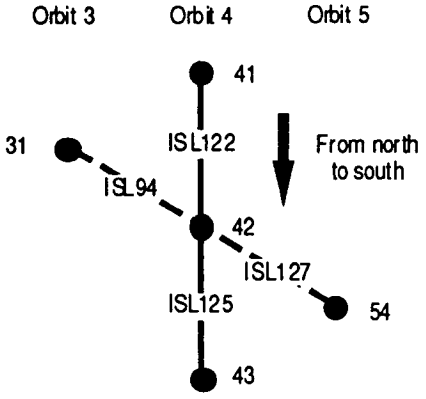


Figure 5 : Active ISLs of satellite 42 while the satellite proceeds southward

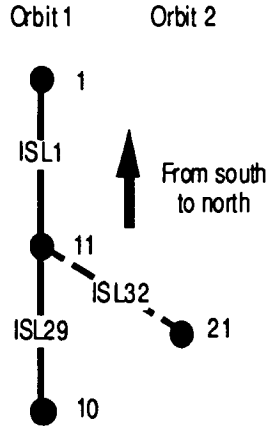


Figure 6 : Active ISLs of satellite 11 while the satellite proceeds northward

ROUTING ALGORITHM

The traffic between any two satellites in the constellation follows a route determined from the Dijkstra algorithm [3, pp 268-270]. The algorithm selects the route with minimum cost, where the route cost is the sum of the cost of all links participating in the considered route. Here, the retained 'link cost' is its length, i.e. the distance between the two linked satellites. Therefore, the algorithm determines the shortest route.

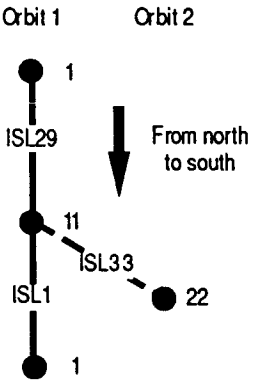


Figure 7 : Active ISLs of satellite 11 while the satellite proceeds southward

HOMOGENEOUS TRAFFIC DISTRIBUTION

This case considers traffic flow between any two satellites in the constellation equal to 1 erlang. This assumption allows to investigate the influence of the routing algorithm on the load of interplane and intraplane ISLs for a non-seam satellite (satellite 42), and for a seam satellite (satellite 11).

ISL load for a non-seam satellite (satellite 42)

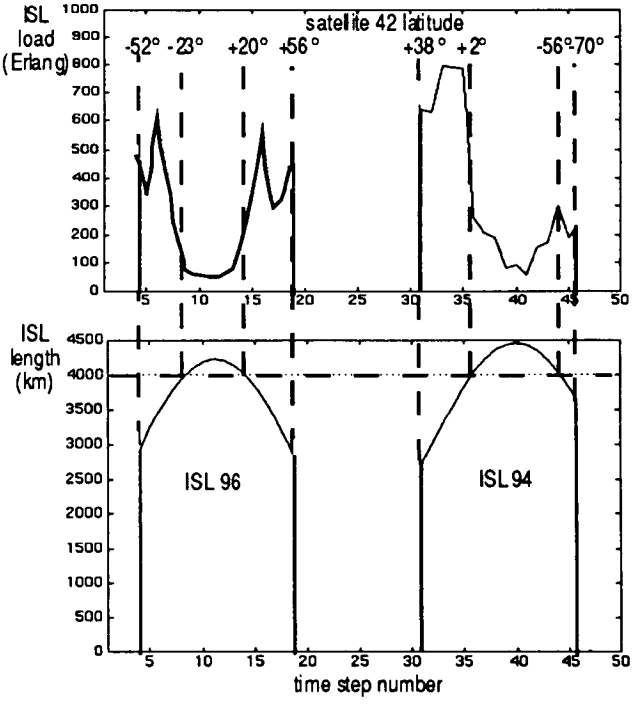


Figure 8 : Load of interplane ISL94 and ISL96 during one orbit period (homogeneous traffic distribution).

Figure 8 and 9 display the load of interplane ISL94 and ISL96 during one orbit period. The lower part of the figure shows at each time step the length of the considered ISL from the time it is switched on to the time it is switched off. The horizontal dashed line at 4000 km indicates the length of an intraplane ISL (which does not depend on time). The horizontal axis indicates the time steps and the corresponding satellite latitude. Each time step corresponds to a 2 minute interval, during which the satellite moves about 7.2 degrees in latitude.

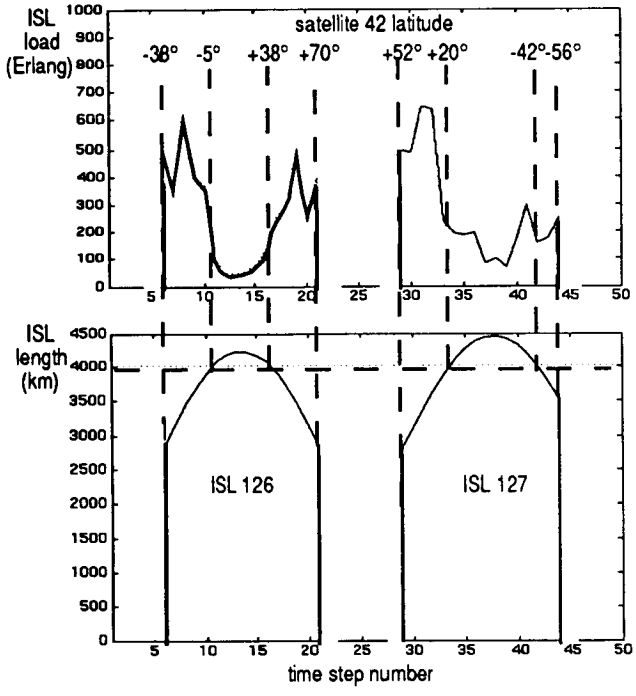


Figure 9 : Load of interplane ISL126 and ISL127 during one orbit period (homogeneous traffic distribution).

Both figures 8 and 9 indicate that interplane ISLs are more loaded when they are shorter than intraplane ISLs. This occurs when the satellite is orbiting over mid-latitude regions. This can be interpreted as a consequence of the routing algorithm selecting as participating links the shortest links : interplane links are selected preferably to interplane links at those latitudes where they are shorter. Also, one can notice that the northbound interplane ISLs have nearly equal peak load at mid-latitudes in the southern and northern hemispheres, while the southbound interplane ISLs are more loaded in the northern hemisphere than in the southern hemisphere.

Figure 10 shows the load of intraplane ISL122 and ISL125 with time. The time diagram of ISL122 displays time shifted variations of the ISL125 curve, with delay equal to the time it takes ISL122 to occupy a geographical latitude span formerly occupied by ISL125. The load appears as a composite traffic with a nearly constant component at about 150 erlangs and a peak-like component, which peaks at about 300 erlangs above the constant component. The peak-like component appears when the satellite travels at mid latitudes.

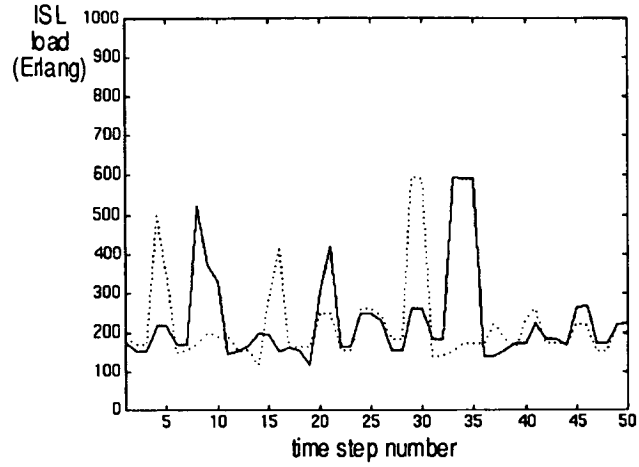


Figure 10 : Load of intraplane ISL122 (solid line) and ISL125 (dotted line) during one orbit period (homogeneous traffic distribution).

The constant component can be interpreted as resulting from the north-south oriented traffic, which is independent of latitude for two reasons :
1- the interplane ISLs are poor candidates for conveying such traffic as they are east-west oriented,
2 - the intraplane ISLs have a constant length whatever the latitude, so none is favoured at any time.

The peak-like component at mid-latitudes can be interpreted as east-west oriented traffic, using alternately east-west oriented interplane and north-south oriented intraplane ISLs. This component reduces near the equator because the interplane ISLs are longer, and therefore are not selected as often by the routing algorithm (the traffic is preferably routed through interplane ISLs at mid-latitudes).

ISL load for a seam satellite (satellite 11)

Figure 11 shows the load of interplane ISL32 and ISL33 over one orbit period. A similar pattern as for the interplane ISL load of a non-seam satellite can be observed (compare with Figure 8). However the average and maximum loads are smaller : indeed a seam satellite participates in fewer routes than a non-seam satellite. This is a general feature of bounded networks, as reported in [4].

Figure 12 shows the load of intraplane ISL1 and ISL29. Here, the pattern is different from that of the intraplane ISL load of a non-seam satellite, as no peak-like traffic appears. Instead, the load is nearly constant over the entire orbit, with a slightly reduced load over the southern hemisphere.

INHOMOGENEOUS TRAFFIC DISTRIBUTION

The traffic distribution is represented by the traffic matrix given in Table 1, where peak traffic is given in erlang

within and between continents. The traffic generated on the earth is allocated to the different satellites according to the fraction of the continent within the footprint of each considered satellite. However no account is made of the time of the day, therefore the results presented here cannot be considered as representative of a real situation. Nevertheless, the procedure allows to investigate how an inhomogeneous traffic distribution impacts the ISL loads.

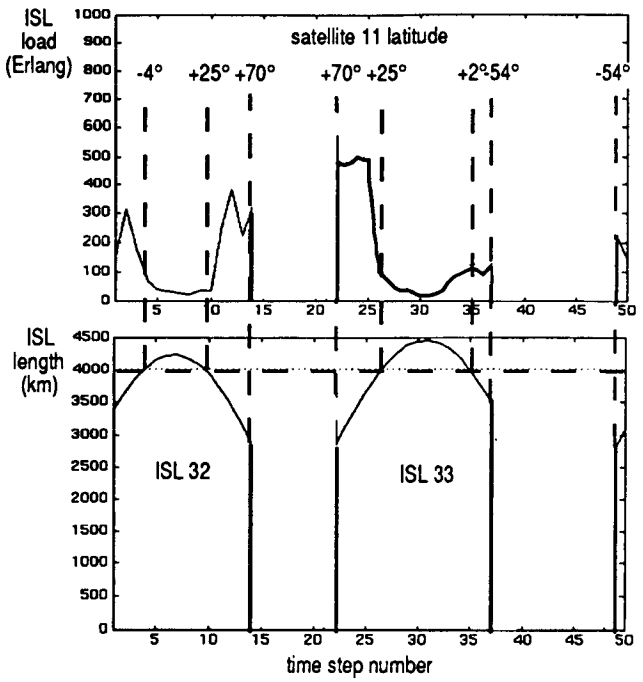


Figure 11 : Load of interplane ISL32 and ISL33 during one orbit period (homogeneous traffic distribution).

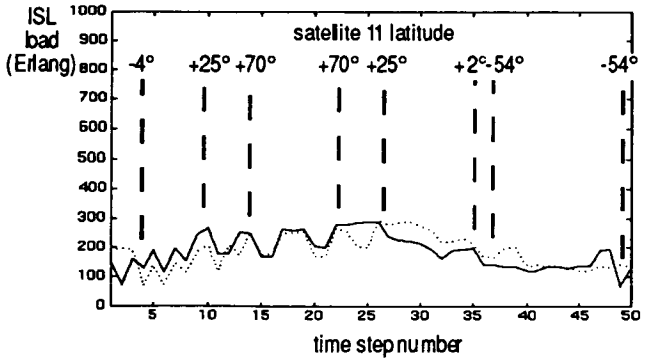


Figure 12 : Load of intraplane ISL1 (solid line) and ISL29 (dotted line) during one orbit period (homogeneous traffic distribution)

ISL load for a non-seam satellite (satellite 42)

Figure 13 to 15 show the traffic load for the ISLs of the non-seam satellite 42. Peak load for interplane ISL94 and ISL96 (Figure 13) occurs while the satellite proceeds respectively over Asia (steps 14-20) and over North America (steps 31-35).

Table 1 : Traffic matrix in Erlang (NA: North America, SA: South America, EU: Europe, AF: Africa, AS: Asia, AU: Australia)

	NA	SA	EU	AF	AS	AU
NA	4122	145	194	97	194	97
SA	145	1746	145	39	39	19
EU	194	145	4122	145	194	48
AF	97	39	145	1746	78	19
AS	194	39	194	78	3298	155
AU	97	19	48	19	155	1765

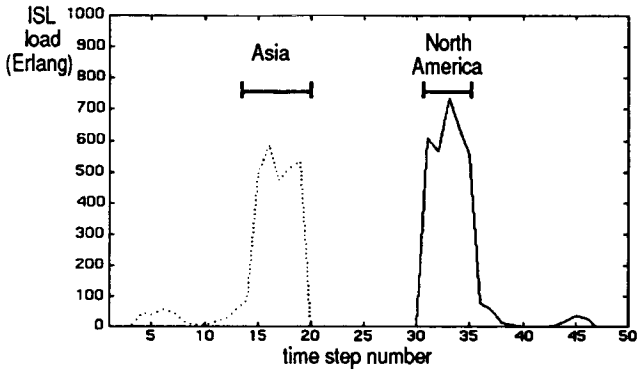


Figure 13 : Load of interplane ISL94 (solid line) and ISL96 (dotted line) during one orbit period (inhomogeneous traffic distribution).

The latter is higher because a larger traffic is generated from North America than from Asia. Peak ISL load occurs when the footprints of both end satellites cover part of a given continent, as most of the traffic is routed within the considered continent. The smaller peak loads observed in Figure 13 correspond to time intervals when satellite 42 does not cover any land mass, but proceeds at mid-latitudes. This means that the ISLs are loaded by long distance traffic routed through satellite 42 in the east-west direction. Similar observations and conclusions can be derived from Figure 14 concerning interplane ISL127 and ISL127.

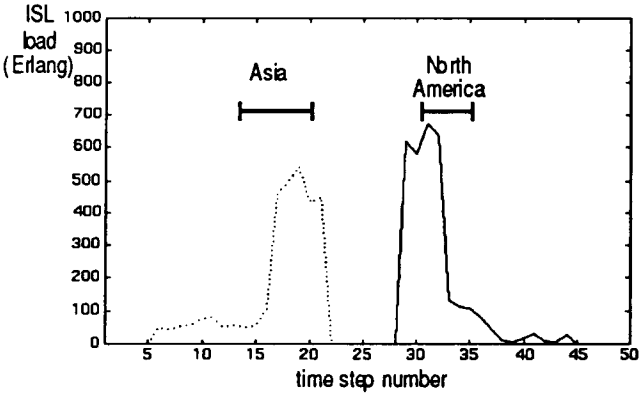


Figure 14 : Load of interplane ISL127 (solid line) and ISL126 (dotted line) during one orbit period (inhomogeneous traffic distribution).

As for the intraplane ISL122 and ISL125, Figure 15 displays a delay between the peak traffic load of those ISLs similar to that observed in the homogeneous traffic distribution case (Figure 10). However the peaks are high only at mid-latitudes over land masses (Asia, North America). No peaks appear over ocean region, contrary to the homogeneous traffic distribution case.

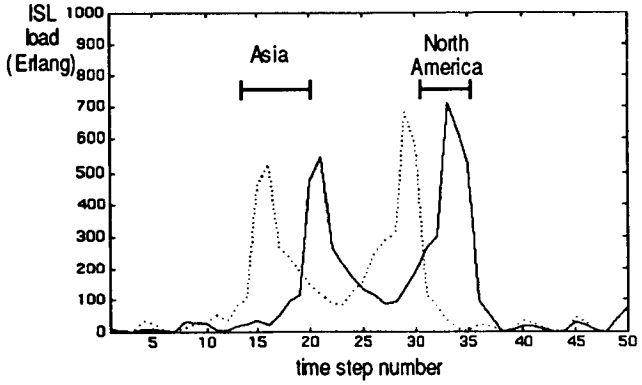


Figure 15 : Load of intraplane ISL122 (solid line) and ISL125 (dotted line) during one orbit period (inhomogeneous traffic distribution).

ISL load for a seam satellite (satellite 11)

Figure 16 and 17 show the traffic load for the ISLs of the seam satellite 11. The high peak observed in Figure 16 for the interplane ISL32 corresponds to local traffic generated in Europe (step 10-14) involving neighbouring satellites. The small peak for ISL33 (step 21-25) results from the satellite flying over part of Asia. The unloaded interval (steps 26-31) corresponds to passage over east Pacific. No traffic is generated in that region, and moreover due to the seam, no traffic is forwarded in the east-west direction through ISL33. The small peak about step 35 is caused by traffic generated in Australia and forwarded through satellites flying over mid-latitudes to North and South America.

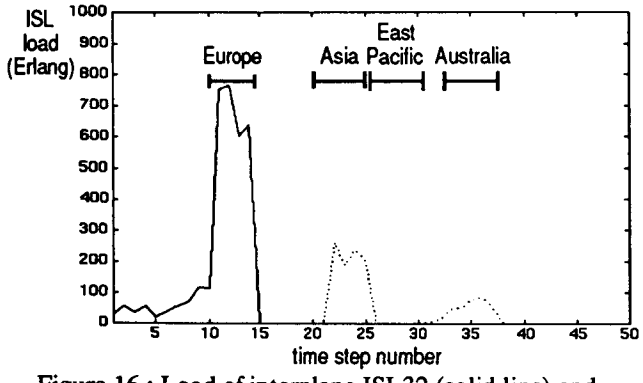


Figure 16 : Load of interplane ISL32 (solid line) and ISL33 (dotted line) during one orbit period (inhomogeneous traffic distribution).

For intraplane ISLs (Figure 17), the load peaks over Europe (steps 10-14). The high load over the north pole (steps 15-20) is traffic between two parts of Asia separated by the seam : the traffic then flows along one side of the seam over the north pole and then is routed to the other side of the seam.

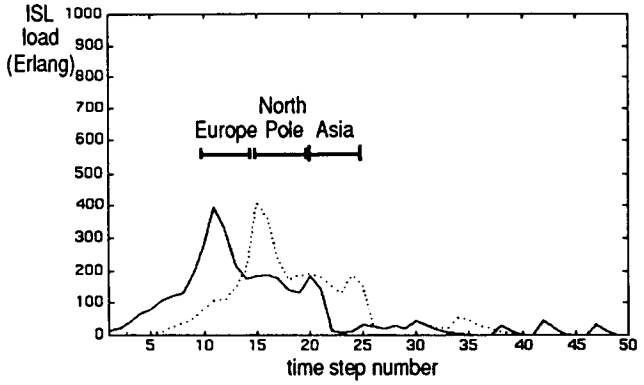


Figure 17 : Load of intraplane ISL1 (solid line) and ISL29 (dotted line) during one orbit period (inhomogeneous traffic distribution).

CONCLUSIONS

The paper has presented major features of traffic routing in an ISL network for a polar satellite constellation. Main results are : when the implemented routing procedure preferably selects short routes, short ISLs are more loaded than others. Different load patterns are observed over ISLs connected to seam and non-seam satellites. ISLs connected to seam satellites are less loaded. Interplane ISLs of non-seam satellites experience high peak load over mid-latitudes, as a result of a more pronounced east-west traffic routing tendency at those latitudes due to the relative shortness of interplane ISLs at those latitudes. The distribution of generated traffic impacts strongly on the ISL load pattern.

REFERENCES

[1] J. Hutcheson and M. Laurin, 'Network flexibility of the IRIDIUM global mobile satellite system', Proc. of the International Mobile Satellite Conference IMSC'95, Ottawa; pp 503-507.

[2] M. A. Sturza, 'Architecture of the TELEDESIC satellite system', Proc. of the International Mobile Satellite Conference IMSC'95, Ottawa; pp 212-218.

[3] M. Schwartz, 'Telecommunication Networks : protocols, modelling and analysis', Addison-Wesley publishing company, 1987.

[4] Lloyd Wood, 'Network performance of non-geostationary constellations equipped wiht intersallite links' MSC Diploma thesis, University of Surrey, 1995

ATM Networking for Future ISL-Based LEO Satellite Constellations

Markus Werner, Cecilia Delucchi, Karina Burchard
German Aerospace Research Establishment (DLR)
Institute for Communications Technology
P.O. Box 11 16, D-82230 Wessling, Germany
Phone: +49 8153 28-{2826,2883} Fax: +49 8153 28-1442
Email: {Markus.Werner,Cecilia.Delucchi}@dlr.de

ABSTRACT

ATM networking for future LEO satellite systems employing intersatellite links (ISLs) is considered. The underlying dynamic network topology is of fundamental importance for the connection-oriented operation of such broadband satellite systems. Therefore, in a first part a pragmatic approach toward the topological design of ISL networks for Walker constellations is presented. Using M-Star as example system, a reference topology is derived and used for further concept implementation and simulation. In a second major part, an overall ATM networking concept for the complete LEO scenario is proposed. Based on that, a dynamic virtual topology approach is developed for the networking and routing within the ISL subnetwork. Basic concepts and schemes as known from terrestrial networks are adapted to the specific dynamic LEO scenario. Finally, some fundamental feasibility considerations as well as first preliminary simulation results are given to emphasize the significance and potential of the proposed concepts.

I. INTRODUCTION

Low Earth Orbit (LEO) satellite constellations are foreseen as strong candidates to provide global mobile and fixed communication services in the future. Among the "first generation big LEOs", Motorola's Iridium [1] will be the only one operating with intersatellite links (ISLs) for the transport of long-distance traffic. Even before the first launches of their satellites, proposals for a second generation of such systems, aiming at global broadband communications, are already on their way. Two reputable system filings at FCC – Teledesic [2] and M-Star [3] – are an early embodiment of the coming era in satellite communications. Both systems are strongly based on the extensive use of an ISL infrastructure that of course imposes high networking complexity but can, on the other hand, provide considerable benefit for truly global broadband communications.

Obviously, ATM is predestined to form the operational basis of such a system, (a) working on a time-varying network topology and (b) facing diverse traffic from a growing spectrum of broadband services. The focus of this paper is on ATM networking concepts for the ISL subnetworks of current and upcoming LEO constellations.

To this end, in Section II we first analyse the impact of satellite dynamics on intra- and interplane ISLs in terms of distance variation, pointing angle variation, and requirements for on/off-switching and acquisition. This is achieved using the Walker-type M-Star constellation as reference. From this knowledge, a straightforward design of potential and reasonable ISL topologies for M-Star is achieved. A reference topology is selected for simulations and performance evaluation.

In Section III an overall framework for ATM-based operation of the considered LEO systems is presented, which is particularly tailored to promote a decomposition approach for the different network segments. From here, further concern is about the ISL subnetwork alone, with some emphasis on routing aspects. Section IV then introduces the proposed ATM networking/routing approach *within* the ISL network in a top-down manner, ranging from basic concepts to algorithmic and optimization details.

Finally, some preliminary simulation results are reported and discussed in Section V. They provide clear indications on the direction future work should reasonably focus on.

II. TOPOLOGICAL DESIGN OF INTERSATELLITE LINK NETWORKS

A. Background and Motivation

Recalling recent years' LEO system proposals and correlated research, one notices that Iridium has for a long time been the only promising candidate employing ISLs and at the same time revealing (step by step) some details on the planned ISL topology. Studies focusing on the routing in the Iridium ISL network have clearly identified the seam between contra-rotating orbits and the on/off-switching of its interplane ISLs as two fundamental drawbacks for the connection-oriented operation [4]. Already before Motorola unveiled its final ISL topology design, these general problems behind the near-polar constellations have stimulated the proposal of a medium inclined Walker-type MEO constellation (for study purposes) employing ISLs [5]. It has been shown that such constellation types generally provide the possibility to set up a number of interplane ISLs that can be maintained *permanently* with acceptable pointing, acquisition and tracking (PAT) requirements and much lower interference problems than their counterparts in Iridium-type constellations. Obviously, this is a highly desirable feature

in the light of real-time connection-oriented services, since it immediately reduces potential problems due to path switching. Considering future goals to route jitter sensitive ATM cell streams across the ISL network, this becomes even more striking.

With the M-Star filing, the first realistic candidate system utilizing the promising combination of Walker orbits and ISLs seems to be on its way. A proper design of the ISL topology to be implemented is a first important step to guarantee efficient ATM networking in the operational system. In the following we present a pragmatic approach to the ISL topology design in Walker constellations, deriving a reference topology for M-Star as an example. Fig. 1 shows the M-Star constellation, and relevant parameters are summarized in Table I in comparison with Iridium and Teledesic.

TABLE I

PARAMETERS OF LEO SATELLITE SYSTEMS EMPLOYING ISLS.

	Iridium	Teledesic	M-Star
Orbit altitude	780 km	700 km	1350 km
Orbit period	100 min	99 min	113 min
Number of satellites	66	840	72
Number of orbits	6	21	12
Inclination	86.4°	98.2°	47°
Intraplane ISLs per sat.	2 perman.	up to 4	2 perman.
Interplane ISLs per sat.	up to 2	up to 4	2 perman.
Primary service	narrowband, voice	broadband, multimedia	broadband, real-time V/D

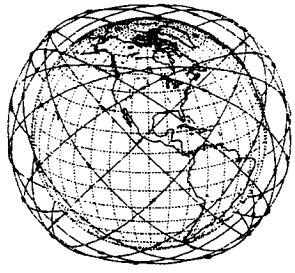


Fig. 1. The M-Star constellation [3].

B. ISL Topology Design for M-Star

A closer look on the planar projection of the 3D constellation, Fig. 2, facilitates the first step in the ISL topology design, which is to identify potential links to be reasonably implemented in the network. Due to the perfect symmetry of the constellation it is sufficient to consider generic types of ISLs between satellite 0 and its neighbours (on "right",i.e. eastward planes) at $t = 0$ as example. These are then applicable to all other satellite pairs correspondingly. Of course, the implementation of the convenient intraplane ISLs (0-1, 0-5) with both constant link distance and fixed pointing angles is beyond question. Then one would intuitively envisage links toward the next neighbours on the adjacent plane, the one ahead (0-6) and the one behind (0-11), as illustrated in Fig. 3. Finally, the same could be applied to partners on the second but next plane (0-12, 0-17). So far, this procedure is generally applicable to Walker constellations *before* considering any system specific constellation parameters.

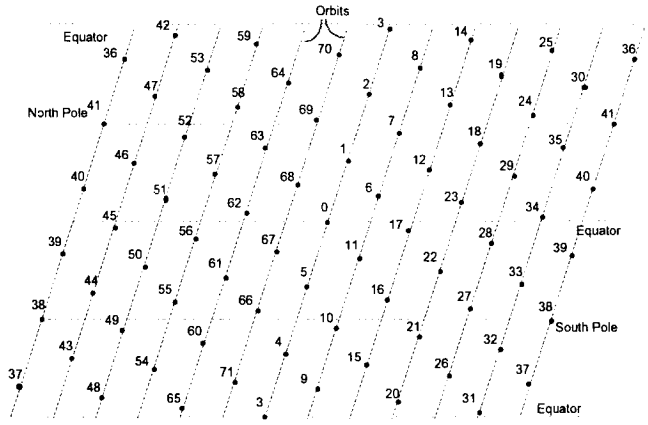


Fig. 2. Planar projection of the M-Star constellation at $t = 0$. Distances and angles are *not* true to scale.

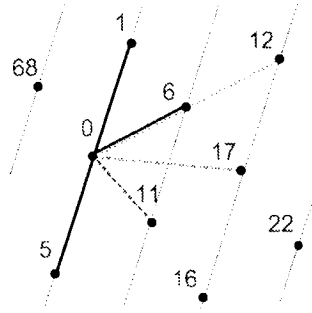


Fig. 3. Potential generic ISLs in a regular Walker constellation.

In fact, the feasibility of all these envisaged links has still to be proven taking into account geometrical and technological constraints for a specific constellation. The diagrams in Fig. 4 display important geometrical data for this purpose. The vertical and horizontal deviation of the pointing angle from flight direction are intuitively called elevation and azimuth in this context. One can notice that both links toward the adjacent orbit show relatively little variation in distance, elevation, and azimuth, with overall slight advantages for ISL 0-6. This means that establishing these ISLs in permanent mode does not introduce severe problems. On the contrary, ISL 0-17 is less attractive but still possible, whereas Earth shadowing completely prevents implementation of ISL 0-12. Simple geometrical considerations yield that requiring a minimum distance of 30 km between any ISL and Earth's surface leads to a maximum allowed length of roughly 8600 km for an ISL – or, equivalently, to a maximum allowed elevation of 34°between the two connected satellites. The dash-dot lines in the diagrams indicate these values. ISL 0-12 is clearly above these limits at all instants of time.

From a networking viewpoint (considering dimensioning,

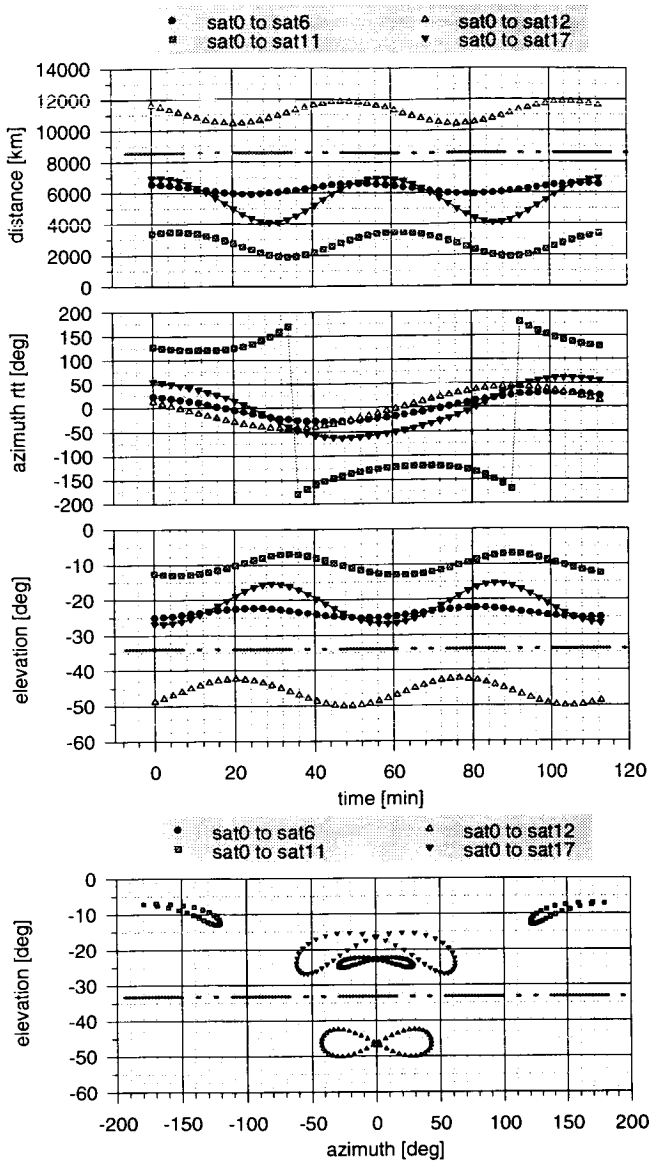


Fig. 4. Geometrical feasibility of ISLs in M-Star: Time variation of (a) ISL distance, (b) vertical pointing (elevation) and (c) horizontal pointing (azimuth); (d) pointing diagram. The dash-dot line represents the upper bound with respect to Earth shadowing.

even distribution of traffic in the network etc.) a higher degree of meshing is of course desirable, but it should be traded off vs. implementation complexity. Moreover, a topology exclusively based on permanent links is very attractive. In the following, we will use as reference topology T1 the case of only implementing generic ISL type 0-6 besides the intraplane ISLs. Fig. 5 displays a snapshot of T1.

III. ATM Networking Concept

The end-to-end connection scenario of ISL-based LEO satellite systems comprises three segments, according to Fig. 6, which illustrates the connection between a mobile

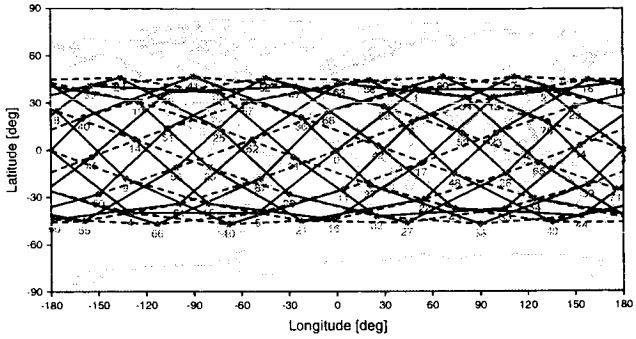


Fig. 5. Considered M-Star ISL topology T1.

user MU_A and a remote fixed partner FU_B: The (1) *Up-/Downlink(UDL)* segment incorporates the respective links between mobile users and satellites as well as the links between satellites and fixed earth stations (*gateways GW*). The gateway stations act as interface between the (2) *Intersatellite Link (ISL)* segment and the (3) *Terrestrial Network Link (TNL)* segment. In the following, the satellites serving both UDL segments will be referenced to as *start* and *end satellites* (of the connection). Considering bidirectional communication, clearly every start satellite always acts simultaneously as end satellite, and vice versa.

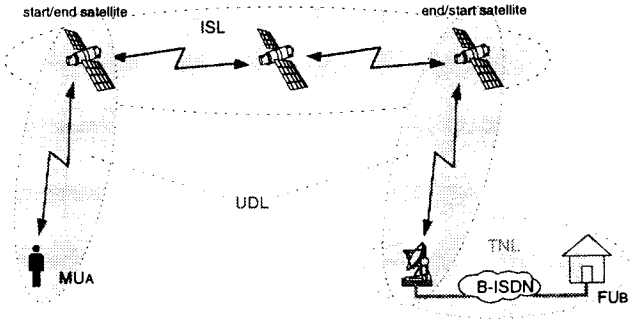


Fig. 6. Segments of a typical end-to-end connection in ISL-based LEO satellite systems.

The application of ATM concepts to satellite systems, especially of the future LEO systems, aims at the exploitation of the powerful features of the connection-oriented ATM layer transport functions on *Virtual Channel (VC)* and *Virtual Path (VP)* level. A virtual path is understood as a logical link between two adjacent ATM nodes. It can assemble a group of virtual channels belonging to different (and potentially multi-service) end user connections that are utilizing the physical channel between these two nodes. The connection-oriented operation mode of virtual connections – established either as *VC Connection (VCC)* or as *VP Connection (VPC)* – ensures that cell sequence is preserved during their lifetime. Classically, one dedicated VCC between two ATM end users is established for the holding time of a call and then used by all related ATM cells. Redundant (standby) paths/VCCs may be provided to allow hitless path-switching

either on-demand (predictable) or as reaction to path failures (unpredictable). In both cases, proper alignment of the ATM cell streams during switch-over is essential for nondisruptive operation of the user connection.

In the considered LEO scenario we face the challenge of a multiply dynamic network topology, extending on both UDL and ISL segments. Inherently, this feature in general requires switching between subsequent different paths during the lifetime of a user connection. In the following, this is referenced to as virtual connection *handover (HO)*.

A. UDL and TNL Segments

The TNL segment is easily separable by assuming that the gateway either acts as central VC/VP switch between the internal (satellite) and the external TNL part or is even implemented as intermediate node between two different ATM networks. The networking of the TNL part itself turns out to be the classical terrestrial B-ISDN one and is not further considered in this paper.

For the UDL routing part, an extension of the *Virtual Connection Tree (VCT)* concept – originally proposed by Acampora et al. [6] for terrestrial cellular ATM networks – is considered. In the original concept one root switch manages a fixed tree serving a certain mobile service area (MSA), Fig. 7(a). The VCT takes care of connection setup and VCI reservation and ensures fast and transparent handovers of roaming users in the MSA. To apply the VCT architecture and methodology to the LEO satellite environment, a dynamic satellite cluster headed by a *Master Satellite* is defined, Fig. 7(b). This satellite cluster builds a time-dependent VCT with spotbeams, ensuring that fast and transparent spotbeam and satellite handovers can be performed [7].

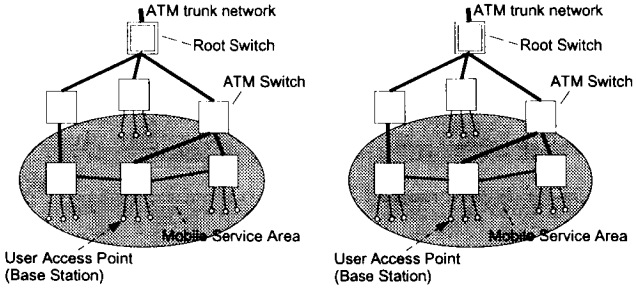


Fig. 7. Virtual Connection Tree (VCT) concept [6] applied to (a) terrestrial and (b) satellite ATM networks.

B. ISL Segment

The multiple ISL topology dynamics is due to the fact that pairs of satellites in adjacent orbits have a varying distance over time or can even lose sight of each other. Topological or connectivity changes may demand rerouting of calls during their lifetime, i.e. on-line switching between subsequent different paths. Fortunately, the "physical" ISL topology dynamics is periodically deterministic with the constellation period. For a specific constellation, this allows to set up off-line a unique time-dependent virtual topology providing continuous operation of end-to-end connections.

We propose an approach where a possible route between a start and an end satellite is modeled as a VPC. Since most satellite pairs are not directly connected by an ISL, intermediate satellites will often be needed to constitute a complete VPC route. Every section of this route, i.e. every ISL between adjacent satellites, in general contains a number of VPs (resulting from the considered and all other "crossing" VPCs), all sections of course using the complete set of VP identifiers (VPIs) independently. The routing function is then incorporated in the respective dynamic VPI translation tables.

The basic idea behind this approach is that end-to-end VCCs sharing the same start and end satellites at arbitrary time can simply be aggregated in one common VPC across the ISL subnetwork. Every transit satellite provides – on the basis of locally available switching tables – pure VP switching functionality between every pair of ISL ports, and thus the whole space segment becomes a pure fast operating cross-connect network. Avoiding any switching on VC level turns out to be especially favourable in the case of many simultaneous low bit rate connections on a trunk line; this situation is very likely to be the dominating one in systems providing primarily voice services.

Fig.8 illustrates the overall concept comprising the integration of UDL and ISL segment. It specifically addresses the problem to guarantee a continuous connection in the case of an ISL subnetwork internal path handover. It has been shown [8] that the operation of both basic types of handover – UDL satellite handover and ISL VPC/path handover – can be managed with the proposed concept but requires sophisticated optimization algorithms and techniques to keep the possible impairments to a minimum.

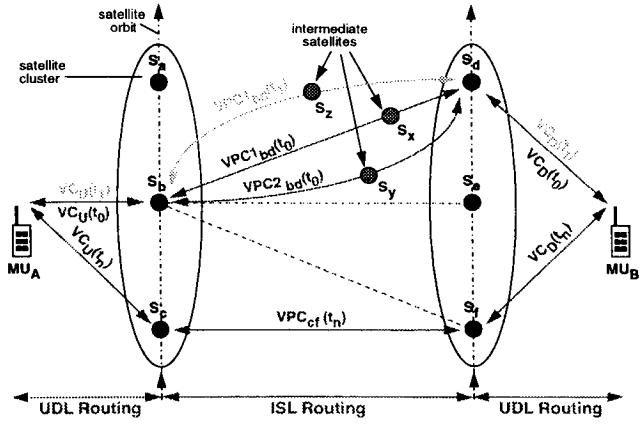


Fig. 8. Virtual connection based routing concept in satellite systems with ISLs. The continuous operation of an existing end user connection is illustrated for the case of VPC handover between the same start/end satellite pair from steps t_0 to t_1.

IV. Dynamic Virtual Topology Set-Up

The complete approach to the operation of the dynamic ISL network is a time-discrete one: choosing a sufficiently small time step, the (continuous-time) dynamic physical topology is modeled as a sequence of quasi-static topology snapshots. Then, the set-up of the VPC-based dynamic *virtual* topology for the ISL network comprises two major steps, namely (1) the set-up of instantaneous virtual topologies for all time-discrete snapshots, each one comprising a set of paths (VPCs) for each start/end satellite pair, and (2) the selection of applicable path sequences over time (from the above sets) due to a parametrical optimization procedure.

1. Set-up of instantaneous virtual topologies. For each of the snapshots we define an instantaneous virtual topology by means of direct VPCs connecting all $n(n-1)/2$ pairs of satellites, n being the total number of satellites in the constellation. In ATM terminology, we build a fully meshed SHVP (single hop virtual path) topology that contains a set of (up to) k alternative paths for each start/end satellite pair. Two reasonable solutions are the following:

(a) All paths are link-disjoint. This reduces the number of alternatives for the latter optimization just from the beginning and at the same time stocks up reasonable backup paths for the operational system. In the case of our reference M-Star topology, $k \leq 4$ since the number of link-disjoint paths is clearly limited by the number of ISL ports of the start/end satellites.

(b) The k "shortest" paths (with respect to a certain cost metric) are in the set. This enhances the flexibility in the latter optimization, which is specifically important when aiming at dimensioning of the operational system. Additionally, corresponding link-disjoint backup paths may be assessed for each of the k paths.

These considerations are illustrated for *one* satellite pair of the considered M-Star ISL topology in Fig.9. Solution (a) extends over two dimensions only and is completely defined by the top "layer" of the picture: For each time step, 4 link-disjoint paths are in the set, and each of them can on principle be chosen as active one, leaving the other three as potential backups. Solution (b) additionally comprises the third (vertical) dimension: Active paths can be chosen from the left "column" per time step, whereas the corresponding (link-disjoint) backup paths are already prepared as well. One should note that any link-disjoint backup (row) path can be physically identical with any of the k shortest (column) paths, as long as they are not in the same row.

2. Path sequence selection. The second major task is now to select (for each satellite pair) appropriate *sequences* of paths from the subsequent instantaneous sets according to a certain optimization function. Three possible candidates are illustrated by the arrows in Fig. 9. Following the solid one simply means that always the instantaneously "best" path, e.g. the one with minimum transfer delay, is chosen. This simple approach seems to be quite promising for the M-Star topology T1 that employs only permanent ISLs, which means that the most critical path switching situations (in terms of delay jitter) are avoided straight from the beginning. The dashed arrow may reflect a path sequence that tries to minimize the worst case handover delay jitter encountered when path switching from one to another step is inevitable, like it is the case in Iridium when ISLs are switched on or off. This approach has been extensively studied and evaluated in [4]. A basic outcome is that sophisticated optimization schemes can significantly reduce the handover delay jitter, but only at the expense of high complexity both before and during system operation. However, this is a general drawback of bounded polar ISL topologies. The dotted arrow finally represents a VPC sequence that guarantees a constant underlying physical path over time. This is for instance possible within the selected non-switched M-Star topology T1, if only k is chosen big enough to include this path in the set in every step.

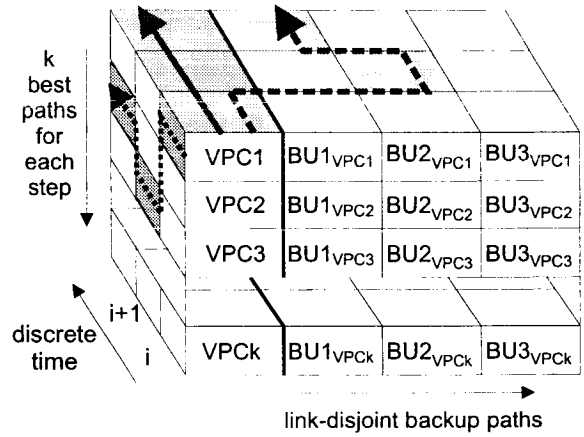


Fig. 9. Illustration of instantaneous VPC set-up and VPC sequence selection over time.

V. Feasibility and Numerical Results

With respect to the fundamental limits of the concept, the major concern must be about the maximum number of VPs required on the worst-case link, considering all ISLs over the whole constellation period. The ATM cell header contains 12 bits in the Virtual Path Identifier (VPI) field, thus allowing a maximum of $2^{12} = 4096$ VPs on a single ISL per step. The number of instantaneously required VPs per link is determined by the number of VPCs using this link. Exact figures for that can not be calculated, since they depend on the features of the network topology and the routing algorithm. However, assuming that only one unique path sequence is chosen for each satellite pair, we find that the number of simultaneous VPCs in the *complete network* is equivalent to the number of start/end satellite pairs, $\frac{N(N-1)}{2}$ (= 2556 for M-Star, $N = 72$). If we further assume that due to the relative symmetry of the selected ISL topology the routing strategy should "distribute" the paths quite evenly in the network consisting of 144 links in total, and that a path consists of roughly 5 links in the average, we can

estimate that the corresponding number of VPs per link is in this optimal case $2556 \cdot 5/144 \approx 90$. In the real scenario, i.e. with non-ideal distribution of the routes, we can of course expect a worst case value which is maybe a factor 2-3 higher, but then still far below the allowed maximum.

Applying the presented concepts and procedures in a simulation package, first preliminary numerical results have been gained. First, evaluating the required VPs/link count, the above estimation is verified: the worst case link carries roughly 230 VPCs in a certain interval, if always the path with shortest geometrical length is selected. Secondly, evaluating the distribution of VPs/link one finds that interplane ISLs in mid-latitudes are prefered for path search due to their relative shortness, and therefore carry more VPs than their counterparts in equatorial regions.

Finally, Figs. 10(a) and (b) show the distribution of delay offset as measured between subsequent steps in lasting connections. For both, Iridium and M-Star, telephony service has been used as simple target profile. Note that all simulations have been run on the basis of link-disjoint path sets only; the k-shortest-path option has not been used so far. The comparison of histograms (a) and (b) clearly stresses a basic advantage of the non-switched M-Star ISL topology over the switched Iridium one: Whereas for Iridium high values of delay offset are observed due to unfavourable forced path switching situations – and this still after a sophisticated optimization scheme! – the M-Star figures turn out to reflect more or less only the distance (and transfer time) difference between two discrete steps. In some cases, path switching may happen between subsequent steps (because the strategy has been to always select the shortest path), but only due to the fact that the former second best path has become slightly better in the new step than the former first choice path. To sum up, the superior performance of M-Star is due to the fact that its topology operates on permanent ISLs only and thus completely avoids unfavourable forced path handover situations as encountered in Iridium.

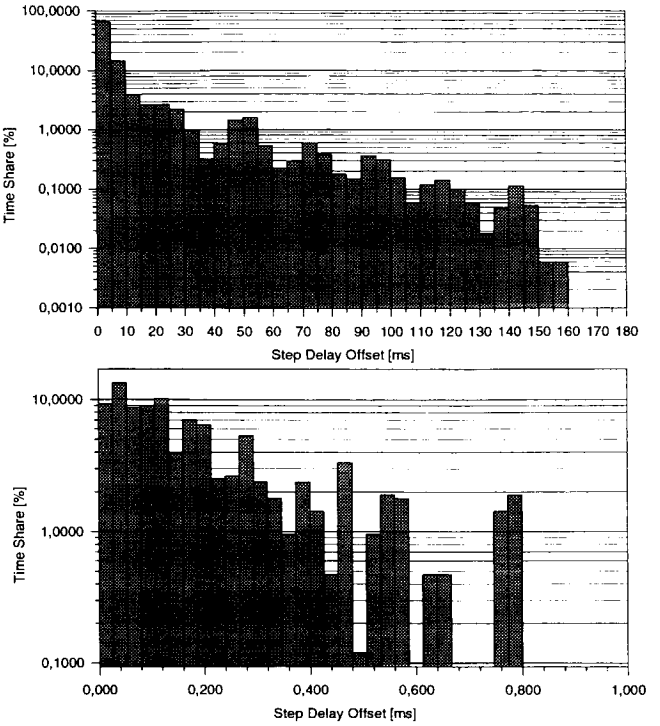


Fig. 10. Distribution of delay offset between subsequent steps for telephony. (a) Iridium, minimize handover delay jitter; (b) M-Star, always select shortest path.

VI. Conclusions and Outlook

The paper has essentially presented a framework for ATM-based networking in future broadband LEO satellite systems employing intersatellite links. Due to the central importance of the underlying dynamic network topology, in a first step a pragmatic approach toward the topological design of ISL networks for Walker constellations has been presented and used for the example system M-Star. In a second step, the overall ATM networking concept has been presented and a virtual topology approach for routing within the ISL subnetwork has been discussed in detail. Preliminary simulation results for the simple case of telephone traffic emphasize the feasibility and the potential of the proposed concepts. Moreover, it is shown that inclined Walker constellations inherently provide some significant advantages over Iridium-type near-polar constellations with respect to ATM networking. The only precondition is that the constellation characteristics are efficiently utilized for an appropriate ISL topology design. Future work shall mainly extend the investigations towards real broadband traffic scenarios.

References

[1] J. Hutcheson and M. Laurin. Network flexibility of the IRIDIUM global mobile satellite system. In *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, pages 503–507, Ottawa, Canada, June 1995.

[2] M. A. Sturza. Architecture of the TELEDESIC satellite system. In *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, pages 212–218, Ottawa, Canada, June 1995.

[3] R. Meidan. M-Star, a global LEO-based broadband service. In *Proceedings COMMSPHERE '97*, page unpaginated, Lausanne, Switzerland, Feb. 1997.

[4] M. Werner, C. Delucchi, H.-J. Vögel, G. Maral, and J.-J. De Ridder. ATM-based routing in LEO/MEO satellite networks with intersatellite links. *IEEE Journal on Selected Areas in Communications*, 15(1):69–82, Jan. 1997.

[5] M. Werner. Analysis of system connectivity and traffic capacity requirements for LEO/MEO S-PCNs. In E. Del Re, editor, *Mobile and Personal Communications, Proceedings 2nd Joint COST 227/231 Workshop*, pages 183–204, Florence, Italy, Apr. 1995. Elsevier.

[6] A. S. Acampora and M. Naghshineh. An architecture and methodology for mobile-executed handoff in cellular ATM networks. *IEEE Journal on Selected Areas in Communications*, 12(8):1365–1375, Oct. 1994.

[7] G. Luton. An ATM based concept for handover operation in LEO/MEO satellite systems. Master's thesis, ENST Toulouse / DLR Oberpfaffenhofen, Toulouse, France / Weßling, Germany, Aug. 1995.

[8] M. Werner. ATM concepts for satellite personal communication networks. In D. W. Faulkner and A. L. Harmer, editors, *ATM, Networks and LANs, Proceedings European Conference on Networks & Optical Communications (NOC '96)*, pages 247–254, Heidelberg, Germany, June 1996. IOS Press. Invited Keynote Paper.

Session 9
Modulation, Coding and Multiple Access I

Session Chairperson—*Marvin Simon*, Jet Propulsion Laboratory, USA
Session Organizer—*Dariush Divsalar*, Jet Propulsion Laboratory, USA

Performance Analysis of Concatenated Reed-Solomon Trellis Coded Modulation over Mobile Satellite Channel

M.G. Françon, M. Bousquet
SUPAERO
10 Avenue E. Belin, 31055 Toulouse Cx4, France
Phone : + 33 5 62 17 80 80 ext. 9525 Fax : + 33 5 62 17 83 45
email : francon@supaero.fr bousquet@supaero.fr

ABSTRACT

Concatenated code consisting of trellis inner codes and Reed-Solomon (RS) outer codes are considered to achieve large coding gains with small bandwidth expansion in the presence of frequency-nonselective slow fading. In this paper, a Rice/Rayleigh-Lognormal channel model is used which is relevant for mobile satellite communications.

Assuming perfect interleaving of the symbols, performances are analysed. The Bit Error Rate (BER) is analytically and semi-analytically (Spline functions) evaluated for RS codes with varying blocklengths and code rates, different elevation angles and fixed mobile velocity and carrier frequency. It is shown that an optimum code rate exists and that optimum code rate decreases with the elevation angle.

Finally, limited interleaving is considered with inner (convolutional) and outer (block) interleavers or only inner interleaver depending on the RS code blocklength used. Computer simulation results are presented which show the penalty in bit energy to noise ratio due to limited interleaving as a function of end to end interleaving time delay.

I. INTRODUCTION

Transmission of both voice and data is now digital. For data file transfer all messages must be delivered error-free while the BER for digitized voice transmission is less critical. Transmission errors on mobile satellite links are mainly caused by variations in received signal referred to as fading.

For such communications over fading channels, errors occur in bursts during fades and induce large performance degradation [1]. Consequently, fading severely degrades the system performance and a considerable amount of coding is needed to reduce the penalty in Energy per Bit to Noise Power Spectral Density Ratio (E_b/N_o). That requires a trade-off between carrier bandwidth and power. On the other hand, the growing needs for transmitting data at higher rates require an efficient use of the radio spectrum. To cope with the power and bandwidth limitations of Mobile Satellite Communication system (MSC), a combination of bandwidth-efficient modulation techniques and forward error correction coding is needed. Trellis Coded Modulation (TCM) processes the best attributes each of the above [2]. However, to achieve high coding gains based on TCMs only, the decoding complexity increases drastically as more trellis states are used [2]. An obvious alternative to enhance the performance of digital communication systems operating over fading channels is to use TCM in a concatenated scheme with RS codes [3] at the expense of an increase in the required bandwidth. In the concatenated scheme considered in this paper simple TCM are used as the inner codes and different RS codes as the outer codes. TCM codes provide satisfactory inner codes for the concatenated scheme.

Concatenated codes with inner codes and RS outer codes on Gaussian channels have been considered in [4]. In this paper, the analysis is carried out over a fading channel which is a combination of Rice/Rayleigh-Lognormal statistics, with shadowing affecting both direct and diffuse components [5-7].

To evaluate performances in terms of BER versus E_b/N_o at the output of the outer decoder assuming ideal interleaving, it is necessary to evaluate the bit error rate P_b at the output of the inner code. For such a channel, P_b cannot easily be obtained analytically and consequently, a semi-analytical method is proposed assuming full interleaving. This semi-analytical procedure first consists in evaluating P_b by simulation with interpolation using second order spline functions in order to have a set of regularly distributed values, and then using this set to obtain analytically the BER at the output of the outer decoder. The validation of this method is made by comparing its BER results to those given by a pure analytical method and pure simulations for simple cases (e.g. QPSK modulation over Rayleigh channel or AWGN

channel). The semi-analytical method shows a good agreement with the two others methods and is used to evaluate BER performances over channels with more complex statistics.

The outline of the paper is as follows. Section II presents the characteristics of the inner and outer codes. The description of the communication system is given in Section III. Section IV describes the semi-analytical method to evaluate bit error rate performances. In Section V, performances are given for concatenated RS-TCM using the semi-analytical method assuming perfect interleaving. In Section VI, performances with limited interleaving are explored by simulation. Finally, in Section VII, the conclusions are drawn.

II. Concatenated Code

The inner code

TCM techniques are strong candidates for the inner code as they include low complexity codes which provide significant coding gain without increasing transmission bandwidth [8]. The TCM used in this paper is Ungerboeck's rate 2/3 coded 8-PSK with 8 states and is denoted by U8. The decoding of the inner code is based on Viterbi algorithm.

The outer code

When choosing candidates for the outer code, one should consider the trend of the inner code to produce short error bursts because of its trellis-like structure, and the requirement for both error correction and detection capabilities to provide reliable data [8]. RS codes are among the best linear block codes for correcting bursts errors. The outer code in an (n,k) RS code with symbols from the Galois field GF(2^m) and minimum Hamming distance d=n-k+1. Each symbol of the outer code is represented by a binary m-tuple based on a certain basis of GF(2^m). RS(31,k), RS(63,k), RS(127,k) and finally RS(225,k) codes are considered in this paper. The outer code is decoded by the Berlekamp-Massey algorithm using the error-only algorithm : the inner code provides no erasure information and, the outer code is designed to correct symbols errors only [11].

III. System and Channel Models

The system under consideration is shown in Figure 1. One or two interleavers are used depending on the RS code blocklength considered. The interleaving process is used to combat the effect of errors bursts and to create a sequence of errors which are approximately independent.

In our analysis, perfect carrier and timing synchronization are assumed and the transmitted symbol rate considered is 9.6 Kbaud.

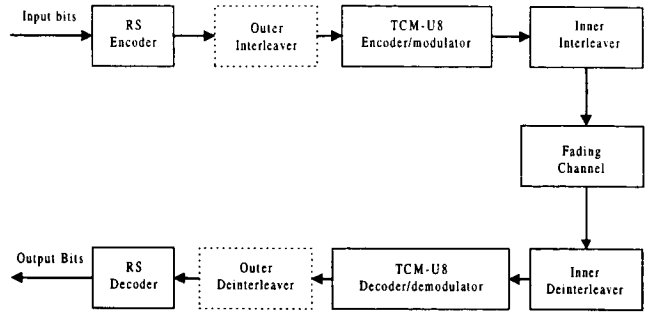


Fig.1 System block diagram

As stated above, the channel model is a combination of Rice/Rayleigh-Lognormal statistics, with shadowing affecting both direct and diffuse components [5-7]. This channel model is suitable for all types of environments (rural, suburban, urban) by tuning the model parameter and by taking into account the elevation angle as well as the mobile velocity and the carrier frequency. The rural environment is considered, with the following statistical channel modeling derived from [5].

The Probability Density Function (PDF) $P_d(r)$ of received signal envelope r is given by

$$P_d(r) = \int_0^\infty p(r|s)p_s(s)ds \qquad (1)$$

In (1) $p(r|s)$ is a Rice PDF conditioned on a certain shadowing s

$$p(r|s) = 2(K+1)\frac{r}{s^2}e^{-(K+1)\frac{r^2}{s^2}-K}I_0\left(2\frac{r}{s}\sqrt{K(K+1)}\right) \qquad (2)$$

where I_0 is the zero order modified Bessel function of the first kind, and K is the so called Rice factor.
The shadowing s is lognormal with PDF

$$p_s(s) = \frac{1}{\sqrt{2\pi}h\sigma s}e^{\frac{-1}{2}\left(\frac{\ln s-\mu}{h\sigma}\right)^2} \qquad (3)$$

where $h = (\ln 10)/20$, μ and $(h\sigma)^2$ are respectively the mean and the variance of the lognormal process.

It is important to note that errors tend to occur in bursts due to the trellis-like structure. In addition, the channel fading process tend to be correlated over a RS code-word span, especially when the data rate is near the coherence bandwidth [9]. To counteract the correlation effect, interleaving is commonly used. Ideally, we should have two stages of interleaving/deinterleaving in order to remove any correlation between successive symbols. The outer block interleaver (shown dashed in Figure 1) is

designed to spread out the TCM decoding error. However, when a powerful burst error correcting code (e.g. a long RS code blocklength) is used as the outer code, then the outer interleaver can be removed. This is useful for applications such as MSC in which time delay allocated to interleaving process is limited. The inner interleaving stage is required to randomize the errors generated by the channel. A convolutional interleaver structure is used for two main reasons [10] : convolutional interleaver have less than half the time delay for the same degree of interleaving as block interleaver (important fact in speech transmission applications), and is easier to synchronize than block interleaver (important fact in fading channels where synchronization must be rapid to recover from deep fades).

IV. Evaluation of Bit Error Rate Performances

The following analysis is partly derived from [11] and perfect interleaving is assumed.

If the decoded bits from the inner decoder are independent, the input RS-symbol error rate P_{si} (an input RS-symbol consists of m bits) is

$$P_{si} = 1-(1-P_b)^m \tag{4}$$

where P_b is the bit error rate at the output of the inner code.

The method usually employed to evaluate P_b when operating over slow fading channel is to average the AWGN channel probability of error over fading channel statistics

$$P_b = \int_0^\infty P_{b/r} P_d(r) dr \tag{5}$$

where $P_d(r)$ represents the fading channel PDF.

If the input RS-symbols are independent, the decoded block error rate at the output of the outer decoder is given by the approximation formula

$$P_{bl} \approx \sum_{e=t+1}^{n} \frac{e+t}{n} \binom{n}{e} P_{si}^{\,e} (1-P_{si})^{n-e} \tag{6}$$

Independent input RS-symbols can be obtained by using interleaving between the inner and the outer codecs (outer interleaver). However, if the outer code is powerful enough, it is able to correct bursts errors coming out from the inner decoder without outer interleaver. The decoded symbol error rate at the outer decoder output is

$$P_s = 1-(1-P_{bl})^{1/n} \tag{7}$$

Finally, the decoded BER is given by

$$BER = 1-(1-P_s)^{1/m} \tag{8}$$

One of the main difficulties with the analytical evaluation of decoded BER is to solve Eq. 5 using Eq. 1 describing the statistical channel model. To face the difficulty, a semi-analytical method is proposed. First Eq. 5 is simulated with interpolation using second order spline functions in order to have a set of regularly distributed values, and then is used to obtain analytically the decoded BER thanks to Eq. 4,6,7 and 8.

The validation of this method is made by comparing its BER results to those given by a pure analytical method and pure simulations for simple cases (e.g. QPSK modulation over AWGN channel or Rayleigh channel). For this two cases, P_b is respectively given by

$$P_b = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_o}}\right) \text{ and } P_b = \frac{1}{2}\left[1-\sqrt{\frac{\overline{E_b/N_o}}{1+\overline{E_b/N_o}}}\right]$$

(the quadratic mean of the channel amplitude is normalized to unity). $\overline{E_b/N_o}$ is the mean bit energy to noise power spectral density ratio.

These results are shown in Figure 2(a) and 2(b).

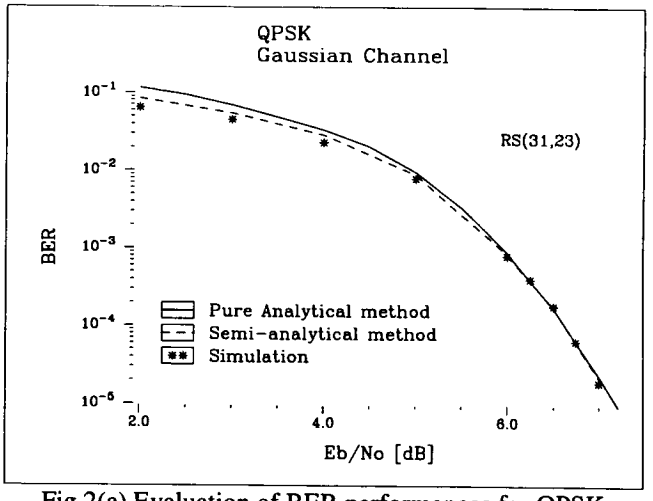


Fig.2(a) Evaluation of BER performances for QPSK modulation over AWGN channel

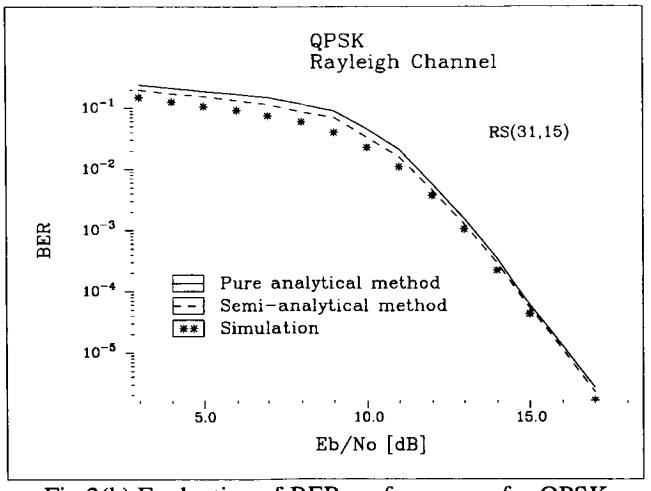


Fig.2(b) Evaluation of BER performances for QPSK modulation over Rayleigh channel

Finally, the semi-analytical method shows a good agreement with the two others methods. This method will be used when considered U8 modulation over the Rice/Rayleigh-Lognormal channel model given by Eq. 1-3.

V. Performance of the Concatenated System with Infinite Interleaving

The following performances are given for the system described in Figure 1 assuming perfect interleaving. Different RS codes with varying blocklength (RS(31,k) up to RS(225,k)) have been investigated in the initial study. For the sake of clarity, only the main results are presented in this paper. Figure 3 shows the required bit energy to noise power density ratio E_b/N_o to achieve BER=10^{-6} for RS(127,k) and different elevation angles. Results are both given for semi-analytical method and simulation.

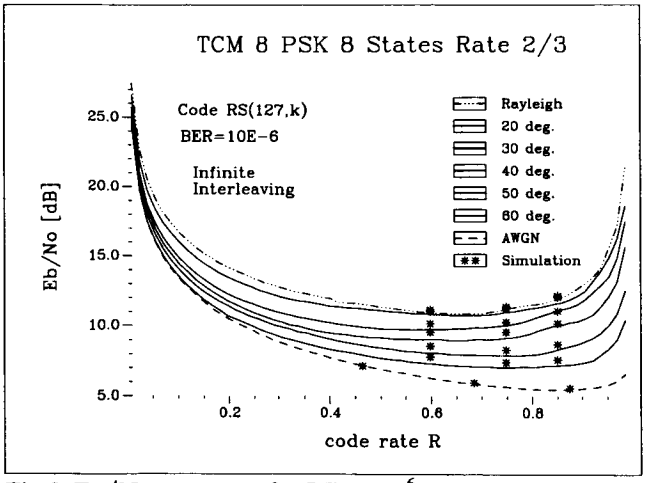


Fig.3 E_b/N_o necessary for BER=10^{-6} with RS(127,k) over Rice/Rayleigh-Lognormal channel

For each elevation angle or type of channel (AWGN or Rayleigh) an optimum code rate R exists which leads to a minimum value of the required E_b/N_o. If the code rate is higher than optimum, the required E_b/N_o increases because the error correction capability of the code is insufficient with respect to the channel behavior. If the code rate is lower, the required E_b/N_o increases again because too much signal energy is wasted by the parity symbols of the code [12]. Comparing the optimum code rate for different channels, we see that the optimum code rate decreases together with the reduction of the elevation angle.

One way to compare RS code performances with different types of modulation, is to evaluate performances at given theoretical spectral efficiency η and BER. For U8 and QPSK modulations without coding, η is equal to 2 bit/s/Hz. Table 1 summarizes for BER=10^{-6}, the minimum E_b/N_o and the corresponding code rate for these modulations (U8 and QPSK) and for two types of channels (AWGN and Rayleigh).

	AWGN				Rayleigh			
	QPSK		U8		QPSK		U8	
	min. E_b/N_0	R	min. E_b/N_0	R	min. E_b/N_0	R	min. E_b/N_0	R
RS(31,k)	7.60	0.68	6.15	0.87	16.75	0.35	12.75	0.61
RS(63,k)	6.85	0.71	5.75	0.87	14.75	0.39	11.55	0.61
RS(127,k)	6.30	0.73	5.45	0.87	13.45	0.43	10.90	0.70
RS(255,k)	5.90	0.73	5.25	0.88	12.55	0.41	10.50	0.72

Table 1 Performances RS-QPSK/U8 for BER=10^{-6}

One can see that for any RS codes considered, the required E_b/N_o is higher and the corresponding code rate R is lower for QPSK modulation than for U8.

The strength of U8 compared to QPSK is again highlighted within Figure 4. This one shows the penalty in E_b/N_o as a function of the spectral efficiency (depending on the code rate) considering RS (63,k) code and BER=10^{-6}. The 0 dB reference corresponds to U8 modulation over AWGN channel with a code rate R=0.6.

It should be noted from Fig. 4 considering the Rayleigh channel, that the change from a spectral efficiency of 1.2 bit/s/Hz to 1.8 bit/s/Hz corresponds to a 3 dB penalty in E_b/N_o with U8, and a 10 dB penalty with QPSK modulation. Consequently, significant spectral efficiency improvement can be achieved with a small penalty in E_b/N_o considering U8 compared to QPSK.

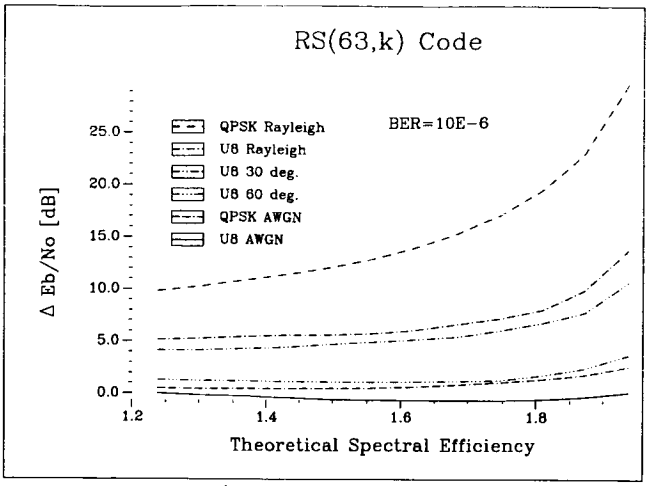


Fig.4 Penalty in E_b/N_o as a function of spectral efficiency for BER=10^{-6}

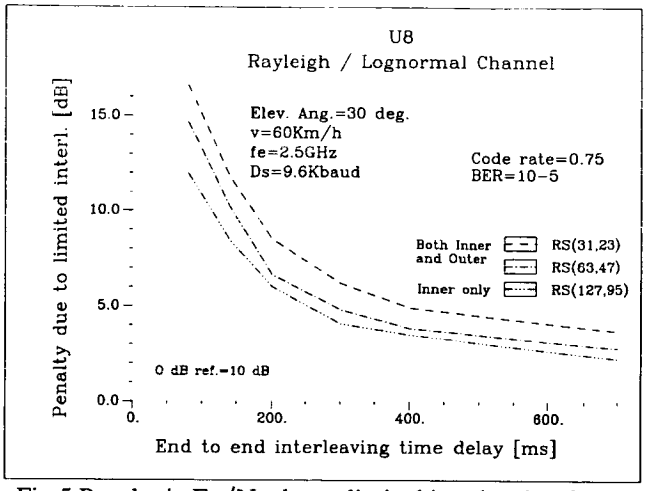


Fig.5 Penalty in E_b/N_o due to limited interleaving for an elevation angle of 30°

VI. Influence of Limited Interleaving

In Section V, full interleaving of code symbols was assumed. When limited interleaving is considered, formulas 4,6,7 and 8 are no longer valid. Therefore, the effect of limited interleaving imposed by time delay constraint should be explored by simulations. The total time delay is the sum of the one originating from outer interleaver (block) and the one from inner interleaver (convolutional) with RS(31,k) and RS(63,k) codes and only the latter with RS(127,k) codes for which no outer interleaver is considered.

Figures 5 and 6 give simulation results with different RS codes. The penalty in E_b/N_o due to limited interleaving is shown as a function of end to end interleaving time delay. This results are given for fixed mobile velocity, carrier frequency, symbol rate and two elevation angles (30 and 60°). The code rate considered here is R=0.75.

These figures show that with decreasing interleaver depth more and more penalty in signal energy has to be paid to achieve a given error probability (BER=10^{-5} is considered here). For short codes a deeper interleaving (resulting in larger time delay) should be used for a given penalty compared to long codes.

Tables 2 and 3 summarize the penalty in E_b/N_o due to interleaving limited to 80,140 and 200 ms with different RS codes, code rates and elevation angles for BER=10^{-5}.

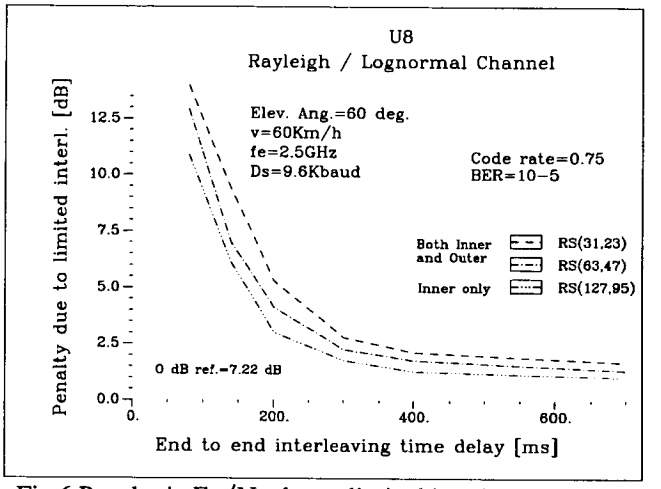


Fig.6 Penalty in E_b/N_o due to limited interleaving for an elevation angle of 60°

	BER=10^{-5}					
	R=0.75			R=0.85		
	RS (31,23)	RS (63,47)	RS (127,95)	RS (31,26)	RS (63,54)	RS (127,108)
80 ms	26.6	24.6	21.9	27.3	25.9	23.3
140 ms	21.8	20.2	18.4	23.1	22.1	20.5
200 ms	18.5	16.6	16.0	20.1	18.5	17.4

Table 2 Required E_b/N_o for limited interleaving with an elevation angle of 30°

	BER=10^{-5}					
	R=0.75			R=0.85		
	RS (31,23)	RS (63,47)	RS (127,95)	RS (31,26)	RS (63,54)	RS (127,108)
80 ms	21.2	20.1	18.1	21.7	21.1	19
140 ms	16.6	14.2	13.3	17.4	16.1	14.8
200 ms	12.5	11.3	10.2	13.9	12.5	11.6

Table 3 Required E_b/N_o for limited interleaving with an elevation angle of 60°

These tables highlight the trade-off between E_b/N_o values, spectral efficiency conditionned by the code rate and end to end interleaving time delay. In any cases, the longer the RS codes considered, the better the performances in terms of E_b/N_o will be, despite the lack of outer interleaver with long code.

VII. Conclusions

In this paper, data transmission over Rice/Rayleigh-Lognormal channel using concatenated RS TCM-U8 was investigated. Considering the difficulties with the theoretical evaluation of BER performances with this type of channel, a semi-analytical method was developed to evaluate BER performances considering full interleaving with different RS codes. This method has shown good agreement with theoretical analysis when available and simulations. It was shown that an optimum code rate exists leading to a minimum value of the required E_b/N_o. This optimum code rate decreases with the elevation angle.

In the final part of the paper the influence of limited interleaving was explored by simulations. Limited interleaving results in a penalty in E_b/N_o which increases as end to end interleaving time delay (or interleaving size) is shorter. For a given acceptable time delay limiting the interleaving depth, an appropriate combination of code rate and RS blocklength can be selected as a function of the affordable value of the required E_b/N_o. On the other hand, when varying the end to end interleaving time delay, an asymptotic behavior of the E_b/N_o penalty is observed.

References

[1] P. Fines and A.H. Aghvami, "A Comparison Study of Low Bit Rate DE-QPSK and TCM-8PSK Fully Digital Demodulators over a Land Mobile Satellite Link", *Globecom'90,* San Diego, California, pp. 401.6.1-401.6.6, Dec. 1990.

[2] G. Ungerboeck, "Channel Coding with Multilevel/ Phase Signals", *IEEE Trans. On Commun.*, vol. COM-28, Jan. 1982.

[3] G.D. Forney, Jr., *Concatenated Codes,* Cambridge, MA: M.I.T. Press, 1966.

[4] R.H. Deng and D.J. Costello, Jr., "High Rate Concatenated Coding System using Bandwidth Efficient Trellis Inner Codes", *IEEE Trans. On Commun.*, vol. COM-37, May. 1989.

[5] G.E. Corazza and F. Vataloro, "A Statistical Model for Land Mobile Satellite Channels and its Application to Nongeostationnary Orbit Systems", *IEEE Trans. On Vehic. Tech.*, vol. VT-43, Aug. 1994.

[6] G.E. Corazza, A. Jahn, and al., "Channel Characterization for Mobile Satellite Communications", *EMPS'95,* Roma, Italy, Oct. 1995.

[7] M.G. Françon and M. Bousquet, "Analysis of Improvement due to Interleaving for Data Transmission over Mobile Satellite Channel with Trellis Coded Modulation System", *EMPS'96,* Roma, Italy, Oct. 1996.

[8] M.L. Moher and J.H. Lodge, "Performance of Concatenated Reed-Solomon Trellis-Coded Modulation over Rician Fading Channels", *IMSC'90,* Ottawa, Canada, pp. 600-604, June 1990.

[9] K.H. Biyari and S.A. Al-Semari, "Concatenated Coding with Inner Trellis Codes for Fading Channels", *Intern. Journ. of Sat. Commun.*, vol. 13, 1995.

[10] A.C.M. Lee and P.J. McLane, "Convolutionally Interleaved PSK and DPSK Trellis Codes for Shadowed Fast Fading Mobile Satellite Communication Channels", *IEEE Trans. On Vehic. Tech.*, vol. VT-39, Feb. 1990.

[11] B. Vucetic, "Bandwidth Efficient Concatenated Coding Schemes for Fading Channels", *IEEE Trans. On Commun.*, vol. COM-41, Jan. 1993.

[12] E. Lutz, "Code and Interleaver Design for Data Transmission over Fading Channels", *Globecom'84.*

Eigen-Based Detection for DS-SS Signals Over Rician Multipath Fading Channels

François Patenaude, John Lodge

Communications Research Centre/Centre de recherches sur les communications
3701 Carling Avenue, Ottawa, Ontario, K2H 8S2, Canada
Phone: 613-990-5878 FAX: 613-998-1686
Email: francois.patenaude@crc.doc.ca

Abstract

The analysis of conventional coherent and differential detection schemes for BPSK Direct Sequence-Spread Spectrum (DS-SS) signals transmitted over Rician multipath fading channels is presented. The theoretical results show that the system average Bit Error Rate (BER) performance is dependent on a given number of eigenvalues and central factors. The theoretical results are compared with the Rayleigh channel theoretical results and against simulation results. Also, eigen-based versions of the detection schemes are presented to show that only a few propagation modes are necessary to obtain an average BER performance close to or even better than that of the full dimension signal. The eigen-based version uses an adaptive rank one modification algorithm with an informatic theoretical criterion to track the dominant modes. Simulation results are presented showing 0 to 1 dB degradation and 0.5 to 2 dB improvement at BER 10^{-3} relative to the conventional coherent and differential detection schemes respectively. It is also shown that practical detectors are still a few dB away from the maximum ratio combining strategy.

Introduction

Analysis of impulse response measurements, for the wideband indoor channel (at 950 MHz and 40 GHz) and for the wideband micro-cellular mobile radio channel at 910 MHz, has shown that the propagation process is essentially dominated by only a few modes [1][2]. The path gains have been shown to be correlated with an overall mean values of ≈ 0.3 and a standard deviation of ≈ 0.2. There are only a small number of equivalent non-correlated paths with significant energy even for impulse responses with a large number of taps, i.e., more than 10. After applying an eigen-transformation on the complex path gain, a study on these non-correlated paths has shown that they are normally distributed, and that the dominant path envelope is well represented by a Rice random process with a K_R factor in the range from −5 dB to −12 dB. The mobile satellite channel is also Rician in many applications. For low elevation angle links, it is also expected that the channel will suffer from time dispersion. Therefore the technique presented here is also of interest for particular mobile satellite links. The eigen-structure detector presented in this paper utilizes linear subspace signal statistics to identify and use the relevant modes of propagation. For each propagation mode, an eigen-structure matched filter is constructed from the RAKE receiver and used in the coherent or differential detection process to form the independent variables representing the diversity. It will be shown through analysis and computer simulation that, for typical multipath channel with relatively long impulse responses, the proposed receiver results in almost equivalent or in some cases, in a reduced average bit error probability relative to conventional methods. The analysis presents the theoretical results for conventional and optimum coherent linear detection, and standard differential detection for a binary DS-SS system. The analysis method is based on an indefinite quadratic form representation of the detector decision variable. Once in such a form, it can be modeled as a sum of weighted non-central chi-square variables for which an expression for the probability of the decision variable being smaller than a threshold exists. The new proposed detector uses an eigen-based algorithm, with a rank-one eigen-decomposition updating over reduced signal subspace at every new received symbol. An information theoretic criterion is used to adaptively determine the number of significant signal dimensions. With this algorithm, the number of orthogonal or uncorrelated paths necessary is either 2 or 3, for both coherent and differential detection. Simulation examples will be given.

System Description

Denote the transmitted signal $s(t)$ at baseband as

$$s(t) = \sum_n b_n f(t - nT), \quad (1)$$

where $b_n \in \{-1,+1\}$ is the independent equiprobable and random transmitted symbol during the n-th signaling interval $nT \le t < (n+1)T$, T is the period of the symbols going into the transmit filter $f(t)$ of duration T_f and of useful bandwidth B_f. The transmit filter $f(t)$ is a wideband filter, having an impulse response equal to a pseudo-noise sequence with rectangular chip of period T_c. The transmitted energy per bit E_b is set to one with an energy constraint on $f(t)$, i.e.,

$$\int_{-\infty}^{\infty} f^2(t)dt = \int_0^{T_f} f^2(t)dt = 1. \quad (2)$$

The transmission of $s(t)$ through a channel with coherence bandwidth smaller than B_f is described by the linear time varying filter with a complex lowpass equivalent impulse response of the form

$$h(t;\xi) = \sum_{d=1}^{D(t)} a_d(t)\delta[\xi - \xi_d(t)], \quad (3)$$

with t the observation time and ξ the excitation delay, $a_d(t)$ the complex path gain and $\xi_d(t)$ the path delay of the d-th path. $D(t)$ is the number of paths at time t. To capture the energy offered by the time dispersion of the channel, the characteristics of the parameters of $h(t;\xi)$ are required. In the analysis that follows, we assume:

i) the complex path gains $a_d(t)$ are complex Gaussian processes,
ii) the complex path gains $a_d(t)$ are constant over a pulse duration T_f,
iii) the path delay $\xi_d(t)$ are changing much more slowly than the $a_d(t)$,
iv) $\max_d\{\xi_d(t)\} << T_f$ or $T_f + \max_d\{\xi_d(t)\} \le T$,
v) the number of paths $D(t)$ is fixed to D.

The condition ii) implies that the fading is slow or moderate and iv) ensures that there is negligible intersymbol interference or implies that a guard interval is used. With these assumptions, the received signal after matched filtering with $f(t)$ is expressed as

$$r(t) = b_n \sum_{d=1}^{D} a_d(t) p[t - \xi_d(t) - nT] + w(t) \quad (4)$$

for $nT \le t < (n+1)T$, where $p(t) = f(t) \otimes f(t)$ and $\otimes$ represents the convolution operation. The random process $w(t)$ with variance N_0 is the result of filtering a white Gaussian noise process filtered with $f(t)$. Making use of assumptions ii) and iii) above, the variable t in the time interval $nT \le t < (n+1)T$ can be fixed to $t_n \equiv nT$ in $a_d(t)$ and $\xi_d(t)$. With the change of variable $\tau = t - nT$, equation (4) is equivalent to

$$r(t_n;\tau) = b_n \sum_{d=1}^{D} a_d(t_n) p[\tau - \xi_d(t_n)] + w_n(\tau) \quad (5)$$

for $0 \le \tau < T$ with $w_n(\tau) \equiv w(\tau + nT)$. In the discrete-time format, (5) can be rewritten in a vector form as

$$\mathbf{r}_n = b_n \mathbf{P}\mathbf{a}_n + \mathbf{w}_n, \quad (6)$$

$\mathbf{a}_n = [a_1(t_n), \ldots, a_D(t_n)]^T$, $\mathbf{P} = \{\mathbf{p}[\xi_1(t_n)], \ldots, \mathbf{p}[\xi_D(t_n)]\}$, with $\mathbf{p}[\xi_d(t_n)] = \{p[\tau_1 - \xi_d(t_n)], \ldots, p[\tau_K - \xi_d(t_n)]\}^T$, and $\mathbf{w}_n = [w_n(\tau_1), \ldots, w_n(\tau_K)]^T$, $\tau_k \equiv kT_c/2$ with $1 \le k \le K$. The variance of the samples of the additive noise process is $E\{|w_n(\tau_k)|^2\} = \sigma_w^2 = N_0$. The value of K is determined by the $\max_d\{\xi_d(t)\}$. The model is also applicable for cases where the RAKE taps are located at the maximum of the average power delay profile. In both cases, the elements of $\mathbf{P}\mathbf{a}_n$ are non-stationary while the elements of $\mathbf{w}_n$ are stationary. The vector samples $\mathbf{r}_n$ are further processed coherently or differentially depending on the system. The probability of bit error will be derived for these detection schemes, by making use of the quadratic form of the decision variable. The theoretical results will serve as reference marks for the basic coherent and differential receiver.

BER Analysis for Coherent Detection

The computation of the average bit error probability P_e for the case of coherent detection is derived. Define $\mathbf{y}_n$ as the vector over which the RAKE receiver output is projected. For the case where the symbols b_n are equiprobables and independent, the exercise is equivalent to finding the probability that the decision variable $U_n < 0$, knowing that $b_n = +1$, has been transmitted. Two coherent schemes can be envisaged: an optimal scheme in the signal-to-noise ratio where $\mathbf{y}_n = \mathbf{R}_w^{-1}\mathbf{P}\mathbf{a}_n$ [3], $\mathbf{R}_w = E\{\mathbf{w}_n\mathbf{w}_n^H\}$, and a conventional scheme, where $\mathbf{y}_n = \mathbf{P}\mathbf{a}_n$. In both cases, the decision variable for the detection of the received vector can be written in quadratic form. The form is

$$\begin{aligned} U_n &= \Re\{\mathbf{y}_n^H \mathbf{r}_n\} \\ &= \frac{1}{2}\left[\mathbf{y}_n^H \mathbf{P}\mathbf{a}_n + \mathbf{a}_n^H \mathbf{P}^T \mathbf{y}_n + \mathbf{y}_n^H \mathbf{w}_n + \mathbf{w}_n^H \mathbf{y}_n\right] \\ &= \frac{1}{2}\begin{bmatrix}\mathbf{a}_n^H \mathbf{P}^T & \mathbf{w}_n^H\end{bmatrix}\begin{bmatrix}2\left(\mathbf{X}^{-1}\mathbf{P}\mathbf{a}_n + \mathbf{X}^{-1}\mathbf{w}_n\right) \\ \mathbf{X}^{-1}\mathbf{P}\mathbf{a}_n\end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix}\mathbf{a}_n^H \mathbf{P}^T & \mathbf{w}_n^H\end{bmatrix}\begin{bmatrix}2\mathbf{X}^{-1} & \mathbf{X}^{-1} \\ \mathbf{X}^{-1} & \mathbf{0}_K\end{bmatrix}\begin{bmatrix}\mathbf{P}\mathbf{a}_n \\ \mathbf{w}_n\end{bmatrix} \\ U_n &= \mathbf{z}_n^H \mathbf{Q}\mathbf{z}_n \end{aligned} \quad (7)$$

with $\Re\{x\}$ the real part of x, $\mathbf{X}=\mathbf{R}_w$ for the optimal scheme and, $\mathbf{X}=\mathbf{I}_K$ for the conventional one. Also,

$$\mathbf{z}_n^{\mathrm{H}}=\left[\mathbf{a}_n^{\mathrm{H}}\mathbf{P}^{\mathrm{T}}\quad \mathbf{w}_n^{\mathrm{H}}\right], \tag{8}$$

and

$$\mathbf{Q}=\frac{1}{2}\begin{bmatrix}2\mathbf{X}^{-\mathrm{i}} & \mathbf{X}^{-1}\\ \mathbf{X}^{-1} & \mathbf{0}_K\end{bmatrix}. \tag{9}$$

The optimum receiver thus uses the information about the noise correlation in the symmetric matrix $\mathbf{Q}$. The matrices $\mathbf{I}_K$, $\mathbf{0}_K$ are respectively the identity matrix and the zero matrix of dimension $K\times K$. For a Rayleigh fading channel, it has been shown by Barrett [4] that the average probability of bit error can be expressed as

$$P_e=\sum_{\beta_d<0}\prod_{\substack{i=1\\ i\neq d}}^{2D}\frac{1}{1-\beta_i/\beta_d}, \tag{10}$$

where β_i are the eigenvalues of $\mathbf{R}_z\mathbf{Q}$ with

$$\mathbf{R}_z\equiv E\left\{\mathbf{z}_n\mathbf{z}_n^{\mathrm{H}}\right\}=\begin{bmatrix}\mathbf{R}_s & \mathbf{0}_K\\ \mathbf{0}_K & \mathbf{R}_w\end{bmatrix}, \tag{11}$$

and $\mathbf{R}_s=E\left\{\mathbf{P}\mathbf{a}_n\mathbf{a}_n^{\mathrm{H}}\mathbf{P}^{\mathrm{T}}\right\}$.

For the Rician channel, the probability of error is more complicated to evaluate. The Rician channel arises in cases where the expected values of some or all elements of $\mathbf{a}_n$ are not zero. Let $\mu^{\mathrm{T}}=E\left\{\mathbf{z}_n^{\mathrm{T}}\right\}=\left[\mathbf{m}^{\mathrm{T}}\quad \mathbf{0}_{1\times K}\right]$ where $\mathbf{m}=\mathbf{P}E\left\{\mathbf{a}_n\right\}$. For this case, the covariance matrix of $\mathbf{z}_n$, denoted Σ_z, is assumed to be semi-positive definite. It is not positive define since it is possible to have models where the number of modes is smaller than the length K of $\mathbf{r}_n$. Thus $\Sigma_z=\mathrm{Cov}\left\{\mathbf{z}_n\right\}\geq 0$, $\mathrm{rank}\left(\Sigma_z\right)=r$ with $K<r\leq 2K$, and $\Sigma_z=\mathbf{B}\mathbf{B}^{\mathrm{H}}$ where $\mathbf{B}$ is a $2K\times r$ matrix. Then, the quadratic form (7) can be rewritten [5]

$$U_n=\sum_{i=1}^{r}\lambda_i\left(Z_i+\frac{b_i}{\lambda_i}\right)^2 \tag{12}$$

where

$$\mathbf{b}^{\mathrm{T}}=\left(b_1,\ \ldots,\ b_r\right)=\mu^{\mathrm{T}}\mathbf{Q}\mathbf{B}\mathbf{V}, \tag{13}$$

and where the λ_i's $i=1,\ldots,r$ and $\mathbf{V}$ are respectively the eigenvalues and eigenvector matrix of $\mathbf{B}^{\mathrm{H}}\mathbf{Q}\mathbf{B}$. The random variables Z_i's are normal with zero mean and unit variance. They are given by

$$\mathbf{Z}_n=\left(Z_1,\ldots,Z_r\right)^{\mathrm{T}}=\mathbf{V}^{\mathrm{H}}\mathbf{B}^{\mathrm{H}}\left(\mathbf{z}_n-\mu\right). \tag{14}$$

Although Σ_z is semi-positive define, the matrix $\mathbf{B}^{\mathrm{H}}\mathbf{Q}\mathbf{B}$ is not, as in the Rayleigh channel case. The quadratic form is thus indefinite and the random variable U_n is now seen as a weighted sum of unit variance random variables with mean b_j/λ_j, or a weighted sum of independent non-central chi-square random variables with centrality factor b_j/λ_j. The characteristic function of U_n given by (12) is known [6] to be

$$\phi_U(s)=\prod_{j=1}^{r}\left(1-i\lambda_j s\right)^{-1}\exp\left\{i\sum_{j=1}^{r}\frac{b_j^2\lambda_j s}{\left(1-i\lambda_j s\right)}\right\}. \tag{15}$$

It has been shown by [7], that

$$\Pr\left[U_n<y\right]=\frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}s^{-1}\Im\left\{\exp\left(-isy\right)\phi_U(s)\right\}ds. \tag{16}$$

Applying (15) in (16) with $y=0$, and proceeding as in [8], it is found that

$$P_e=\frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}\frac{\sin\theta(u)}{u\rho(u)}du, \text{ where} \tag{17}$$

$$\theta(u)=\sum_{j=1}^{r}\left[\tan^{-1}\left(\lambda_j u\right)+\frac{b_j^2 u}{\lambda_j\left(1+\lambda_j^2u^2\right)}\right],$$

$$\rho(u)=\prod_{j=1}^{r}\left(1+\lambda_j^2u^2\right)^{1/2}\exp\left\{\sum_{j=1}^{r}\frac{\left(b_j u\right)^2}{\left(1+\lambda_j^2u^2\right)}\right\}. \tag{18}$$

For most practical channels, the ∞ upper limit of integration in (16) can easily be replaced by a finite value without sacrificing the accuracy.

BER Analysis for Differential Detection

For differentially encoded symbols, the decision variable at the receiver at time n is given by

$$U_n=\Re\left\{\mathbf{r}_{n-1}^{\mathrm{H}}\mathbf{r}_n\right\}. \tag{19}$$

The sign of U_n is indicative of the transmitted bit. Assuming that the information bits were differentially encoded at the transmitter, the probability of bit error for U_n is again expressed in a quadratic form. Proceeding as in (7), the decision variable becomes

$$U_n=\mathbf{z}_n^{\mathrm{H}}\mathbf{Q}\mathbf{z}_n, \tag{20}$$

where

$$\mathbf{z}_n^{\mathrm{H}}=\left[\mathbf{a}_n^{\mathrm{H}}\mathbf{P}^{\mathrm{T}}\quad \mathbf{a}_{n-1}^{\mathrm{H}}\mathbf{P}^{\mathrm{T}}\quad \mathbf{w}_n^{\mathrm{H}}\quad \mathbf{w}_{n-1}^{\mathrm{H}}\right], \text{ and}$$

$$\mathbf{Q}=\frac{1}{2}\begin{bmatrix}\mathbf{B} & \mathbf{B}\\ \mathbf{B} & \mathbf{B}\end{bmatrix}, \text{ with}$$

$$\mathbf{B}=\begin{bmatrix}\mathbf{0}_K & \mathbf{I}_K\\ \mathbf{I}_K & \mathbf{0}_K\end{bmatrix}.$$

The matrix $\mathbf{R}_z$ is know and given by

$$\mathbf{R}_z = E\{\mathbf{z}\mathbf{z}^H\} = \begin{bmatrix} \mathbf{R}_s & \kappa\mathbf{R}_s & \mathbf{0}_{2K} & \\ \kappa\mathbf{R}_s & \mathbf{R}_s & & \\ \mathbf{0}_{2K} & & \mathbf{R}_w & \mathbf{0}_K \\ & & \mathbf{0}_K & \mathbf{R}_w \end{bmatrix}, \quad (21)$$

where the value of κ is given by

$$\kappa = \left[R_g(t) \overset{D-1}{\otimes} R_g(t) \right]_{t=T}, \quad (22)$$

with $\overset{D-1}{\otimes}$ denoting the $(D-1)$-fold convolution. The quantity $R_g(t)$ is the normalized (by the variance) autocorrelation function of the complex orthogonal gains $\mathbf{V}_s^H \mathbf{P}\mathbf{a}_n$, with $\mathbf{V}_s$ the eigenvectors of $\mathbf{R}_s$. For simplicity we have set the shape of $R_g(t)$ to be the same for all complex orthogonal gains. The procedure to calculate the probability of error takes into account the interpath interference, the fading variations, the correlation between the path gains and the additive colored noise.

The probability of bit error for the Rayleigh channel is then computed with expression (10), using the new matrices defined above. For the Rician channel, the procedure is similar as well with $\mathbf{\Sigma}_z$ having a form like (21) and $\mu = E\{\mathbf{z}_n\} = [\mathbf{m}^T \quad \mathbf{m}^T \quad \mathbf{0}_{1\times K} \quad \mathbf{0}_{1\times K}]$. Note that the dimensionally of the problem is now $4K$ instead of $2K$ for the full rank $\mathbf{\Sigma}_z$.

Eigen-Based Detection

The eigen-based version of the coherent and differential detection scheme is partially based on a rank one modification of the eigen-decomposition [9]. The full algorithm has already been presented in [1] and will be summarized here for completeness. The autocorrelation matrix $\mathbf{R}_z$ at time n is recursively estimated (dropping the z subscript for clarity)

$$\mathbf{R}_n = \alpha\mathbf{R}_{n-1} + (1-\alpha)\mathbf{r}_n\mathbf{r}_n^H, \quad (23)$$

where $0 < \alpha < 1$ is a weighting factor for the newly received vector. Assuming a known eigen-decomposition at time $n-1$, $\mathbf{R}_n$ is given by

$$\begin{aligned} \mathbf{R}_n &= \alpha\mathbf{V}_{n-1}\Lambda_{n-1}\mathbf{V}_{n-1}^H + (1-\alpha)\mathbf{r}_n\mathbf{r}_n^H \\ &= \mathbf{V}_{n-1}\left[\alpha\Lambda_{n-1} + (1-\alpha)\mathbf{c}_n\mathbf{c}_n^H\right]\mathbf{V}_{n-1}^H, \mathbf{c}_n = \mathbf{V}_{n-1}^H\mathbf{r}_n. \quad (24) \end{aligned}$$

Removing the modulation from inside the bracket, the matrix becomes

$$\mathbf{R}_n = \mathbf{V}_{n-1}\mathbf{G}_n\left[\alpha\Lambda_{n-1} + (1-\alpha)\mathbf{q}_n\mathbf{q}_n^T\right]\mathbf{G}_n^H\mathbf{V}_{n-1}^H, \quad (25)$$

where

$$\mathbf{q}_n = \mathbf{G}_n^H\mathbf{c}_n \text{ and } \mathbf{G}_n = \text{diag}\left(c_1/|c_1|, \cdots, c_K/|c_K|\right).$$

Making the lower energy dimension spherical, the matrix is well approximated by

$$\mathbf{R}_n \approx \mathbf{V}_{n-1}\mathbf{G}_n\mathbf{H}_n\left[\alpha\mathbf{D}_{n-1} + (1-\alpha)\mathbf{s}_n\mathbf{s}_n^T\right]\mathbf{H}_n^T\mathbf{G}_n^H\mathbf{V}_{n-1}^H, \quad (26)$$

with $\mathbf{s}_n = \mathbf{H}_n^T\mathbf{q}_n$ and $\mathbf{H}_n$ is a Householder transform constructed with the $K-r$ values of $\mathbf{q}_n$. If the diagonal elements (eigenvalues) of the inner bracket are denoted by $\mathbf{D}_n$ and the eigenvalues by $\mathbf{U}_n$, then the new eigenvectors are

$$\mathbf{V}_n \approx \mathbf{V}_{n-1}\mathbf{G}_n\mathbf{H}_n\mathbf{U}_n. \quad (27)$$

For the differential detection scheme, the decision variables U_n^D are giving by

$$U_n^D = \Re\left[\sum_{i=1}^{M_n} \frac{\lambda_{n,i} - \lambda_{n,M+1}}{\lambda_{n,i}} \mathbf{v}_{n-1,i}^H \mathbf{v}_{n,i}\right], \quad (28)$$

with $\lambda_{n,i}$'s the M_n largest eigenvalues of $\mathbf{D}_n$. The value M is given by the Maximum Description Length (MDL) information theoretic criteria [10], and $\lambda_{n,M+1}$ is the average of the $K-M$ smallest eigenvalues of $\mathbf{D}_n$. For the coherent detection scheme, the decision variable depends on the method used to estimate the complex path gains. Here, the receiver of Fig.1 was investigated. The transmitter inserts a reference symbol every $r-1$ symbols to enable the receiver to get samples of the channel. The receiver first reduces the dimension of the problem from K to D and then extracts the channel state information from these D orthogonal paths.

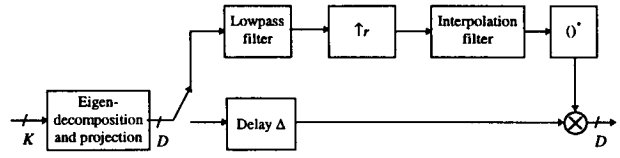


Figure 1. Reference symbol eigen-based coherent detector.

The algorithm is the same as presented above with the decision variable given by

$$U_n^C = \Re\left[\sum_{i=1}^{M_n} \hat{g}_{n,i}^* \mathbf{v}_{n,i}^H \mathbf{r}_n\right], \quad (29)$$

where $\hat{g}_{n,i}$ is the estimated orthogonal channel state information (upper branch of Fig. 1) for the i-th mode of $\mathbf{D}_n$. The value of M_n in this case is set to the nearest integer according to the filtered MDL criterion. Both U_n^D and U_n^D have been found mainly by empirical techniques but based on basic combining methods.

Theoretical and Simulation Results

To give an example of the performance of the eigen-based detection schemes, two channels with short and long delay spread were used. The transmitted m-sequence has 127 chips and the symbol rate is 9.6 baud. The short-delay channel has $K = 14$ paths with

$\max_d\{\xi_d(t)\} = 5$ μs, while the long-delay channel has $K = 37$ paths with $\max_d\{\xi_d(t)\} = 12.5$ μs. The received signal is sampled at two samples per chip for processing by the RAKE receiver.

The coherent detection results are presented in Fig. 2 for the Rayleigh channel and in Fig. 3 for the Rice channel where the dominant path has a Rician factor K_R of -10 dB. It can be seen that for the conventional scheme performance is better than for the eigen-based version but by only 1 dB for the Rayleigh channel at BER 10^{-3}. The complexity of the eigen-based version is less since only D paths needs to be tracked instead of K. The eigen-decomposition complexity is about KD^2. For the Rician channel, the eigen-based approach has a performance similar to the conventional structure, for P_e smaller than 10^{-2}. In this last channel, both detectors based on reference symbols have a performance significantly worse than a receiver with perfectly known channel state information (CSI).

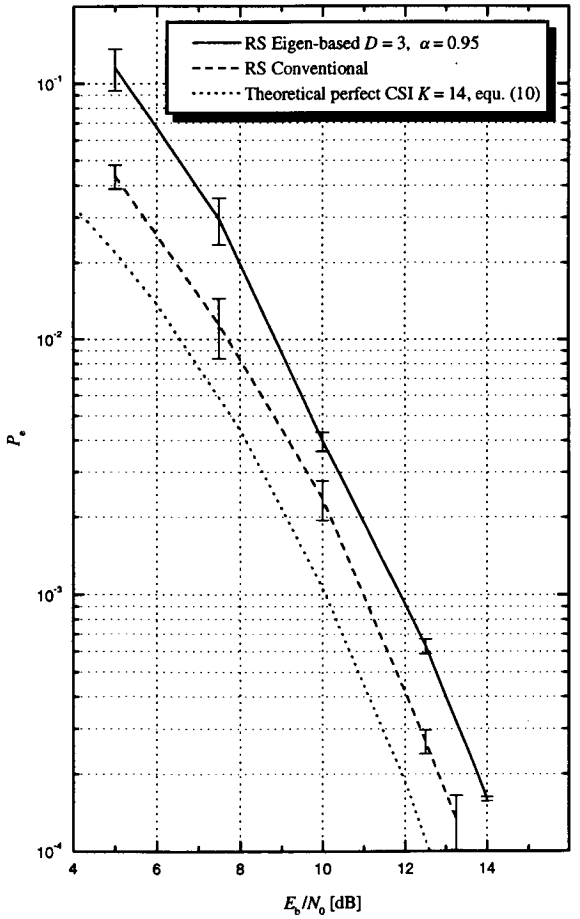


Figure 2. P_e for reference symbol (RS) based coherent detection with Rayleigh short delay channel and normalized fading rate of 0.0025.

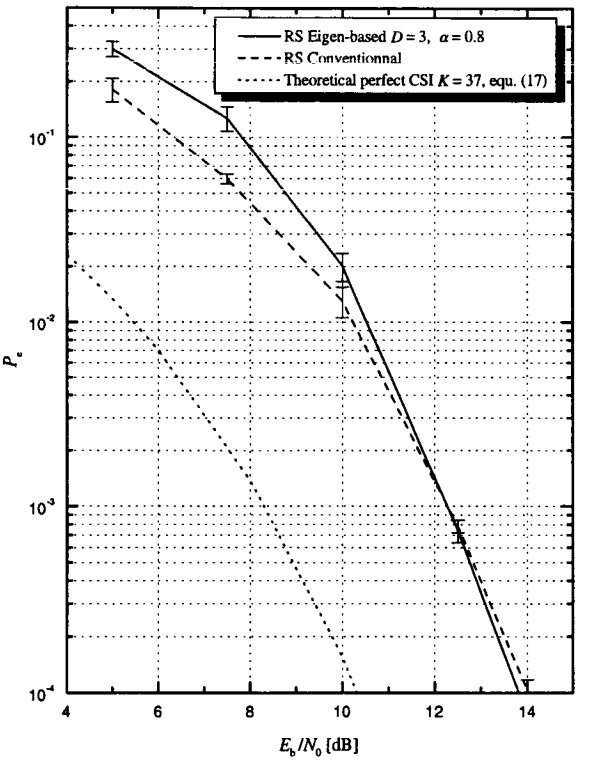


Figure 3. P_e for reference symbol (RS) based coherent detection with first path Rice $K_R = -10$ dB long delay channel and normalized fading rate of 0.025.

The differential detection results are now presented in Figures 4 and 5. Here the eigen-based approach performs better than the conventional detector by 1 dB or more, for most of the range of P_e. The theoretical results for coherent detection with differential decoding have been included has well to indicate the potential performance improvement that could still be obtained with a smarter detector. The complexity of the eigen-based detector also on the order of KD^2 compared to K for conventional differential detection. Note the excellent agreement between the theoretical results and the simulated results in Fig. 5. The theoretical computation were the most numerically sensitive due to the relatively large value of K. The truncation of the integral upper limit is negligible as well.

Conclusion

This paper has presented theoretical results for conventional maximum ratio coherent detection and conventional differential detection for binary DS-SS signals over multipath Rician fading channels. It was also shown by computer simulation that an eigen-based version of the coherent and differential detection receivers can perform almost as well, or better in some cases, as the conventional counterparts when the

channel can be represented by a small number of orthogonal paths. Mobile satellite SS signals transmitted over channels with time dispersive properties can therefore benefit from the presented techniques. The complexity of the eigen-based detectors is not drastically larger when 2 or 3 equivalent paths are used.

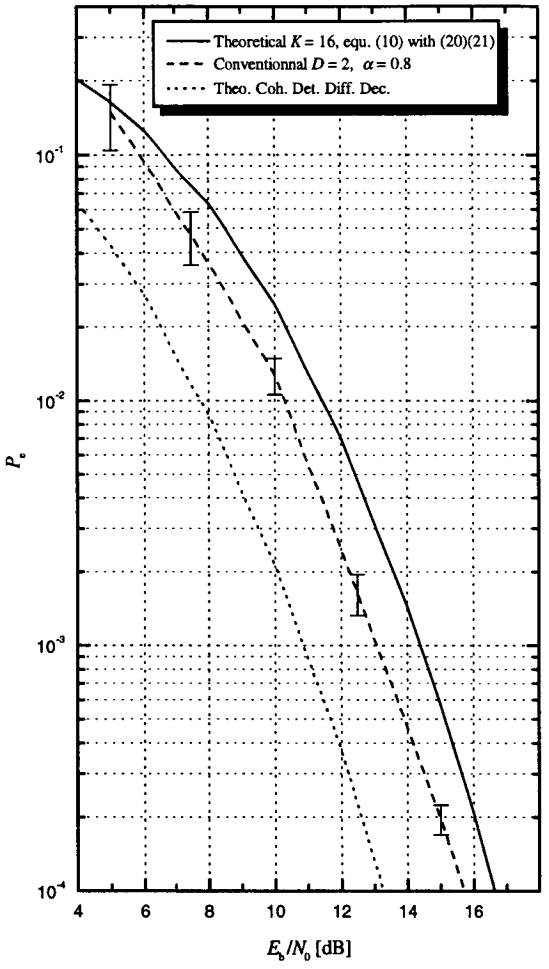


Figure 4. P_e for differential detection with Rayleigh short delay channel and normalized fading rate of 0.01.

References

[1] F. Patenaude, J.H. Lodge, and J.-Y. Chouinard, "Temporal Correlation Analysis of Frequency Selective Fading Channels," *Proceedings of the Sixth Annual International Conference on Wireless Communications (Wireless'94)*, Calgary, Canada, vol. 1, pp. 134-139, July 1994.

[2] F. Patenaude, J.H. Lodge, and J.-Y. Chouinard, "Eigen-Analysis of Wideband Fading Channel Impulse Responses," Submitted to the *IEEE Transactions on Vehicular Technology*, July 1996.

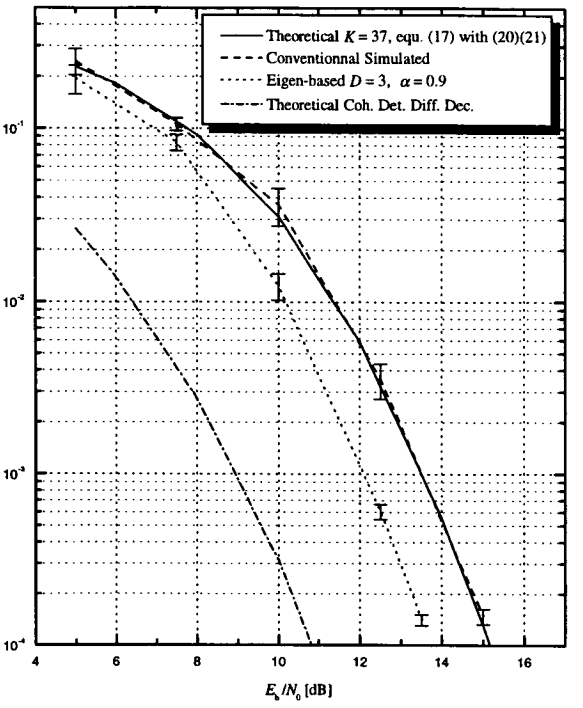


Figure 5. P_e for differential detection with first path Rice $K_R = -10$ dB long delay channel and normalized fading rate of 0.025.

[3] H. Stark, and J.W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice-Hall, 1994.

[4] M.J. Barrett, "Error Probability for Optimal and Suboptimal Quadratic Receivers in Rapid Rayleigh Fading Channels," *IEEE Jour. Select. Areas. Commun.*, vol. 5, no. 2, pp. 302-304, Feb. 87.

[5] A.M. Mathai, and S.B. Provost, *Quadratic Forms in Random Variables Theory and Applications*, Marcel Dekker Inc, 1992.

[6] G.L. Turin, "The Characteristic Function of Hermitian Quadratic Forms in Complex Normal Variables," *Biometrika*, vol. 47, pp. 199-201, 1960.

[7] J. Gil-Pelaez, "Note on the Inversion Theorem," *Biometrika*, vol. 38, pp. 481-482, 1951.

[8] J.P. Imhof, "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, vol. 48, pp. 419-426, 1961.

[9] R.D. DeGroat, and R.A. Roberts, "Efficient, Numerically Stabilized Rank-One Eigenstructure Updating," *IEEE Trans. Acoust., Speech. Signal Processing*, vol. 38, no. 2, pp. 301-316, Feb. 1990.

[10] M. Wax, et I. Ziskind, "Detection of the Number of Coherent Signals by the MDL Principle," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 1190-1196, August 1989.

A New Channel Estimator for Narrowband Satellite Mobile Communications

Jan Erik Håkegård, Marie-Laure Boucheret

Ecole Nationale Supérieure des Télécommunications (ENST), site de Toulouse
10 Avenue Edouard Belin, BP 4004
31 028 Toulouse-FRANCE
Phone: +33 - 562 17 29 84 FAX: +33 - 562 17 29 89
email: haakegaa@tlse.enst.fr, bouchere@tlse.enst.fr

ABSTRACT

A new CSI estimator is proposed, which in multi-user narrowband communications increases the capacity of satellite mobile communication systems. The impact of the CCI on the CSI estimates is reduced by the means of embedded orthogonal pilot sequences. The multi-user metric, which is optimal in the ML sense with perfect CSI, then assures a good performance for any SIR. The receiver performs well even if the users are not perfectly symbol synchronous. This approach is best suited to handle small number of users due to a high complexity and spectral efficiency aspects. New frequency reuse schemes adapted to the receiver are proposed.

1. INTRODUCTION

In satellite mobile communications, multiple spot beams covering the intended service area are necessary to reduce the required transmit power per channel. Both fixed contiguous beams and agile beams capable of being pointed anywhere within the coverage area are strategies studied for satellite mobile systems [1]. In such systems, frequency reuse (FR), where users located within different beams reuse the same frequency slot, is a key consept. The advantage of FR is that it increases the bandwidth utilization efficiency, the drawback is that it introduces co-channel interference (CCI). With a contiguous circular fixed-beam organization, the FR distance, i.e., the minimum distance which allows the same frequency to be reused, decides the beam size and the reuse pattern [2]. It depends on the allowable interference level in the link budget and on the angular separation between the main beam and the allowable side-lobe level. In both the contiguous fixed-beam case and the agile beam approach, FR can be further facilitated by means of active interference suppression (AIS). The principle of AIS is to place nulls in the antenna diagram in the direction of the interferers by the means of digital signal processing, and hence increase the co-channel isolation.

Both the maximization of the FR distance and the AIS are techniques which aim to minimize the interference level prior to the decision device. In this paper, we propose a receiver structure which performs well at any signal to interference ratio (SIR). Recently, some concepts from multi-user communication have been investigated in the context of coded transmission over fading channels with space diversity [3]. With perfect knowledge of the channel state information (CSI), i.e., the superposition of various propagation effects like fading and shadowing, the multi-user receiver does not exhibit an irreducible error rate, denoted an error floor. Hence, the error probability can be made arbitrary small by increasing the transmitted power. With estimated CSI however, this is no longer the case. With a Rayleigh fading channel, conventional CSI estimators, based on for instance embedded pilot symbols or pilot tones, delivers under severe CCI unreliable CSI estimates which prevent the multi-user receiver to perform properly.

The CSI estimator proposed in this paper reduces, and ideally eliminates, the impact of the CCI on the CSI estimates. It is based on orthogonal pilot sequences and interpolation. Simulation results illustrate that the resulting error floor is well below the bit error rates (BERs) of interest. This approach is best suited to handle a small number of users. It incites consequently a new beam organization adapted to this receiver, where no longer separate co-channel beams, but rather groups of closely located co-channel beams are uniformly distributed within the service area. A relatively high intra group interference level is taken care of by digital signal processing, and the inter group in-

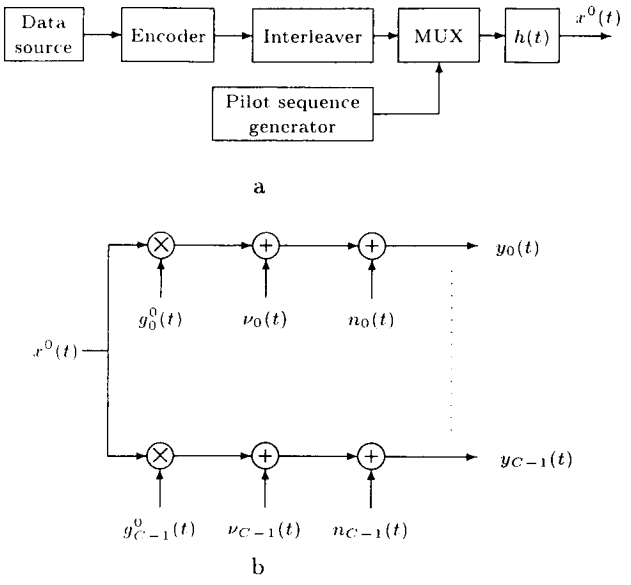


Figure 1: *The system model. a: transmitter, b: channel*

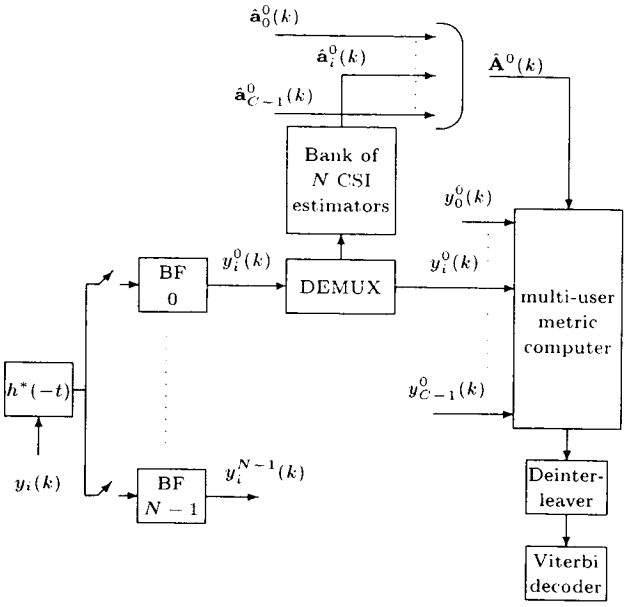


Figure 2: *The receiver model.*

terference level will be smaller than for a conventional beam organization.

The outline of the paper is as follows. In section 2, the system model is defined and the notations are established. The multi-user metric computer is described in section 3, and in section 4 the CSI estimator is presented. The performance of this receiver structure is evaluated in section 5 by computer simulations, and finally, in section 6 some conclusions of this study are drawn.

2. The System Model

The system model is illustrated in the figures 1 and 2. In the transmitter, sequences of M_p pilot symbols are periodically inserted into the interleaved coded data sequence. Hence, the transmitted signal is formatted into frames of length $M = M_p + M_d$, where M_d is the number of data symbols per frame. The number of users allocated to the same frequency band is N, of which one is the user of interest and $N-1$ are interferers. Without loss of generality, we consider the 0'th user to the one of interest. The number of receiving antennas is C. To avoid that a higher number of branches entails an increase in total received power, with an obvious gain in performance, we consider *normalized diversity* [4]. This consists of splitting the total energy among the diversity branches for the useful signal as well as for the interferers. The distance between the antennas is assumed large enough to assure that the fading processes of the different branches are uncorrelated. The signal captured by the i'th antenna can be written as:

$$y_i(t) = g_i^0(t)x^0(t) + \nu_i(t) + n_i(t) \quad (1)$$

where $\nu_i(t)$ denotes the sum of the interfering signals, themselves affected by fading:

$$\nu_i(t) = \sum_{n=1}^{N-1} g_i^n(t)x^n(t) \quad (2)$$

$x^0(t)$ is the transmitted useful signal, and $x^n(t)$, $n = 1,\ldots,N-1$ are the transmitted interfering signals. The useful and the interfering signals are assumed to be coded q-PSK, i.e., $x^n(k) \in \mathcal{X}_q\{e^{j\frac{2\pi i}{q}} : i = 0,\ldots,q-1\}$ for all n. $g_i^0(t)$ and $g_i^n(t), n = 1,\ldots,N-1$, are the fading processes affecting the useful and interfering signals, respectively. After matched filtering and optimal symbol sampling with respect to the signal of interest, the received signal is passed through a beamformer (BF). The output of the BF takes on the following form:

$$y_i^0(k) = g_i^0(k)x^0(k) + \sum_{n=1}^{N-1} b_i^{0n} g_i^n(k)z^n(k) + n_i(k) \quad (3)$$

where $g_i^n(k), n = 0,\ldots,N-1$, are zero mean unit variance complex Gaussian variables and b_i^{0n} are the BF coefficients. In the case of symbol synchronous users, the signal to n'th interferer ratio is given by:

$$\Lambda_i^n = \frac{1}{|b_i^{0n}|^2} \quad (4)$$

The total signal to interference ratio (SIR) for the i'th branch is equal to:

$$\Lambda_i = \frac{1}{\sum_{n=1}^{N-1}(\Lambda_i^n)^{-1}} \tag{5}$$

The white additive noise $n_i(k)$ is a zero mean complex Gaussian variable with variance:

$$E[|n_i(k)|^2] = SNR^{-1} \tag{6}$$

In the general case, the signals are not symbol synchronous, and the interfering signals will be affected by inter-symbol interference (ISI). $z^n(k)$, $n = 1, \ldots, N-1$, are samples of the interfering signals after matched filtering and sampled with timing offsets τ_n and carrier offsets ϕ_n, and take on values from the set:

$$\mathcal{S}_n = \left\{ s = \sum_{l=-L}^{L} x^n(k-l) f(lT - \tau_n) e^{j\phi_n} \right\} \tag{7}$$

where $f(t) = h(t) * h^*(-t)$ is a real unit-energy pulse such that $f(lT) = 1$ for $l = 0$ and 0 for $l \neq 0$. The received signals of the C diversity branches are organized in the vector $\mathbf{y}^0(k) = [y_0^0(k), \ldots, y_{C-1}^0(k)]^T$. Defining the vector $\mathbf{n}(k)$ in the same manner, and $\mathbf{a}^{0n}(k) = [b_0^{0n} g_0^n(k) \cdots b_{C-1}^{0n} g_{C-1}^n(k)]^T$ the incoming signal vector can be expressed as:

$$\mathbf{y}^0(k) = \mathbf{A}^0(k)\mathbf{z}(k) + \mathbf{n}(k) \tag{8}$$

where $\mathbf{z}(k) = [x^0(k)\ z^1(k) \cdots z^{N-1}(k)]^T$ and $\mathbf{A}^0(k)$ is a $C \times N$ matrix whose n-th column is $\mathbf{a}^{0n}(k)$. The rows of $\mathbf{A}^0(k)$ are denoted $\mathbf{a}_i^0(k)$, and contain the outputs of the bank of N CSI estimators following the 0'th BF. In practical systems, transmission from interferers can be considered as independent, even if in reality they are TCM-encoded. Under mild assumptions on the TCM scheme involved, the random variables $z^n(k)$ are assumed uniformly distributed over the set $\mathcal{S}_n$ [5].

The i'th receiver branch contains N BFs. The outputs of the BFs can be organized in the vector $\mathbf{y}_i(k) = [y_i^0(k) \cdots y_i^{N-1}(k)]^T$ given by:

$$\mathbf{y}_i(k) = \mathbf{B}_i\mathbf{G}_i(k)\mathbf{z}(k) + \mathbf{n}_i(k) = \mathbf{A}_i(k)\mathbf{z}(k) + \mathbf{n}_i(k) \tag{9}$$

where $\mathbf{B}_i$ is the $N \times N$ beamforming network gain matrix with ones on the diagonal and $\mathbf{G}_i(k)$ is the diagonal matrix containing the fading samples. $\mathbf{B}_i$ is unknown due to uncertainties about the exact position of the mobile and/or the radiation pattern of the satellite antenna. The pilot symbols and the data symbols are separated by a demultiplexer, and the pilot symbols are used in the CSI estimator. A metrics computer exploits the fact that all the components of $\mathbf{y}^0(k)$ and $\hat{\mathbf{A}}^0(k)$ are available. The output of the metrics computer is fed to the Viterbi decoder.

3. The multi-user metrics computer

In the multi-user metrics computer, all available information is exploited. For each $x^0(k) \in \mathcal{X}_q$, we define the set of vectors $\mathcal{S}(x^0(k))$ as all vectors $\mathbf{z}(k)$ for which the first element is equal to $x^0(k)$, and the remaining components take on values from the signal sets $\mathcal{S}_n$. Let $E_{\tilde{\mathbf{z}}(k)}$ denote expectation with respect to $\tilde{\mathbf{z}}(k) \in \mathcal{S}(\tilde{x}^0(k))$. A likelihood index equivalent to the log-likelihood function is given by:

$$\begin{aligned} \lambda(\mathbf{y}^0(k), \tilde{x}^0(k)) &= \log E_{\tilde{\mathbf{z}}(k)} \left\{ e^{-|\mathbf{y}^0(k) - \mathbf{A}^0(k)\tilde{\mathbf{z}}(k)|^2} \right\} \\ &= \log E_{\tilde{\mathbf{z}}(k)} \left\{ e^{\Omega(\tilde{\mathbf{z}}(k))} \right\} \end{aligned} \tag{10}$$

where:

$$\Omega(\tilde{\mathbf{z}}(k)) = 2\mathrm{Re}\left\{\tilde{\mathbf{z}}^\dagger(k)\mathbf{A}^{0\dagger}(k)\mathbf{y}^0(k)\right\} - \tilde{\mathbf{z}}^\dagger(k)\mathbf{H}(k)\tilde{\mathbf{z}}(k) \tag{11}$$

The superscript "$\dagger$" denotes conjugate transpose. The matrix $\mathbf{H}(k) = \mathbf{A}^{0\dagger}(k)\mathbf{A}^0(k)$ is the instantaneous $N \times N$ correlation matrix of the fading vectors. Using the approximation $\log \sum_i z_i \approx \max_i z_i$, for sufficiently high SNR, the branch metrics (10) can by approximated by the simpler metric [3]:

$$\lambda(\mathbf{y}^0(k), \tilde{x}^0(k)) = \max_{\tilde{\mathbf{z}}(k) \in \mathcal{S}(\tilde{x}^0(k))} \{\Omega(\tilde{\mathbf{z}}(k))\} \tag{12}$$

4. The CSI estimator

Each diversity branch contains N^2 CSI estimators. In conventional CSI estimators based on pilot symbols and interpolation, an incoming pilot symbol is multiplied by the complex conjugate of its transmitted value. An estimate of the fading sample affecting the pilot symbol is then obtained. The fading samples affecting the data symbols are then obtained by interpolation [6]. This estimator will be called a *single-user CSI estimator (SE)* in this paper.

For the sake of simplicity, we omit the index i identifying the i'th diversity branch. First we assume that all the signals are symbol synchronous. This is not very realistic as the mobiles do not have a common clock, and there is no way to tell them exactly at which time to start the transmission. However, it simplifies our task, and simulation results will illustrate the impact of errors due to asynchronous users. Furthermore, the signals are assumed to be frame synchronous. All the mobiles transmit a pilot symbol sequence of length N at instants $((k - N/2)T, \ldots, (k + N/2 - 1)T)$. Let the vectors $\mathbf{x}(k - N/2), \ldots, \mathbf{x}(k + N/2 - 1)$, where $\mathbf{x}(k) = [x^0(k) \cdots x^{N-1}(k)]^T$, contain the pilot sequences. The pilot sequences of the different users are assumed to be orthogonal. $\mathbf{A}(k)$ can then be estimated as:

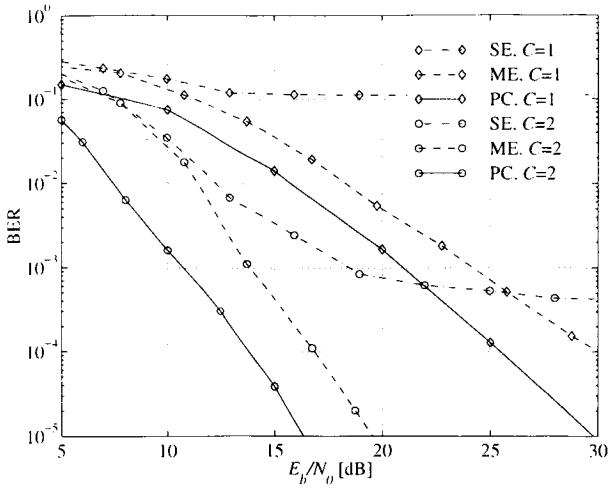


Figure 3: *The BER for the multi-user receiver for the diversity orders $C = 1$ and 2 with the ME and the SE as CSI estimator, and with perfect CSI (PC). Linear interpolation is used. SIR = 10 dB, $f_dT = 0.01$, $M = 5$ (SE) and 6 (ME), $N = 2$, $N_x \times N_y = 20 \times 20$, code U8.*

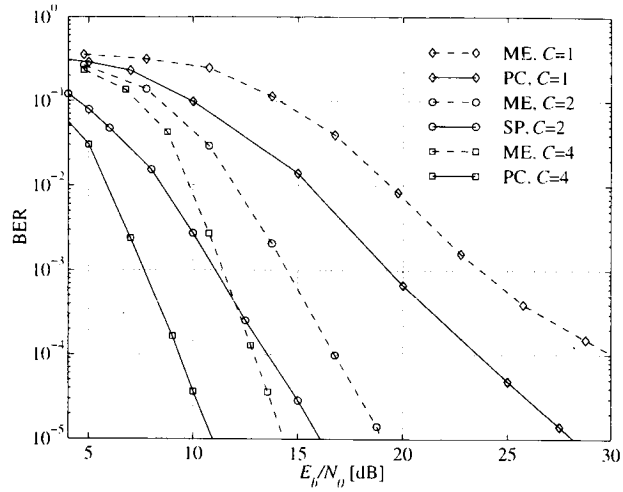


Figure 4: *The BER for the multi-user receiver for the diversity orders $C = 1$, 2 and 4. The CSI is estimated with the ME (or MME) and linear interpolation. The results with perfect CSI (PC) is also shown. SIR = 0 dB, $f_dT = 0.01$, $M = 6$, $N = 2$, $N_x \times N_y = 20 \times 20$, code U8.*

$$\begin{aligned}\hat{\mathbf{A}}(k) &= \frac{1}{N}\sum_{l=k-N/2}^{k+N/2-1}\mathbf{y}(l)\mathbf{x}^{\dagger}(l) \\ &= \frac{1}{N}\mathbf{B}\sum_{l=k-N/2}^{k+N/2-1}\mathbf{G}(l)\mathbf{x}(l)\mathbf{x}^{\dagger}(l) \\ &\quad + \frac{1}{N}\sum_{l=k-N/2}^{k+N/2-1}\mathbf{n}(l)\mathbf{x}^{\dagger}(l) \qquad (13)\end{aligned}$$

Defining the matrices:

$$\mathbf{N}(k) = \frac{1}{N}\sum_{l=k-N/2}^{k+N/2-1}\mathbf{n}(l)\mathbf{x}^{\dagger}(l) \qquad (14)$$

and

$$\Delta\mathbf{G}(l,k) = \mathbf{G}(l) - \mathbf{G}(k) \qquad (15)$$

the estimated value of $\mathbf{A}(k)$ can be written as:

$$\hat{\mathbf{A}}(k) = \mathbf{A}(k) + \Delta\mathbf{A}^{s}(k) + \mathbf{N}(k) \qquad (16)$$

where

$$\Delta\mathbf{A}^{s}(k) = \frac{1}{N}\mathbf{B}\sum_{l=k-N/2}^{k+N/2-1}\Delta\mathbf{G}(l,k)\mathbf{x}(l)\mathbf{x}^{\dagger}(l) \qquad (17)$$

The elements of $\mathbf{N}(k)$ are zero mean Gaussian variables with variance $(N\cdot\mathrm{SNR})^{-1}$. $\Delta\mathbf{A}^{s}(k)$ is due to the error we make assuming that the fading processes are constant over the duration of the pilot sequences. The mean square error (MSE) of this estimator as a function of the Doppler spread of the fading, and of the relative synchronization between the users is developed in [7]. This estimator will be denoted *multi-user CSI estimator (ME).*

As noted above, the beamforming network gain matrix $\mathbf{B}$ is unknown. We assume however that it is constant during a communication. It can therefore be estimated in the beginning of a communication in the following way:

$$\hat{\mathbf{B}} = \frac{1}{K}\sum_{l=1}^{K}\hat{\mathbf{A}}(k)\hat{\mathbf{G}}^{-1}(k) \qquad (18)$$

Once a reliable estimate of $\mathbf{B}$ is obtained, only the $\mathbf{G}(k)$ matrix, or equivalently the diagonal elements of $\mathbf{A}(k)$, need to be estimated. If the non-diagonal elements of $\mathbf{B}$ satisfy $|b^{ij}| < 1$ for all $j \neq i$, the MSE of the CSI estimates corresponding to the interfering signals will be reduced. This CSI estimator will be denoted *modified multi-user CSI estimator (MME).*

5. Simulation results

The performance of the multi-user receiver with the SE, the ME and the MME is assessed by computer simulations. The normalized Doppler spread of the fading is $f_dT = 0.01$, and a block interleaver of size 20×20 is used. The channel filter used is a 3rd order Butterworth filter, which is commonly used in the lit-

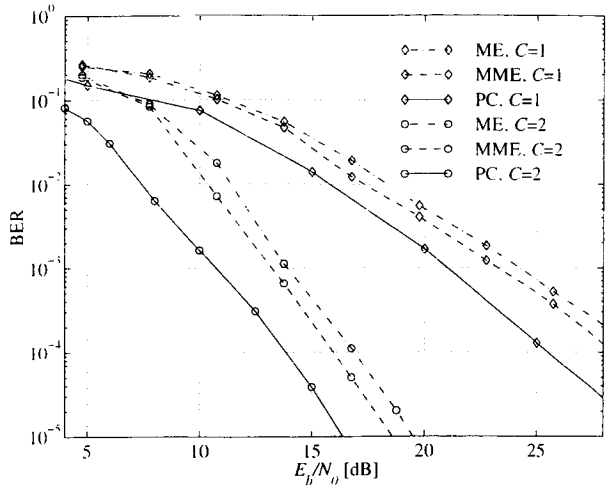


Figure 5: *The BER for the multi-user receiver for the diversity orders $C = 1$ and 2 with the MME and the ME as CSI estimator, and with perfect CSI (PC). Linear interpolation is used. SIR = 10 dB, $f_dT = 0.01$, $M = 6$, $N = 2$, $N_x \times N_y = 20 \times 20$, code U8.*

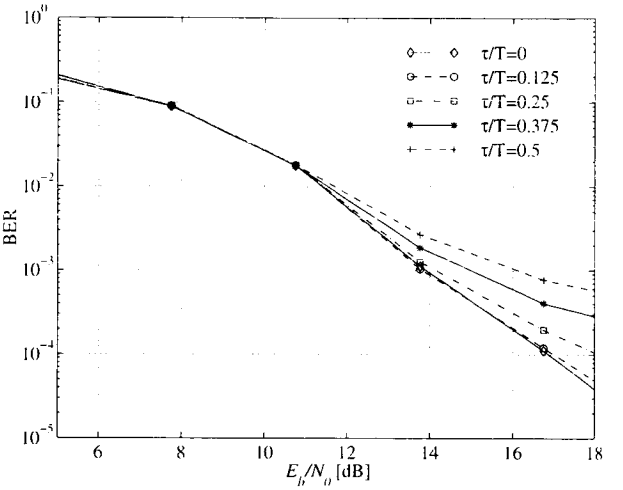


Figure 6: *The BER for the multi-user receiver as a function of the relative delay τ/T between the users. The CSI is estimated by the ME and linear interpolation. $C = 2$, $N = 2$, SIR = 10 dB, $f_dT = 0.01$, $M = 6$, $N_x \times N_y = 20 \times 20$, code U8.*

erature (see e.g. [8]). The fading samples affecting the data symbols are estimated by means of a linear interpolator. It is not optimal, but for low SIRs and/or low SNRs, its loss with respect to more complex interpolators like the Wiener interpolator, is small. The code used is Ungerboeck's 8 state 8PSK TCM scheme (U8) [9]. In the simulations, the frame length is chosen equal to six, i.e., two pilot symbols for four data symbols. This relatively high ratio of pilot symbols reduces the effective power spent in information bearing symbols by about 1.76 dB. Simulation results show however, that even if longer frames, i.e., $M = 10$, are rewarded by a 0.5-1 dB gain at low E_b/N_0, the penalty is more severe error floors.

In figure 3, the performances of the multi-user receiver with the SE and ME are compared for SIR = 10 dB. It is clear that the ME allows the multi-user receiver to operate at low SIRs where the interference prevents the SE to provide reliable CSI estimates. In figure 4, the SIR is 0 dB, which corresponds to the worst case for the multi-user metrics computer. The loss due to the non-perfect CSI is about 4 dB at an BER of 10^{-3}. An error floor is only apparent for high E_b/N_0, and is due to the interpolator.

In figure 5, the MME and the ME are compared for SIR = 10 dB. For the MME, the $\mathbf{B}$ matrix is assumed perfectly estimated. The gain obtained by estimating $\mathbf{B}$ is small, especially if the estimate is affected by errors. For higher SIRs, the estimates of the non-diagonal elements of $\mathbf{A}(k)$ using the ME will have a larger MSE with respect to the MME, but the effect of the interfering signals in the metric computation will be smaller. For lower SIR, the picture is inverted. The difference in MSE is smaller, but the influence of errors is larger. For SIR = 0 dB, there is no difference in performance between the MME and the ME. In any case, the extra computation involved in estimating $\mathbf{B}$ seems not be justified.

It is of interest to see what happens when the relative delay between the users is fractions of a symbol duration. We assume that the delay can be kept smaller than half a symbol duration, which also makes it possible to keep the users frame synchronous. In figure 6, the BER is tabulated for different delays τ/T for 2 users and $C = 2$. For $\tau < T/8$, the degradation in performance is very small, and becomes only distinguishable at BERs around 10^{-4} and below. Even with $\tau = T/2$ the loss is small for low E_b/N_0. If the delay τ/T can be kept inferior of 0.25, and this should not be a big problem, it is fair to say that the performance loss with respect to perfect symbol synchronization is moderate.

6. Conclusions

A new strategy to combat CCI in narrowband satellite mobile communications has been proposed. Instead of minimizing the interference level by maximizing the FR distance, or by means of AIS, small groups of two or four co-channel beams are located close together,

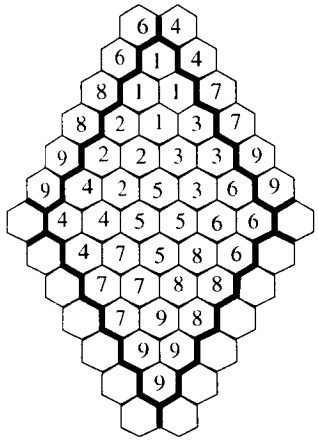


Figure 7: *Example of a frequency reuse pattern adapted to the multi-user receiver. N=4.*

entailing an important mutual CCI (see figure 7). By the means of orthogonal pilot sequences, the impact of the CCI on the CSI estimates is significantly reduced. Due to the enhanced CSI estimates, the multi-user metrics are close to those corresponding to perfect CSI, for which they are optimal. As the distances between the groups are increased with respect to a uniform co-channel beam distribution, the effect of other interferers will be decreased. This benefit can be used to increase the capacity of the system.

Another possible approach is to reuse the same frequency slot within a beam. If the same frequency can be reused two or four times, this leads directly to a gain in capacity by a factor two or four.

It is impossible to obtain symbol synchronous transmissions. The receiver proves however to be robust in the sense that a relative delay smaller than $0.25T$ does not significantly degrade the system performance.

References

[1] G. Björnström, "Digital Payloads: Enhanced Performance Through Signal Processing," *ESA Journal*, Vol. 17, 1993.

[2] W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw-Hill Book Company, 1989.

[3] J. Ventura-Traveset, G. Caire, E. Biglieri, "A multi-user approach to combating co-channel interference in narrow band mobile communications," *The 7th Tyrrhenian International Workshop on Digital Communications*, Italy, Sept., 1995.

[4] J. Ventura-Traveset, G. Caire, E. Biglieri and G.Taricco, "Impact of diversity reception on fading channels with coded modulation. Part III: Co-Channel Interference," submitted to *IEEE Trans. Commun.*, Nov. 1995.

[5] G. Caire, G. Taricco, J. Ventura-Traveset, E. Biglieri, "A multi-user approach to narrowband communications," submitted to *IEEE Trans. Inform. Theory*, Dec. 1995.

[6] J. K. Cavers, "An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels," *IEEE Trans. Veh. Technol.*, Vol. 40, No. 4, Nov. 1991.

[7] J. E. Håkegard, M.-L. Boucheret, "A Fading Channel Estimator adapted to Multi-user Receivers in Narrowband Satellite Mobile Communication," accepted for publication at *International Wireless and Telecommunications Symposium/Exhibition (IWTS'97)*, Shah Alam, Malaysia, May 14-16 1997.

[8] G. M. Vitetta, D. P. Taylor, "Maximum Likelihood Decoding of Uncoded and Coded PSK Signal Sequences Transmitted over Rayleigh Flat-Fading Channels," *IEEE Trans. Commun.*, Vol. 43, No. 11, Nov. 1995.

[9] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, Vol. IT-28, Jan. 1982.

Robust Block Coded M-PSK Modulation For The Gaussian and a Rayleigh Fading Channel

E. Husni and P. Sweeney
Centre for Communication Systems Research
University of Surrey
Guildford GU2 5XH, United Kingdom
Phone: +44 (0) 1483 259800 FAX: +44 (0) 1483 259504
email: e.husni@ee.surrey.ac.uk and p.sweeney@ee.surrey.ac.uk

ABSTRACT

In this paper, construction of partitioned Reed Solomon coded modulation (RSCM) which is robust for the additive white Gaussian noise channel and a Rayleigh fading channel is investigated. By matching configuration of component codes with the channel characteristics, it is shown that this system is robust for the Gaussian and a Rayleigh fading channel.

This approach is compared with not-partitioned RSCM, a Reed Solomon code combined with an MPSK signal set using Gray mapping; and block coded MPSK modulation (BCM) using binary codes, Reed Muller codes, (RMCM). All codes use hard decision decoding. Simulation results for these schemes show that RSCM based on set partitioning performs better than not based on set partitioning and RMCM accross a wide range of conditions.

The novel idea here is that in the receiver, we use a rotated 2^{m+1}-PSK detector if the transmitter uses a 2^m-PSK modulator.

I. INTRODUCTION

Much work has been done on design of efficient coded modulation schemes for improving the performance of digital transmission systems since the publication of Ungerboeck's paper [1] for trellis coded modulation (TCM) and Imai and Hirakawa paper [2] for BCM. Recently, the increasing interest for digital mobile radio or indor wireless systems has led to the consideration of coded modulation design for combatting fading channels.

In a number of previous papers [7]-[9], codes were designed for the Rayleigh fading channel so as to maximize their diversity by not using coded modulation techniques for the Gaussian channel.

Here, an alternative approach for combatting the Rayleigh fading channel is proposed. The coded modulation system is based on partitioned BCM using Reed Solomon codes which is optimum BCM for the Gaussian channel. By using different configuration of component codes which is matched with the channel characteristic, it is shown that this approach yields a better coding gain over a Gaussian and Rayleigh channel as compared to the previous approach.

There are several reasons for using Reed Solomon codes, such as:

- These codes are maximum distance separable codes, and hence, they make highly efficient use of the redundancy
- Reed Solomon codes are burst error correcting codes, which are suitable for non-Gaussian channels.
- Reed Solomon codes provide a wide range of code rates that can be chosen such that the coded scheme has bandwidth efficiency compatible with the reference uncoded system.

For each code, we use (n,k) primitive Reed Solomon codes over $GF(2^q)$ having code symbol length $n = 2^q - 1$, minimum Hamming distance $(n-k+1)$ and error correcting capability $(n-k)/2$.

In partitioned RSCM, each of the m bits defining an MPSK symbol, where $M = 2^m$ is coded and decoded by different Reed Solomon codecs. The set partitioning principle is applied to define subsets with distances Δ_i, (i = 1 to m) that are nondecreasing with i. Each of the m bits defines a subset and is decoded in multistage decoding schemes. This method was first proposed by Cusack [3], who used Reed Muller codes and a QAM signal set. Later Sayegh [4] generalised Cusack's work to other signal constellations by using known binary codes. Here, we use Reed Solomon codes combined with an MPSK signal set.

In section II, we define a baseline coded modulation approach to which the new coded modulation system is compared. The baseline system uses a Reed Solomon code which is combined with an MPSK signal set using Gray code mapping, this approach is called not-partitioned RSCM.

Specific designs of partitioned RSCM are given in section III. Performance analysis over the Gaussian channel and the Rayleigh fading channel are dealt with in section IV. Finally, conclusions are given in section V.

II. NOT-PARTITIONED REED SOLOMON CODED MODULATION

Here, we address the issue of designing these schemes based on maximizing the time diversity (the effective length) of the code. We consider two methods generalising this approach.

A. Method 1

The first method is that a Reed Solomon code, defined over $GF(2^{ym})$, are mapped to the signal points of a 2^m-PSK signal set such that each symbol of the code consists of the concatenation of y channel symbols. In this combination the code rate is chosen such that the rate of the coded scheme is the same as the uncoded one (usually 2^{m-1}-PSK). This method is based on Jamali and Le-Ngoc's work in [7]. The proposition in [7] indicates that the effective order of time diversity in such a mapping is at least d, the minimum Hamming distance of the Reed Solomon code.

As an example we consider a Reed Solomon (63, 42) code, defined over $GF(2^6)$, combined with 8-PSK, and hence, each code symbol consists of two concatenated 8-PSK symbols. The rate of the code is 2/3 which translates into 2 bits/s/Hz throughput, equivalent to that of uncoded QPSK.

B. Method 2

The second method is that a Reed Solomon code, defined over $GF(2^z)$, having code rate $R_c = m/(m+1)$ combined with a 2^{m+1}-PSK signal set. In this combination the code rate is chosen such that the rate of the coded scheme is the same as the uncoded one (2^m-PSK). In this case there is no relationship between the MPSK signal set used for modulation and finite field over which the code is defined.

For example, a (30, 15) Reed Solomon code, defined over $GF(2^5)$, is combined with QPSK signalling. The overall coded QPSK throughput is comparable to that of uncoded BPSK, i.e., 1 bit/s/Hz.

III. PARTITIONED REED SOLOMON CODED MODULATION

The block coded modulation encoder consists of m Reed Solomon codes (called Reed Solomon component codes); this is illustrated in Fig. 1. Binary digits are assigned to each point in the signal space according to Ungerboeck's set partitioning scheme [1]. We now form an array of v·n columns and m rows, where v denotes bit length of one symbol and n coded symbol length.

Each column of the array will correspond to one point in the signal space, with the bit in the first row corresponding to the leftmost digit and the bit in the last row corresponding to the rightmost digit in the representation of the signal space points. The array will be transmitted one column at a time, each column being represented by the corresponding signal space point.

The array contains $m \cdot n \cdot v$ bits of which $v \cdot n \cdot m \cdot R_c$ bits are information bits. Denoting the number of information bits in ith row by $v \cdot k_i$, we can write

$$v \cdot k_1 + v \cdot k_2 + \cdots + v \cdot k_m = v \cdot m \cdot n \cdot R_c$$

For a given rate R_c, the values of k_i's are chosen subject to the above conditions in such a way as to maximise the minimum Euclidean distance between the codewords of the code.

A. Multistage Decoding

A multistage decoding approach has been used for partitioned RSCM. Multistage decoding of multilevel trellis modulation codes has been recently studied and analyzed in a number of papers [5, 6]. The main case of interest here is using a block encoder and block decoding algorithm for each component code of a multilevel modulation code. The complete explaination of multistage decoding algorithm was mentioned in [10].

The novel idea here is that in the receiver, a rotated 2^{m+1}-PSK detector will be used if the transmitter uses a 2^m-PSK modulator. This is illustrated in Figure 2(a), a detector of a QPSK modulator, and Figure 2(b), a detector of an 8-PSK modulator. This ensures that the received

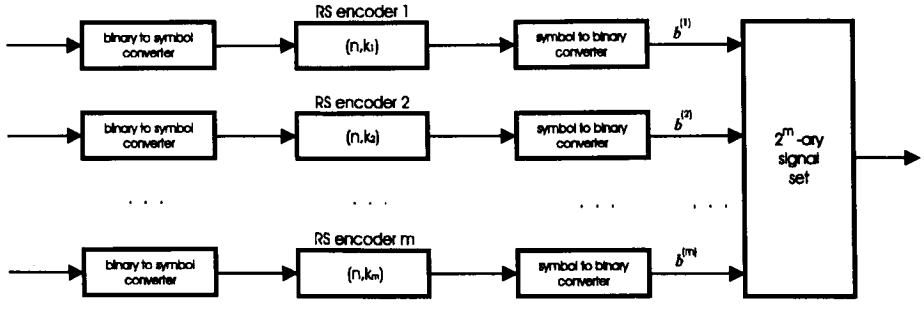


Fig. 1 A Reed Solomon coded modulation based on set partitioning encoder

level does not fall on a decision boundary when decoding any of the bits in the symbol. It can be seen that there are two signal points of the rotated 2^{m+1}-PSK signal set in each signal point's region of the 2^m-PSK signal set.

The block diagram for multistage decoding of partitioned RSCM using 8-PSK is illustrated in Fig. 3. Either one or two iterations may be used for decoding. In the first iteration, level 1 is first decoded, then level 2 is estimated based on the result of level 1. Finally level 3 is determined based on levels 1 and 2. For the second (optional) iteration, level 1 is again decoded using the result of levels 2 and 3 from the first iteration. The corresponding is done for levels 2 and 3.

B. Distance Considerations

The miminum squared Euclidean distance of binary block coded 8-PSK scheme is obtained as

$$D^2_{min} = \min\left(\Delta_0^2 d_1, \Delta_1^2 d_2, \Delta_2^2 d_3\right) \qquad (1)$$

Thus in binary block coded scheme design, the minimum Hamming distance of each component code can be determined by set $\Delta_0^2 d_1 \cong \Delta_1^2 d_2 \cong \Delta_2^2 d_3$.

The optimum values for minimum Hamming distance of each component code of partitioned RSCM can not be determined. Firstly Reed Solomon codes are non-binary codes so minimum Hamming distance is not a binary measure and Equation (1) is no longer valid for this code. Secondly in multistage decoding the first stage decoding gives an output codeword to the second stage decoding, and so on. Thirdly Equation 1 is valid for a Gaussian channel only. Thus the second stage decoding depends on and takes advantages from the first stage decoding, and so on. Consequently, equation 1 can not be used to determine exactly the minimum Hamming distance of each component code.

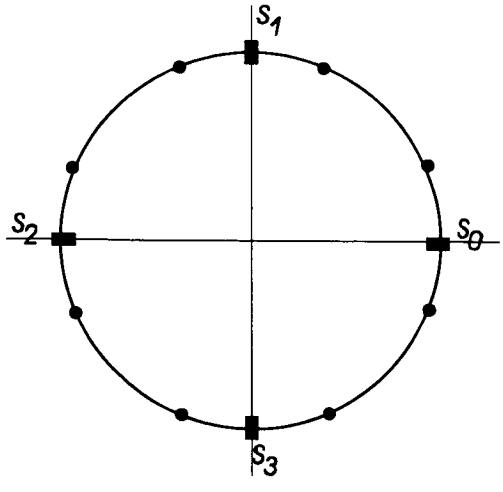


Fig. 2(a). A rotated 8-PSK signal set for a QPSK modulator

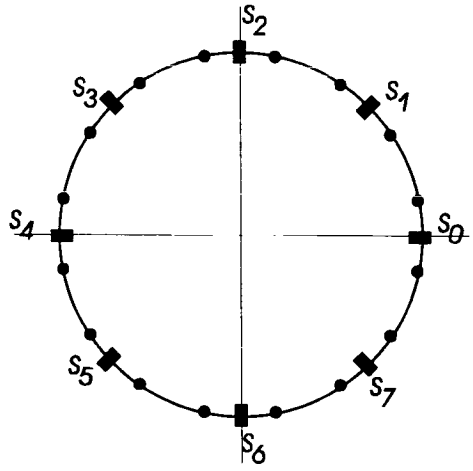


Fig. 2(b). A rotated 16-PSK signal set for an 8-PSK modulator

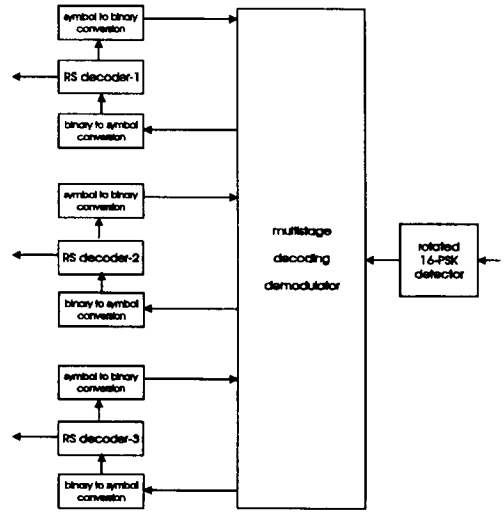


Fig. 3. Multistage decoding of RS coded 8-PSK modulation based on set partitioning.

IV. ERROR PERFORMANCE

In this section, the error performance of partitioned RSCM over the Gaussian channel and the Rayleigh fading channel are compared with not-partitioned RSCM and RMCM using computer simulations.

BCM using Reed Muller codes require the same approches. At the first approach, Reed Muller code having code rate $R_c = {}^{m}/_{(m+1)}$ combined with a 2^{m+1}-PSK signal set. In this combination the code rate is chosen such that the rate of the coded scheme is the same as the uncoded one (2^m-PSK). At the second approach, Reed Muller codes are used for component codes. It uses similar multistage decoding procedure for decoding the received codewords.

A. Over the Gaussian Channel

Reed Solomon codes provide a wide range of code rates, thus there are many configurations of component codes for RSCM based on set partitioning. Table 1 gives a list of good codes for each specified coded symbol length for coded 8-PSK modulation and Table 2 for coded QPSK modulation.

level	k	
	n = 63	n = 127
1	7	11
2	59	119
3	61	125
throughput (bits/s/Hz)	2.01	2.007
information-bit length	(7+59+ 61) ·6 = 762	(11+119 +125)·7 = 1785

Table 1. Configuration of good codes for RS coded 8-PSK modulation for coded symbol length of 63 and 127.

level	k	
	n = 31	n = 63
1	7	11
2	25	53
throughput (bits/s/Hz)	1.03	1.01
information-bit length	(7+25)·5 = 160	(11+53)·6 = 384

Table 2. Configuration of good codes for RS coded QPSK modulation for coded symbol length of 31 and 63.

Figs. 4-5 show block error probability for Reed Solomon coded QPSK modulation and coded 8-PSK modulation. In these figures, the error performances are compared with those of some uncoded reference modulation systems for transmitting the same (or almost the same) number of information bits.

We can see that for coded QPSK modulation, partitioned RSCM has improvement on not-partitioned RSCM by an amount approximates equivalently to doubling code length. For coded 8-PSK modulation at block error probability of 10^{-4}, set partitioning produces 1 dB more coding gain than doubling the code length.

Fig. 6 shows bit error rates of various coded modulation schemes for coded QPSK modulation. Reed Muller codes used here are however more complex to decode than the Reed Solomon codes used. At bit error probability of 10^{-4}, partitioned RSCM has 0.75 dB more coding gain than not-partitioned RSCM with the same code length. It appears that for coded QPSK modulation, Reed Muller coded modulation based on set partitioning is worse than Reed Muller coded modulation not based on set partitioning. For example, at bit error rate 10^{-4}, Reed Muller coded modulation based on set partitioning, length 256, is 0.5 dB worse than Reed Muller coded modulation not based on set partitioning, length 128.

Fig. 7 shows bit error rates of various coded modulation schemes of coded 8-PSK modulation. For partitioned coded modulation, it turns out that all codes have about the same performances. At a bit error rate of 10^{-4}-10^{-5}, partitioned coded modulation has at least 1 dB coding gain over not-partitioned coded modulation for the same code length.

In this channel, it appears that one iteration and two iterations multistage decoding have about the same performances. The results was reported in [10]. In [6] it was claimed that in the AWGN channel, two iterations will give better performance than one iteration if there is an interleaver between the coded bits of all stages or we use modified generalized-minimum-distance-decoding. They used a three dimensional block interleaver for three levels coded modulation with memory cells as many as coded bit length in every dimension. Every memory cell contains one coded bit from every level. The coded bits of the first level are always written in the cells corresponding to the direction of the first dimension. The coded bits of the second level are written in according to the direction of the second dimension. The corresponding is done for the third level. This has not yet been attempted for our schemes.

B. Over the Rayleigh Fading Channel

In this subsection, we analyse the error performance of partitioned RSCM compared with not-partitioned RSCM and RMCM over the nonselective slow Rayleigh fading channel. Here 'slow' means that the fading bandwidth is small compared to the signal bandwidth so that the receiver will be able to track the phase variations.

For all schemes of Reed Solomon coded modulation, the code symbols are interleaved before modulation in order to destroy the memory of the fading channel.

Figs. 8 and 9 show bit error rates of various coded modulation schemes for coded QPSK and 8-PSK modulation. We can see that in this channel, RSCM schemes have a large coding gain to Reed Muller coded modulation schemes. We also see that the error performances of partitioned RSCM using 8-PSK are better than not-partitioned RSCM at high BER, and finally they become the same at low BER; and the error performances of partitioned RSCM using QPSK are better than those of not-partitioned RSCM. Therefore, the error performances partitioned RSCM are never worse than those of not-partitioned RSCM for the same code length.

This scheme is different from the Gaussian channel in that good configurations have the same component codes. Therefore, it seems that all levels have the same distance. In other words, the fading phase is a uniformly distributed random process. By matching configuration of component codes to the channel characteristic, it is shown that partitioned RSCM can be robust codes for the Rayleigh fading channel.

The error performances of RSCM not based on set partitioning using the first approach whose code symbol consists of one channel symbol can be seen in [7]. These schemes are not simulated because they have limited configurations.

In this channel, two iteration multistage decoding must be used. This because the good codes have the same component codes, thus the first level component code also needs decoded codewords of other levels. the complete results were given in [10].

V. CONCLUSION

In this paper, construction of partitioned Reed Solomon coded modulation (RSCM) which is robust for the additive white Gaussian noise channel and a Rayleigh fading channel is proposed. By matching configuration of component codes with the channel characteristics, it is shown that this system is robust for the Gaussian and a Rayleigh fading channel.

Its error performances were compared with those of not-partitioned RSCM and coded MPSK modulation using binary codes, Reed Muller codes. It appears that partitioned RSCM performs better than not-partitioned RSCM and RMCM over the Gaussian and a Rayleigh fading channel.

REFERENCES

[1] G. Ungerboeck, *Channel coding with multilevel/phase signals*, IEEE Trans. Inf. Theory, IT-28, pp. 55-67, Jan. 1982.

[2] H. Imai and S. Hirakawa, *A new multilevel coding method using error correcting codes*, IEEE Trans. Inf. Theory, IT-23, pp. 371-376, May 1977.

[3] E.L. Cusack, *Error control codes for QAM signalling*, Electron. Lett., 20, pp. 62-63, Jan. 1984.

[4] S.I. Sayegh, *A class of optimum block codes in signal space*, IEEE Trans. Commun., COM-30, pp. 1043-1045, Oct. 1986.

[5] T. Takata, S. Ujita, T. Kasamiand S. Lin, *Multistage decoding of multilevel block M-PSK modulation codes and its performance analysis*, IEEE Trans. Inf. Theory, 39, pp. 1204-1218, July 1993.

[6] T. Woerz and Fazel, *Comparison of different decoding strategies for block coded modulation*, in Proceedings of 1992 URSI Int. Symp. Signals, Sys. and Electronics, Paris, pp. 117-120, Sept., 1992.

[7] S.H. Jamali, *Coded modulation techniques for fading channels*, Klower Academic Publishers, 1994.

[8] E. Biglieri, G. Caire, G. Taricco and J. Ventura-Traveset, *Coding and modulation for the fading channel: the quest for robustness*, in Proceedings of the Fifth ESA Intl. Workshop on DSP techniques Applied to Space Communication, Barcelona, pp. 8.1-89, Oct., 1996.

[9] E. Zehavi, *8-PSK trellis codes for a Rayleigh channel*, IEEE Trans. on Comm., 40, pp. 873-884, May 1992.

[10] E. Husni and P. Sweeney, *Block coded MPSK modulation using Reed Solomon codes*, in Proceedings of the Fifth ESA Intl. Workshop on DSP techniques Applied to Space Communication, Barcelona, pp. 1-8, Oct., 1996.

[11] P. Sweeney, *Error control coding an introduction*, Prentice Hall, 1991.

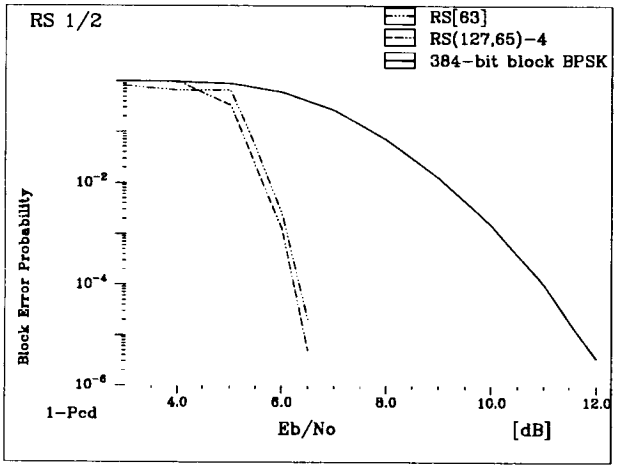


Fig. 4. Error performances of coded QPSK modulation; partitioned RSCM **RS[63]** of length 63 listed in Table 2 and not-partitioned RS(127, 65) with information-bit length of 65·7 = 455.

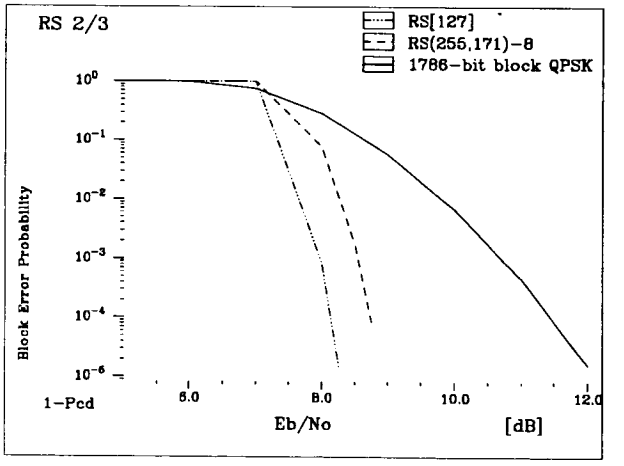


Fig. 5. Error performances of coded 8-PSK; partitioned RSCM **RS[127]** of length 127 listed in Table 1 and not-partitioned RS(255,171) with information-bit length of 171·8 = 1368.

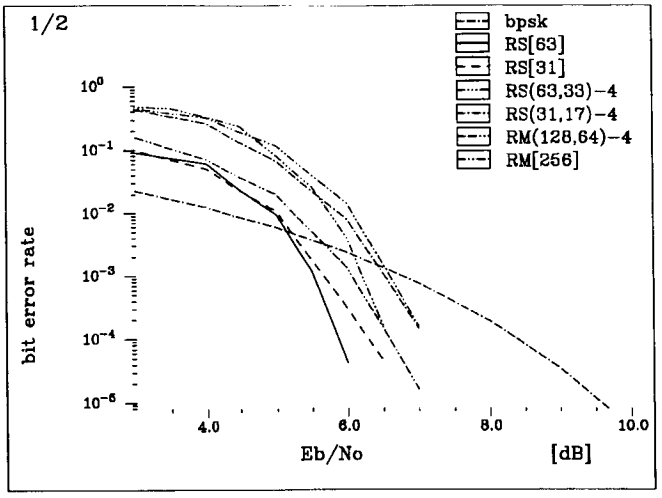


Fig. 6. Error performances of coded QPSK modulation; partitioned RSCM: **RS[63]** of length 63 listed in Table 2, **RS[31]** of length 31 listed in Table 2, not-partitioned: RS(63,33), RS(31,17), and partitioned RMCM **RM[256]** of length 256, not-partitioned 3rd-order RM(128,64).

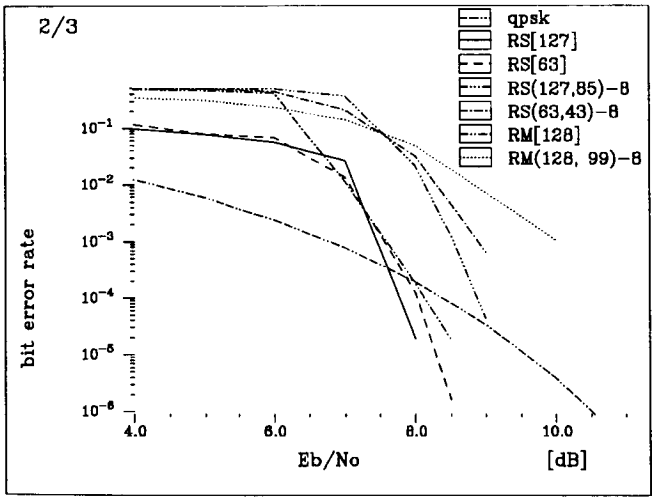


Fig. 7. Error performances of coded 8-PSK modulation; partitioned RSCM: **RS[127]** of length 127 listed in Table 1, **RS[63]** of length 63 listed in Table 1, not-partitioned: RS(127,85), RS(63,43), and partitioned RMCM **RM[128]** of length 128, not-partitioned 4th-order RM(128,99).

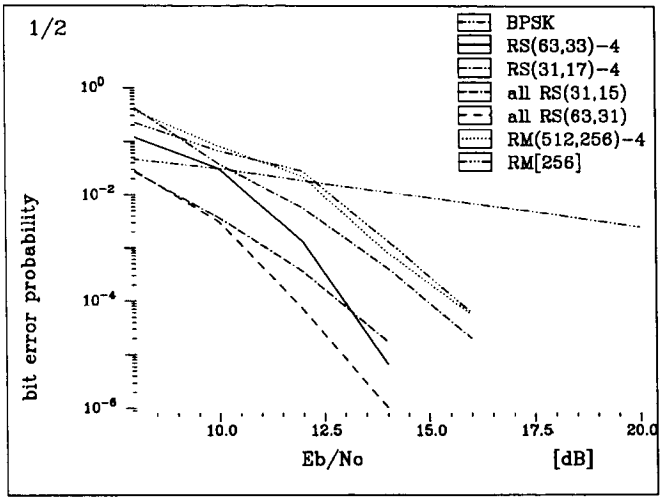


Fig. 8. Error performances of coded QPSK modulation over a Rayleigh fading channel; partitioned RSCM: **all RS(63,31)** of length 63 whose component codes are all RS(63,31), **all RS(31,15)** of length 31 whose component codes are all RS(31,15), not-partitioned: RS(63,33), RS(31,17), and partitioned RMCM **RM[256]** of length 256, not-partitioned 4th-order RM(512, 256).

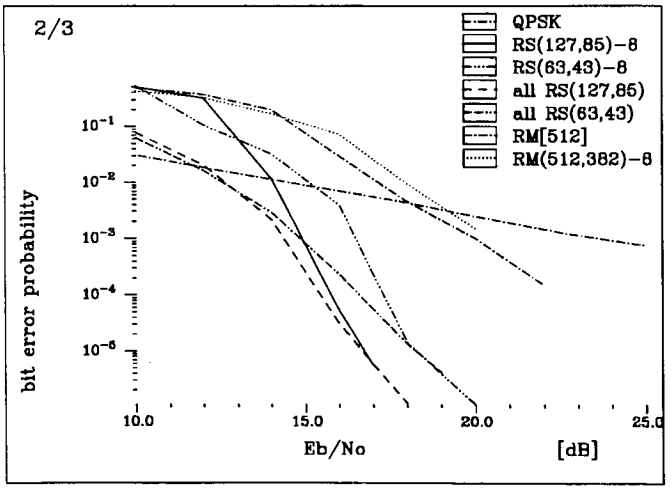


Fig. 9. Error performances of coded 8-PSK modulation over a Rayleigh fading channel; partitioned RSCM: **all RS(127,85)** of length 127 whose component codes are all RS(127,85), **all RS(63,43)** of length 63 whose component codes are all RS(63,43), not-partitioned: RS(127,85), RS(63,43), and partitioned RMCM **RM[512]** of length 512, not-partitioned 5th-order RM(512, 382).

A Comparative Study of Co-Channel Interference Suppression Techniques*

Jon Hamkins
hamkins@jpl.nasa.gov
818-354-4764

Ed Satorius
satorius@bvd.jpl.nasa.gov
818-354-5790

Jet Propulsion Laboratory
4800 Oak Grove Dr.
Pasadena, CA 91109-8099

Gent Paparisto
paparist@milly.usc.edu
213-740-4673

Andreas Polydoros
andreas@solar.usc.edu
213-740-4667

Electrical Engineering — Systems
University of Southern California
Los Angeles, CA 90089-2565

Abstract

We describe three methods of combating co-channel interference (CCI): a cross-coupled phase-locked loop (CCPLL); a phase-tracking circuit (PTC), and joint Viterbi estimation based on the maximum likelihood principle. In the case of co-channel FM-modulated voice signals, the CCPLL and PTC methods typically outperform the maximum likelihood estimators when the modulation parameters are dissimilar. However, as the modulation parameters become identical, joint Viterbi estimation provides for a more robust estimate of the co-channel signals and does not suffer as much from "signal switching" which especially plagues the CCPLL approach. Good performance for the PTC requires both dissimilar modulation parameters and a priori knowledge of the co-channel signal amplitudes. The CCPLL and joint Viterbi estimators, on the other hand, incorporate accurate amplitude estimates. In addition, application of the joint Viterbi algorithm to demodulating co-channel digital (BPSK) signals in a multipath environment is also discussed. It is shown in this case that if the interference is sufficiently small, a single trellis model is most effective in demodulating the co-channel signals.

Introduction

Co-channel interference (CCI) can be a serious impairment to any communication system. In this paper, we first focus on separating two CCI FM signals of the form

$$r(t) = A_1 \cos(\omega_1 t + \theta_1(t)) + A_2 \cos(\omega_2 t + \theta_2(t)), \quad (1)$$

where $\theta_i(t) = k_i \int_{-\infty}^{t} m_i(s)ds$ and $m_i(t)$ are the modulating signals. (Subsequently, we will consider the demodulation of co-channel digitally modulated (BPSK) signals in the presence of multipath.) As evident from the model described by Equation (1), each FM signal alone has a constant envelope and an instantaneous frequency which is proportional to its modulating signal; on the other hand, $r(t)$ has a widely varying envelope and an instantaneous frequency which contains large spikes. A conventional FM receiver containing a single PLL may be able to correct for the varying envelope with the use of a hard limiter, but the spikes will remain and $m_1(t)$, $m_2(t)$, and $m_1(t) + m_2(t)$ are each unrecoverable. Ideally, $m_1(t)$ and $m_2(t)$ are recovered separately, with no crosstalk between the two.

There are several approaches to CCI suppression which are based on either linear or non-linear processing. Linear processing methods encompass narrowband linear filtering as well as adaptive filtering techniques which suppress interference based on statistical differences between the interference and desired signal components. Examples of adaptive interference suppression filtering techniques include linear prediction error filtering as well as "blind" adaptive processing wherein linear filter coefficients are adjusted to enhance certain properties of the desired signal (thereby suppressing the interference).

Non-linear techniques are usually based upon the "capture" effect wherein the strongest signal is enhanced by some type of non-linearity at the expense of weaker signals which are suppressed. Since these techniques are dependent on the relative amplitudes of the various signal components, they can be utilized for separating multiple components with highly overlapped spectra. Examples include the FM limiter/discriminator, PLL, PTC [3], CCPLL [2,6], and maximum likelihood (ML)-based techniques [1,5].

CCPLL and PTC Methods

The CCPLL and PTC techniques fall within a class of

*The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

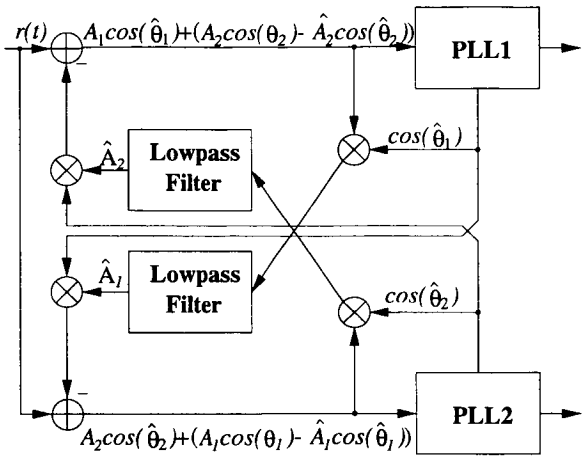
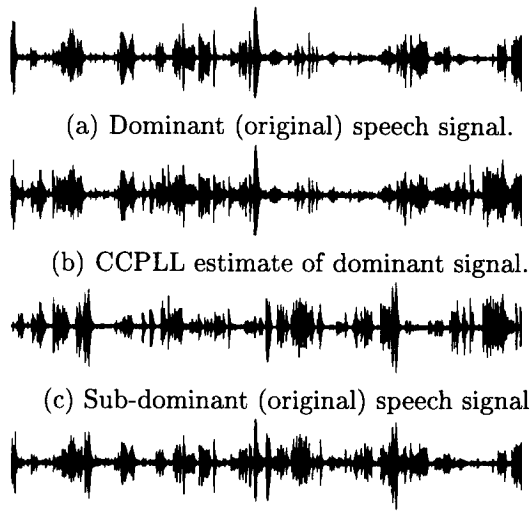


Figure 1: CCPLL block diagram.



(a) Dominant (original) speech signal.

(b) CCPLL estimate of dominant signal.

(c) Sub-dominant (original) speech signal.

(d) CCPLL estimate of sub-dominant signal.

Figure 2: Performance of CCPLL estimator on FM voice signals, each with frequency deviation of 12kHz.

methods that exploit the capture property of the PLL to separate co-channel signals. The CCPLL comprises two PLLs: one used to capture the dominant signal and the second to capture the sub-dominant (its architecture can easily be extended to separate multiple signals). A simplified block diagram, based on the feed-forward, difference-amplitude tracking topology developed in [7], is illustrated in Figure 1. We note that both PLLs are second-order, incorporating proportional plus integral control, and both lowpass filters, used in generating the amplitude estimates, are second-order digital Butterworth designs.

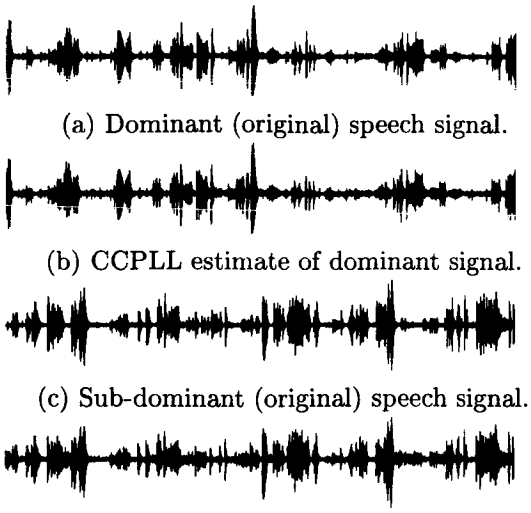
Although this method has been proposed for separating a wide variety of signals including CW, FM, AM and even digital signals, e.g., BFSK, we have found that for cases of co-channel signals with comparable modulation parameters the CCPLL becomes less effective. As an illustrative example, consider the case of two synthesized voice signals (approximately 3.7 kHz bandwidth) that are FM-modulated. The parameters from Equation (1) are $A_1 = 1$, $A_2 = 0.5$, $k_1 = k_2 = 2\pi(12\text{kHz})$, and $\omega_1 = \omega_2 = 0$ (i.e., baseband). Thus the weaker (sub-dominant) signal is 6 dB below the dominant. The signal was sampled at 131kHz. Using the CCPLL design rules developed in [6], we obtain sample computer simulation results as depicted in Figure 2 for a three second segment. Both the original and CCPLL-reconstructed estimates (demodulated by the PLLs) for the dominant and sub-dominant synthesized voice waveforms are displayed.

Generally the CCPLL is able to separate the voice waveforms reasonably well in this example. However, close inspection of the various waveforms reveals certain features that are captured either by the wrong PLL or by both PLLs simultaneously. An example of the latter is the distinct pulse train occurring in the dominant waveform (Figure 2(a)) at about the middle of the three second segment (between approximately 1.4 and 1.5 seconds). The dominant CCPLL estimate (Figure 2(b)) captures this feature but unfortunately so does the sub-dominant (Figure 2(d)). Vice versa at the end of the segment (between approximately 2.8 and 2.9 seconds) the amplitude burst in the sub-dominant waveform (Figure 2(c)) is captured only by the dominant CCPLL estimate.

This phenomenon of "signal switching" is created by a number of factors including instantaneous frequency crossing which causes the individual PLLs to lock onto and track the wrong signals. This problem is less severe as the occurrence of instantaneous frequency crossings is reduced. This situation can be simulated using the same synthesized voice waveforms considered above in Figure 2 but using different FM modulation parameters. For example, if we increase the FM deviation frequency of the sub-dominant from 12 kHz to 24 kHz (with all other modulation parameters the same) and correspondingly open up the PLL bandwidths (again following the CCPLL design rules provided in [6]), we obtain much better voice separation as seen in Figure 3.

Now the objectionable signal switching has been significantly reduced. Of course in practice the presence of instantaneous frequency crossings in the input co-channel data cannot be controlled. Thus the utility of the CCPLL will generally be limited unless some means of compensation can be developed to mitigate the deleterious effects of instantaneous frequency crossings. Possible methods include the inclusion of Kalman tracking techniques that will track previous PLL phase estimates and use these estimates to prevent signal phase switching. This is an area of current research.

The PTC [3] also comprises two PLLs but both are



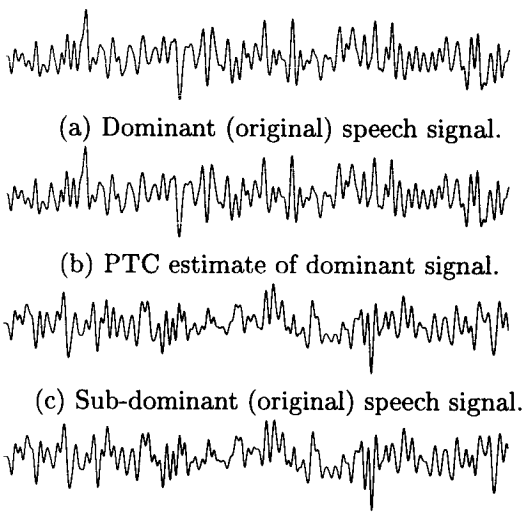
(a) Dominant (original) speech signal.

(b) CCPLL estimate of dominant signal.

(c) Sub-dominant (original) speech signal.

(d) CCPLL estimate of sub-dominant signal.

Figure 3: Performance of CCPLL estimator on FM voice signals, with frequency deviations of 12kHz and 24kHz.



(a) Dominant (original) speech signal.

(b) PTC estimate of dominant signal.

(c) Sub-dominant (original) speech signal.

(d) PTC estimate of sub-dominant signal.

Figure 4: PTC performance on FM voice signals.

used to capture and lock onto the dominant signal, i.e., the first PLL captures the dominant while the second further smoothens the dominant phase estimate as well as inverts its polarity (sign) so that it can be subtracted from the IF input thereby recovering the sub-dominant.

The PTC provides an interesting alternative to the CCPLL in that it reduces the complicated nonlinear dynamics inherent in the CCPLL. However, a key limitation is that amplitude adjustment is required to obtain optimal dominant signal cancellation at the IF. This is typically performed manually which is a clear limitation in operational scenarios — especially in an amplitude-fading environment.

An example showing PTC performance, when the amplitudes of the co-channel signals are presumed known, is presented in Figure 4 for a very short (50 msec) simulated segment. In this example two synthesized voice signals (approximately 3 kHz bandwidth) are FM-modulated and linearly combined using the following parameters from Equation (1): $A_1 = 1$, $A_2 = 0.5$, $k_1 = 2\pi(70\text{kHz})$, $k_2 = 2\pi(14\text{kHz})$, $\omega_1 = 2\pi(455\text{kHz})$, and $\omega_2 = 2\pi(450\text{kHz})$. Thus, the sub-dominant IF bandwidth is considerably smaller than the dominant IF bandwidth (approximately 25% of the dominant bandwidth), and the IF carrier frequencies are chosen differently. So although there is still significant spectral overlap between the two signals, this example is not as stressing as the previous examples (Figures 2 and 3).

Both the original and PTC-reconstructed estimates (demodulated by the PLLs) for the dominant and sub-dominant synthesized voice waveforms are displayed. As can be seen there is virtually perfect agreement between the originals and their estimates. Additional simulation experiments with comparable co-channel signal IF bandwidths reveal that the PTC remains very effective in signal separation under the assumption of constant and known signal amplitudes. The next step in our research is to incorporate some form of dominant amplitude estimation into the PTC structure.

Maximum Likelihood Separation

In the case of no interference ($A_2 = 0$ in Equation (1)), Cahn [1] derived an approximation

$$\hat{r}(t) = A_1 \cos(\omega_1 t + \hat{\theta}_1(t))$$

of $r(t)$ by minimizing $\int ||\hat{r}(t) - r(t)||^2 dt$ subject to the constraint that within successive time intervals of duration Δt, $\hat{\theta}''(t)$ is a constant and equal to either $-C$ or C. That is, the second derivative of the phase is quantized and used to derive $\hat{\theta}(t)$, which in turn is used to determine $\hat{r}(t)$. Achieving the closest approximation of $r(t)$ requires the proper estimation (quantization) of $\hat{\theta}''(t)$. Since $\hat{\theta}(t)$ depends not only on $\hat{\theta}''(t)$ but also on $\hat{\theta}''(s)$, $s < t$, the estimator needs to contain memory. For example, each state of a 16-state trellis corresponds to a quantized value ($\pm C$) of $\theta''(t)$, $\theta''(t - \Delta t)$, $\theta''(t - 2\Delta t)$, and $\theta''(t - 3\Delta t)$, where Δt is the sampling period.

The Viterbi algorithm chooses one of two values for $\hat{\theta}''(t)$ in precisely the same way that a maximum likelihood sequence estimator chooses between demodulated bits 0 or 1. Thus, this is a maximum likelihood estimate of $r(t)$, provided the constraint $\theta''(t) = \pm C$ is an accurate one.

We extend Cahn's idea to the case of $A_2 \neq 0$ by jointly estimating $\theta_1(t)$ and $\theta_2(t)$. That is, $r(t)$ is estimated by

$$\hat{r}(t) = A_1 \cos(\omega_1 t + \hat{\theta}_1(t)) + A_2 \cos(\omega_2 t + \hat{\theta}_2(t)), \tag{2}$$

and maximum likelihood estimation is performed as before. The new trellis size for joint estimation is the square of Cahn's trellis size. However, the increase in computing power since Cahn's 1974 paper has allowed us to simulate three bits of memory for each signal (64 states) with excellent results.

The algorithm is implemented in a discrete environment. We define the discrete versions of the functions in the usual way, i.e., $r[n] = r(n\Delta t)$, $\theta[n] = \theta(n\Delta t)$, etc. To determine $\hat{r}[n]$, the following update equations are used, for $i = 1, 2$:

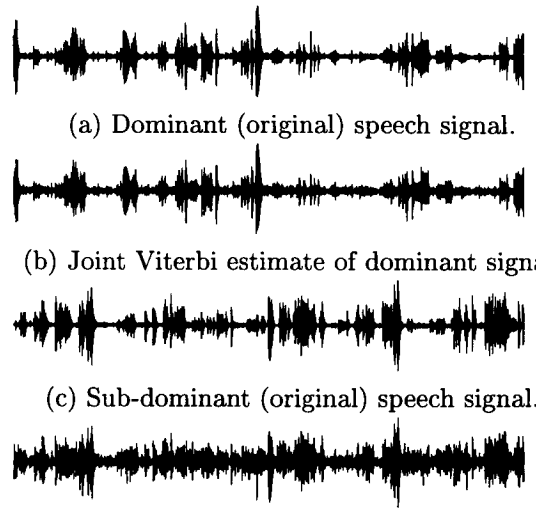
$$\begin{aligned} \hat{\theta}_i'[n] &= \hat{\theta}_i'[n-1] + \hat{\theta}_i''[n]\Delta t \\ \hat{\theta}_i[n] &= \hat{\theta}_i[n-1] + \hat{\theta}_i'[n]\Delta t + \hat{\theta}_i''[n](\Delta t)^2/2, \end{aligned}$$

and $\hat{r}[n]$ is determined from Equation (2).

Because of the quite general modeling of the received waveform, this technique is potentially applicable to a wide class of CCI problems, including both analog signals (FM, PM) and digital signals (FSK, MSK, PSK, QAM, etc.)— indeed, any signal which can be written in the form of Equation (1). The Viterbi algorithm merely attempts to approximate the received signal by the sum of two waveforms; it requires no knowledge of the modulation or timing of the waveforms. Incorporation of this knowledge allows improved performance for specific applications such as BPSK/BPSK interference, which will be discussed in the following section.

Figure 5 indicates the performance of the joint Viterbi algorithm on co-channel FM voice signals. The modulating signals are synthesized voice, and the parameters chosen are identical to those used in Figure 2. That is, the subdominant signal power is 6dB below the dominant signal power, the frequency deviation is 12kHz for each signal, and there is no carrier offset for either signal, i.e., we assume the carrier has already been removed. The received signal was again sampled at 131kHz. As can be seen, nearly all features of each signal are recovered and separated into the dominant and subdominant signals, with virtually no signal switching occurring. This is especially significant in view of the fact that the modulating signals are statistically identical, and equal frequency deviations are used. When real voice signals are used, both signals are easily intelligible, with the dominant signal suffering almost no degradation in sound quality. The quality of the subdominant signal is somewhat degraded.

Figure 6 indicates the performance of the joint Viterbi algorithm on BFSK co-channel interfering signals. The modulating signals are determined from independent,



(a) Dominant (original) speech signal.

(b) Joint Viterbi estimate of dominant signal.

(c) Sub-dominant (original) speech signal.

(d) Joint Viterbi estimate of sub-dominant signal.

Figure 5: Performance of joint Viterbi estimator on CCI FM voice signals.

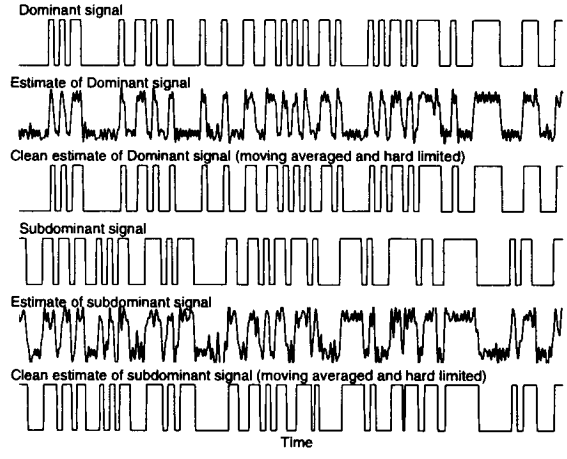


Figure 6: Joint Viterbi estimation of CCI BFSK signals.

uniformly distributed bit streams, each at a baud rate of 10,000 bits per second. A rectangular pulse shape is used, and the interfering signals were offset by 1/4 bit. The parameters used in Equation (1) are $A_1 = 1$, $A_2 = 0.5$, $k_1 = k_2 = 2\pi(5\text{kHz})$, and $w_1 = w_2 = 0$. The received signal was sampled at 40kHz, i.e., there were 4 samples per bit. As can be seen, the raw output of the Viterbi algorithm recovers much of the information of the original signals; when a moving average and hard limiter are applied the original and estimated signals match almost perfectly.

In practice, the receiver does not have a priori knowledge of A_1 and A_2, and these quantities must also be estimated at the receiver. We used the LMS algorithm to adaptively estimate A_1 and A_2. To improve perfor-

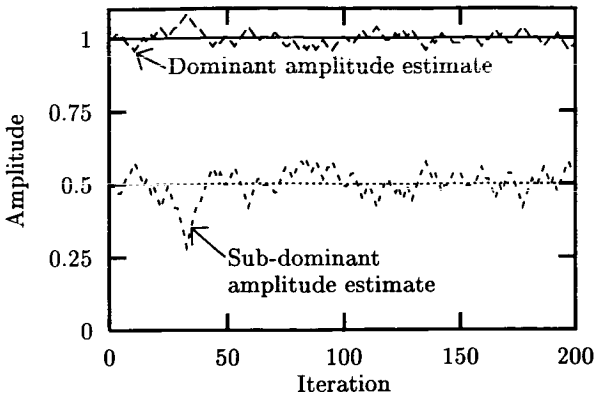


Figure 7: Amplitude estimates for two statistically identical FM voice waveforms. The true amplitudes were constant, as shown.

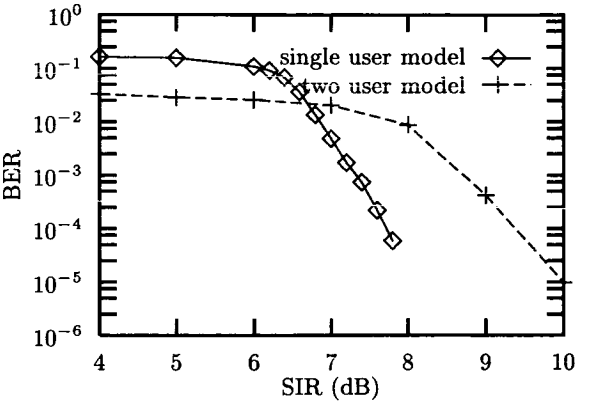


Figure 8: Static channel bit error rate results for both a joint and single trellis demodulation algorithm.

mance, the estimates were updated in per-survivor processing (PSP) fashion, i.e., separate estimates were stored at each state of the trellis, instead of one estimate at each time step [4]. We found that the estimates accurately track the true values, as shown in Figure 7. (In this case, the estimate itself is quite good but can be improved further by applying a moving average to the estimate.) As a result, there was virtually no difference in the quality of the $\hat{r}(t)$ estimate when working with known amplitudes or estimated amplitudes. This method also tracks amplitudes which vary slowly in time.

It is clear how this joint estimator may be extended to more than two interfering signals with the addition of more trellis states. Since the trellis size grows exponentially in the number of interfering signals, there is a practical limit to the model, however.

Maximum Likelihood Demodulation of BPSK Signals in the Presence of Multipath and Co-Channel Interference

Here we briefly discuss the application of ML-based methods to the demodulation of BPSK signals in the presence of multipath and co-channel BPSK interference. In particular, we consider further the joint trellis algorithm originally presented in [5] as it applies to co-channel BPSK demodulation and channel estimation in the presence of static and fading multipath. This algorithm represents a significant departure from the joint Viterbi algorithm presented above in that it estimates the data symbols directly as opposed to the (continuous-time) phase waveforms. The underlying system assumed here is modeled as a baud-spaced, or T-spaced, baseband digital communication system, i.e., the BPSK data symbols for both the desired signal and interferer are transmitted in a T-spaced data sequence. The corresponding envelope of the received signal is formed by convolving both desired and interference BPSK waveforms with generally different T-spaced channel impulse responses (CIRs) and then linearly combining the convolved waveforms. It is assumed (at least initially in our studies) that no additive noise is present and that the received desired signal and interferer are synchronous and that perfect symbol timing is available. Performance metrics are the bit error rate (BER) of the desired signal and the received signal-to-interference power ratio (SIR).

Two simulation examples are considered here. The first corresponds to a static multipath environment wherein both the desired and interference multipath channels are modeled by a T-spaced CIR with two equal non-zero taps. The tap value for the desired multipath channel is unity whereas the corresponding value for the interference channel is varied to achieve different values of SIR. Plots of BER versus SIR are presented in Figure 8 corresponding to both a joint trellis algorithm (with 2 states allocated each to the desired signal and interferer for a total of $2 \times 2 = 4$ states) and a single trellis algorithm (modeling only the desired signal with a total of 2 states). Both algorithms are initialized with a single (2 state) trellis which is trained using a known, desired signal sequence. The purpose of this training interval is to obtain an estimate of the desired multipath taps. Then, in the case of the joint trellis algorithm, the trellis is expanded to 4 states after the desired signal training sequence in order to estimate the interference multipath channel (in the blind) and in the process demodulate the desired BPSK signal.

As is seen from Figure 8, the joint trellis algorithm provides a BER $< 3\%$ at all SIRs between 4 and 10 dB. However, once the SIR exceeds 7 dB, the single trellis outperforms the joint trellis and so it is better to discard the states allocated to the interferer once its power

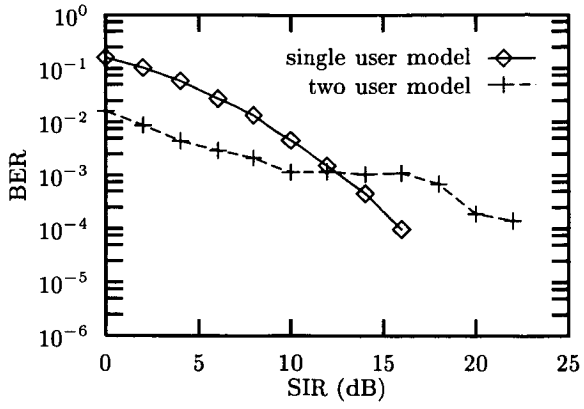


Figure 9: Fading channel bit error rate results for both a joint and single trellis demodulation algorithm.

falls below a certain (threshold) level. This behavior is a consequence of the initialization process described above, i.e., once the interference power becomes too small the resulting estimate of the interference multipath channel (derived blindly by the joint trellis) is too noisy to yield good demodulation of the desired signal and hence the single trellis algorithm provides the best performance.

Similar results are obtained in a simulated slow Rayleigh fading multipath environment wherein both the desired signal and interference channels are modeled by independent, three tap (T-spaced) channels. Each tap is fading independently with the same average power. The average tap power is unity for the desired signal channel and is varied for the interference channel to achieve different values of average SIR. Plots of BER versus SIR are presented in Figure 9 corresponding to both a joint trellis algorithm (with 4 states allocated each to the desired signal and interferer for a total of $4 \times 4 = 16$ states) and a single trellis algorithm (modeling only the desired signal with a total of 4 states). Both algorithms are initialized as described above (but using 4 states per co-channel signal instead of 2 states). Again it is seen that the single trellis algorithm outperforms the joint trellis when the SIR exceeds a certain level (12 dB in this case).

Conclusions

The CCPLL is extremely effective at separating an FM voice signal from an unmodulated carrier signal, but it suffers from occasional "signal switching" if the subdominant signal has modulation parameters very similar to those of the dominant signal. The PTC also separates FM voice signals very well for the cases tested, but in its present state, it is of limited practical utility because a priori knowledge or a manual estimate of the dominant signal amplitude is required. Two versions of a joint Viterbi algorithm were presented. The first is based on Cahn's approach [1] and can be applied to a wide variety of analog and digital signals. Results with this algorithm using a 64-state trellis revealed very good demodulation of digital BFSK signals and furthermore yield co-channel FM voice separation without the "signal switching" associated with the CCPLL. The second type of joint Viterbi algorithm considered [5] yields direct data symbol estimates in multipath environments. The application of this algorithm to both static and fading multipath channels reveals that once the interference power falls below a certain threshold level, it is better to discard the joint trellis architecture and just model the desired signal states.

References

[1] C. R. Cahn, "Phase tracking and demodulation with delay," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 1, pp. 50-58, Jan. 1974.

[2] F. Cassara, H. Schachter, "Suppression of interchannel interference in FM receivers," Polytechnic Report, POLY EE79-056, Polytechnic Institute of New York, Farmingdale, NY, July 1979.

[3] G. Myers, "Multiple reuse of an FM band," US Patent No. 489958, August 22, 1989.

[4] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: a general approach to MLSE in uncertain environments," *IEEE Trans. Comm.*, vol. 43, no. 2/3/4, pp. 354-364, Feb./Mar./Apr. 1995.

[5] A. Polydoros, G. Paparisto, "Per-survivor processing for joint data/channel estimation in multipath fading and co-channel interference channels," Proceedings of MILCOM'95 (Classified), La Jolla, CA, Nov. 5-8, 1995.

[6] S. Say, "Vector-locked loop interference canceller," Ph.D. dissertation, Polytechnic Institute of New York, June 1985.

[7] G. Zimmerman, "Applications of frequency modulation interference cancellers to multi-access communication systems," Ph.D. Dissertation, California Institute of Technology, Pasadena, CA, 1990.

Turbo Trellis Coded Modulation with Iterative Decoding for Mobile Satellite Communications

D. Divsalar and F. Pollara
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Dr., Pasadena, CA 91109
Phone: 818-354-4287 FAX: 818-354-6825
email: dariush@shannon.jpl.nasa.gov

Abstract— **In this paper, analytical bounds on the performance of parallel concatenation of two codes, known as turbo codes, and serial concatenation of two codes over fading channels are obtained. Based on this analysis, design criteria for the selection of component trellis codes for MPSK modulation, and a suitable bit-by-bit iterative decoding structure are proposed. Examples are given for throughput of 2 bits/sec/Hz with 8PSK modulation. The parallel concatenation example uses two rate 4/5 8-state convolutional codes with two interleavers. The convolutional codes' outputs are then mapped to two 8PSK modulations. The serial concatenated code example uses an 8-state outer code with rate 4/5 and a 4-state inner trellis code with 5 inputs and 2×8PSK outputs per trellis branch. Based on the above mentioned design criteria for fading channels, a method to obtain the structure of the trellis code with maximum diversity is proposed. Simulation results are given for AWGN, and an independent Rayleigh fading channel with perfect Channel State Information (CSI).**

I. Introduction

Trellis coded modulation (TCM) proposed by Ungerboeck in 1982 [1] is now a well-established technique in digital communications. Since its first appearance, TCM has generated a continuously growing interest, concerning its theoretical foundations as well as its numerous applications, spanning high-rate digital transmission over voice circuits, digital microwave radio relay links, and satellite communications. In essence, it is a technique to obtain significant coding gains (3-6 dB) sacrificing neither data rate nor bandwidth.

Turbo codes represent a more recent development in the coding research field [2], which has raised large interest in the coding community. They are *parallel concatenated convolutional codes* (PCCC) whose encoder is formed by two (or more) *constituent* systematic encoders joined through one or more interleavers. The input information bits feed the first encoder and, after having been scrambled by the interleaver, they enter the second encoder. A codeword of a parallel concatenated code consists of the input bits to the first encoder followed by the parity check bits of both encoders. Analytical performance bounds for PCCC with uniform interleaver and maximum likelihood receiver were obtained in [3], and [6] for AWGN channel, and in [16] for Rayleigh fading channel with binary modulation.

The (suboptimal) iterative decoding structure [15] is modular, and consists of a set of concatenated decoding modules, one for each constituent code, connected through the same interleavers used at the encoder side. Each decoder performs weighted soft decoding of the input sequence. Bit error probabilities as low as 10^{-6} at $E_b/N_0 = -0.6$ dB have been shown by simulation [11] using codes with rates as low as 1/15. Parallel concatenated convolutional codes yield very large coding gains (10-11 dB) at the expense of a data rate reduction, or bandwidth increase.

In [4] we merged TCM and PCCC in order to obtain large coding gains and high bandwidth efficiency. In [14] and [13] we suggested merging TCM with the recently discovered serial concatenated convolutional codes (SCCC) [12], adapting the concept of iterative decoding used in parallel concatenated codes. We refer to the concatenation of an outer convolutional code with an inner TCM as serial concatenated TCM (SCTCM).

For parallel concatenated trellis coded modulation (PCTCM), also addressed as "turbo TCM", a first attempt employing the so-called "pragmatic" approach to TCM was described in [5]. Later, turbo codes were embedded in multilevel codes with multistage decoding [7]. Recently [8], punctured versions of Ungerboeck codes were used to construct turbo codes for 8-PSK modulation. In [4] a different approach to construct PCTCM was proposed. Results in [4] show that the performance of the proposed codes is within 1 dB from the Shannon limit at bit error probabilities of 10^{-7} over AWGN channels.

In this paper we used turbo trellis coded modulation and serial trellis coded modulation as discussed above, over fading channels for mobile satellite communications. For fading channels, we assume Rayleigh fading. Rician fading is actually a better model for mobile satellite communications since there is LOS (line-of-sight), but when an omni-directional antenna is used and LOS is blocked by trees, poles, or buildings Rayleigh fading can be used as a worst-case scenario.

II. Analytical Bounds on the Performance of Codes over Fading Channels

Consider an (n, k) block code C with code rate $R_c = k/n$ and minimum distance h_m. An upper bound on the conditional bit-error probability of the block code C over fading channels, assuming coherent detection, maximum likelihood decoding, and perfect Channel State Information (CSI) can

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

be obtained in the form

$$P_b(e|\rho) \leq \sum_{h=d_{min}}^{n} \sum_{w=1}^{k} \frac{w}{k} A_{w,h}^{C} Q\left(\sqrt{2R_c E_b/N_0 \sum_{i=1}^{h} \rho_i^2}\right) \quad (1)$$

where E_b/N_0 is the signal-to-noise ratio per bit, and $A_{w,h}^{C}$ for the block code C represents the number of codewords of the block code with output weight h associated with an input sequence of weight w. $A_{w,h}^{C}$ is the input–output weight coefficient (IOWC). The Q function can be represented as [17]

$$Q(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} e^{-\frac{x^2}{2\sin^2\theta}} d\theta \leq \frac{1}{2} e^{-\frac{x^2}{2}} \quad (2)$$

To obtain the unconditional bit error rate, we have to average over the joint density function of fading samples. For simplicity assume independent Rayleigh fading samples. This assumption is valid if we use an interleaver after the encoder and a deinterleaver before the decoder. Thus the fading samples ρ_i are independent identically distributed (i.i.d.) random variables with Rayleigh density of the form

$$f(\rho) = 2\rho e^{-\rho^2}$$

Using (2) and results in [10], by averaging the conditional bit error rate over fading we obtain

$$P_b(e) \leq \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \sum_{h=d_{min}}^{n} \sum_{w=1}^{k} \frac{w}{k} A_{w,h}^{C} \left[\frac{\sin^2\theta}{\sin^2\theta + R_c E_b/N_o}\right]^h d\theta \quad (3)$$

We can further upper bound the above result and obtain [10]

$$P_b(e) \leq \frac{1}{2} \sum_{h=h_m}^{n} \sum_{w=1}^{k} \frac{w}{k} A_{w,h}^{C} \left[\frac{1}{1 + R_c E_b/N_o}\right]^h \quad (4)$$

Extension of results to independent Rician fading is straightforward (see for example [10]). All these results apply to convolutional codes as well, if we construct an equivalent block code from the convolutional code. Obviously results apply also to concatenated codes including parallel and serial concatenations. As soon as we obtain the input–output weight coefficients $A_{w,h}^{C}$ for a particular code we can compute the performance.

III. Parallel Concatenated Convolutional Codes

The structure of a parallel concatenated convolutional code (PCCC) or "turbo code" is shown in Fig. 1. Figure 1 refers to the case of two convolutional codes, code C_1 with rate $R_c^1 = p/q_1$, and code C_2 with rate $R_c^2 = p/q_2$, where the constituent code inputs are joined by an interleaver of length N, generating a PCCC, C_P, with rate $R_c = \frac{R_c^1 R_c^2}{R_c^1 + R_c^2}$. Note that N is an integer multiple of p. The input block length $k = N$, and the output codeword length $n = n_1 + n_2$ as shown in Fig. 1.

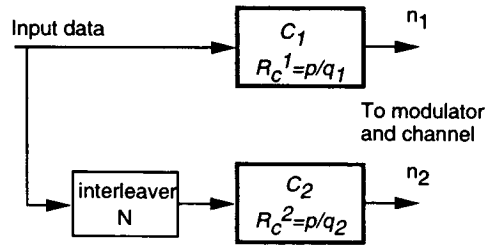


Fig. 1. Parallel Concatenated Convolutional Codes (PCCC).

A. Computation of input–output weight coefficient (IOWC) $A_{w,h}^{C_P}$ for PCCC (turbo codes)

Uniform Interleaver. A crucial step in the analysis of concatenated codes and in particular PCCC consists of replacing the actual interleaver that performs a permutation of the N input bits with an abstract interleaver called a uniform interleaver [3], defined as a probabilistic device that maps a given input word of weight w into all distinct $\binom{N}{w}$ permutations of it with equal probability $p = 1/\binom{N}{w}$. An example for $N = 4$, $w = 2$ is shown in Fig. 2

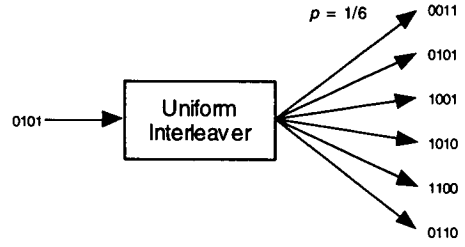


Fig. 2. The action of a uniform interleaver of length 4 on sequences of weight 2

Using the concept of uniform interleaver, i.e., averaging the $P_b(e)$ over all possible interleavers, we can obtain $A_{w,h}^{C_P}$ for turbo codes.

With the knowledge of the $A_{w,h_1}^{C_1}$ for code C_1, and $A_{w,h_2}^{C_2}$ for code C_2, using the concept uniform interleaver, IOWC $A_{w,h}^{C_P}$ for PCCC can be obtained as follows. The main property of the uniform interleaver is that it transforms an input block of weight w at the input of the encoder C_1 into all its distinct $\binom{N}{w}$ permutations. As a consequence, each input block of code C_1 of weight w, through the action of the uniform interleaver, enters the encoder C_1 generating $\binom{N}{w}$ input-words of code C_2. Thus, the number $A_{w,h_1,h_2}^{C_P}$ of codewords of the PCCC with output weights h_1,and h_2 associated with an input sequence of weight w is given by

$$A_{w,h_1,h_2}^{C_P} = \frac{A_{w,h_1}^{C_1} \times A_{w,h_2}^{C_2}}{\binom{N}{w}}$$

where $A_{w,h_1,h_2}^{C_P}$ is related to $A_{w,h}^{C_P}$ as

$$A_{w,h}^{C_P} = \sum_{\substack{h_1,h_2: \\ h_1+h_2=h}} A_{w,h_1,h_2}^{C_P}$$

Example 1. Consider a rate 1/2 PCCC formed by two identical 4-state convolutional codes: Code C_1 with rate 2/3 and code C_2 with rate 1/1 (this is obtained by not sending the systematic bits of the rate 2/3 C_2 convolutional code). The inputs of encoders are joined by a uniform interleaver of lengths $N = 50$, 100 and 256. Both codes are systematic and recursive, and are shown in Fig. 3. Using the previously outlined analysis for PCCC, we have obtained the bit-error probability bounds shown in Fig. 3. The performance is shown both for AWGN and Rayleigh fading channels.

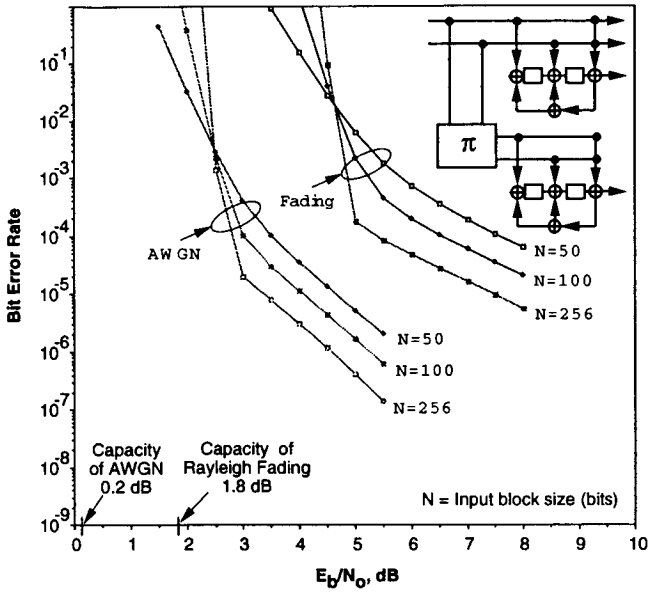


Fig. 3. Performance of rate 1/2 PCCC over AWGN and Rayleigh Fading Channels

The reason for such a good performance of turbo codes is that the coefficients $\frac{A^{C_P}_{w,h}}{N}$ decrease with interleaver size. For large interleavers the maximum component of $\frac{A^{C_P}_{w,h}}{N}$ or equivalently $\frac{A^{C_P}_{w,h,h_2}}{N}$, over all input weights w and output weights h_1 and h_2, is proportional to N^{α_M}, with corresponding output weights $h_1(\alpha_M)$, and $h_2(\alpha_M)$. If both convolutional codes are recursive (i.e., the output weight due to input weight one is very large) then $\alpha_M \leq -1$. (This occurs for $w = 2$.) Any other choice of encoders results in $\alpha_M \geq 0$. When α_M is negative we say that we have "interleaving gain". The negative value of α_M implies that the exponents of N in the bit error rate expression are always negative integers. Thus, for all $h = h_1 + h_2$, the coefficients of the exponents in h decrease with N, and we always have an *interleaving gain* [9].

Define $d_{i,f,eff}$ as the minimum weight of codewords of a recursive code C_i, $i = 1, 2$ generated by weight-2 input sequences. We call it the effective free Hamming distance of a recursive convolutional code. To maximize the *interleaving gain*, i.e., minimize N^{α_M} corresponding to output weight $h_1(\alpha_M)$, and $h_2(\alpha_M)$ we should maximize the $d_{i,f,eff}$, $i = 1, 2$. The sum $d_{1,f,eff} + d_{2,f,eff}$ represents the effective free distance of the turbo code [9] [11]. Thus, substituting the exponent α_M into the expression for bit error rate approximated by keeping only the term of the summation in h_1, and h_2 corresponding to $h_1 = h_1(\alpha_M)$, and $h_2 = h_2(\alpha_M)$, yields

$$\lim_{N\to\infty} P_b(e) \simeq BN^{-1}\left[\frac{1}{1 + R_c\frac{E_b}{N_0}}\right]^{d_{1,f,eff}+d_{2,f,eff}} \quad (5)$$

where B is a constant independent of N.

IV. Parallel Concatenated Trellis Coded Modulation

The basic structure of parallel concatenated trellis coded modulation is shown in Fig. 4.

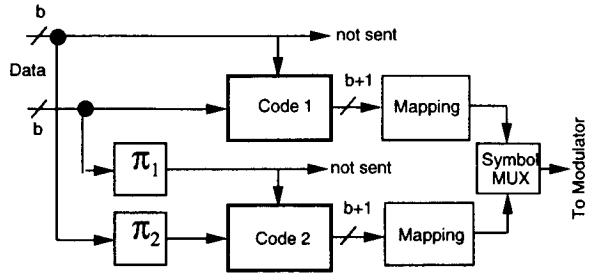


Fig. 4. Block Diagram of the Encoder for Parallel Concatenated Trellis Coded Modulation.

This structure uses two rate $\frac{2b}{2b+1}$ constituent convolutional codes. The first most significant output bits of each convolutional code are only connected to the shift register of the TCM encoder and are not mapped to the modulation signals. The last $b + 1$ least significant output bits however are mapped to the modulation signals. This method requires at least two interleavers. The first interleaver permutes the b least significant input bits. This interleaver is connected to the b most significant bits of the second TCM encoder. The second interleaver permutes the b most significant input bits. This interleaver is then connected to the b least significant bits of the second TCM encoder.

A. Design Criteria for PCTCM over Rayleigh Fading Channels

To extend the asymptotic results we obtained for binary modulation to M-ary Modulation (e.g. MPSK), let $\mathbf{x}_i$ represent the sequence of M-ary output (complex) symbols $\{x_{i,j}\}$ of trellis code i ($i = 1, 2$). Complex symbols have unit average power. Let $\mathbf{x}'_i$ represent another sequence of the output symbols $\{x'_{i,j}\}$ for $i = 1, 2$. Then the above asymptotic result should be modified to

$$P_b(e) \simeq BN^{-1}\prod_{n_1\in\eta_1}\left[\frac{1}{1 + |x_{1,n_1} - x'_{1,n_1}|^2 R_c\frac{E_b}{4N_0}}\right]\times$$

$$\prod_{n_2\in\eta_2}\left[\frac{1}{1 + |x_{2,n_2} - x'_{2,n_2}|^2 R_c\frac{E_b}{4N_0}}\right]$$

where, for $i = 1, 2$, η_i is the set of all n_i with the smallest cardinality $d_{i,f,eff}$ such that $x_{i,n_i} \neq x'_{i,n_i}$. Then $d_{i,f,eff}$ represents the minimum (M-ary symbol) Hamming distance of

trellis code i ($i = 1, 2$) corresponding to input Hamming distance 2 between binary input sequences that produce $d_{i,f,eff}$. The $d_{i,f,eff}$, $i = 1, 2$ is also called the minimum diversity of trellis code i. We note that the asymptotic result on the bit error rate is inversely proportional to the product of the squared Euclidean distances along the error event paths which result in $d_{i,f,eff}$ i=1,2. Therefore the criterion for optimization of the component trellis codes is to maximize the minimum diversity of the code and then maximize the product of the squared Euclidean distances which result in minimum diversity.

B. *2 bits/sec/Hz PCTCM with 8PSK for AWGN and Fading Channels*

The code we propose has $b = 2$, and employs 8PSK modulation in connection with two 8-state, rate 4/5 constituent codes. The selected code uses reordered mapping: If b_2, b_1, b_0 represents a binary label for natural mapping for 8PSK, where b_2 is the MSB and b_0 is the LSB, then the reordered mapping is given by $b_2, (b_2 + b_1), b_0$. The effective Euclidean distance of this code is $\delta^2_{f,eff} = 5.17$ (unit-norm constellation is assumed), using two interleavers.

The structure of this code is shown in Fig. 5, and its BER for AWGN and Rayleigh fading channels in Fig. 6.

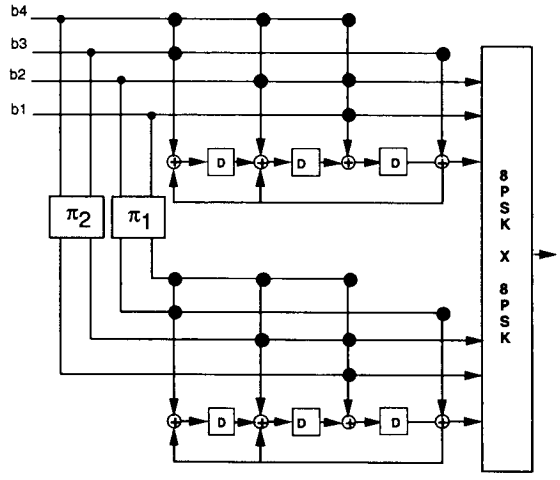


Fig. 5. Parallel Concatenated Trellis Coded Modulation, 8PSK, 2 bits/sec/Hz.

V. Serially Concatenated Convolutional Codes

The structure of a serially concatenated convolutional code (SCCC) is shown in Fig. 7. Figure 7 refers to the case of two convolutional codes, the outer code C_o with rate $R_c^o = q/p$, and the inner code C_i with rate $R_c^i = p/m$, joined by an interleaver of length N bits, generating an SCCC C_S with rate $R_c = k/n$. Note that N must be an integer multiple of p. The input block size is $k = Nq/p$ and the output block size of SCCC is $n = Nm/p$.

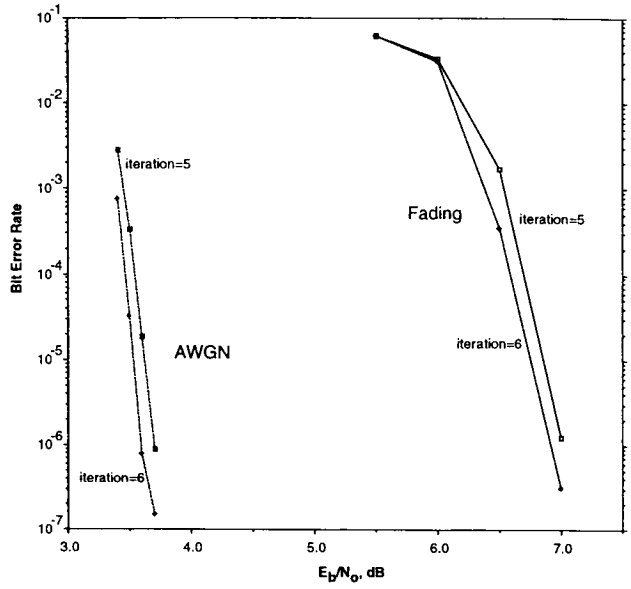


Fig. 6. BER Performance of Parallel Concatenated Trellis Coded 8PSK, 2 bits/sec/Hz.

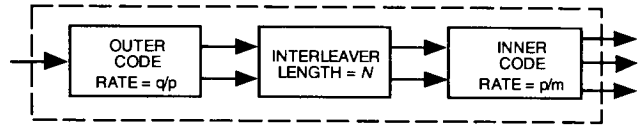


Fig. 7. Serial Concatenated Convolutional Codes (SCCC).

A. *Computation of input–output weight coefficient (IOWC) $A^{Cs}_{w,h}$ for SCCC*

Using the concept of uniform interleaver, i.e., averaging $P_b(e)$ over all possible interleavers, we can obtain $A^{Cp}_{w,h}$ for serial concatenated codes. $A^{Cs}_{w,h}$ is the number of codewords of the SCCC with weight h associated with an input word of weight w. A similar definition applies to input–output weight coefficients (IOWC) of the outer code denoted by $A^{Co}_{w,l}$ and to IOWC of the inner code denoted by $A^{Ci}_{l,h}$.

With the knowledge of the $A^{Co}_{w,l}$ for the outer code, $A^{Ci}_{l,h}$ for the inner code, and using the concept of uniform interleaver, the IOWC $A^{Cs}_{w,h}$ for SCCC can be obtained as follows. We recall that a uniform interleaver transforms a codeword of weight l at the output of the outer encoder into all its distinct $\binom{N}{l}$ permutations. As a consequence, each codeword of the outer code C_o of weight l, through the action of the uniform interleaver, enters the inner encoder generating $\binom{N}{l}$ codewords of the inner code C_i. Thus, the number $A^{Cs}_{w,h}$ of codewords of the SCCC of weight h associated with an input word of weight w is given by

$$A^{Cs}_{w,h} = \sum_{l=0}^{N} \frac{A^{Co}_{w,l} \times A^{Ci}_{l,h}}{\binom{N}{l}}$$

Example 2. Consider a rate 1/2 SCCC formed by a 4-state

convolutional code C_o with rate 1/2 and an inner 2-state convolutional code C_i with rate 1/1 (this is obtained by not sending the systematic bits of the rate 1/2 C_i convolutional code). The two codes are joined by a uniform interleaver. Input blocks of length $N = 50, 100$ and 256 were considered. The outer code is a nonrecursive code, the inner code is systematic and recursive, and the generators are shown in Fig. 8. Using the previously outlined analysis for SCCC, we have obtained the bit-error probability bounds shown in Fig. 8. The performance was obtained both for AWGN and Rayleigh fading channels. Comparing to Fig. 3, the performance of SCCC is better than PCCC both over AWGN and fading channels.

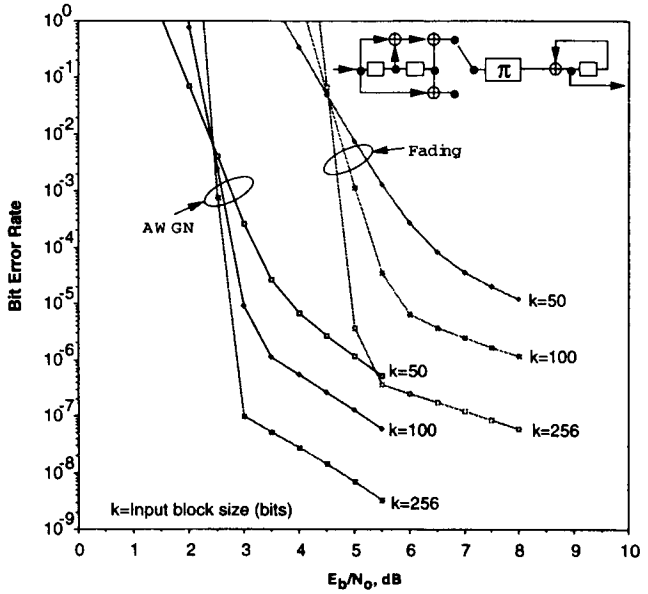


Fig. 8. Performance of rate 1/2 SCCC over AWGN and Rayleigh Fading Channels

For large interleavers the maximum component of $\frac{A_{w,h}^{C_S}}{N}$ over all input weights w, and output weights h is proportional to N^{α_M} with corresponding output weights $h(\alpha_M)$. If the inner convolutional code is recursive (i.e., with feedback) then $\alpha_M = -\left\lfloor \frac{d_f^o+1}{2} \right\rfloor$ where d_f^o is the free (minimum) distance of the outer convolutional code.

The value of α_M shows that the exponents of N are always negative integers. Thus, for all h, the coefficients of the exponents in h decrease with N, and we always have an "interleaving gain".

Define $d_{f,eff}^i$ as the minimum weight of codewords of the inner code generated by weight-2 input sequences. We obtain a different weight $h(\alpha_M)$ for even and odd values of d_f^o. For even d_f^o, the weight $h(\alpha_M)$ associated to the highest exponent of N is given by

$$h(\alpha_M) = \frac{d_f^o d_{f,eff}^i}{2}$$

Substituting the exponent α_M into the expression for bit error rate, approximated by only the term of the summation in h corresponding to $h = h(\alpha_M)$, yields

$$\lim_{N\to\infty} P_b(e) \simeq B_{even} N^{-d_f^o/2} \left[\frac{1}{1 + R_c \frac{E_b}{N_0}} \right]^{\frac{d_f^o d_{f,eff}^i}{2}} \tag{6}$$

where B_{even} is a constant independent of N.

For d_f^o odd, the value of $h(\alpha_M)$ is given by

$$h(\alpha_M) = \frac{(d_f^o - 3)d_{f,eff}^i}{2} + h_m^{(3)} \tag{7}$$

where $h_m^{(3)}$ is the minimum weight of sequences of the inner code generated by a weight-3 input sequence.

Thus, substituting the exponent α_M into the expression for bit error rate approximated by keeping only the term of the summation in h corresponding to $h = h(\alpha_M)$ yields

$$\lim_{N\to\infty} P_b(e) \simeq B_{odd} N^{-(d_f^o+1)/2} \left[\frac{1}{1 + R_c \frac{E_b}{N_0}} \right]^{\frac{(d_f^o-3)d_{f,eff}^i}{2} + h_m^{(3)}} \tag{8}$$

where B_{odd} is a constant independent of N.

VI. Serial Concatenated Trellis Coded Modulation

The basic structure of serially concatenated trellis coded modulation is shown in Fig. 9.

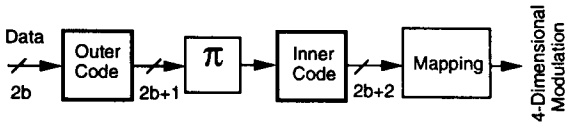


Fig. 9. Block Diagram of the Encoder for Serial Concatenated Trellis Coded Modulation.

We propose a novel method to design serial concatenated TCM for Rayleigh fading channels, which achieves b bits/sec/Hz, using a rate $2b/(2b+1)$ non-recursive binary convolutional encoder with maximum free Hamming distance as outer code. We interleave the output of the outer code with a random permutation. The interleaved data enters a rate $(2b+1)/(2b+2)$ recursive convolutional inner encoder. The $2b+2$ output bits are mapped to two symbols belonging to a 2^{b+1} level modulation (four dimensional modulation). In this way, we are using $2b$ information bits for every two modulation symbol intervals, resulting in b bit/sec/Hz transmission (when ideal Nyquist pulse shaping is used) or, in other words, b bits per modulation symbol. For the AWGN channel the inner code and the mapping are jointly optimized based on maximizing the effective Euclidean distance of the inner TCM. The optimum 2-state inner trellis code is shown in Fig. 10. The effective Euclidean distance of this code is 1.76 (for unit norm constellation) and its minimum M-ary Hamming distance is 1.

A. Design Criteria for SCTCM over Rayleigh Fading Channels

To extend the asymptotic results obtained for binary modulation to to M-ary modulation (e.g., MPSK), criteria simi-

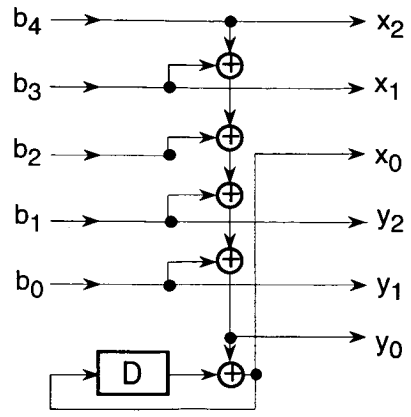


Fig. 10. Optimum 2-state inner trellis encoder for SCTCM with 2× 8PSK Modulation.

lar to those discussed for parallel concatenated trellis coded modulation (PCTCM) are now applied to serial concatenated trellis coded modulation (SCTCM). The interleaving gain is still $N^{-\lfloor (d_f^o+1)/2 \rfloor}$, however now the minimum diversity is $\frac{d_f^o d_{f,eff}^i}{2}$ for even d_f^o, and $\frac{(d_f^o-3)d_{f,eff}^i}{2} + h_m^{(3)}$ for odd d_f^o, where $d_{f,eff}^i$ represents the minimum (M-ary symbol) Hamming distance of the inner trellis code corresponding to input Hamming distance 2 between binary input sequences to the trellis code that produce $d_{f,eff}^i$. Therefore the criterion for optimizing the inner trellis code in SCTCM is to maximize the minimum diversity of the code and then maximize the product of the squared Euclidean distances which result in minimum diversity. For odd d_f^o, first we maximize $d_{f,eff}^i$, then among the codes with maximum $d_{f,eff}^i$, we maximize $h_m^{(3)}$, the minimum (M-ary symbol) Hamming distance of the inner trellis code corresponding to input Hamming distance 3 between binary input sequences to the trellis code that produce $h_m^{(3)}$. As is seen from the previous results, large d_f^o produces large interleaving gain and diversity.

B. Design Method for Inner TCM

The proposed design method is based on the following steps:

1. The well known set partitioning techniques for Rayleigh fading channels using multidimensional signal sets are used (see for example [10] and the references therein).
2. The input labels' assignment is based on the codewords of the parity check code $(2b+1, 2b, 2)$ and its set partitioning, to maximize the quantities described in the design criteria subsection. The assignment of codewords of the parity check code to the 4-dimensional signal points is not arbitrary. We would like somehow to relate the Hamming distance between input labels to the Euclidean distance between corresponding 4-dimensional signal points, under the constraint that the minimum Hamming distance between input labels for parallel transitions be equal to 2. To do so: Assign the b most significant bits of the input label to the first constellation with 2^{b+1} points by retaining only the b most significant bits of the Gray code mapping for the constellation. Use the same assignment for the b least significant bits of the input labels; the middle bit in the input label represents the overall parity check bit.
3. A sufficient condition to have very large output Euclidean and M-ary symbol Hamming distances for input sequences with Hamming distance 1, is that all input labels to each state be distinct.
4. Assign pairs of input labels and 4-dimensional signal points to the edges of a trellis diagram based on the design criteria in subsection VI-A.

To illustrate the design methodology we developed the following examples.

C. Examples of the Design Methodology

Example 1: Set partitioning of 2×8PSK and input labels' assignment.

Let the eight phases of 8PSK be denoted by $\{0, 1, 2, 3, 4, 5, 6, 7\}$. Consider the 2×8PSK signal set $A_0 = [(0,0), (1,3), (2,6), (3,1), (4,4), (5,7), (6,2), (7,5)]$. Each element in the set has two components. The second component is 3 times the first one modulo 8. Also consider the 2× 8PSK signal set $B_0 = [(0,0), (1,5), (2,2), (3,7), (4,4), (5,1), (6,6), (7,3)]$. Each element in the set has two components. The second component is 5 times the first one modulo 8. For these sets, the Hamming distance between elements in each set is 2, and the minimum of the product of square Euclidean distances is the largest possible.

The following sets are constructed from A_0 and B_0 as: $A_2 = A_0 + (0,2)$, $A_4 = A_0 + (0,4)$, $A_6 = A_0 + (0,6)$, $A_1 = B_0 + (0,1)$, $A_3 = B_0 + (0,3)$, $A_5 = B_0 + (0,5)$, $A_7 = B_0 + (0,7)$, where addition is component-wise modulo 8. Map the first and last 2 bits of input labels to the 8PSK signals as $\{00, 00, 01, 01, 11, 11, 10, 10\} \Rightarrow \{0, 1, 2, 3, 4, 5, 6, 7\}$.

The fifth bit for the input label is the parity check bit. Use an even parity check bit for signal sets A_0, A_4, A_1, A_5 and an odd parity check bit for signal sets A_2, A_6, A_3, A_7. This completes the input label assignments to signal sets.

Now the Hamming distance between input labels for each set A_i i=0,1,2,...,7, is at least 2 and the corresponding M-ary Hamming distance between signal elements in each set is 2. Consider a 4-state trellis code with full transition. Assign A_0, A_2, A_4, A_6, to the first state, and A_1, A_3, A_5, A_7 to the second state, and permutations of these sets to the third and fourth states. This completes the input label and 2×8PSK signal set assignments to the edges of the 4-state trellis. Therefore the minimum Hamming distance of the 4-state trellis code is 2. At this point to obtain a circuit that generates this trellis we need to use an output label. We used reordered mapping as it was discussed before to obtain the circuit for the encoder.

The implementation of the 4-state inner trellis code is

shown in Fig. 11. The ROM maps 32 addresses in the range of 0 to 31 to a single output. The 32 binary outputs can be summarized in hex as 3A53ACC5.

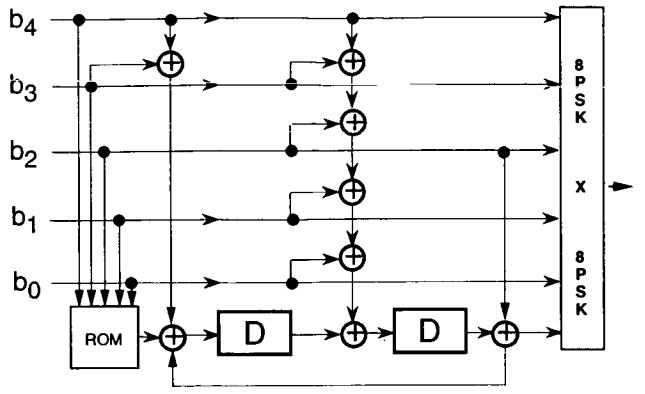


Fig. 11. 4-state inner trellis encoder for SCTCM with 2×8PSK modulation for Rayleigh fading.

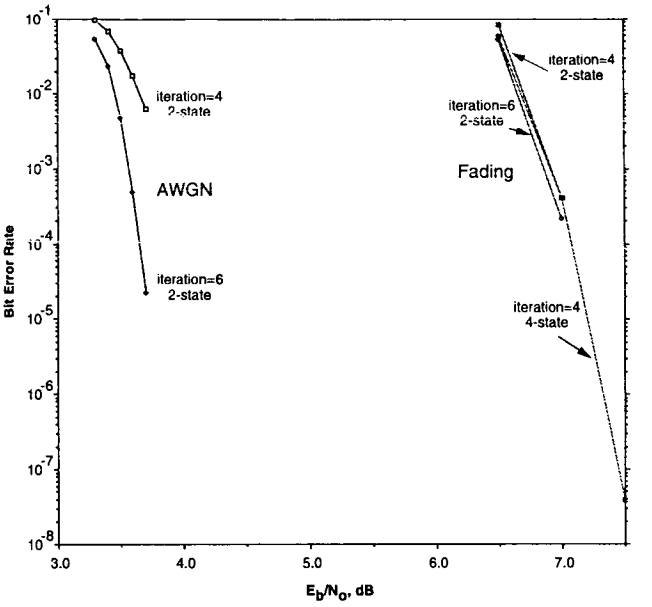


Fig. 12. Performance of Serial Concatenated Trellis Coded Modulation, 8-state outer, 2-state or 4-state inner, with 2×8PSK, 2 bits/sec/Hz

VII. Simulation of Serial Concatenated Trellis Coded Modulation with Iterative Decoding

In this section the simulation results for serial concatenated TCM, with 2×8PSK over the Rayleigh fading channel are presented. For SCTCM with 2×8PSK, the outer code is a rate 4/5, 8-state nonrecursive convolutional encoder with $d_f^o = 3$, and the inner code is the 4-state TCM designed for $2 \times 8PSK$ in subsection VI-C. The bit error probability vs. bit signal-to-noise ratio E_b/N_o for various numbers of iterations is shown in Fig. 12. The performance of the inner 2-state code is also shown in Fig. 12. This example demonstrates the power and bandwidth efficiency of SCTCM, over a Rayleigh fading channel at low BERs.

References

[1] G. Ungerboeck, "Channel coding with multilevel phase signaling", *IEEE Trans. Inf. Th.*, vol.IT-25, pp.55-67, Jan. 1982.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding: Turbo Codes," *Proc. 1993 IEEE International Conference on Communications*, Geneva, Switzerland, pp. 1064–1070, May 1993.

[3] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes", *IEEE Trans. on Inf. Theory*, March 1996.

[4] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, " Parallel Concatenated Trellis Coded Modulation", proceedings of ICC'96 , June 1996.

[5] S. LeGoff, A. Glavieux, and C. Berrou, "Turbo Codes and High Spectral Efficiency Modulation", *Proceedings of IEEE ICC'94*, May 1-5, 1994, New Orleans, LA.

[6] D. Divsalar, S. Dolinar, F. Pollara, and R. J. McEliece, "Transfer Function Bounds on the Performance of Turbo Codes," *Proceedings of IEEE Micom'95*, Nov. 5-8,1995, San Diego, CA.

[7] L.U. Wachsmann, and J. Huber, "Power and Bandwidth Efficient Digital Communication Using Turbo Codes in Multilevel Codes," *European Transactions on Telecommunications*, vol. 6, No. 5, Sept./Oct. 1995, pp. 557–567.

[8] P. Robertson, and T. Woerz, "Novel Coded modulation scheme employing turbo codes," Electronics Letters, 31st Aug. 1995, Vol. 31, No. 18.

[9] S. Benedetto and G. Montorsi, "Design of Parallel Concatenated Convolutional Codes," *IEEE Transactions on Communications*, vol. 44, no. 5, pp. 591–600, May 1996.

[10] D. Divsalar, and M. K. Simon, " Design of Trellis Coded MPSK for Fading Channels: Performance Criteria, and Set Partitioning for Optimum Code Design," IEEE Trans. Commun., Vol. 36, No. 9, pp. 1004-1012, Sept. 1988.

[11] D. Divsalar and F. Pollara, "On the Design of Turbo Codes", *The Telecommunications and Data Acquisition Progress Report 42-123, July–Sept. 1995*, Jet Propulsion Laboratory, Pasadena, California, pp. 99-120, Nov. 15, 1995.

[12] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial Concatenation of Interleaved Codes: Performance Analysis, Design, and Iterative Decoding," *The Telecommunications and Data Acquisition Progress Report 42-126, April–June 1996*, Jet Propulsion Laboratory, Pasadena, CA, Aug. 15, 1996. *http : //edms – www.jpl.nasa.gov/tda/progress_report/42 – 126/126D.pdf*

[13] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial Concatenated Trellis Coded Modulation with Iterative Decoding: Design and Performance," Submitted to IEEE Comm. Theory Mini Conference 97, (Globecom 97).

[14] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, " Serial Concatenated Trellis Coded Modulation with Iterative Decoding," IEEE ISIT 97, June 29 - July 4, 97, Ulm, Germany.

[15] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Soft-Output Decoding Algorithms in Iterative Decoding of Turbo Codes," *The Telecommunications and Data Acquisition Progress Report 42-124, Oct.–Dec. 1995*, Jet Propulsion Laboratory, Pasadena, CA, pp. 63–87, Feb. 15, 1996.

[16] E. K. Hall, and S. G. Wilson, "Design and Analysis of Turbo Codes on Rayleigh Fading Channels," IEEE JSAC special issue, submitted Sept. 96.

[17] J. Craig, " A new, simple and exact result for calculating error probability for two-dimensional signal constellation," *Proceedings of IEEE Milcom'91*, 1991.

Session 10
Terminal Antennas

Session Chairperson—*Shingo Ohmori*, Communications Research Laboratory, Japan
Session Organizer—*John Sydor*, Communications Research Centre, Canada

Optimizing Smart Antenna Characteristics for CDMA Mobile Satellite Communications

S. H. H. Lim, L. P. Sabel and C. Brander
Institute for Telecommunications Research
University of South Australia
The Levels, South Australia 5095, Australia
stanley@spri.levels.unisa.edu.au

ABSTRACT

Significant performance improvement can be obtained by suppressing undesired co–channel interferers within a CDMA mobile satellite communications system. A smart antenna is proposed as an approach to suppress the unwanted interference. The characteristics of the smart antenna are determined with the use of satellite constellation information. We consider smart antenna techniques at the mobile and concentrate on the performance analysis of the down–link noting that commensurate performance improvements are likely to occur on the up–link.

I. INTRODUCTION

In a CDMA system, each co–channel user is ideally separated through code orthogonality. In reality the user codes are not completely orthogonal and hence the co–channel interference suppression capability is limited. Consequently, the bit error rate (BER), or the frame error rate, increases as the number of co–channel users increases. This applies to both terrestrial and satellite communication systems using CDMA [1], [2]. Through the use of a RAKE receiver structure, CDMA systems provide a method of achieving spatial diversity. For satellite systems spatial diversity is provided through the visibility of multiple satellites at the user terminal. However, as the number of visible satellites increases, the interference received at the mobile terminal also increases. This is particularly the case when the mobile terminal uses an omni–directional (hemispherical beam pattern) antenna. The increase in interference power level over any coverage footprint is directly proportional to the number of users in each beam and the number of beams.

In this paper, we analyse the suitability of smart antennas for the mobile terminal to combat the system performance degradation caused by co–channel interference, especially in a region where there is a high likelihood of coverage by several satellites. In particular we determine the trade–offs between the antenna beam pattern characteristics and the performance improvement offered. Once suitable beam pattern requirements have been determined potential antenna realisations can be examined.

The paper will be structured as follows: Section II will provide a brief overview of the CDMA interference problem in a mobile satellite context. In Section III we model the effect of co–channel interference on $E_b/(N_0+I_0)$ as a function of the antenna characteristics. We examine the effects of varying parameters such as main beamwidth (and gain) and null depth, considering a range of conditions such as the number of visible satellites and their relative spatial positioning. We will demonstrate that considerable performance improvement can be obtained using the beamforming capabilities of smart antennas. Preliminary results show that the beamwidth and null depth requirements are quite modest considering the performance improvements offered. Clearly, increased performance in terms of receiver $E_b/(N_0+I_0)$ can be transformed into increased system capacity. In Section IV we determine suitable antenna beam pattern characteristics through an analysis of the satellites' spatial distribution. Section V will briefly discuss possible realisations of the antenna. Finally, conclusions will be drawn in Section VI.

II. CDMA Interference Analysis

Consider a block diagram of a CDMA system in its most general form [3] shown in Figure 1 where $s_i(t)$ is the modulated signal for user i, with symbol duration T seconds. For simplicity let us assume that the pseudo random code sequence, $c_i(t)$ has a chip period defined as $T_c = T/L$. In order to calculate the orthogonality of the code, it is necessary to know the cross correlation between the spreading sequences. The cross–correlation between the spreading sequences for user i and user j is defined as

$$R_{ij} = \int_{t_0}^{t_0+T} c_i(t)c_j(t)\,dt \qquad (1)$$

where $|R_{ij}| \leq 1$ and R_{ij} is a periodic corrrelation function.

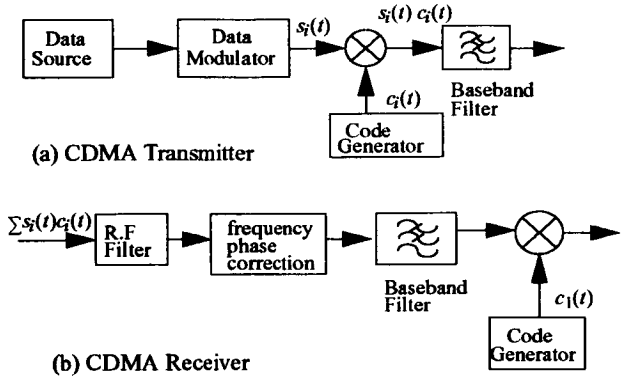


Figure 1. Basic CDMA Transmitter and Receiver.

Assume that $\sum s_i(t)\ c_i(t)$ is received at the receiver, where $s_i(t)$ are the transmitted data signals for i=1 to N active users, assuming no receiver noise, no distortion introduced at the output of the baseband matched filter, and perfect spreading code, frequency and phase synchronisation. The wanted information signal for user 1 is recovered by multiplying by an exact replica of the spreading sequence which yields

$$\textit{wanted signal, } y_W(t) = s_1(t) \int_{t_0}^{t_0+T} c_1^2(t)\ dt \qquad (2)$$

$$\textit{interference signal, } y_I(t) = \sum_{i=2}^{N} s_i \int_{t_0}^{t_0+T} c_1(t)c_i(t)\ dt \qquad (3)$$

Equation (3) gives the multiuser interference signal in the presence of N–1 interferers. The average interference power per symbol is

$$I = 10log_{10}(E[y_I^2(t)]) = 10log_{10}(\sum_{i=2}^{N} E[R_{1i}^2]C_i) \qquad (4)$$

where C_i =E[$s^2_i(t)$] is the energy of interferer i.

Here we assume that there are 128 Walsh code channels available for any forward traffic. Theoretically, the Walsh sequences are orthogonal to each other within the designated forward traffic channel. Orthogonality in this case means the presence of other users in the same channel of a given satellite beam will not contribute to any interference. However it was found from our simulation that due to baseband filtering, these Walsh sequences introduce a correlation which corresponds to approximately 2.7 random interferers. This approximation of the number of interferers comes as a result of normalising the energy in the 128 chips to unity. The simulation also revealed that there is very little if any orthogonality of these codes between beams due to asynchronous detection of these interfering signals at the user end. This would imply that if two beams overlap, which is the case for satellite diversity scheme proposed, the user terminal simply sees an additional N interferers.

Considering this, we now determine the interference power density in the forward link of the CDMA satellite communication system. This interference may come from within the satellite as a result of imperfect orthogonality of codes or purely from the other visible satellites. The latter will contribute a significant portion towards the overall interference power density.

In the presence of only one satellite, the interference power density within the same satellite cell can be written as

$$I_0 = P^r_{min} + G_r + 10log_{10}(2.7N/128)–10log_{10}(B) \qquad (5)$$

where P^r_{min} is the minimum receive power at the mobile station, G_r is the mobile station receive antenna gain, N is the number of active users and B is the spreading bandwidth.

The contribution to the interference power density by the undesired satellites is therefore given by

$$I^e_0 = P^r_{min} + G_r + 10log_{10}(SN)–10log_{10}(B) \qquad (6)$$

and is directly proportional to the number of active users where S is the number of interfering satellites, each assumed to have N active users.

The total interference is therefore given by the addition of (5) and (6) as

$$I^m_0 = P^r_{min} + G_r + 10log_{10}(2.7N/128 + SN)–10log_{10}(B) \qquad (7)$$

which takes into consideration the contribution of additional interference power density from both within and other interfering satellites illuminating the same footprint. This is valid only for an omnidirectional receive antenna with gain G_r.

Figure 2 gives a comparison of interference power density level as a function of the number of users as calculated using (7). As the number of interfering satellites increases, the interference power density level rises. Also indicated on the plot is the constant receiver thermal noise power density, N_0 [2]. The system $E_b/(N_0+I_0)$ is clearly limited by noise power density in the absence of any other interfering satellite(s). Also, it is worth noting that the interfering satellites have limited influence on the system $E_b/(N_0+I_0)$ if the number of users in the system is low. The system only becomes interference limited when the number of users as well as the number of interfering satellites reaches a certain level. For simplicity, we can assume that when the interference power density becomes equal or greater than N_0, the system changes from noise power density limited to interference power density limited.

From this argument, it makes sense to configure the mobile terminal antenna so as to steer nulls toward the interfering satellites and provide some gain in the direction of the desired satellite. There is potential to increase the system $E_b/(N_0+I_0)$ by just nulling the interference power from undesired satellite(s).

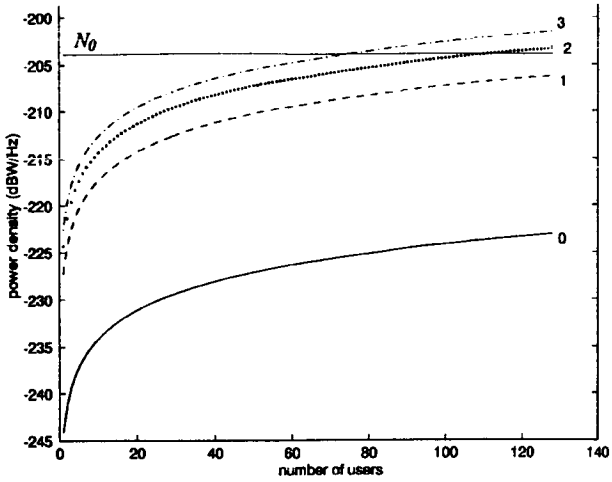


Figure 2. Interference Power density vs number of users for 0, 1, 2 and 3 interfering satellite(s).

III. Performance against Antenna Radiation Pattern

In this section, a comparison of a number of possible antenna patterns is carried out to investigate the system $E_b/(N_0+I_0)$ as a function of the number of users. Then, we interpret the results in order to give a comparison for the presumed case of 64 users taken as the expected number of users and that of 128 users which is the maximum system capacity. The degradation of $E_b/(N_0+I_0)$ as the number of users increases from 64 to 128 as a function of the receive antenna gain will be pointed out.

The interference power density analysis in three different scenarios was investigated. $E_b/(N_0+I_0)$ is used as a parameter to measure and compare the system performance. The presence of two, three and four visible satellites give rise to the three scenarios. For each of these scenarios, three different types of receive antenna patterns are considered. All receive antenna patterns considered allow coherent combining gain but only *Case* 2 and *Case* 3 allow interference reduction. Coherent combining gain is an inherent property of a RAKE receiver. From [2], the coherent combining gain is given as 2.5 dBi. These three receive antenna patterns are:

Case 1 : *Traditional Omnidirectional*

The interference power density is calculated using (7). The definition of omni for this case refers to a beam which is wide enough to capture all satellites in sight with equal gain in each of the satellite directions.

Case 2 : *Single Beam and Multiple Nulls*

This is applicable only if more than two satellites are visible at any instant. The general expression for determining the interference power density is

$$I'_0 = 10^{\left[\frac{I_0(X)}{10}\right]} + (U-X)10^{\left(\frac{I_0^e - G_r + G_r^n}{10}\right)} \qquad (8)$$

where X is the number of the satellites at which the beam is pointed (2 in this case), G^n_r is the depth of null pointed to unwanted satellite(s) and U is the number of visible satellites.

The single beam is always considered to cover the two satellites used for the provision of diversity. Therefore, identical results in $E_b/(N_0+I_0)$ is expected in the scenario of only two visible satellites. The receiver structure consists of an m–element antenna, a beamformer and a multifinger RAKE receiver.

Case 3 : *Two Independent Beams with Multiple Nulls*

In this case, there are two independent beams each with multiple nulls. Each independent beam directs its gain to the desired satellite and nulls to the undesired satellite(s). Therefore, the number of nulls required is always one less than the number of visible satellites at the time. The interference power density can be calculated using

$$I'_0 = L[(10^{\left[\frac{I_0(1)}{10}\right]} + (U-1)10^{\left(\frac{I_0^e - G_r + G_r^n}{10}\right)})] \qquad (9)$$

where L is the number of independent beams (2 in this case). The receiver structure may consists of n–element antenna, two individual beamformers and two single finger RAKE receiver.

The above three antenna configurations differ at the RF front end. The total signal and noise plus interference is then delivered to the common RAKE receiver. The signals from the individual front end receivers in *Case* 3 are added in the power sense.

Variation of Receiver Antenna Gain

Receive antenna gain of 2.6 dBi is used as an absolute minimum in the simulation. This value has also been used in the link budget analysis [2]. Wherever applicable, the null is kept constant at an absolute level of –5 dBi independent of the main beam gain.

One can observe the changes of $E_b/(N_0+I_0)$ as a function of the number of users for the three different receive antenna patterns under the different scenarios. The $E_b/(N_0+I_0)$ for two, three and four visible satellites are plotted as a function of number of users for different values of receive antenna gain ranging from 2.6 dBi to 5.6 dBi in steps of 1 dBi in Figure 3, 4 and 5 respectively. The crosses on these Figures indicate a reference point of 50% channel capacity for *Case* 1. Solid line represents G_r=2.6 dBi, dashed line represents G_r=3.6 dBi, dotted line represents G_r=4.6 dBi and dash–dotted line represents G_r=5.6 dBi. The three curves of each line type correspond to the use of three different receive antenna patterns, *Case* 1 being the lowest curve and *Case* 3 the highest.

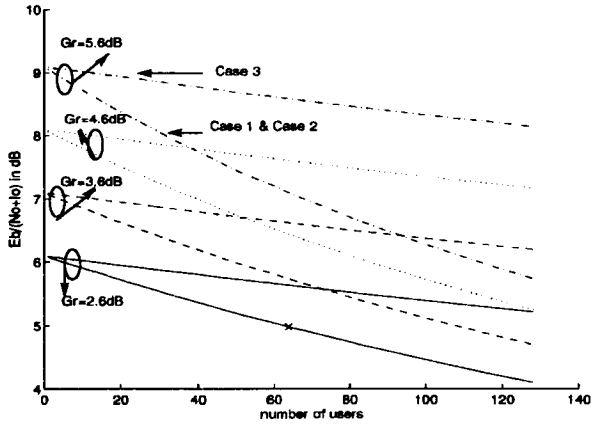


Figure 3. $E_b/(N_0+I_0)$ as a function of number of users and antenna parameter (see text) (two satellites).

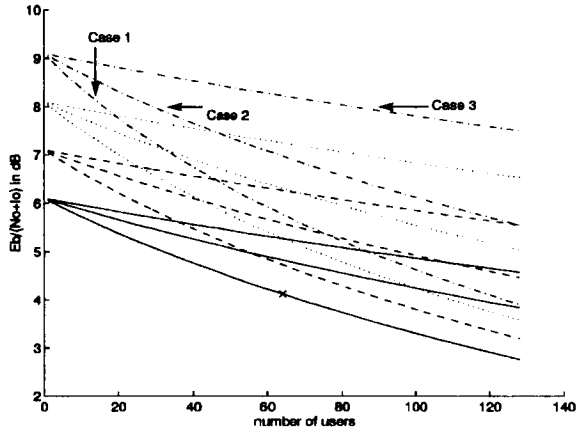


Figure 4. $E_b/(N_0+I_0)$ as a function of number of users and antenna parameter (see test) (three satellites).

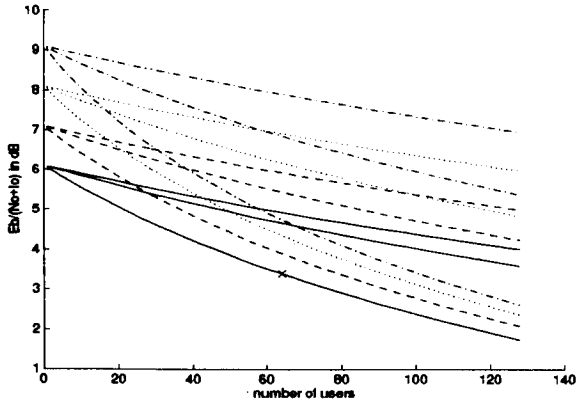


Figure 5. $E_b/(N_0+I_0)$ as a function of number of users and antenna parameter (see text) (four satellites).

From the above figures, if the required $E_b/(N_0+I_0)$ is 5 dBi, the traffic channels for *Case* 1 will be 64, 30 and 20 in the presence of one, two and three interfering satellites. The actual capacity will be half of this number because of satellite diversity.

Null Sensitivity of Shaped Antenna Beam

This section investigates the null depth required to obtain the optimum benefit in the system $E_b/(N_0+I_0)$. By increasing the null depth pointing to any unwanted signal sources, further interference power reduction can be achieved. However it is essential to determine the minimum null depth required to make the system work, and not make unreasonable demands on the antenna design.

With receive antenna gain ranging from 2.6 to 5.6 dBi, there appears to be an upper bound on $E_b/(N_0+I_0)$ as illustrated by Figure 6. For Case 2, null depth of greater than 10 dBi has little influence on the $E_b/(N_0+I_0)$. The upper bound limit is about 6.7 dBi with G_r=5.6 dBi and slightly above 4 dBi with G_r=2.6 dBi. Similarly, we have the upperbound value of 6 dBi (G_r=2.6) and 9 dBi (G_r=5.6) respectively for Case 3, regardless of the number of visible satellites, as the null depth increases to 20 dBi and below.

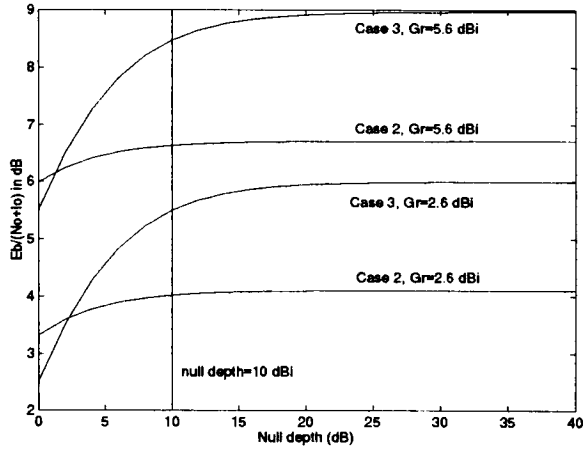


Figure 6. Comparison of $E_b/(N_0+I_0)$ vs null depths (3 satellites visible).

The null depth requirement varies depending on the number of interferers, in this case the number of active users in the visible satellites. Even with a shallow null, quite a good performance improvement results. A 10 dBi null is perhaps a fair compromise between performance improvement and antenna realisation. The effect of this is shown in Figures 7, 8 and 9 for a system with 128 users/channels.

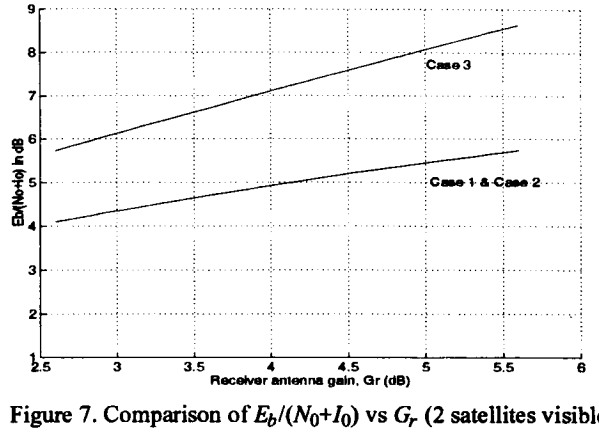


Figure 7. Comparison of $E_b/(N_0+I_0)$ vs G_r (2 satellites visible)

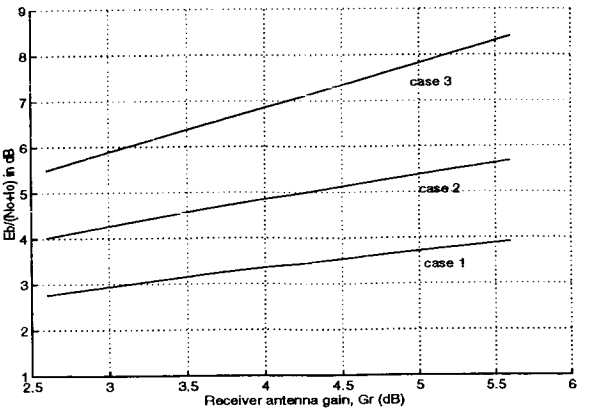


Figure 8. Comparison of $E_b/(N_0+I_0)$ vs G_r (3 satellites visible).

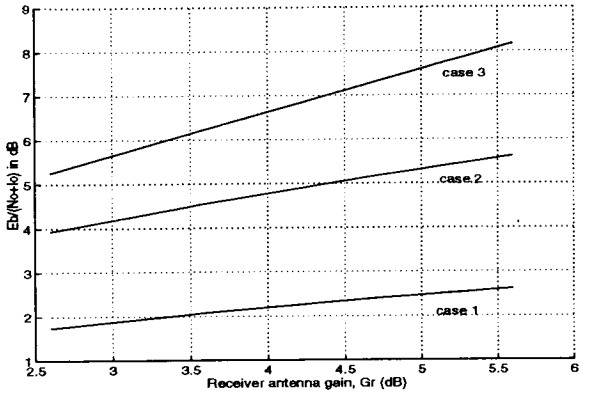


Figure 9. Comparison of $E_b/(N_0+I_0)$ vs G_r (4 satellites visible).

Clearly, *Case* 3 has interference rejection capability superior to the other two cases. But it is also more complex in its realisation. The degradation of $E_b/(N_0+I_0)$ for *Case* 1 is quite significant in the presence of three and four visible satellites. In such situations, *Case* 2 can be introduced to achieve stable demodulation and acquisition.

IV. Desired Mobile Satcom Antenna Characteristics

The desired antenna characteristics for a mobile satellite handset are very similar to those of a terrestrial. These characteristics can be described by its physical and electrical requirements. The physical requirements are small in size, light in weight and compact while the electrical requirements may include high efficiency, sufficient bandwidth, simple and low loss impedance matching network, sufficient gain, matched polarisation, little degradation due to near–in objects and health hazard free. But with the presence of a LOS signal path in satellite communication systems, higher gain and null steering antennas are attractive.

In this context, the antenna characteristics associated with beam–shaping requirements were derived purely based on knowledge of the satellite spatial distribution. The satellite which has the shortest slant range is selected as the reference satellite in every instance. This reference satellite also has the highest elevation angle amongst all visible satellites. Depending on the number of satellites it is desired to cover, the requirements vary.

In the presence of only two satellites, the antenna radiation pattern can be omnidirectional, or shaped if higher gain is desired. However, in the presence of more than two satellites, the beam must be configured to cover only the desired two satellites for diversity purposes. It appears that there is no need for variation in elevation beamwidth but a variation of azimuth beamwidth and a tracking mechanism is necessary. Adelaide, South Australia was selected as the ground station where all these simulated data were obtained. Each sample contains information about the satellite(s)'s spatial location at that instant and at an interval step of 20 seconds. 8000 samples corresponding to a time interval of 44 hours were used. The selection of the number of samples is of no significance but a large enough number of samples is required to give an adequate picture of the actual scenario and a reliable statistic to work with. Also from the simulation, an approximation for the percentage of time as a function of number of visible satellites was determined. Adelaide (34.767S, 139.133E) will see two satellites at all times, three satellites 91.7% of the time and four satellites 32.7% of the time. This pattern applies not only to Adelaide but also to any place with the latitude of 34.767 degree either south or north.

The azimuth angle is measured with respect to the north pole. Zero degrees azimuth means straight north from the desired ground station, and 180 degrees means straight south from the desired ground station.

Figure 10 indicates the percentage of time when two satellites are within the shaped antenna beam of various azimuth beamwidths and a fixed elevation beamwidth of 80° (elevation angle of 10° to 90°). It is seen that an azimuth beamwidth of 120° to 125° sees two satellites at all times, assuming ideal tracking.

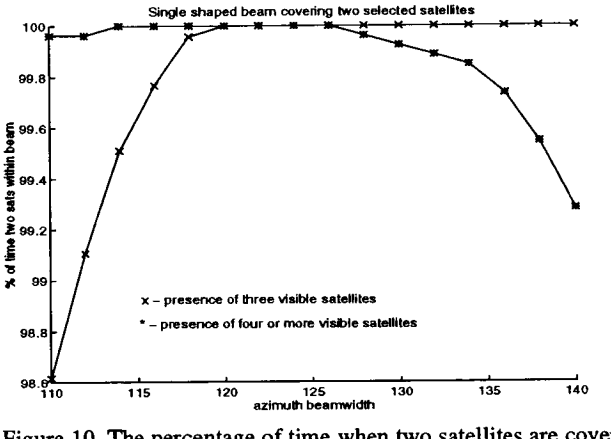


Figure 10. The percentage of time when two satellites are covered by the shaped beam as a function of azimuth beamwidth at an elevation beamwidth of 80 degrees.

V. Smart Antenna Realisation

The quadrifilar helix has always been a potential candidate as a radiating element for satellite related applications due

to its inherent circularly polarised (CP) waves. Various array configurations (of helices or other CP elements) could produce good approximations to the antenna patterns required if there were no space limitations.

Calculations indicate that a 4–element ring array with a diameter of about 0.08 wavelength is sufficient to control the azimuth beam, and a 3 element vertical array with about 0.2 wavelength spacing can produce the required elevation patterns. At 2.5 GHz, this would be 1 cm diameter and 6 cm long – a tolerable size and shape, not too different from the generally accepted monopole. The antenna (and probably any antenna for this purpose) would need to be extended from the case to be clear of head blockage, and oriented vertically in operation.

The crossed slot is a suitable radiating element which would fit around this shape (4 elements on about 3 cm circumference) if the length of the slot could be made less than 1 cm, which is about 1/12 wavelength. Such a slot could be matched, but would have narrow bandwidth. A way of making the slots resonant is to load them with dielectric.

We are investigating a structure shown in figure 11 where the radiating slots and associated feed networks are printed on a thin low permittivity sheet, which is then rolled around a high permittivity rod. Some preliminary calculations indicate that this structure can produced a promising approximation to the required beam pattern.

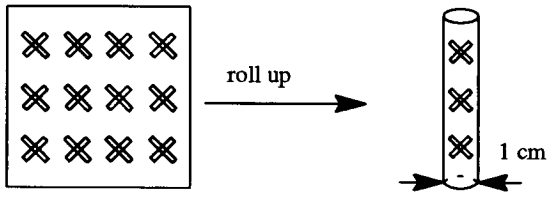


Figure 11. A possible structure for realising a smart antenna for mobile satellite handsets.

The same array may serve as the transmit antenna if a proper coupling technique is employed. Alternatively, a separate transmit antenna could be printed on the same substrate with or without beam shaping.

VI. Conclusion

Table 1.1 provides a comparison of $E_b/(N_0+I_0)$ for the three different antenna radiation pattern cases discussed above as a function of the receive antenna gain and the null depth in the presence of four visible satellites where 64 users/channels is assumed. A similar performance comparison was given in Table 1.2 assuming that the system is fully loaded, that is 128 users/channels. One important point to note here is that no null steering is required as long as the maximum sidelobe level of the antenna radiation pattern relative to the main lobe is under the required null depths. With the null depth requirement values obtained from our study, we can conclude main lobe tracking without null steering is sufficient.

If 5 dBi $E_b/(N_0+I_0)$ is taken as a criterion [2], beamforming at the mobile terminal is of significant benefit in the presence of more than two satellites, or even with only two satellites if more than 50% capacity which equates to 25% capacity when two satellite diversity reception is used. The proposed minimum antenna beamwidth in the azimuth plane was found to be 120° and 80° coverage in the elevation plane. For these specified beamwidths and the proposed realisation, a gain of 5.6 dBi is within reach.

	Gain (dBi)	**2.6**	**3.6**	**4.6**	**5.6**
Case 1	$E_b/(N_0+I_0)$	3.5	4	4.4	4.8
Case 2	*Null depth*	10	10	8	8
	$E_b/(N_0+I_0)$	4.8	5.7	6.3	6.9
Case 3	*Null depth*	10	10	8	8
	$E_b/(N_0+I_0)$	5.8	6.8	7.7	8.7

Table 1.1 — Comparison of $E_b/(N_0+I_0)$ with main beam gain and null depth for *Case* 1, 2 and 3. (64 users/channel, 4 visible satellites)

	Gain (dBi)	**2.6**	**3.6**	**4.6**	**5.6**
Case 1	$E_b/(N_0+I_0)$	1.75	2.1	2.4	2.6
Case 2	*Null depth*	10	10	8	8
	$E_b/(N_0+I_0)$	3.9	4.5	5	5.5
Case 3	*Null depth*	10	10	8	8
	$E_b/(N_0+I_0)$	5.2	6.1	6.8	7.8

Table 1.2 — Comparison of $E_b/(N_0+I_0)$ with main beam gain and null depth requirement for *Case* 1, 2 and 3 (128 users/channel, 4 visible satellites)

It is evident that there is significant improvement in an interference limited scenario with the use of either *Case* 2 or *Case* 3 antenna configuration. *Case* 2 is proposed here due to its relative ease of realisation given the antenna design requirements determined from the study.

Future work will include realisation of the proposed antenna and incorporation of the real antenna radiation pattern in our simulation. A real antenna may not have the same gain in the direction of the two wanted satellites, or the same nulls depth in the direction of each of the unwanted ones, but should still be capable of providing the $E_b/(N_0+I_0)$ enhancement required.

REFERENCES

[1] E. Hirshfield, *The Globalstar System*, Applied Microwave & Wireless, pp. 26–41, Summer 1995.

[2] K.G.Johannsen, *Mobile P–Service Satellite System Comparison*, International Journal of Satellite Communications, vol. 13, pp. 453–471.

[3] R.A.Scholtz, *The Spread Spectrum Concept*, IEEE Transactions on Communications, Vol. COM–25, No. 8, pp. 748–755, August 1977

Performance of Antennas Array for Mobile Communications Receivers

E. Del Re, L.Pierucci, M. Rupi
Dipartimento di Ingegneria Elettronica
Via S. Marta 3, 50139 Firenze - ITALY
Tel: + 39 55 4796271 - Fax: + 39 55 4796485
e-mail: marilli@tmed.die.unifi.it

Abstract

The rapidly increasing demands in Personal Mobile Communications have determined interest in new antennas architectures.
The described antenna system is realized for a scenario of satellite and terrestrial cellular integrated systems.
The work presents linear and planar array systems capable to estimate the DOA (Direction Of Arrival) of maximum power path and to adaptively suppress the interference signals using beamforming techniques also in a multipath environment with Doppler effects.
The use of planar arrays permits to follow the satellite for a longer time than linear arrays, and it is particularly useful in LEO (Low Earth Orbit) satellite systems, in which the satellite platform is in Line Of Sight for a brief time in a day.
The manuscript describes the performance of the MUSIC (MUltiple SIgnal Classification), Minimum Norm, Frost and Howells-Applebaum algorithms explained in literature applied to the Third Generation UMTS (Universal Mobile Telecommunication System) systems.

1. Introduction

The analyzed antenna systems are directive arrays based on adaptive techniques recovering directions of impinging waves in different satellite/radio mobile environment, and realizing beamforming techniques which deal with interference signals.
For the third generation UMTS systems, the personal communication services are conceived in order to work with support of satellite platforms integrated with terrestrial base stations. The array systems presented in this paper are suitable to be implemented at the satellite ground stations and/or cellular base stations, and supply the following advantages with respect to the omnidirectional ones, actually used:

- they are directive and offer the capability to select the active user;
- they can tracking the satellite, and especially the planar array can be used to scan the main beam towards any point in the space for a time sufficiently long to follow also satellites in LEO orbit.

The main goals of antennas array introduction are in localizing the satellite station in order to establish a link with mobile users so as to radiate density power flux directly, obtaining the following advantages:

- reducing power requirement of the transmitted antenna;
- attenuating interference signals towards other users or other communications systems.

The antenna array permits the implementation of a new multiplex technique called Adaptive Space Division Multiple Access (A-SDMA), which raises channel capacity via an increase of frequency re-use.
Such innovative array architectures, in fact, are very suitable to be implemented to the future communication systems, due to their capability to maximize the antenna gain along directions corresponding to that of desired signal and to realize an angular discrimination of beams determining an increase of system traffic capacity by a more efficient re-use of frequency techniques.
The work presents planar array systems capable to estimate the DOA of maximum power path and multiple replica of desired signal. Besides, this contribution considers, for the linear array, the introduction of beamforming techniques which deal with co-channel interference and propagation impairment such as selective fading or multipath phenomenon.
The Multiple Signal Classification (MUSIC) and Minimum Norm algorithms have been developed for DOA estimation in the case of planar arrays. Besides, beamforming algorithms, as Howells-Applebaum and Frost, are considered for linear arrays. These two aspects, DOA estimation and beamforming techniques, are strongly related, due to the necessity to know, for beamforming algorithms, the steering vector corresponding to the desired direction.
In literature these algorithms are used for uncorrelated signals, and this property is not present in the

multipath signals, as in the Personal Mobile Communications ones. The Spatial Smoothing method is used to decorrelate the real signals and therefore it is applied to restore the original capability of the above algorithms.
The DOA estimation algorithms and Spatial Smoothing method are developed for the planar array case. In the planar array the sensors are located along a rectangular grid and this kind of bidimensional array permits to determine symmetrical shapes of pattern with low lateral lobes.
It is important to outline that the linear arrays have only one constraint, corresponding to azimuth angle, and thus can localize all the incident signals coming from directions located in the plane orthogonal to the same array. The linear antenna, then, are not capable to discriminate signals with an elevation angle respect to such orthogonal plane (they have some blind incident angles).
The planar arrays, instead, can be used to scan the main beam of the antenna towards any point in the space and permit to follow the satellite for a time more long than the linear array. The second property is important in the case of LEO orbit, when the satellite is in sight only few minutes, while it is necessary to elaborate a high number of data.
The paper is organized as follow: the next section explains the modified DOA smoothed algorithms applied for the planar array; Section 3 presents the performance of DOA estimation algorithms for planar array in the case of satellite and cellular terrestrial communication channel, and of beamformer smoothed methods for linear arrays in the case of terrestrial environment. Finally, the conclusions are provided.

2. DOA estimation algorithms

Linear and planar array systems capable to estimate the DOA of maximum power path in order to compensate the uncorrelated co-channel interference and the multiple replica of the desired signal with Adaptive Beamforming techniques are analyzed [5].
The MUSIC or Minimum Norm [1] [2] algorithms have been developed for DOA estimation, in companion to smoothing technique [3], and then their information has been used by the beamforming algorithms.

DOA for planar arrays

Let us consider an equally spaced planar array of $N \times N$ sensors, in which impinge a plane wave propagating from L point sources from different directions. Individual radiators can be positioned along a square grid to form a planar array. The elements of the array outputs can be represented as

$$x_{n,m}(t) = \\ = \sum_{i=1}^{L} s_i(t) e^{jkd((n-1)sin(\theta_i)cos(\varphi_i)+(m-1)sin(\theta_i)sin(\varphi_i))} \tag{1}$$

for $n, m = 1, ..., N$, where d is the inter element distance and $k = \frac{2\pi}{\lambda}$ is the propagation factor. Numerating by column the element sensors until $N_T = N \times N$, it is possible to write the elements of the output vector, for $i = 1, ..., N_T$, as [1]

$$x_i(t) = \sum_{m=1}^{L} s_m(t) e^{j\boldsymbol{q}_m^T \boldsymbol{r}_i} + n_i(t) \\ \boldsymbol{q}_m(\lambda, \theta, \varphi) = k[sin\theta_m cos\varphi_m, sin\theta_m sin\varphi_m, cos\theta_m]^T \tag{2}$$

where $\boldsymbol{r}_i = (x_i, y_i, 0)$ is the position vector of ith element array and n_i is the AWGN noise component. It is possible to define a $N^2 \times N^2$ correlation matrix that, with incoherent signal sources and AWGN components noise, uncorrelated with the signal and each other, is a diagonal matrix. By means of eigendecomposition methods, it is possible to calculate an angular spectrum that tends to have large peaks at the angular frequencies corresponding to the plane wave arrival directions. From the correlation matrix we can proceed evaluating the DOA by searching for the peaks present in the bidimensional functions $P_M(\theta, \varphi)$ or $P_{MN}(\theta, \varphi)$ for the MUSIC or Minimum Norm methods respectively.

$$P_M(\theta, \varphi) = \frac{1}{\sum_{i=L+1}^{N} |\boldsymbol{a}(\theta, \varphi)^H \boldsymbol{v}_i|^2} \\ P_{MN}(\theta, \varphi) = \frac{1}{|\boldsymbol{a}(\theta, \varphi)^H \boldsymbol{e}|^2} \tag{3}$$

where $\boldsymbol{v}_i$ is the eigenvector of the signal subspace of the autocorrelation matrix, $\boldsymbol{e}$ is a vector which lies in the signal subspace, has the first element unitary and has minimum Euclidean norm (hence the name), and $\boldsymbol{a}(\theta, \varphi)$ is an $N \times N$ matrix (with $B = N - 1$)

$$\boldsymbol{a}(\theta, \varphi) = \\ \begin{bmatrix} 1 & .. & e^{jkdBsin(\theta)sin(\varphi)} \\ .. & .. & .. \\ e^{jkdBsin(\theta)cos(\varphi)} & .. & e^{jkdB(sin(\theta)cos(\varphi)+sin(\theta)sin(\varphi))} \end{bmatrix} \tag{4}$$

The main drawback of the classical DOA estimation and beamforming algorithms is the forced condition that the source signals are no coherent or uncorrelated, too restricting hypothesis for Personal Mobile Communications.

Considering a planar array system that can be used to link the satellite with the mobile user, the received signals are subjected to multipath and Doppler effects. In a real scenario, in fact, some physical propagation phenomenon, like reflections, diffraction or scattering that attenuate and delay the original transmitted signal, has to be considered. In addition, the wideband channel characteristics decompose the desired information signal into the sum of correlated multipath contributes unresolvable by the conventional DOA estimation algorithms and can be overcome using the spatial smoothing methods.

The Spatial Smoothing method uses spatial averaging techniques to decorrelate the signals and therefore it is applied to restore the original capability of the above algorithms. Such method gives the necessary information to the beamforming algorithms so as to permits to shape the antenna pattern in order to insert nulls near the angular directions corresponding to a coherent and co-channel signals (Frost and Howells-Applebaum algorithms).

For the planar array, the spatial smoothing method has been modified in order to extend the algorithms to the bidimensional case. Considering a square window of size m that shifts on the planar array along horizontal and vertical directions, a $M = (N-m+1)^2$ sub-planar array of length $m \times m$ is obtained.

It is possible to represent the output signal with the $m \times mL$ matrix, in which each $m \times m$ block, $\boldsymbol{A}^i_{m,m}$, represents the response of the sensors to signal ith, for $i = 1, ..., L$ (where $b = m-1$).

$$\boldsymbol{A}^i_{m,m} = \begin{bmatrix} 1 & .. & e^{jkdbsin(\theta_i)sin(\varphi_i)} \\ .. & .. & .. \\ e^{jkdbsin(\theta_i)cos(\varphi_i)} & .. & e^{jkdb(sin(\theta_i)cos(\varphi_i)+sin(\theta_i)sin(\varphi_i))} \end{bmatrix} \tag{5}$$

Let the matrix elements be expressed separately with different letters:

$$\boldsymbol{A}_{m,m} = \left[\begin{array}{ccc|c|ccc} a^{(1)}_{11} & .. & a^{(1)}_{1m} & .. & a^{(L)}_{11} & .. & a^{(L)}_{1m} \\ a^{(1)}_{21} & .. & a^{(1)}_{2m} & .. & a^{(L)}_{21} & .. & a^{(L)}_{2m} \\ .. & .. & .. & & & & \\ a^{(1)}_{m1} & .. & a^{(1)}_{mm} & .. & a^{(L)}_{m1} & .. & a^{(L)}_{mm} \end{array}\right] \tag{6}$$

The matrix of signals received by the sub planar array corresponding to the ij position of the window, has an expression such as

$$x^{i,j} = \begin{bmatrix} x^{i,j}_{11} & x^{i,j}_{12} & \dots & x^{i,j}_{1m} \\ x^{i,j}_{21} & x^{i,j}_{22} & \dots & x^{i,j}_{2m} \\ \dots & \dots & \dots & \dots \\ x^{i,j}_{m1} & x^{i,j}_{m2} & \dots & x^{i,j}_{mm} \end{bmatrix} \tag{7}$$

$$x^{i,j}_{k,q} = a^{(1)}_{k,q} s_1 a^{(1)}_{i,j} + a^{(2)}_{k,q} s_2 a^{(2)}_{i,j} + \dots + a^{(L)}_{k,q} s_L a^{(L)}_{i,j}$$

Finally, we have a $[x^{i,j}]_{m\times m}$ matrix, with $i,j = 1,\dots,M$. Considering the covariance matrix of the ij sub-planar array $\boldsymbol{X}^{i,j} = E[\boldsymbol{x}^{i,j}\boldsymbol{x}^{i,j^H}]$ it is possible to define also for bidimensional array, the spatially smoothed covariance matrix as a weighted sum of the covariance matrix of each subplanar array. In order to obtain estimates of the DOA for the planar array, a signal subspace algorithm, such as MUSIC or Minimum Norm, can employ this matrix. The main problem is to determine the covariance matrix, which can be only estimated.

3. Simulation Results

The simulations consider a 450 Kbit/s QPSK 900 MHz and 1800 MHz signal in an integrated Terrestrial and Satellite environment respectively for linear and planar arrays; the characteristics of channel attenuation, delay and Doppler spread follow the GSM Rec.05.05. version 3.7.0 for the terrestrial case (urban area and hilly terrain), and the requirements of DLR Final Report for the satellite case [4]. The adopted channel model is a discrete system with two paths, the direct one and the reflected one for the satellite case, and three paths with two reflected signals for terrestrial environment.

More specifically, Rayleigh Fading Channel for the terrestrial case is considered, and Rayleigh-Rice distribution in a shadowed and unshadowed environment for the desired signal and Lognormal Fading Channel for the reflected signal has been adopted for the satellite case.

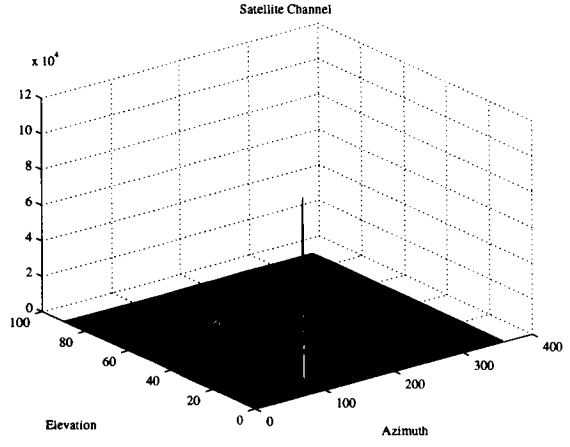
In presence of multipath, the performance of standard DOA algorithms is often unsatisfactory; therefore the Spatial Smoothing Method has been used.

For direct path in satellite transmission, it is supposed the absence of Doppler shift, due to the deterministic motion of the satellite platform. In addition, the Doppler frequency shift for the lognormal reflected path has not been considered. On the contrary, in the terrestrial case, the continuous and unpredictable variation of the mobile user must be considered.

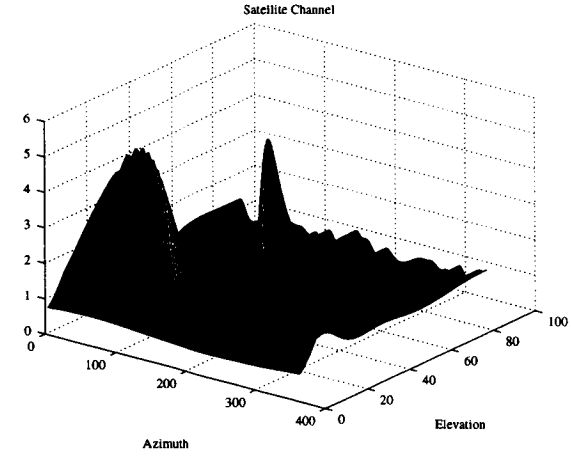
In the following figures , a $N \times N$ equally spaced planar array system, with $N = 12$ elements, and seven subplanar arrays of 6×6 elements are considered.

The first three graphics (Fig.1(a), Fig.1(b) and Fig.2) show the Rayleigh-Rice Fading satellite channel; it has been considered an E_b/N_0 of 30 dB, with two incident signals incoming from the bidimensional directions $[(\theta_i, \varphi_i)]^2_{i=1} = [(10°, 100°); (40°, 200°)]$.

All of these graphics, highlight the capacity of the classic method to estimate the direct signal only, not giving any information about the direction of arrival estimation of the second reflected path. On the contrary, the smoothed method may discriminate all the incident signals, without to maintain however



(a)



(b)

Figure 1: *Rayleigh Fading Satellite Channel: Minimum Norm (a) classic, (b) Smoothed.*

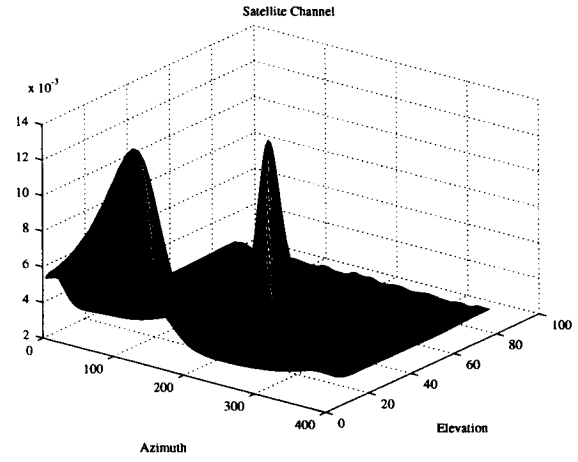


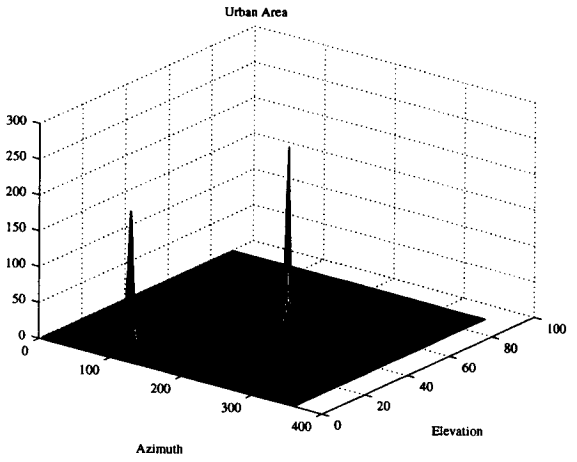
Figure 2: *Rice Fading Satellite Channel: MUSIC Smoothed*

the information about the difference between the main power of the desired signal and the reflected path.

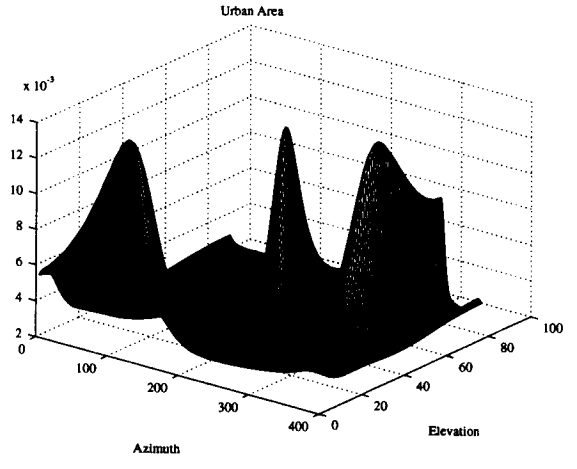
The last set of simulations for the planar arrays, consider the terrestrial environment. The Figg.3(a) 3(b), regards an urban area with MUSIC and Smoothed MUSIC algorithms; the Figg.4(a) 4(b) analyze the propagation signal on hilly terrain with Minimum Norm and Smoothed Minimum Norm methods. Now, three incident coherent waves, impinging from the directions of $[(\theta_i, \varphi_i)]_{i=1}^{3} = [(10°, 100°); (60°, 200°); (80°, 300°)]$ are considered, with an E_b/N_0 value of 10 dB and 20 dB respectively. In this case, the results show the capacity of the classical algorithms to determine the first two significant signals, for the urban area, in which these signals are similarly attenuated by the channel, and the desired high powered signal only for the hilly terrain. The smoothed methods, again, result in a total discrimination of the incident paths.

The two algorithms (MUSIC and Minimum Norm) present the same behaviour in the angular estimation, highlighting higher variance and eigen-spectrum values of Minimum Norm with respect to MUSIC. As a consequence of the results, a shared use of the classic and smoothed techniques can provide the exact direction for all incident paths and the maximum power of the desired signal.

The DOA direction estimate is naturally the input to adaptive beam forming techniques, which tend to compensate the multipath impairment, attenuating the interference signals. The Fig.5 shows the performance of Howells-Applebaum smoothed algorithm [5] in the case of urban area, for a linear array of $N = 16$ elements, and a sublinear array of size $m = 4$, with multipath signals, impinging the array from the direction $[\theta_1, \theta_2] = [-20°, 30°]$, and uncorrelated co-channel interference, coming from $\theta_3 = -50°$.

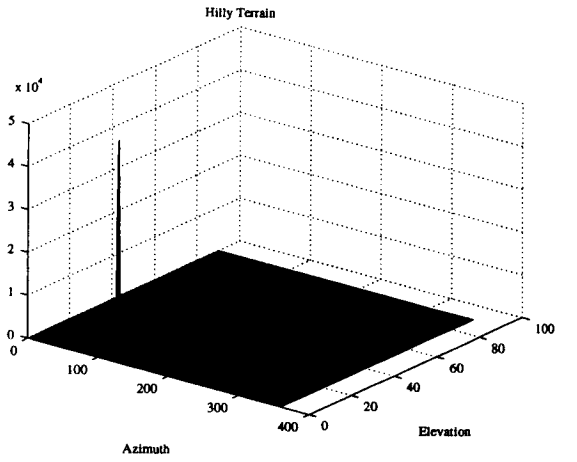


(a)

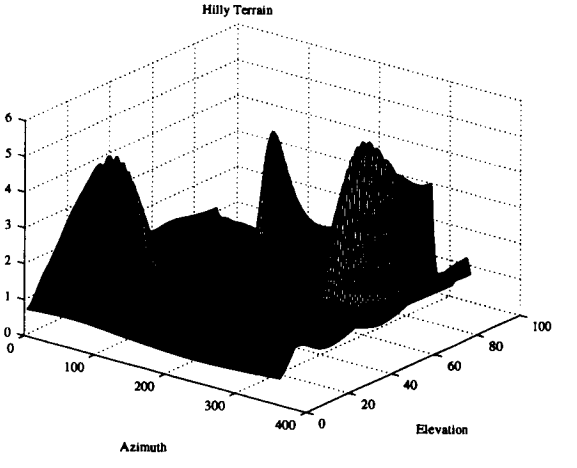


(b)

Figure 3: *Urban Area: MUSIC classic (a) and Smoothed (b).*



(a)



(b)

Figure 4: *Hilly Terrain: Minimum Norm classic (a) and Smoothed (b).*

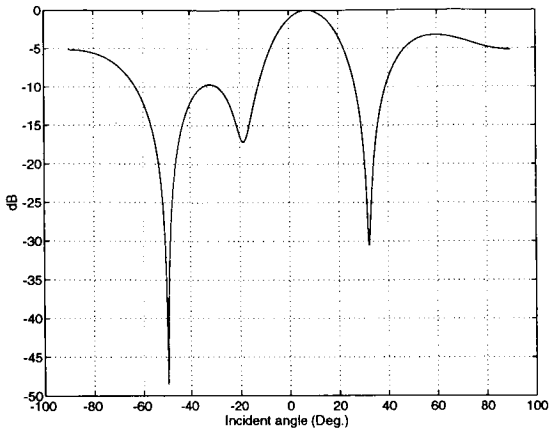


Figure 5: *Urban Area: Howells-Applebaum Smoothed.*

The desired signal comes from the broad side direction, and in the other incident angles, the beam forming technique inserts nulls in the antenna pattern.

Conclusions

In this paper we have investigated the performance of adaptive arrays in an integrated terrestrial satellite system. The proposed planar array is suitable to be implemented at the satellite ground stations and/or cellular base stations to adaptively track the satellite (also in the LEO orbit) and/or communicate to the mobile user. Therefore, the received multipath signals arrive at the base station within a given beam width, and the results show the capability of adaptive array , using DOA estimation plus beam forming techniques, to combat fading and suppress interference signals. Considering the carrier frequency of $1.8GHz$ and an inter element distance of $\frac{\lambda}{2}$, the size of planar array is about $1.5m. \times 1.5m.$, which is tolerable dimension for cellular systems.
As show by the results, the smoothed version of these algorithms resolve the true directions of all the incident signals without discriminating between the direct power path and the replica signals power. Furthermore, to estimate the maximum power path it is instead necessary to use the standard DOA algorithms. Once the DOA estimation is obtained, the beam forming technique attenuate all the type of interferences, improving the acquisition of the received signal with an enhancement of the signal carrying more information.

References

[1] **R.T. Compton Jr.** *Adaptive antennas.* Prentice-Hall, 1988.

[2] **B. P. Ng and M. H. Er and C. Kot.** *A MUSIC approach for estimation of directions of arrival of multiple narrowband sources.* Signal Processing, 40, 1994.

[3] **S. Haykin.** *Adaptive filter theory.* Prentice-Hall International Editions, 1991.

[4] **T.-J. Shan and M. Wax and T. Kailath.** *On spatial smoothing for direction-of-arrival estimation of coherent signals.* IEEE Transections on Acoustics, Speech, and Signal Processing, August 1985.

[5] **E. Lutz and A. Jahn.** *Propagation data and channel model for lms systems.* Technical report, DLR - German aerospace research establishment Institute for communications technology, January 1995.

A Ku-Band DBS Airborne Antenna System for Commercial Airlines

W. Horton Prather
EMS Technologies, Inc.
660 Engineering Drive Norcross, GA 30092 USA
Phone: 770-263-9200 FAX: 770-263-9200
email: prather_h@elmg.com

ABSTRACT

Live video reception from direct broadcast system (DBS) satellites is being planned for future inflight entertainment systems. Some of the most challenging system requirements are those of the satellite tracking antenna which must provide high G/T required for reception of dual polarization, wideband compressed video signals, while also providing a low profile, accurate tracking, and low production cost. This paper describes the recent development of such a system by EMS Technologies, Inc. for use in commercial airline entertainment systems.

INTRODUCTION

The Ku-Band DBS Airborne Antenna System can provide reception of live television signals or broadband data onboard an aircraft in flight. The dual polarization antenna has been successfully demonstrated using signals from both DirecTV® and Echostar satellites.

The system consists of the Receive Antenna Subsystem (RAS), Antenna Positioner Subsystem (APS), and Antenna Control Subsystem (ACS). A block diagram of the system is shown in Figure 1.

The RAS is a dual polarized, waveguide slot array which receives both left hand and right hand circularly polarized signals in the 12.2 to 12.7 GHz frequency range. The antenna provides a G/T of greater than 10 dB/°K with low axial ratio. The dual polarized waveguide slot array uses EMS' patented antenna construction to achieve a high performance producible design.

The APS is an elevation over azimuth mechanical positioner subsystem which provides accurate pointing and tracking of the RAS beam peak with respect to the DBS satellites during aircraft flight. The APS/RAS comprises the Main Antenna Assembly (MAA) which mounts to the outside of the aircraft, requiring only mounting bolt holes and a 1 in. diameter penetration for cables and coax. This antenna assembly fits within a radome which is only about 6 inches high and 36 inches in maximum width.

The ACS provides the pointing and tracking commands necessary for the APS. It operates autonomously with no inputs required from the aircraft navigational system. A GPS receiver internal to the ACS provides information on the location and velocity vector of the aircraft. The antenna for the GPS receiver is mounted under the radome, just aft of the MAA. Quartz inertial rate sensors are used to provide information on the attitude in space of the aircraft by continuously measuring its pitch, roll and yaw. Using this information, the ACS continuously calculates the pointing vector to the satellite constellation. Data between the ACS and APS is interchanged via a RS-422 data bus. Although not required for operation in the aircraft, an RS-485 data port is also provided to allow commands to, or receipt of built-in-test data from, the ACS. The ACS is mounted within the fuselage of the aircraft.

Figure 2. Shows a photograph of the system hardware. The Main Antenna Assembly (MAA), consisting of the Receive Antenna Subsystem (RAS) and Antenna Positioner Subsystem (APS) is shown sitting beside the Antenna Control Subsystem (ACS). The small box sitting in front is the GPS antenna assembly which is mounted under the radome with the MAA. The cables from the MAA containing power, data and rf signals extend to a connector under the radome which penetrates the fuselage.

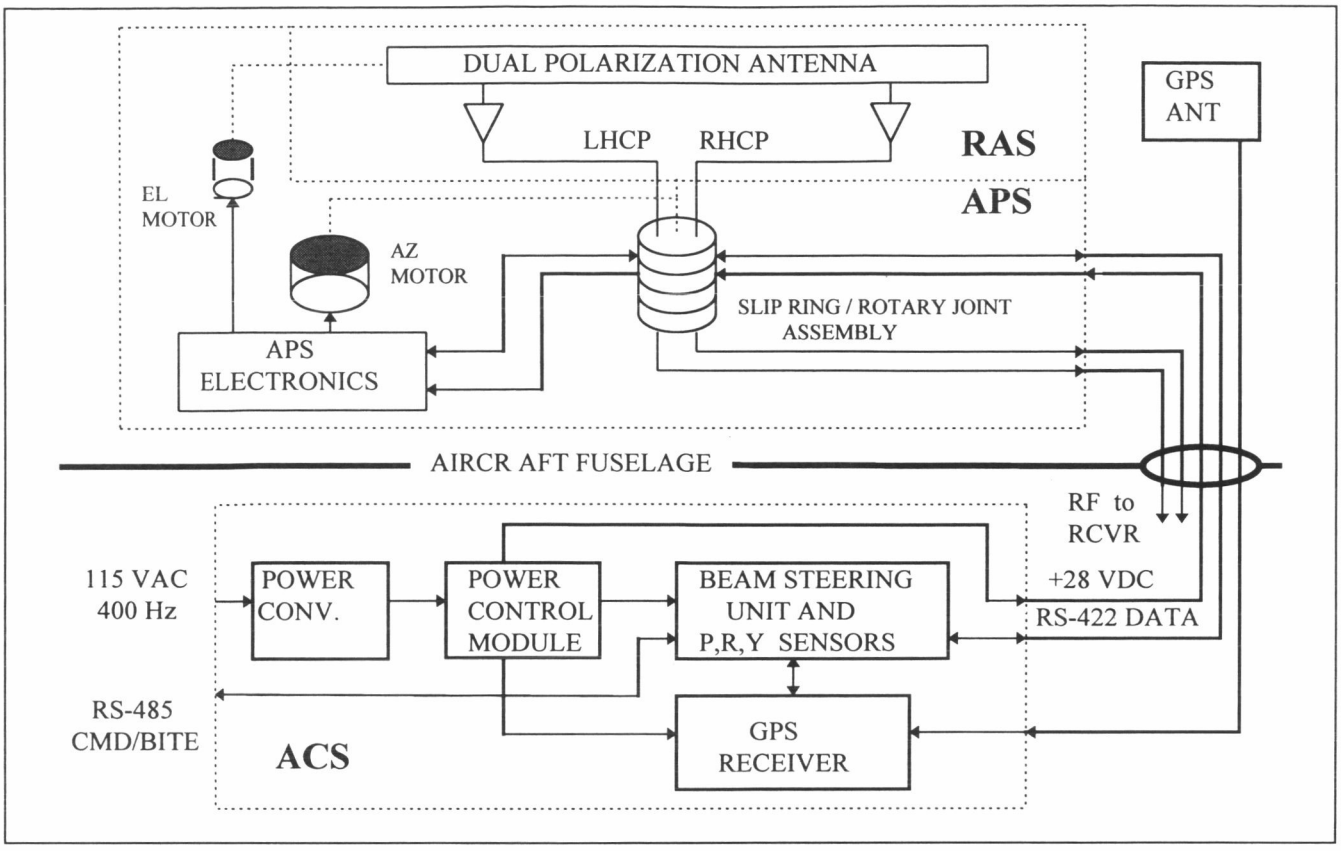


FIGURE 1. BLOCK DIAGRAM, Ku-BAND DBS AIRBORNE ANTENNA SYSTEM

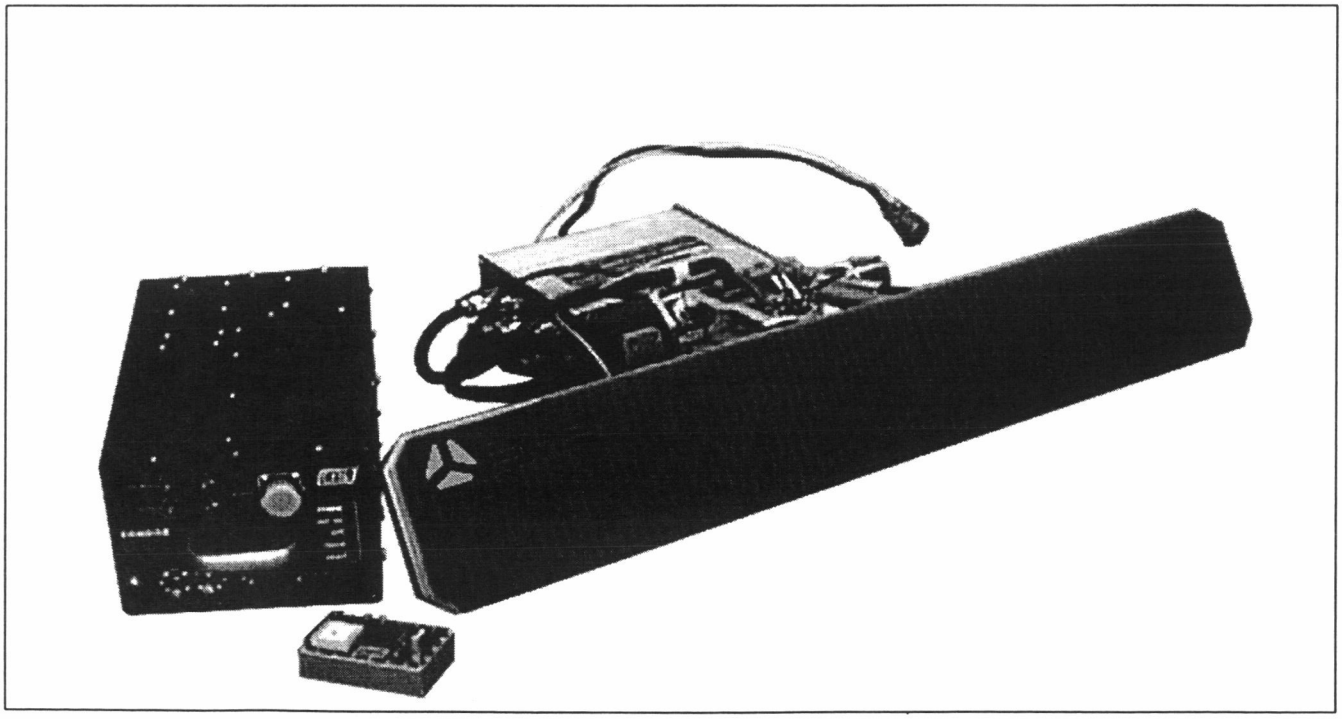


FIGURE 2. PHOTOGRAPH, Ku-BAND DBS AIRBORNE ANTENNA SYSTEM

Active Integrated Antenna Arrays for K/Ka-Band Portable Communication Terminals

D. Roscoe, A. Ittipiboon, P. Lafleur, and M. Cuhaci
Directorate of Antennas and Integrated Electronics
Communications Research Centre
3701 Carling Avenue
Ottawa, Ontario, Canada
K2H 8S2

INTRODUCTION

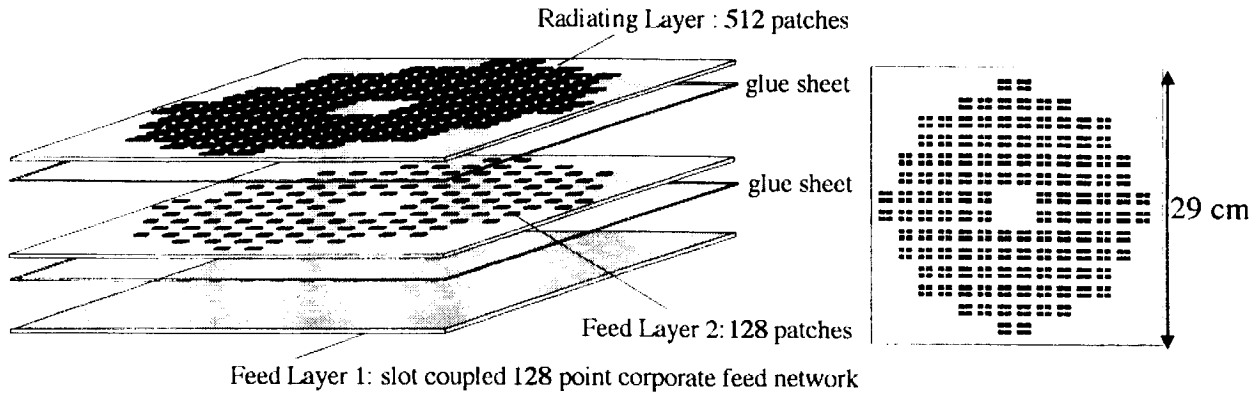
A variety of antenna candidates are being proposed for application to the portable personal communication terminal operating at 20 GHz (receive) / 30 GHz (transmit). These architectures range from integrated arrays to deployable reflector systems. Presently, the portable terminal developed within the Communications Research Centre (CRC) utilizes a deployable reflector system. Using a 0.5W or 1.0W power amplifier, an EIRP of 39.3 dBW or 42.3 dBW, respectively, was achieved. Although the deployable reflector is viable, alternate architectures have been investigated to yield a more efficient and less obtrusive portable antenna. The integrated antenna array approach permits the integration of low to medium power amplifiers which can be distributed within an array. Spatial power combining is used to achieve the relatively high EIRP required. As compared to the single power amplifier used in conjunction with the reflector, the spatial power combining approach is more efficient as there are less combining losses. In addition, an integrated antenna array can be designed to be low profile and compact.

Two low profile active integrated antenna arrays utilizing different technologies have been developed to determine the feasibility of each technology as related to the application at hand. Each antenna array incorporates novel radiating elements detailed in [1]. An antenna array utilizing multilayer microstrip technology was designed for the receive band of 20 GHz. The resulting structure is planar and has low noise amplifiers distributed throughout the array. The other architecture, developed for the 30 GHz transmit band, implements cavity antenna technology where a thick ground plane is available for device integration and heat dissipation. The resulting structure is low profile and incorporates 18 medium power amplifiers (device P_{out} ~300mW).

This paper presents two array architectures, implementing different antenna technologies, which could be used in a portable communications terminal. The paper will emphasize the transmit array (30 GHz).

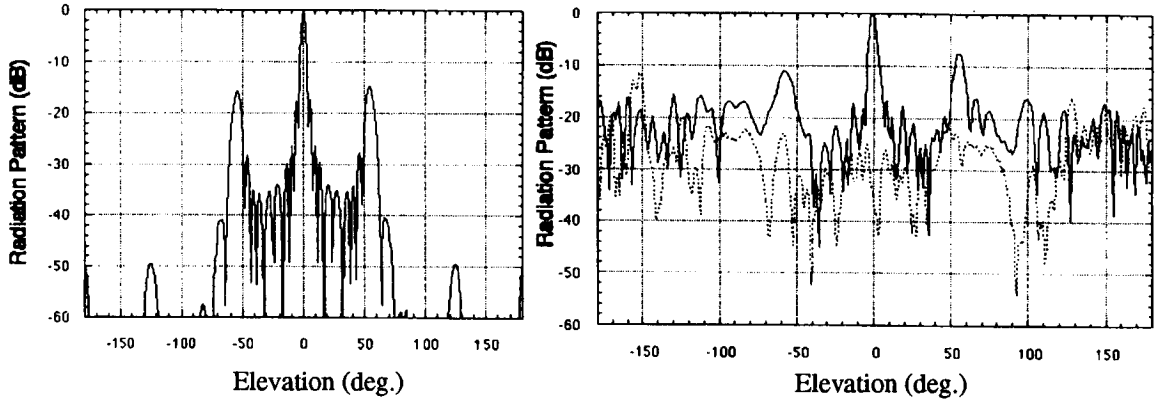
SECTION I : 20 GHz Receive Array

The receive array architecture is based upon a multilayer microstrip patch antenna element [1]. The Quad EMC patch is a multilayer microstrip patch configuration developed to reduce the number of feed points required within an array. A 4:1 reduction has been achieved where the developed array has 512 patches on the top surface, but requires only 128 feed points. The reduction simplifies and reduces the feed network, and hence reduces the feed network losses and provides the necessary real-estate for device integration. The complete array architecture is presented in figure 1. The array is a three layer structure where energy is coupled through slots in the common ground plane between the feed network (feed layer 1) and the layer of single patches (feed layer 2). These patches then distribute the energy to groups of four patches located directly above the single patches of feed layer 2 (electromagnetic coupling is utilized throughout). Design details of the array architecture can be found in [2].



a) Exploded view b) top view

Figure 1 : Multilayer microstrip patch array architecture.



a) Simulated b) Measured

Figure 2: Simulated and measured radiated patterns (H-plane) of passive multilayer microstrip array.

A corporate feed network was implemented in this particular design. LNAs were integrated for every cluster of 8 quad EMC microstrip patch elements (i.e. 32 radiating elements). For optimal G/T performance, an LNA located at the feed point of each quad EMC element would be required. However, this is both costly and impractical. It has been determined that for this particular an element-to-device ratio of 8:1 provides the optimum tradeoff between performance and cost. A printed bias network has been integrated on the same surface as the microstrip feed network.

The array architecture itself was designed so that the radiated pattern would be rotationally symmetric, having a half power beamwidth (HPBW) of 3°. The sidelobe level (SLL) specification was ≤ -15 dB. The array architecture presented in figure 1 satisfies the specifications where the grating lobe level was the limiting factor (figure 2a). A passive array prototype was fabricated and tested. The measured radiated pattern is presented in figure 2b). The recorded cross-pol. on boresight is ~-40dB. It can be seen that the measured radiated pattern agrees with the predicted with the exception of the SLL. The fabricated array has grating lobes located ±60° > -15dB.

SECTION II : 30 GHz Transmit Array

Antenna Element

The transmit array architecture developed is based upon a cavity type microstrip element as presented in [1]. This element is amenable to device integration since the thick metallic ground plane incorporated within its design allows for easy mounting. In addition, this design functions as a heat sink for the integrated power amplifiers. The element is depicted in figure 3. The cavity is formed from a thick ground plane and a microstrip element is used to feed the cavity. Coupling from the feed line to the microstrip patch occurs through a slot in the thin ground plane within the cavity, while device integration can be achieved using the thick ground plane outside the cavity. The patch within the cavity is the feed for the cavity. Cavity radiators employing rectangular or circular microstrip patch feed elements have been developed. The effective aperture area of this antenna is outlined by the circumference of the cavity. The cavity diameter is not limited to a specific size, but is typically between $1.5\lambda_0$ and $2.0\lambda_0$. For this particular design, a diameter of $1.70\lambda_0$ was selected to obtain an optimum element pattern. The achievable gain of the cavity element is a function of the cavity diameter, but is typically 13-15 dBi. The radome is utilized within the element design to produce a uniform distribution across the cavity aperture and functions as a matching network to free space. The overall height of this antenna element is approximately $\lambda_0/2$. At 30 GHz, this antenna element is approximately .5cm in height and 1.7cm in diameter.

The cavity antenna element depicted in figure 3 was fabricated and measured. The measured return loss over the band (29.0 - 30.0 GHz) was $\leq$ -10 dB and the measured radiated pattern at 29.5 GHz is presented in figure 4. The measured peak gain (13.2 dBi, feed losses included) varied by 2 dB over the band of interest while the cross-pol. component was $\leq$ 25dB below the co-pol. component.

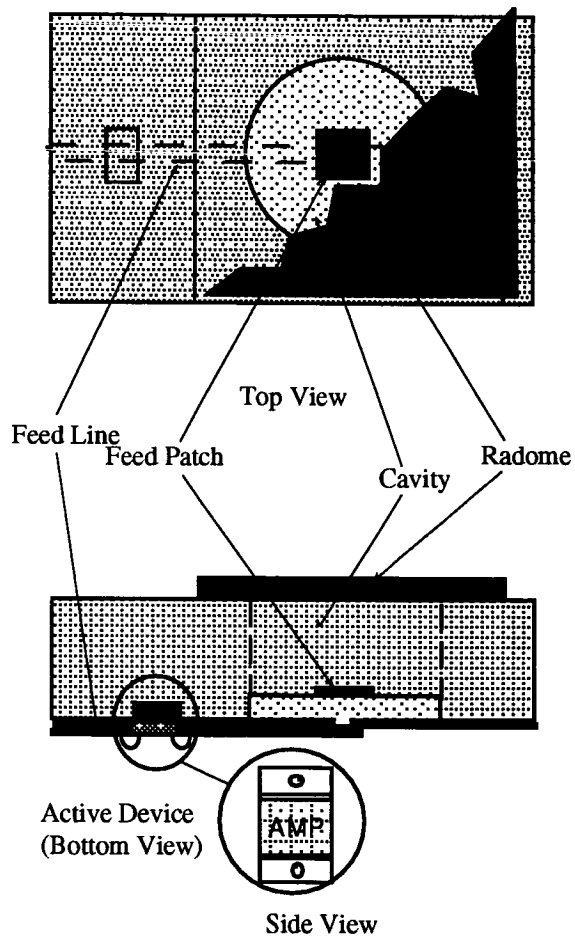


Figure 3: Cavity antenna element amenable to device integration.

Array Architecture

To improve the efficiency of the transmit antenna, an array design with distributed power amplifiers was pursued. This approach utilizes spatial power combining techniques to minimize power combining losses. The most efficient architecture would incorporate a power amplifiers located at the input of each cavity element. However, as with the receive array, this is too costly. It was determined that an array architecture utilizing 64 cavity elements with sixteen 1/2 Watt power amplifiers would generate the required EIRP

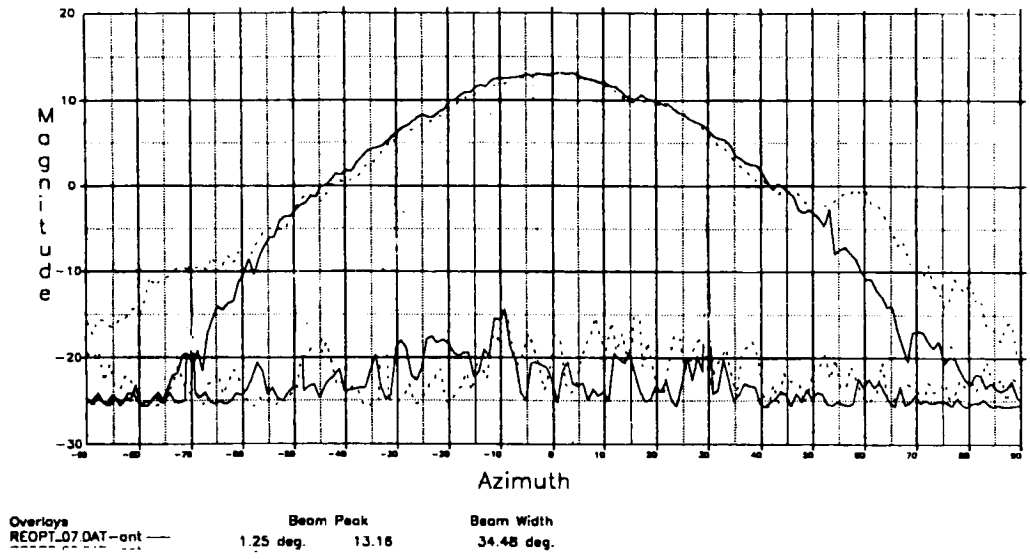


Figure 4 : Measured far field pattern of cavity element, f = 29.5 GHz, E-plane ———, H-plane - - - - - -

of 40dBW. The 64 cavity elements are arranged in 4 rings of radii 2 λ_0, 4.5 λ_0, 7 λ_0, and 12.3 λ_0, containing 4, 8, 16, and 36 cavity elements, respectively. The array architecture is presented in figure 5. A corporate feed network was designed for the array geometry such that 16 packaged amplifiers could integrated within the feed network. Each amplifier feeds 4 cavity elements. Two additional pre-amplifiers were also integrated close to the array input for a total of 18 amplifiers. The feed network was fabricated on Rogers 5880, (t = .010", ε_r = 2.2) and Wilkinson power dividers were used at each split. The total loss between the output of each amplifier and input of each cavity radiator was determined to be 2.2 dB. A passive array was fabricated and measured to verify the array architecture and feed network. A measured pattern is presented in figure 6. As can be seen, the measured radiated pattern has sidelobe levels greater than expected, however, the half-power beamwidth and overall array pattern characteristics are as expected.

The power amplifiers to be integrated are 1/2 Watt devices purchased from Northrop-Grumman. The devices were packaged and characterized to determine gain, phase, and output power. The variation between the 18 amplifiers were ±2.9dB gain, ±20.6deg. phase, and ±1.1dBm output power. Based upon the characterized parameters, the array pattern was re-calculated and is presented in figure 7. As can be seen, the variation between integrated devices should not impact the array pattern. The initial stage of device integration is presented in figure 5b). The power supply for the 18 amplifiers was designed to be a single layer located directly above the feed network layer. The bias leads shown in figure 5b) are connected vertically to the power supply board located above the devices.

The active array has been fabricated and the measured results are to be presented at the conference.

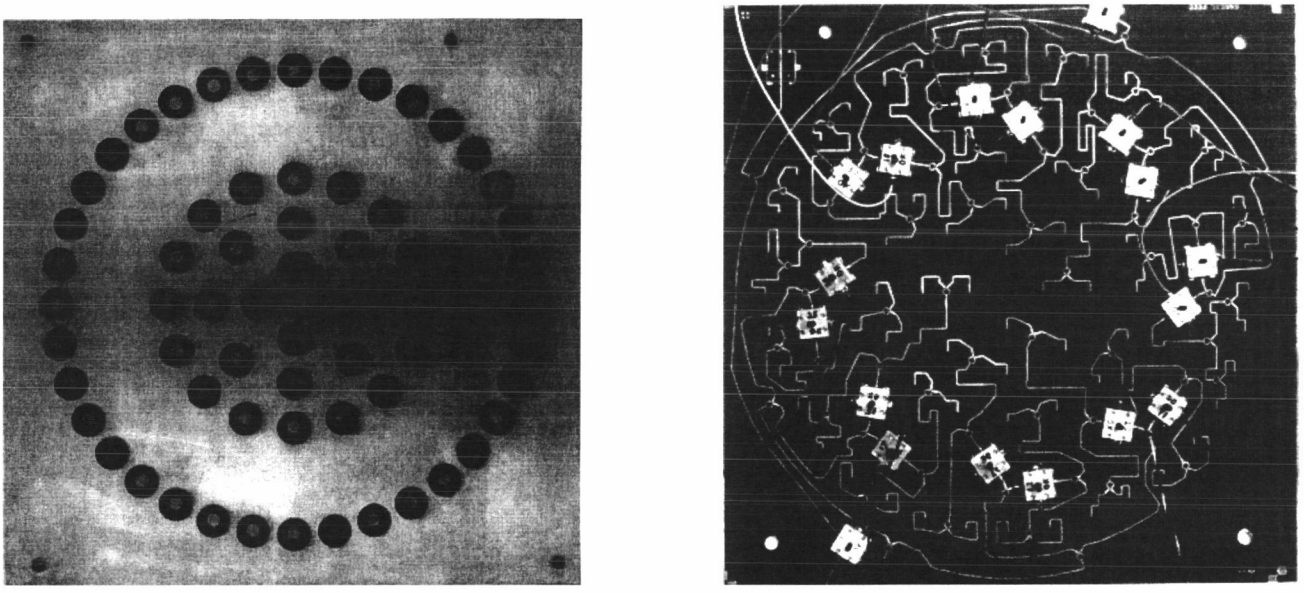
Summary

Two array architectures utilizing various antenna technologies have been presented. A multilayer microstrip architecture at 20 GHz and a microstrip-cavity architecture at 30 GHz have been fabricated and the measured results of the passive arrays have been presented. The measured results of the 30 GHz active array are to be presented at the conference.

REFERENCES

[1] D. Roscoe, L. Shafai, A. Ittipiboon, M. Cuhaci, and H. Moheb, "*Novel Low Profile Antenna Candidates for EHF Portable Terminals*", International Mobile Satellite Conference (IMSC) 1995, Ottawa, Canada, pp. 318-323, June 6-8, 1995.

[2] P. Lafleur, D. J. Roscoe, and M. Cuhaci, "*A 20 GHz Active Integrated Multilayer Microstrip Patch Array for Portable Communication Terminals*", accepted 1997 IEEE AP-S International Symposium and URSI Radio Science Meeting.



a) Top view of 64-element cavity array

b) Device integration within corporate feed network

Figure 5 : Fabricated array architecture.

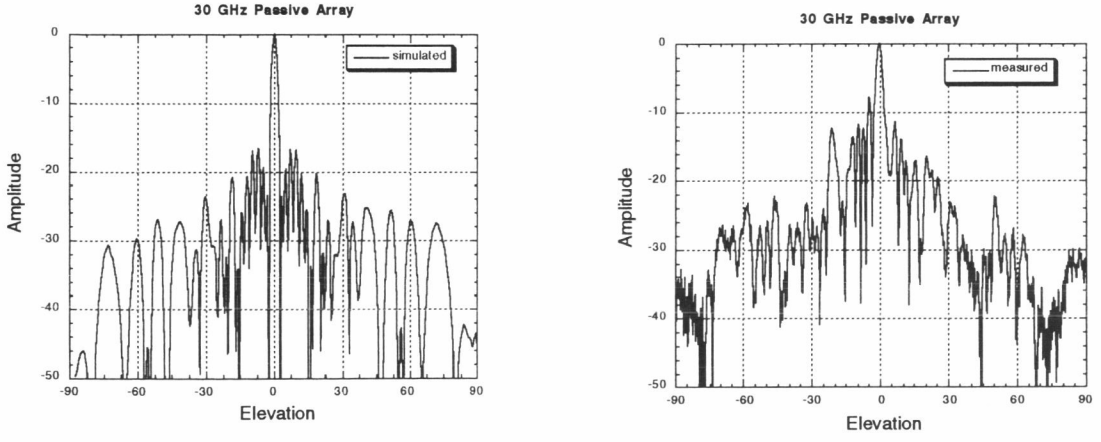


Figure 6 : Simulated and measured radiated pattern (H-plane) of passive array, f = 29.5 GHz.

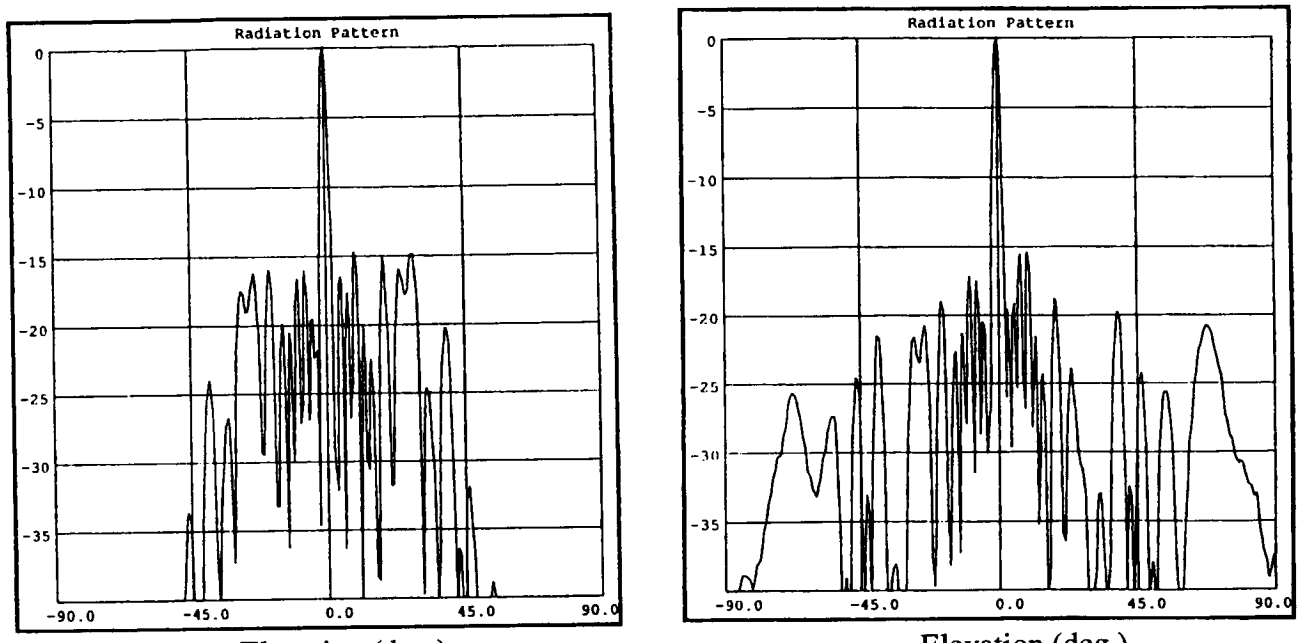


Figure 7 : Simulated array pattern incorporating measured device characteristics.

Dielectric Resonator Antennas for Mobile Satellite Applications

Aldo Petosa, David Roscoe, Apisak Ittipiboon, Michel Cuhaci
Communications Research Centre
3701 Carling Ave.
Ottawa, Ontario, Canada
K2H 8S2
Phone: 613-991-9352 Fax: 613-990-8369
email: aldo.petosa@crc.doc.ca

ABSTRACT

This paper presents various dielectric resonator antennas developed at the Communications Research Centre which have potential applications in mobile satellite communication terminals.

INTRODUCTION

There are many terminal systems being developed for mobile satellite communication applications in which antenna technology forms an integral part. Examples include vehicle-mounted, hand-held, aeronautical, and briefcase terminals. Antenna requirements for each system vary in terms of gain (omnidirectional to highly directive), polarization (linear or circular), frequency (L-Band to Ka-Band), and operational bandwidth (between 1% and 10%). To date, various antenna technologies such as helical, microstrip, and whip antennas have been used in these systems. Helical antennas are sometimes used for high-gain applications requiring circular polarization (CP). These antennas are relatively inexpensive and have wide bandwidth performance. One disadvantage of helical antennas is their size. The length of the helix determines the gain, and thus for high-gain applications it is difficult to maintain a low profile without going to a large array of low-gain helical antenna elements. Microstrip antenna technology is a suitable candidate for some applications in that it is a low cost, low profile technology. However, microstrip antennas suffer from inherent bandwidth limitations and their physical size becomes large at low frequencies. Whip antennas are often used in hand-held sets. Although they are inexpensive and mechanically simple, they are easily susceptible to damage and are not low profile.

Dielectric resonator antennas (DRAs) are being investigated as an alternative technology to the conventional antennas mentioned above. In the following sections, the DRA will be introduced, and several examples will be presented to demonstrate its capability for various mobile satellite terminal applications.

DIELECTRIC RESONATOR ANTENNA OVERVIEW

Dielectric resonator antennas are resonant radiators fabricated from low-loss dielectric materials which are formed into any one of a number of shapes including rectangular, circular disk, ring, cross, triangular, or hemispherical. The resonant frequency of the DRA is a function of its size, aspect ratio, and the permittivity of the dielectric material. DRAs offer increased design flexibility to the antenna engineer since many parameters, such as the dielectric constant and the physical dimensions, can be adjusted to obtain optimum performance for a given application. Coupling to the DRAs can be achieved using a number of techniques including probe coupling, slot coupling, and proximity coupling from microstrip lines. DRAs have been found to offer several advantages over microstrip patch antennas, such as wide impedance bandwidth, low loss, and high radiation efficiency. A good

introductory overview of DRAs is given in [1].

Research has been carried out at the Communications Research Centre (CRC) over the last several years on dielectric resonator antennas (DRAs) of various shapes, sizes, and feeding configurations. This research has been focused on the characterization of individual elements at various frequencies from L-Band to Q-Band. Considerable work has also been invested in designing various linear and planar arrays for moderate to high gain applications [2-11]. The next section presents some of these DRA elements and arrays which may be suitable candidates for mobile satellite applications.

DRA CONFIGURATIONS

This paper presents a small sampling of the DRAs being investigated at the CRC. The configurations were chosen to demonstrate the versatility of DRAs and includes single elements, linear arrays, and planar arrays, operating in various frequency bands from L-Band to Ka-Band. The five DRA configurations selected for this paper are:

- a compact DRA for hand-held applications
- a linear DRA array for low-elevation coverage
- a sub-array of sequentially rotated cross DRAs, for wideband CP operation
- a DRA-fed cavity antenna for high gain CP operation
- a polarization agile DRA for polarization diversity applications

COMPACT DRAs

Work has been carried out to explore the potential advantages of using DRAs for hand-held sets operating at L-Band for applications such as GPS receivers. Antenna requirements for such applications include a compact design and omni-directional coverage. The DRA can be made compact by using a material of high permittivity. The disadvantage is the accompanying decrease in operational bandwidth. An alternative approach involves the placement of a short-circuit wall, as shown in Figure 1 [12]. The wall allows for a theoretical reduction to half the size for a given frequency for certain modes. In practice, the size reduction is not half, since the dimensions of the wall are kept to a minimum and some fringing still occurs. Table 1 shows the reduction in resonant frequency when a short circuit is placed against existing DRAs. Up to a 37% reduction in frequency was observed. Alternatively, for a given frequency, the size of the DRA could be similarly reduced by the use of a shorting wall. (The accompanying reduction in bandwidth can be offset by introducing an air gap between the DRA and ground plane [12]).

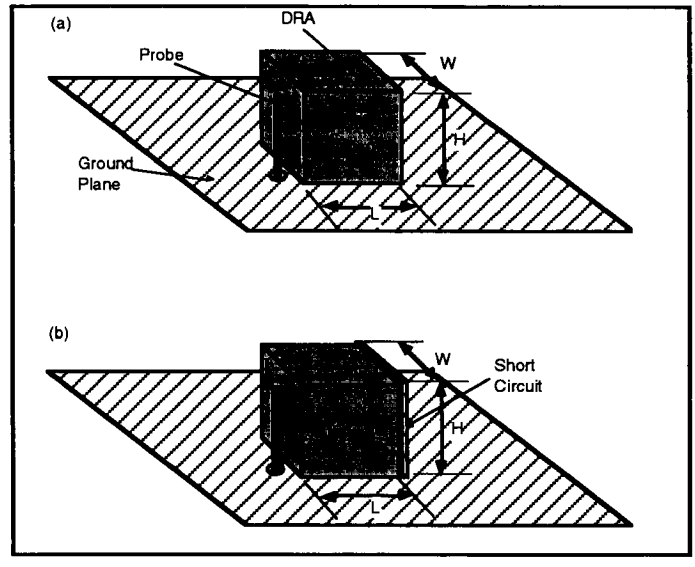


Figure 1. Compact DRA element using a shorting wall.

DRA	Frequency (GHz)	Bandwidth (%)
1(a)	1.98	9.5
1(b)	1.24	5.6
2(a)	1.76	3.2
2(b)	1.34	1.5

Table 1. Reduction in resonant frequency using a shorting wall.

DRA 1: L=2.75 cm, W=2.75 cm, H=2.85 cm, $\varepsilon_r = 12$

DRA 2: L=2.50 cm, W=2.50 cm, H=1.60 cm, $\varepsilon_r = 25$

(a) denotes DRA without shorting wall
(b) denotes DRA with shorting wall

LINEAR DRA ARRAY FOR LOW-ELEVATION COVERAGE

For applications such as vehicle-mounted terminals, low-elevation coverage is usually required so that the antenna peak radiation is aligned with the satellite. It is also important for the antenna to be low profile for aerodynamic and aesthetic reasons. The antenna shown in Figure 2 is a linear array of DRAs in which one element is driven by a feed probe while the other elements are coupled electromagnetically. The effect of these parasitic elements is not only to increase the gain, but to scan the peak of the beam closer to the horizon. Figure 3 shows a comparison of the radiation pattern of a single DRA with that of the linear array of Figure 2. The array demonstrates an increase in gain of about 7 dB at 30 degrees above the horizon. Linear arrays of parasitically coupled DRAs have been demonstrated both at L-Band and K-Band and offer potential solutions for low-elevation coverage applications [12-13].

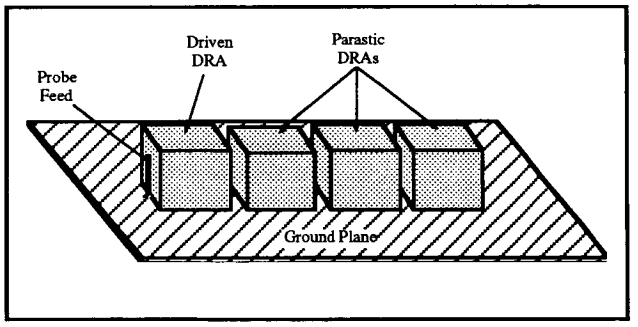


Figure 2. Linear DRA array.

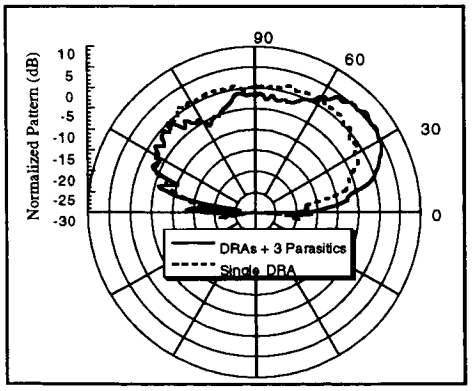


Figure 3. Comparison of H-plane patterns of single DRA with linear DRA array at K-band.

ARRAY OF SEQUENTIALLY ROTATED DRA CROSS ANTENNAS

For many applications, the antennas must radiate a CP pattern. Conventional antennas such as helices or microstrip antennas can be used, however, helices are not low profile and microstrip antennas have limited bandwidth. A potential alternative is the cross DRA [14, 15]. A single cross DRA has been shown to generate a CP pattern from a single feed point with a 4% bandwidth over which the axial ratio is less than 3 dB. This is superior to the 1% bandwidth typically achieved by single-point-fed microstrip patches. The CP bandwidth of the cross DRA can be further improved by forming a sub-array of four crosses as shown in Figure 4 where sequential rotation is used in the feed design [16]. The measured axial ratio versus frequency is shown in Figure 5 while a typical pattern is shown in Figure 6. The CP bandwidth is extended from 4% to 16% while maintaining a compact low-profile structure.

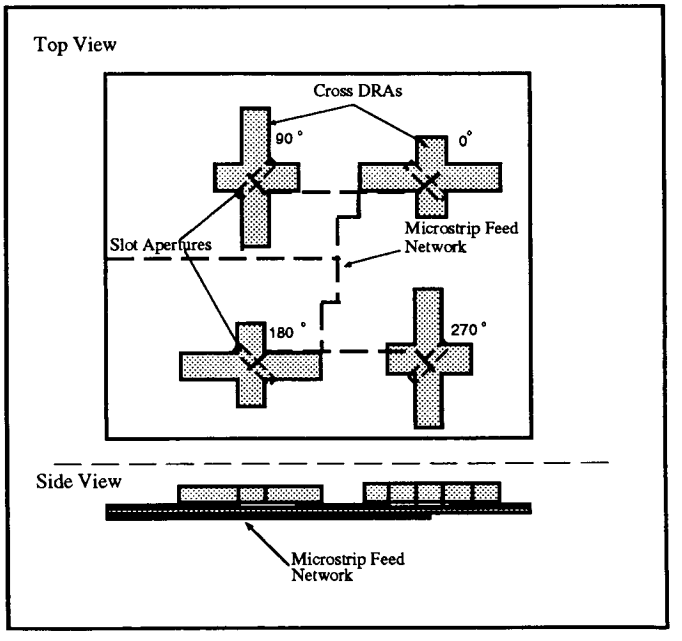


Figure 4. Array of cross DRAs fed by sequential rotation.

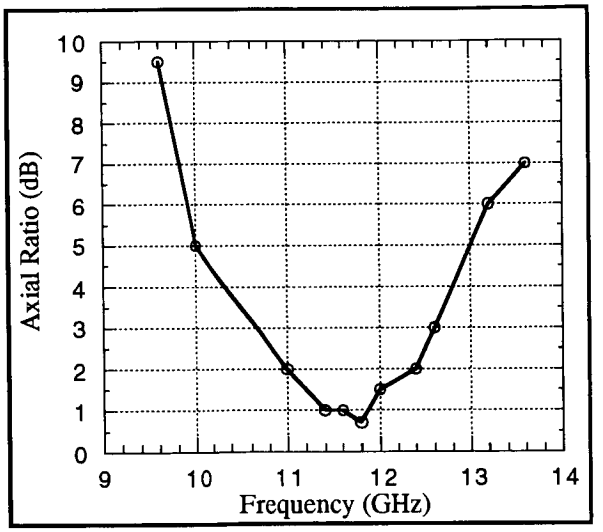


Figure 5. Axial ratio versus frequency of the cross DRA array.

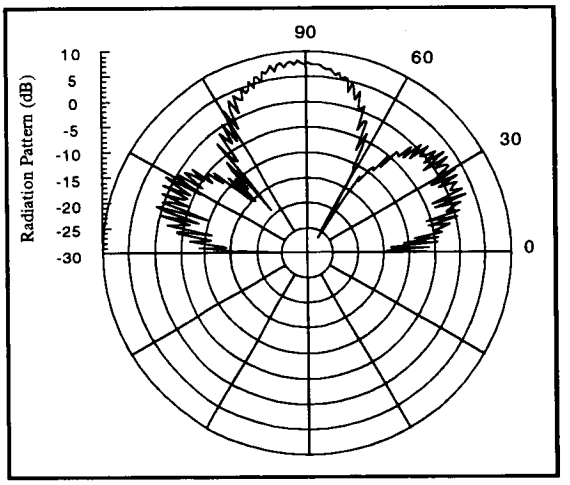


Figure 6. Radiation pattern at 11.4 GHz of the cross DRA array.

DRA-FED CAVITY ANTENNA

One of the challenges of designing printed antennas at higher frequencies is that of increased copper and surface wave losses. This will lower the radiation efficiency of the antenna and will thus impact on the system performance in terms of higher noise temperature or lower EIRP. DRAs do not suffer from copper or surface wave losses and are thus much more efficient at higher frequencies. A medium-gain cavity element has been developed for K- and Ka-bands which uses a cross DRA as the feed element as shown in Figure 7 [17, 18]. The cavity, which is on the order of two wavelengths in diameter is used to reduce the feed losses which would result in arraying together lower gain elements. The cross DRA is used to produce a CP radiation pattern. A typical radiation pattern is shown in Figure 8. Cross-fed cavities have been designed both at K and Ka Band. The gain of these antennas is a function of the cavity diameter and gains of over 13 dB_{ic} with a 3 dB axial ratio bandwidth of 4% have been measured for cavities with a two wavelength diameter. These cavity antennas are currently being used as elements in a large planar array configuration for a high gain portable briefcase terminal [19].

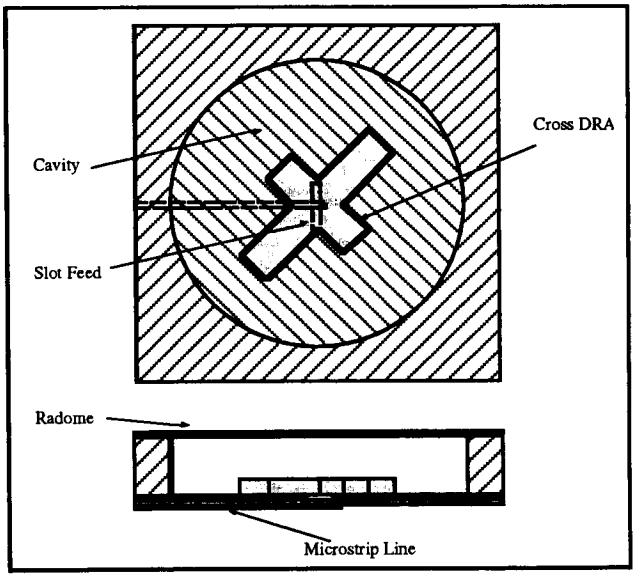


Figure 7. Cross DRA-fed cavity antenna.

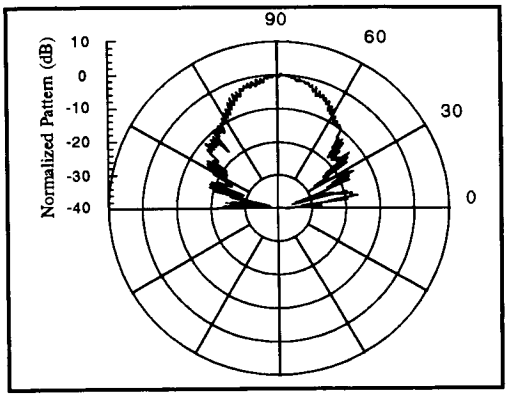


Figure 8. Normalized radiation pattern of the cross DRA-fed cavity at 29.8 GHz.

POLARIZATION AGILE DRA

Some preliminary work has been carried out on using low-loss microwave ferrite materials to fabricate DRAs. These antennas, called ferrite resonator antennas (FRAs), have the advantage that their permeability can be varied by the application of a dc magnetic biasing field which allows for the control of certain radiation characteristics [20-23]. For the circular FRA in Figure 9, the application of a magnetic bias field has resulted in an antenna with polarization agility. When no bias field is applied, the FRA radiates like a simple cylindrical DRA, producing a linearly polarized pattern. However, when a dc magnetic field is applied, the tensor permeability of the ferrite is altered such that two orthogonal modes of the antenna are generated which are in phase quadrature, thus producing a circular polarized pattern. The patterns at 7.2 GHz without and with magnetic bias are shown in Figure 9. A 3 dB boresight axial ratio was maintained over a bandwidth of 1.5% when the magnetic bias was applied. By turning on or off the bias, the polarization of the antenna can be switched between linear and circular, which may prove useful in polarization diversity applications.

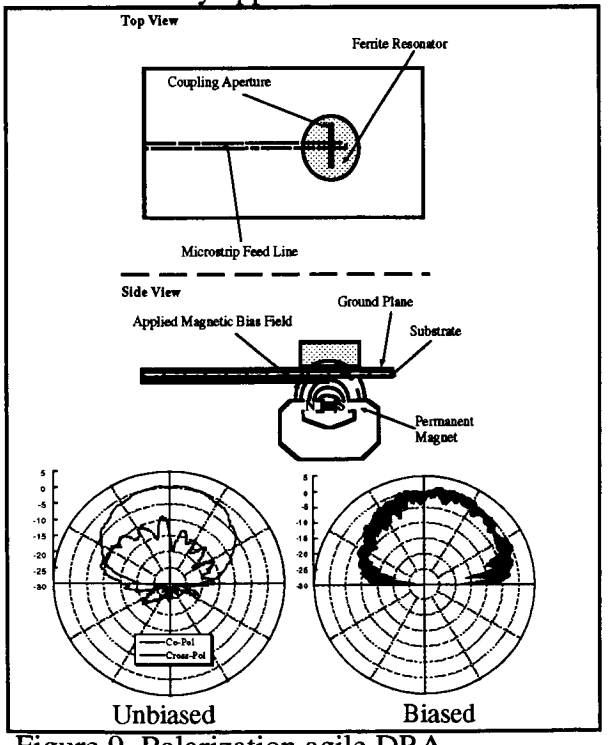


Figure 9. Polarization agile DRA.

SUMMARY

Research carried out by the CRC over the last few years has demonstrated that DRAs offer several attractive features including wide impedance bandwidth, high radiation efficiency, compact structures, and versatility in design. In many cases, DRAs are found to offer superior performance to microstrip antennas. These features make DRAs potential antenna candidates for various mobile satellite communication applications.

REFERENCES

[1] **R.K. Mongia and P. Bhartia**, "Dielectric Resonator Antennas - A Review and General Design Relations for Resonant Frequency and Bandwidth", *Int. Journal of Microwaves and Millimeter-Wave Computer-Aided Engineering*, Vol. 4, No. 3, 1994, pp. 230-247.

[2] **A. Petosa, D.J. Roscoe, A. Ittipiboon, and M. Cuhaci**, "Antenna Research at the Communications Research Centre," *IEEE Antennas and Propagation Magazine*, Vol. 38, No. 5, pp. 7 - 18.

[3] **A. Petosa, R.K. Mongia, A. Ittipiboon, and J.S. Wight**, "Investigation on a Microstrip-Fed Series Array of Dielectric Resonator Antennas," *IEE Electronic Letters* , Vol. 31, No. 16, Aug. 95, pp. 1306-1307.

[4] **A. Petosa, A. Ittipiboon, M. Cuhaci, and R. Larose**, "Bandwidth Improvement for a Microstrip-Fed Series Array of Dielectric Resonator Antennas," *IEE Electronic Letters*, March 96, Vol. 32 No. 7 pp. 608-609.

[5] **A. Petosa, A. Ittipiboon, M. Cuhaci, and R. Larose**, "Low Profile Phased Array of Dielectric Resonator Antennas," *IEEE International Symposium Phased Array Systems and Technology* , Boston, MA., Oct. 1996, pp. 182-185.

[6] **A. Petosa, M. Cuhaci, A. Ittipiboon, N.R.S. Simons, and R. Larose**, "Microstrip-Fed Stacked Dielectric Resonator Antenna," *ANTEM-96*, Montreal, Canada, Aug. 96, pp. 705-708.

[7] **A. Ittipiboon, A. Petosa, D. Roscoe, and M. Cuhaci**, "An Investigation of a Novel Broadband Dielectric Resonator Antenna," *IEEE Antennas and Propagation Symposium Digest*, July, 96, Baltimore, MA, pp. 2038-2041.

[8] **M.G. Keller, M. Fleury, E. Philippouci, A. Petosa, and M.B. Oliver**, "Circularly Polarized Dielectric Resonator Antenna Array," *URSI-Digest*, July 96, Baltimore, MA.

[9] **Z. Fan, Y.M.M. Antar, A. Ittipiboon, and A. Petosa**, "Parasitic Coplanar Three-Element Dielectric Resonator Antenna Subarray," *Electronic Letters*, Vol. 32, No. 9, pp. 789-790.

[10] **M. Cuhaci, A. Ittipiboon, A. Petosa, D. Roscoe, and N. Simons**, "Dielectric Antennas as an Alternative Technology for PCS Applications," *Canadian Conference on Electrical and Computer Engineering*, Calgary, Canada, May 1996, pp. 867-870.

[11] **M. Cooper, A. Petosa, A. Ittipiboon, and J.S. Wight**, "Implementing Dielectric Resonator Antennas at L-Band for Communication Applications," *Proceedings Wireless 96 Conference* Calgary, Canada, May 1996.

[12] **M. Cooper, A. Petosa, A. Ittipiboon, and J.S. Wight**, "Investigation of Dielectric Resonator Antennas for L-Band Communications," *ANTEM-96*, Montreal, Canada, pp. 167-170.

[13] **A. Petosa, R.K. Mongia, A. Ittipiboon, and J.S. Wight**, "Experimental Investigations on Feed Structures for Linear Arrays of Dielectric Resonator Antennas," *IEEE AP-S Conference 95*, Newport Beach, California, pp. 1982-1985.

[14] **A. Ittipiboon, D. Roscoe, R.K. Mongia, and M. Cuhaci**, "A Circularly Polarized Dielectric Guide Antenna With A Single Slot Feed," *ANTEM-94 Digest*, Ottawa, Canada, Aug. 94, pp. 427-430.

[15] **K.P. Esselle**, "Circular polarized dielectric resonator antennas: analysis of near and far fields using the FDTD method," *IEEE Symposium on Antennas and Propagation Digest*, Newport Beach, CA, 1995, URSI pp. 28.

[16] **A. Petosa, A. Ittipiboon, and M. Cuhaci**, "An Array of Circular-Polarised Cross Dielectric Resonator Antennas," *IEE Electronic Letters*, Vol. 32, No. 19, Sept. 96, pp. 1742-1743.

[17] **J. Carrie, N.R.S. Simons, A. Ittipiboon, D.J. Roscoe, A. Sebak, and L. Shafai**, "A Ka-Band Circularly Polarized Dielectric Resonator Modelled Using the Transmission-Line Matrix Method," *ANTEM-96*, Montreal, Canada, pp. 709-712.

[18] **J. Carrie, K. Esselle, D.J. Roscoe, A. Ittipiboon, A. Sebak, and L. Shafai**, "A K-Band Circularly Polarized Cavity Backed Dielectric Resonator," *IEEE AP-96 Symposium*, Baltimore, MD., pp. 734-737.

[19] **D.J. Roscoe, M. Cuhaci, A. Ittipiboon, L. Shafai,** "EHF Active Integrated Arrays Designed for Device Integration," *ANTEM-96*, Montreal, Canada, Aug. 1996, pp. 555-558.

[20] **A. Petosa, R.K. Mongia, M. Cuhaci, and J.S. Wight**, "Magnetically tunable ferrite resonator antenna," *IEE Electronic Letters*, Vol. 30, No. 13, June 1994, pp. 1021-1022.

[21] **A. Petosa, R.K. Mongia, A. Ittipiboon, and J.S. Wight**, "Experimental Investigation on a Magnetically Tunable Ferrite Resonator Antenna," *ANTEM-94 Digest*, Ottawa, Aug. 1994, pp. 697-700.

[22] **A. Petosa**, "Ferrite and Dielectric Antennas for Personal Communications," Carleton University, Ph.D Dissertation, 1995.

[23] **A. Petosa, R.K. Mongia, A. Ittipiboon, and J.S. Wight**, "Switchable LP/CP Ferrite Disk Resonator Antenna," *IEE Electronic Letters*, Vol. 31, No. 3, February, 1995, pp. 163-164.

Session 11
Spacecraft Technologies

Session Chairperson—*Dimitri Antsos*, Jet Propulsion Laboratory, USA
Session Organizer—*Dan Rascoe*, Jet Propulsion Laboratory, USA

Spaceborne L/S-band LNA Family Using MHMIC and MMIC Technology

C.Gingras, L. Hotte, J.F. Rivard, A. Rhodes, B. Stoute
Spar Aerospace Ltd
21025 Trans-Canada Highway
Ste-Anne-de-Bellevue, QC, Canada, H9X 3R2
Tel: (514) 457-2150 Fax: (514) 457-2724

ABSTRACT

A redundant low-noise amplifier has been developed by Spar Aerospace Ltd for applications on L/S-band multiple-beam feed array satellite systems. The main features are low overall noise figure (1.1 dB, including RF redundancy switch), high RF gain (70 dB) and very low mass (170 grams). Advanced design techniques have been applied to simultaneously achieve low input VSWR (1.25:1) and optimum noise figure, thereby eliminating the need for an input isolator. The LNA is a fully redundant unit with prime and redundant amplification chains, and internal RF and DC redundancy switches. This LNA is suited for mounting on the back side of antenna feed elements, for best antenna system performance. Efficient beam forming in the system is realized since the LNA units have very similar gain and phase performance over a wide operating temperature range. The LNA power consumption is typically 240 mW and is powered through a single low-voltage supply. This LNA has been designed to be manufacturable at minimal recurring costs for moderate-to-high volumes, as required for multi-beam antenna systems. This is achieved by making use of highly repeatable hybrid (MHMIC) technology and careful selection of packaging techniques. Costs associated to testing have been kept low by minimizing the tuning and by automation / batching during formal acceptance testing. This LNA, operating at L-band, is currently being supplied for a major mobile satellite program. An alternate version of this LNA using custom-designed MMICs is under development for similar S-band applications.

INTRODUCTION

Satellites for global and local mobile communications are currently being planned with L and S-band uplinks. These satellites, in many cases, use multi-beam systems which are formed from a large number of antenna radiating elements. A low-noise amplifier (LNA), with sufficient gain, is required for each antenna element in order to define the system noise figure before the beam forming network. Typically, the LNA unit is mounted on the back side of the antenna feed array, behind the radiating element, to minimize input losses. Additional requirements of amplitude and phase tracking with time and temperature are needed between all the LNAs in order to preserve the integrity of each formed beam over the operational life of the spacecraft. Redundancy of the LNA is also required in some systems where each beam is formed from a small number of elements.

KEY REQUIREMENTS

Key requirements for an LNA suited for these spaceflight applications are thus high gain, low noise figure and very low mass. Gain and phase tracking over flight sets of large number of units is also needed, therefore requiring use of technologies with repeatable performance, such as MHMIC (miniaturized hybrid microwave integrated circuit) and MMIC (monolithic microwave integrated circuit). The LNA envelope must be sufficiently small for mounting on the back side of antenna feed element. To minimize mass of harness and DC-to-DC converters, a single rail low-voltage power supply interface is used, which is compatible with secondary spacecraft buses.

A further key requirement is minimal recurring unit cost, as these units are needed in large numbers. This requires the use of components which provide high ratios of function to cost and sophisticated assembly and test processes to reduce recurring costs while maintaining the high quality required for spaceflight hardware.

Table I summarizes the key performance specifications for a redundant LNA suitable for the above applications:

Table I: Key LNA Performance Specifications.

Requirement Parameter	Value
Frequency Range (L-band) (S-band)	1626.5 to 1660.5 MHz 1980.0 to 2020.0 MHz
Noise Figure (at +25°C)	1.1 dB max.
Return Loss (input & output)	18 dB min.
Gain	70 dB min.
Gain Flatness	0.5 dB max.
Gain Tracking	1.0 dB p-p max.
Phase Tracking	10 deg. p-p max.
DC Input Voltage	+3.7 Volt
DC Power Consumption	320 mW max.
Weight	185 g. max.
Operating Life	12-15 years

The LNA block diagram is illustrated in Figure 1. While the LNA has been developed as a fully redundant unit, a simplified non-redundant unit could be easily configured, with a noise figure performance of 0.75 dB at +25°C.

DESIGN APPROACH / IMPLEMENTATION

Main design drivers are low unit mass, small envelope, low noise figure, high gain, and low cost. The design approach adopted for each module is targeted to meet these requirements as well as the key performance specifications of Table I.

Amplifier Technology

The input and output amplifiers have been designed using MHMIC technology. Discrete pHEMTs are used, which provide very low noise figure and very high gain at L/S-band, from a low drain-to-source voltage and current. With such devices, a total of five stages (three for the input, two for the output) are needed to achieve the required gain. The packaging of these hybrid amplifiers is performed individually to mitigate risk and for modularity.

These amplifiers are also currently being developed in an MMIC version. These will be packaged in an identical manner as the MHMIC amplifiers, for interchangeability.

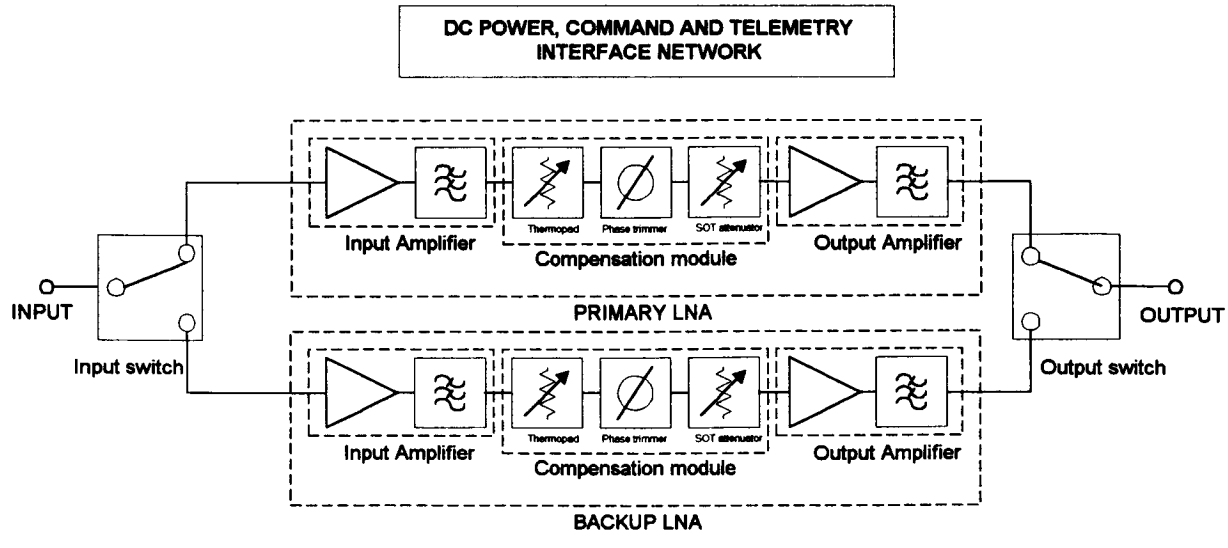


Figure 1: LNA Unit Block Diagram

Input Amplifier

The front-end input amplifier is designed to provide simultaneous noise and input match. This eliminates the need of an input isolator, with inherent mass savings and improvement in unit noise figure.

This is implemented by using both inductive series feedback and load pull on the first stage. Out-of-band stability is ensured by resistively loading the FET with a frequency-selective network which is transparent in-band. A low-loss input matching network is implemented on the MHMIC substrate.

For the MMIC version of this amplifier, a similar off-chip input matching network is used. As for the MHMIC, the MMIC input amplifier is a narrowband design which only covers one specific band (L or S).

Output Amplifier

This general-purpose amplifier is implemented at L-band in MHMIC technology. A wider band MMIC version is currently being developed to cover both L and S bands.

RF Switches

For this fully redundant unit, internal input and output RF switch designs using a miniature electromechanical switch are used to minimize unit mass and envelope.

This solution has been preferred over the more conventional external coaxial input switch approach, which provides lower losses, with the drawback of being much bulkier.

Compensation Module

A passive compensation module serves three purposes: 1) absolute gain set, 2) absolute phase set and 3) temperature compensation of the unit gain.

Unit gain adjustment is performed at amplifier assembly by selection of the appropriate attenuator. Phase setting is performed on a trombone section. Temperature compensation of the gain is achieved with thermistor attenuator.

This module, realized in standard MIC technology, does not require sealing and therefore is easily accessible during unit alignment. It is located between the input and the output amplifiers, to preserve unit noise figure and linearity performance.

DC Power Interface Circuit

This PCB (Printed Circuit Board) is located in a separate compartment of the LNA housing. It provides DC routing of a redundant secondary bus voltage to the appropriate LNA amplifier chain. It also provides an interface for command and telemetry signals between the unit and the spacecraft.

LNA Housing

The housing is machined from solid aluminum, with two back-to-back compartments for the DC and RF circuits. RF and DC interfaces are through SMA and micro-miniature connectors respectively.

Isometric view of the LNA unit is illustrated in Figure 2.

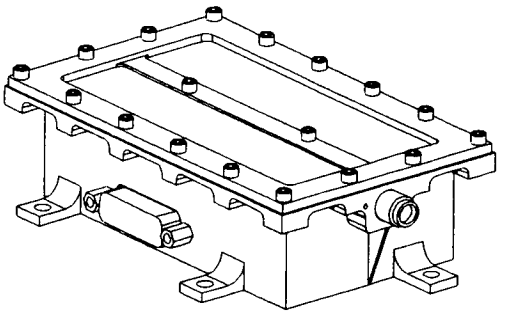


Figure 2: Isometric View of LNA Unit

UNIT PERFORMANCE

The measured LNA RF performance is illustrated in Figures 3 to 6, for RF gain, noise figure, input and output return loss, respectively. A temperature range of -40°C to +60°C is used.

Gain, gain flatness, noise figure and return loss specifications are all satisfied, and DC power consumption is only 230 mW.

Comparison of gain and phase measurement between the prime and redundant channels provides the confidence that the tracking specifications can be satisfied without any form of active compensation, for a large number of LNAs.

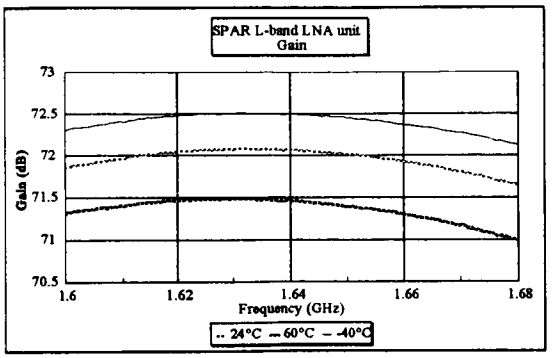


Figure 3: RF Gain of LNA

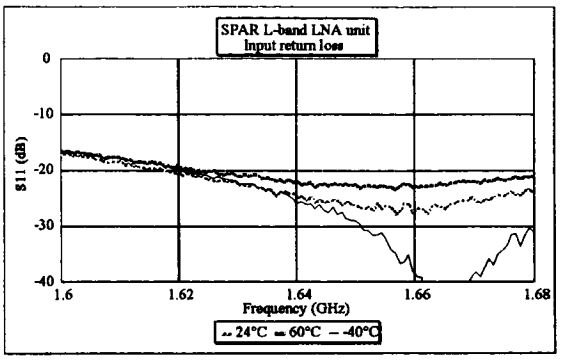


Figure 5: Input Return Loss

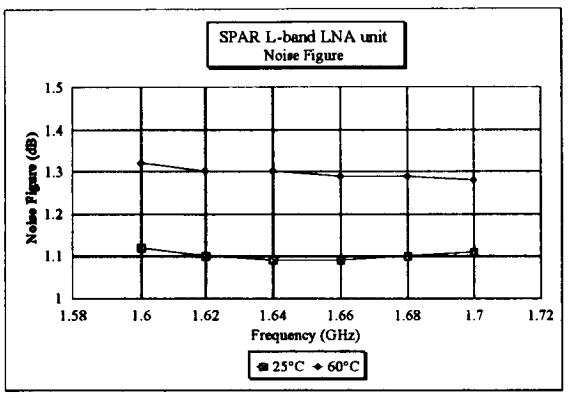


Figure 4: Noise figure Performance

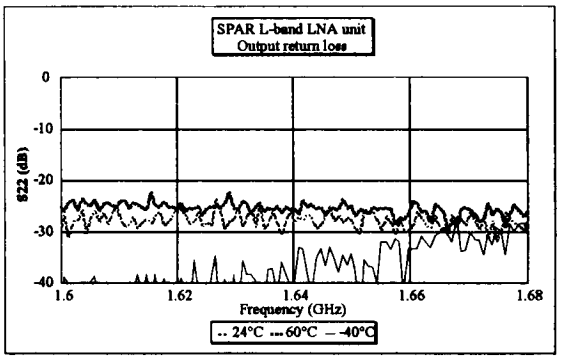


Figure 6: Output Return Loss

Weight has been measured on the breadboard unit. With the estimated necessary additions for a flight unit (strapping, shielding), the total weight is expected to be approximately 170 grams.

These results show that all LNA key performance specifications are satisfied with this unit.

CONCLUSIONS

A low-noise amplifier has been developed by Spar Aerospace Ltd for L/S-band multi-beam feed array satellite systems. This high-gain, light weight fully redundant unit can be mounted on the back side of the antenna feed element, to provide best system performance. MHMIC technology, using discrete pHEMTs, is used for the amplifiers, and a MMIC version is currently under development.

ACKNOWLEDGEMENTS

The majority of this work has been performed under contract no. 67CRC-5-1608/01-ST of the International Mobile Satellite Communications (IMSC) Program, managed by the Canadian Communications Research Centre (CRC) on behalf of Canadian Space Agency (CSA).

Low Cost Large Space Antennas

Artur B. Chmielewski, Robert Freeland
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91009
Phone: 818-354-0255
Fax: 818-393-5011
e-mail: abc@jpl.nasa.gov

ABSTRACT

The mobile communication community could significantly benefit from the availability of low-cost, large space-deployable antennas. A new class of space structures, called inflatable deployable structures, will become an option for this industry in the near future. This new technology recently made significant progress with respect to reducing the risk of flying large inflatable structures in space. This progress can be attributed to the successful space flight of the Inflatable Antenna Experiment in May of 1996, which prompted the initiation of the NASA portion of the joint NASA/DOD coordinated Space Inflatables Program, which will develop the technology to be used in future mobile communications antennas along with other users.

The NASA/DOD coordinated Space Inflatables Program was initiated in 1997 as a direct result of the Inflatable Antenna Experiment. The program adds a new NASA initiative to a substantial DOD program that involves developing a series of ground test hardware, starting with 3 meter diameter units and advancing the manufacturing techniques to fabricate a 25 meter ground demonstrator unit with surface accuracy exceeding the requirements for mobile communication applications. Simultaneously, the program will be advancing the state of the art in several important inflatable technology areas, such as developing rigidizable materials for struts and tori and investigating thin film technology issues, such as application of coatings, property measurement and materials processing and assembly techniques. A very important technology area being addressed by the program is deployment control techniques. The program will sponsor activities that will lead to understanding the effects of material strain energy release, residual air in the stowed structure, and the design of the launch restraint and release system needed to control deployment dynamics. Other technology areas directly applicable to developing inflatable mobile communication antennas in the near future are analytical performance prediction tools, configuration studies and miniaturizing the inflation systems.

LARGE ANTENNAS ARE NEEDED

Large space-deployable reflector antennas are needed for a variety of applications that require structures up to 30 meters in diameter for RF operation between 0.3 and 88 GHz. The classes of applications include mobile communications, earth observation radiometry, active microwave sensing, orbiting very long baseline interferometry (OVLBI) and DOD space-based passive and active sensors. A number of mobile communications system concepts are based on L-band, 1.5 GHz with aperture sizes from 10 to 20 meters and Ka-band, 20 to 30 GHz with apertures from 4 to 8 meters. Earth observation radiometry can utilize reflectors from 15 to 30 meters in diameter for operation somewhere between 1.4 and 60 GHz. Active microwave sensing is based on planar arrays up to possibly 3 by 16 meters for operation from 1 to 90 GHz. The next generation OVLBI after the VLBI Space Observatory Program (VSOP), which was recently launched, will need a reflector somewhere between 15 and 30 meters in diameter for RF operation between 21 and 88 GHz. Current DOD concepts for space based radar are based on structures up to 25 meters in diameter for RF operation from 1.5 to 8.5 GHz.

USER CONCEPT SELECTION CRITERIA

Recent constraints on the availability of resources for these classes of applications within NASA, the Science Community, the commercial sector and even the DOD, have resulted in very stringent user application requirements. The real key is affordable antennas that are mechanically reliable with high mechanical packaging efficiency and aperture precision while maintaining low mass and long term dimensional stability. In addition to designing for these challenging criteria, meaningful demonstrations of such concept capabilities will have to be accomplished to attract any kind of serious user interest. This means that (a) low-cost and lightweight hardware can only be verified by actually building large flight-quality antennas, (b) deployment reliability can only be demonstrated with large flight-type hardware in a realistic service environment, (c) useable reflector surface precision will have to be validated in a realistic service environment and (d) mechanical packaging efficiency can only be validated with full scale, flight configured hardware.

NEW CLASS OF SPACE STRUCTURES

Fortunately for the users, a relatively new class of deployable space structures has recently emerged that has tremendous potential for accommodating the current user criteria. Even though large inflatable space structures have been around for over 40 years, i.e., Echo I, series 1959-64, only recently have a very few organizations learned how to design and manufacture thin membrane structures with enough geometric precision to be seriously considered for specific classes of applications.

Significant technology achievements have been made by both L'Garde, Inc. of Tustin, California and Contraves of Zurich, Switzerland. L'Garde, Inc. has successfully built and flown a large number of small inflatable space structures over a period of 25 years. In this same time frame and in parallel, other concepts for much larger size inflatable structures for solar concentrators, reflector antennas and solar cell support structures were under development. The culmination of the latter effort was the NASA sponsored In-Space Technology Experiments Program (IN-STEP) Inflatable Antenna Experiment. Meanwhile, in the time frame of 1975 to 1985, Contraves developed concepts for large offset reflector antennas intended for mobile communications applications and other applications. A 10 by 12-meter proof of concept antenna reflector structure was designed, built and deployed on the ground. The surface accuracy of a few mm RMS for such a new and unique structure of this size was considered by many a major advancement of this technology. Other organizations which have demonstrated a technology capability for inflatable space structures include SRS, ILC Dover, Thiokol, and Aerospace Recovery Systems.

IAE

The NASA Office of Aeronautics and Space Technology initiated IN-STEP specifically to sponsor the verification and validation of new innovative and high payoff technologies in the space environment. The potential of this new class of space structures for a number of meaningful applications resulted in the selection of the L'Garde, Inc. inflatable deployable reflector antenna concept for an experiment.

The basic objective for the experiment was to validate and characterize the functional mechanical performance of a large inflatable low cost deployable reflector antenna structure in a real orbital operational environment. This objective when evaluated with respect to the L'Garde, Inc. technology data base for the inflatable antenna concept, their manufacturing facilities, and the capability of the Spartan spacecraft, resulted in the selection of a 14 meter diameter reflector structure with three 28 meter long struts. It is interesting to note that, this size structure is reasonably close to what is needed for several classes of application, in particular, mobile communications. The experiment requirements were to (a) develop a large low cost flight quality reflector antenna structure, (b) validate mechanical packaging efficiency, (c) demonstrate deployment reliability and (d) measure surface precision on orbit. The IAE was successfully flown on the Space Shuttle STS-77 on May 29, 1996.

The overall experiment was successful. New, unique and low cost space structures technology was demonstrated on orbit by (a) building a large inflatable space antenna structure on the order of about $1,000,000, (b) demonstrating extremely efficient mechanical packaging by stowing a 14 by 28 meter inflatable structure in a container the size of an office desk, (c) manufacturing an offset membrane reflector structure with a surface precision on the order of a few mm RMS and (d) demonstrating the robustness of deployment for this new class of structure. The results of this experiment were used specifically to establish the technology database and were the basis of a technology road map for the continued development of this type of space structure.

MOBILE COMMUNICATIONS ANTENNA NEEDS

Current concepts for L-band mobile communications are based on using multiple beam offset reflector antennas in the size range of 8 to 20 meters in diameter. Some system concepts utilize numerous spacecraft, while all concepts are based on low cost antennas. In addition to the requirement for low cost, antenna structures must also be lightweight with mechanical packaging techniques that will enable the use of different size and shaped spacecraft. Without such physical capabilities the antenna stowed configuration could drive the spacecraft design and the size of the launch vehicle shroud, which have major inpact on system cost.

SPACE INFLATABLES PROGRAM

The program was created as the logical follow-on to the successful flight of the Inflatable Antenna Experiment that was launched in May 1996. The program is designed to be compatible with a large DOD investment in several different inflatable technology programs. The two largest elements of the DOD programs are a two year Large Inflatable Structure Program funded by the Phillips Laboratory and a Rigidified Inflatable & Materials Development Program, both of which are executed by L'Garde, Inc., Tustin, Ca. The NASA portion of the joint program is designed to leverage off, complement, broaden the scope and accelerate the DOD inflatable technology development program. The results of the joint program will significantly reduce the risk of using the inflatable technology for NASA, DOD and commercial missions such as mobile communications.

The technology investment strategy for the program was based on the inputs from the users. A Quality Function Deployment (QFD) analysis was conducted to determine where NASA money should be invested to maximize the leverage from the DOD program. As a result a robust two year program was created. The program will address all of the fundamental technological areas for inflatables and advance the technology to the point where 25 meter class apertures supported by large rigidizable structures can be baselined for long life space missions.

PROGRAM DESCRIPTION

The Program consists of 10 major activities as described in the following sections:

Deployment
The purpose of this research and development is to understand and characterize the fundamental factors influencing deployment dynamics and therefore the reliability of inflatable structures in space. Upon reaching the vacuum of space, the folded inflatable membranes are subject to the pressure of internal residual gas. The effect of this gas expansion in conjunction with strain energy release of flexible materials and thin films could lead to uncontrolled deployments. To address these phenomena several sub-scale drop tests will be performed to evaluate the newest predictive tools, packing methods and release and ejection mechanisms under simulated microgravity conditions. Thermal vacuum tests will provide additional information regarding different designs of vent valves, vent techniques and temperature effects on different deployment techniques.

Support Structures
The purpose of this research is to understand the whole new class of inflatable rigidizable structures in space. There are several different types of materials and rigidization techniques that are under consideration for applications.

The methods of rigidification under study are: gel impregnation, cold rigidization, UV curing, yielding of aluminum laminate and foam injection. Other advanced concepts such as gas curable epoxies and radiation curing will also be considered for far term applications. The materials that offer the most promise in the short term are gel and rigidified aluminum. An extensive material properties program will be implemented to establish the data base of stress/strain/temperature properties. Large struts, tori and beams will be constructed and tested in thermal vacuum chambers. Thermal vacuum tests will significantly improve the data base of the rigidification process with respect to temperature, pressure and different types of inflatant gases. The optimal ways of joining inflated members will be studied to arrive at designs of tube connectors, joints and saddles.

The support structures technology area is currently being substantially supported by the DOD.

Membranes
The success of inflatables in space in the long run depends on the properties of materials used. This task will investigate a number of candidate materials and their properties, such as stress/strain/temperature characteristics, non-linear modulus, creep, coefficient of thermal expansion, and radiation tolerance. The most promising materials will be tested for compatibility with fabrication and processing techniques, and deposition of optical coatings, flexible interconnects and electronic elements. A number of environmental tests will be conducted to allow more accurate performance predictions of the selected thin films. The films will be subject to accelerated tests of UV radiation, high energy electrons, low energy protons and combined effects. In addition, mid-LEO orbit AO-erosion life tests will be conducted. A comprehensive property data base will be created for selected materials to serve as a universal reference for future inflatable structures. Compatible adhesives for joining membranes and methods for forming doubly curved surfaces will also be investigated.

Analytical Tools
A significant contribution to faster, better, cheaper missions will come from extensive analytical modeling to shorten the design, test and fabrication cycle. This activity includes dynamic and static structural performance prediction, thermal models, reflector performance predictions and ground based simulation methods. It is important that all the existing specialized codes such as the Finite Element Analysis of Inflatable Membranes (FAIM), TRASYS and IMOS be made compatible with each other and work in tandem to allow the engineers several iterations of the preliminary design before any prototype hardware is built and tests are conducted. All codes will reflect the latest knowledge and refinements obtained from other tasks of the program.

Manufacturing and Assembly
This extremely important activity will be entirely funded by the DOD component of the program in the next two years. There will be synergism of the manufacturing tasks with the NASA sponsored work. The manufacturing tasks will depend to a great extent on the materials development results and the characteristics of the resulting products.

System Testing
This is a very important area of inflatable technology. Inflatables are a new class of structures that are very difficult and sometimes impossible to test on the ground. The sizes of the structures considered for proof of concept demonstrations are not compatible with ground facilities of aerospace companies. The testing itself imposes many new challenges: large vacuum tanks, large dynamic test facilities, methods of relating results of these tests to space conditions or conducting partial deployment tests on KC-135 aircraft in simulated microgravity conditions, vent test techniques verification, support structure performance, surface measurement equipment and most of all deployment control technique verification.

Inflation Systems
One of the major benefits of the inflatable structures is their low launch volume and mass. While it is clear that thin membranes have low mass and can be packaged into a small container, the inflatable structural systems do not consist of membranes only. From the system point of view a significant part of the

total mass is attributed to the inflation system. Consequently, a miniature inflation system is equally important. The inflation system is also a major contributor to the mechanical reliability of the structural system and a participant in control of the deployment. Studies will be conducted to determine if an inflation system with its cumbersome tanks and pressure regulation system could not be replaced by some other alternatives. One alternative is on-board gas generation, using chemical reaction between liquids or liquids and solids. Such a gas generation system would likely consist of a reaction chamber, liquid low pressure tank, and low pressure valving.

Interfaces
This technology area deals with the interfaces between the inflatable and the rest of the spacecraft. These include the mechanical and electrical interfaces that are involved with the launch, deployment and post deployment. This technology area is considered to be mission specific. Only a small portion of the funds will be invested here to ensure there are no formidable problems with a spacecraft.

Active Shape Control
Active shape control can contribute to further dimensional precision of inflatable structures. Smart materials can be integrated into the structure to adjust the geometry either in real time through feedback control or before the rigidification takes place. This approach could compensate for many variables in materials, processes, fabrication, space effects or temperature. This technique will probably be required if the inflatables are to be used as high frequency antennas.

Requirements and Applications
This is a very important area of the inflatable technology development since the application requirements will focus and drive the program elements. One of the primary considerations of this task is to collect the user requirements. Based on these requirements conceptual designs will be developed for the primary customers of the technology. These conceptual designs will help focus the trade studies considering both inflatables and mechanically deployed structures and different approaches for achieving inflatable structure configurations. The trade studies, for example, will point out: the benefits and application limitations of gel rigidizables and aluminum rigidizables, and cover contamination requirements, range of radiation exposures, atomic oxygen environments, and lifetime performance. The trade studies will also contribute to investment decisions depending on relative importance of different aspects of the inflatable program.

CONCLUSION

The joint NASA/DOD Space Inflatables Program will develop inflatable structures technology to the point of demonstrating a ground based technology readiness level for 25 meter antenna structures within two years. It is expected that these antenna structures will have reflector surface precision on the order of 2 mm RMS and thus will make inflatables a viable technology for mobile communications.

However, before this inflatable technology will be ready for application to the commercial community it will have to be demonstrated by a long duration, operational, large size space antenna validation. It is expected that hardware development for such a demo will start in 1999 to accommodate a flight in the year 2001 time frame. If such a demo is successful the commercial applications will follow. Even though this timetable points out that we will not see inflatable mobile communication antennas in space before the year 2003, the high payoff of decreasing by orders of magnitude the cost, mass and launch volume of large space structures may be worth waiting for.

ACKNOWLEDGMENT

The work described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

BIBLIOGRAPHY

1. Ciesluk, Jr., W. J., Gaffney, L. M., Hulkower, N. D., Klein, L., Krueger, P. E., Oliver, P. A., and Pomponi, R. A. "Survey of the Mobile Satellite Communications Industry," MTR 92B0000059, MITRE, Bedford, MA, Apr. 1992. (This document was prepared for authorized distribution. It has not been approved for public release.)

2. Ciesluk, Jr., W. J., Gaffney, L.M., Hulkower, N. D., Klein, L., Oliver, P. A., Pavloff, M. S., Pomponi, R. A., and Welch, W. A. "An Evaluation of Selected Mobile Satellite Communications Systems and Their Environment," MTR 92B0000060, MITRE, Bedford, MA, Apr. 1992. (This document was prepared for authorized distribution. It has not been approved for public release.)

3. Sue, M. K., editor, *Personal Access Satellite System (PASS) Study*, JPL Internal Document D-7382, Jet Propulsion Laboratory, Pasadena, CA, Sept. 1990.

4. Naderi, F., editor, *Land Mobile Satellite Service (LMSS): A Conceptual System Design and Identification of the Critical Technologies*, JPL Publication 82-19, Jet Propulsion Laboratory, Pasadena, CA, Feb. 1982.

5. Satter, Celeste M. and Freeland, Robert E., "Inflatable Structures Technology Applications and Requirements," AIAA Paper 95-3737, presented at the Space Programs and Technologies Conference, Huntsville, AL, September 26-28, 1995.

6. Freeland, R. E. and Bilyeu, G., "IN-STEP Inflatable Antenna Experiment," IAF Paper 92-0301, presented at the 43rd Congress of the International Astronautical Federation, Washington, D.C., Aug. 28-Sept. 5, 1992.

7. Freeland, R. E., Bilyeu, G., and Veal, G. R., "Validation of a Unique Concept for a Low-Cost, Light-Weight, Space-Deployable Antenna Structure," IAF Paper 93-I.1.204, presented at the 44 Congress of the International Astronautical Federation, Graz, Austria, Oct. 16, 1993.

8. Veal, G., and Freeland, R., "IN-STEP Inflatable Antenna Description," AIAA Paper 95-3739, presented at the Space Programs and Technologies Conference, Huntsville, AL, Sept. 26-28, 1995.

9. Freeland, R. E., Bilyeu, G. D., and Veal, G. R., "Development of Flight Hardware for a Large, Inflatable-Deployable Antenna Experiment," IAF Paper 95-1, 5.0.1, presented at the 46th Congress of the International Astronautical Federation, Oslo, Norway, October 2-6, 1995.

10. Freeland, Robert, Bard, Steven, Veal, Gordon, Bilyeu, Gayle, Cassapakis, Costa, Campbell, Thomas and Bailey, M. C., "Inflatable Antenna Technology with Preliminary Shuttle Experiment Results and Potential Applications," presented at the 18th Annual Meeting and Symposium of the Antenna Measurement Techniques Association, September 30-October 3, 1996, Seattle, Washington.

11. Bernasconi, M. C., "Development of a 2.8-m Offset Antenna Reflector using Inflatable, Space-Rigidified Structures Technology," Proc. 2nd ESA Workshop on Mechanical Technology for Antennas, Noordwijk, The Netherlands, EA SP-261, 1986, pp. 31-39.

12. Reibaldi, G., Hammer, J., Bernasconi, M. C., Pagana, E., "Inflatable Space-Rigidized Reflector Development for Land Mobile Missions," AIAA Com. Sat. Syst. Conf., Proceedings (1986), pp. 533-538.

13. Bernasconi, M. C., Pagana, E., and G. G. Reibaldi, "Large Inflatable Space Rigidized Antenna Reflectors: Land Mobile Services Development," IAF-87-315, 38th IAF Congress, Brighton, United Kingdom, October 1987.

14. Bernasconi, M. C., "Inflatable Space Rigidized Structures Scale Model Reflector," Proc. 39th Int. Astronautical Federation Congress, Bangalore, India, Paper IAF 88-049, 1988.

L/S-Band Tradeoff for Mobile Communications Payload Applications

Eric Amyotte
Spar Aerospace Limited
21025 Transcanada Hwy.,
Ste-Anne-de-Bellevue, Quebec
Canada H9X 3R2
Tel: (514) 457-2150
Email:eamyotte@spar.ca

Dr. Harry J. Moody
President
Coronation Communications Inc.
4470 Coronation Ave.
Montreal Quebec
Canada H4B 2C4
Tel: (514) 488-3439

ABSTRACT

This paper presents the results of a trade-off study performed to highlight the advantages and disadvantages of spectrum utilization for S-Band versus L-Band in Mobile Communication Systems.

Spectrum availability at L-Band is restricted to 34 MHz of bandwidth which is segmented and congested. The S-Band is uncongested with more total contiguous bandwidth available.

The Mobile Communication System used as a basis of comparison is designed to accommodate large numbers of channels ranging from 10,000 to 20,000 duplex channels from a geostationary orbit with On-Board-Processing.

The beam coverage strategy including the number of beams, frequency reuse and number of frequency segments required as a function of the available spectrum will be presented. The configurations studied shall also include the impact of implementing digital beam forming.

Other considerations such as Power Handling, Passive Inter Modulation (PIM) and Multipactor risk will also be taken into account.

INTRODUCTION

L-Band is segmented into maritime, aeronautical and land usage. This means that a service provider needs some spectrum across the band to provide service in each category. On the other hand at S-Band the three services can be mixed and included in a contiguous frequency segment. This has significant impact on the satellite design particularly the passive intermodulation design and the digital processing load for satellites which include on board digital processing. In addition, because L-Band is already used by other systems, it is congested. The L/S band spectrum assumed for this study is:

L-Band Tx: 1525.0-1559.0 MHz
Rx: 1626.5-1660.5 MHz

S-Band Tx: 2170-2200 MHz
Rx: 1980-2010 MHz

Unless specified otherwise, the trade-off presented in this paper is based on the L and S band Tx carrier frequencies.

TARGET CAPACITY

The current generation of mobile communication systems typically provides support for 10,000 to 20,000 duplex channels. It is more than likely that these requirements will be increased for future systems.

Flexibility

To actually support the number of channels desired with a non-uniform traffic distribution it is necessary to design the system for a higher capacity. The extra capacity provides for traffic flexibility allowing higher than average traffic in regions with high traffic demand. The total capacity, including flexibility, will be referred to as C.

The spectrum utilization efficiency depends on the resolution with which the extra capacity can be allocated to the beams. If the extra capacity can be allocated in smaller sub-bands, less extra capacity is required to support a given geographical traffic distribution.

COVERAGE STRATEGY

Available Spectrum

The total spectrum available at L-Band is typically lower than that available at S-Band. It is assumed in this paper that the available spectrum is about 5 MHz at L-Band and 6 to 12 MHz at S-Band. The total available spectrum will be referred to as S_t.

Spectrum required per User

The spectrum required for each user depends on the access scheme employed. For the purposes of a trade-off between L/S band, it is not relevant. For simplicity, it is assumed that the equivalent of 10 kHz of bandwidth is required to support a single user. The spectrum required by a single user will be referred to as S_u.

Frequency reuse plan

The frequency reuse factor is obtained from the desired capacity and the number of users supported by a single use of the total available bandwidth. The frequency reuse factor will be referred to as F.

$$F = C/(S_t/S_u)$$

For example, the frequency reuse factor required to support a capacity of 15,000 channels with 6 MHz of spectrum and 10 kHz per user is 25. If the spectrum available to the L-Band system is less than that of the S-Band, a higher frequency reuse factor will be required to support the same capacity.

The number of beams per cluster and the frequency reuse factor determine the total number of beams required. The number of beams per cluster N_{bc} will be chosen such that the desired C/I performance can be achieved. The values of N_{bc} yielding an optimal hexagonal grouping of the beams are given by:

$$N_{bc} = 3, 4, 7, 9, 12, 13, 16, 19, 21, \ldots$$

It should be noted that, for example, an optimal layout that uses 7 beams per cluster would yield a higher C/I than any layout with 8 beams per cluster.

Antenna beam parameters

The total number of beams N_b and the size of the coverage area determines the beam spacing B_s and hence the beam overlap level. It should be noted that the minimum number of beams is limited by the gain requirements; beam spacing has to be small enough to meet the gain requirements over the coverage area. The distance between co-channel beams D_f is obtained by multiplying the square root of the number of beams per cluster by the beam spacing B_s.

$$D_f = B_s * \sqrt{N_{bc}}$$

To obtain a good C/I performance it is necessary to maintain a sufficient distance between co-channel beams. When a sufficient distance is maintained the C/I performance of the system is limited by the sidelobe level.

A smaller number of beams per cluster N_{bc} requires a larger beam spacing Bs to maintain C/I performance, hence the total number of beams is limited. **Figure 1** shows the maximum number of beams that can be fitted in a coverage area of 65 deg^2, with a 12-m aperture reflector antenna, while maintaining about 20 dB of co-channel isolation (single entry) as a function of N_{bc} for L-Band and S-Band. The increased number of beams results in smaller cell size for each beam and a higher cross over gain. This type of highly overlaping beam grid can be generated more efficiently using digital beam forming techniques but can also be achieved using traditional analog techniques.

The minimum distance between co-channel beams, for a given C/I performance, depends on the cell size (obtained from the beam spacing) and the half-power beamwidth (HPBW).

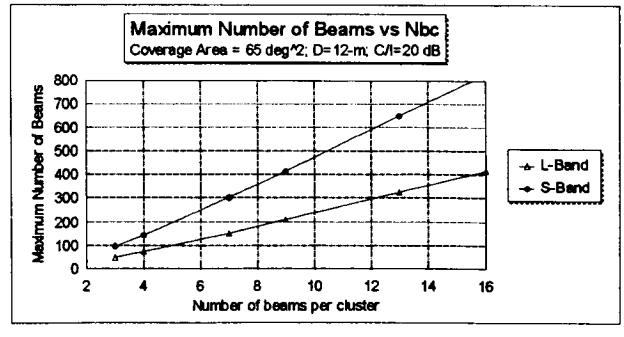


Figure 1 Maximum Number of L/S-Band beams

Assuming the same antenna aperture, the L-Band Tx beamwidth is 41% wider than the S-Band beamwidth. This yields a larger minimum co-channel beam spacing for the L-Band system, therefore limiting the frequency reuse factor achievable in a given area.

Figure 2 shows the maximum frequency reuse factor that can be supported in a coverage area of 65 deg^2, using a 12-m antenna, while maintaining about 20 dB of co-channel isolation (single entry) as a function of N_{bc} for L-Band and S-Band.

As observed in **Figure 2**, The maximum frequency reuse factor achievable at L-Band is about half of that which can be achieved at S-Band for a given area and aperture size. It is also observed that a frequency reuse factor of 30 may not be achieved at L-Band over a coverage area of 65 deg^2.

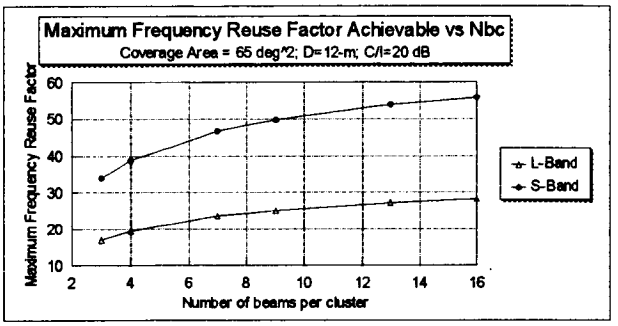


Figure 2 Maximum Frequency Reuse Factor

For a system requiring a high capacity, thus a high frequency reuse factor, over a small coverage area, this can be a major drawback for L-Band. This lower achievable frequency reuse factor combined with the narrower available bandwidth further limits the capacity of the L-Band system. If the requirements of the mobile communication system are such that capacity is not limited by the spectrum bandwidth, including the frequency reuse factor, then L-Band is suitable.

Number of frequency segments

The number of frequency segments used is independent of the number of beams per cluster. Narrower frequency segments allow for the allocation of the capacity according to the geographical traffic distribution, with a finer resolution. This results in less capacity wastage.

Gain and Link

The peak gain obtained with the same antenna aperture is 3.0 dB higher for S-Band Tx than for L-Band Tx. This is however compensated by the propagation losses term $(\lambda/4\pi R)^2$ as shown in the simplified link budget in **Table 1**. It is therefore necessary to use the same aperture size to obtain the same peak SNR for L and S-Band. It should be noted that the same handset G/T is assumed for L-Band and S-Band in **Table 1**.

For a given coverage area and antenna aperture size, the use of the same number of beams at L and S band results in a higher relative overlap level between the L-Band beams. This means that the L-Band Edge-of-Coverage (EOC) and average SNR are higher than those obtained at S-Band. This higher SNR can allow for a reduction of the L-Band transmitted power while maintaining the same link margin as the S-Band system.

Table 1 Simplified L/S band Link Budget

Return Uplink			
	Units	S-Band	L-Band
Frequency	MHz	1,985.00	1,645.00
Mobile EIRP	dBW	-2.00	-2.00
Propagation Losses	dB	-190.33	-188.70
Fade Margin	dB	-10.00	-10.00
G/T	dB/K	17.00	15.37
C/No	dB-Hz	43.27	43.27
Bit Rate	kbps	5.60	5.60
Eb/No	dB	5.79	5.79
Forward Down Link			
	Units	S-Band	L-Band
Frequency	MHz	2,175.00	1,542.00
Satellite EIRP/User	dBW	39.00	36.02
Propagation Losses	dB	-191.12	-188.14
Fade Margin	dB	-10.00	-10.00
G/T	dB/k	-23.00	-23.00
C/No	dB-Hz	43.48	43.48
Bit Rate	kbps	5.60	5.60
Eb/No	dB	6.00	6.00

Figure 3 shows the total transmitted power reduction allowable for an L-Band system that uses the same number of beams as an S-Band system, while maintaining the same link margin. When power control is implemented, the total transmitted power is determined from the average Tx gain.

For the same number of beams, if downlink power control is implemented, the savings realized on the total L-Band transmitted power correspond to the curve labeled "*Avg*" in **Figure 3**.

For a given transmitted power level, twice as many S-Band beams have to be generated to get the same performance as that obtained at L-Band with the same aperture size. The number of beams, for mobile communication systems, is generally determined from the required frequency reuse factor and C/I requirements.

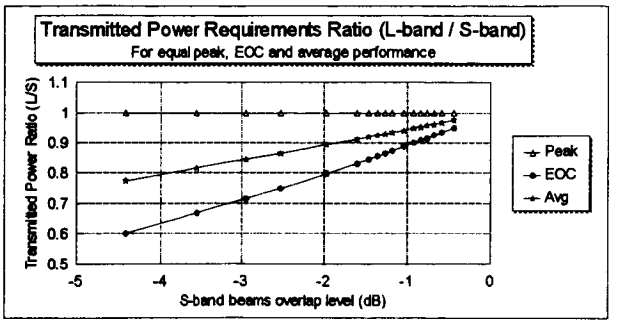


Figure 3 Transmitted Power Ratio (L/S)

ANTENNA DESIGN

Reflector size

It is assumed, for all comparisons presented in this paper, that the L-Band and S-Band systems use the same reflector aperture size of 12-m, a focal length of 7.8-m and an offset of 2.11-m.

Number of elements

Generating beams over a service area with a given reflector geometry requires the multi-element feed array size to be about the same at L-Band or S-Band. The size of the feed array is driven by the angular size of the coverage area and the focal length of the reflector.

The maximum separation between the feed elements is constrained by the generation of grating lobes in the feed radiation pattern which will exist within the spatial solid angle subtended by the reflector. It has been shown that in order to provide closely-spaced beams without the generation of grating lobes over the reflector surface, the inter-element spacing must be less than:

$d_{el}=\lambda/(2 \sin \theta)$

where θ is the half-angle subtended by the reflector

Since, for a given aperture size, the element spacing solely depends on the wavelength and the size of the L/S band feed array are about the same, the S-Band feed is going to necessitate about twice the number of elements (ratio of the wavelengths squared).

SSPA considerations

A reflector antenna system is generally combined with a Hybrid Matrix Power Amplifier (HMPA) scheme to provide flexibility while keeping the power rating of the SSPAs to a minimum. With this approach, the number of SSPAs required (not including redundancy) is equal to, or greater than, the number of elements. Hence, the number of SSPAs required at S-Band is about double the number of SSPAs required at L-Band.

Cost savings can thus be realized with an L-Band system. However, since the number of SSPA is half, the power rating of each must be double to maintain the aggregate EIRP. This will offset some of the savings realized since going from a single FET SSPA to a two FET SSPA increases the cost per unit by about 20%.

Power Handling

Because of the higher power rating of each SSPA at L-Band, the L-Band power handling requirements for the Tx chains are increased by up to a factor of 2. The impacts on Passive Intermodulation and Multipactor performance are significant.

Passive Inter Modulation (PIM)

The PIM concern arises from intermodulation products of the Tx carriers falling in the Rx band. The frequency of an intermodulation product can be obtained from the following:

$F_{pim} = m * F_1 \pm n * F_2$

The order of PIM is obtained from $O=m+n$.

Since the L-Band system has carriers scattered over 34 MHz, the lowest PIM order for L-Band is 5th, and is obtained as follows:

3*1559 - 2*1525 = 1627 MHz in Rx [1626.5,1660.5]

Since the S-Band system uses a contiguous spectrum, the lowest PIM order for S-Band depends on the spectrum bandwidth used by the system. If a total of 12 MHz is used the PIM order is 31st, and is obtained as follows:

16*2170 - 15*2182 = 1990 MHz in Rx [1980,1992]

If a total of 6 MHz is used the PIM order is 63rd, and is obtained as follows:

32*2170 - 31*2176 = 1984 MHz in Rx [1980,1986]

A lower order will result in much higher PIM levels. The relative level of the PIM products as a function of their order can be obtained from the following empirical equation.

$$P_L(dB) - P_S(dB) \cong 60\ Log(O_S/O_L)$$

Assuming 31st order for S-Band and 5th order for L-Band, the L-Band PIM level is expected to exceed the S-Band one by about 48 dB for the same Tx carrier power.

As shown in the Power Handling section, the L-Band Tx chains will have to support power levels that are nearly twice as high (3 dB) as those required for the S-Band chains. The PIM level does not generally vary one-for-one with the power of the Tx carriers. An increase of 3 dB in PIM levels for an increase of 1 dB in transmitted power is typical.

The much lower PIM order combined with the higher power handled by the L-Band Tx chains makes the L-Band PIM a very serious concern. The difference is such that trying to use a common Tx/Rx antenna is deemed very risky at L-Band, while it is recommended at S-Band. This single reflector and feed array configuration results in significant mass and cost savings for the S-Band system.

Multipactor

For a given size and because of multipactor breakdown, S-Band components such as BPF and OHM offer a peak power handling capability approximately 3 dB higher than that of similar L-Band components. This results from the linear frequency dependency of the multipactor threshold voltage.

It was also shown that the power rating of each L-Band Tx chain is about 3 dB higher than that implemented in the S-Band system. This higher power rating combined to the frequency dependency of the multipactor threshold yields a total advantage of about 6 dB to the S-Band system.

Alternatively, the S-Band component size can be reduced, for mass savings, if the additional multipactor margin of the S-Band system is not required.

Filter Considerations

The S-Band frequency plan allows for a narrower bandpass and a higher spacing between the Tx and Rx bands. The S-Band filter design requires less poles to achieve the same performance. It should be noted that if the single reflector configuration is chosen for S-Band, the filter rejection specification will be increased by about 50 dB. The realization of such a filter may require the same number of poles as that of the L-Band filter.

OBP DESIGN

In order to address aeronautical, maritime and land service using a congested L-Band spectrum, it is required for each beam to access slices of spectrum scattered over the entire 34 MHz bandwidth. This constraint increases the complexity of the OBP. The entire bandwidth needs to be digitized and processed, but only a fraction is allocated to the satellite system. The rest of the spectrum is discarded.

The number of beams required for the S-Band system is likely to be about double the number of L-Band beams. This is offset by the fact that each S-Band beam will be supported by a significantly narrower bandwidth (3 to 6 times), ultimately resulting in a lower OBP complexity for the S-Band system.

This more complex L-Band OBP is likely to consume a significantly higher level of DC power than its S-Band equivalent.

SUMMARY (EXAMPLE)

Table 2 summarizes the L/S band key requirements for a system supporting 12,000 channels over a coverage area of about 65 deg^2 (solid angle) with a 12-m aperture antenna.

There are four sections in **Table 2**:

1. Basic Requirements and Characteristics
2. Coverage Strategy
3. Payload Configuration
4. Risk Comparison

In this example, it should be noted that the maximum capacity of an L-band system using 10 kHz per user and 5 MHz of spectrum is of 14,000 channels (compared to 33,600 channels for S-band) due to the maximum frequency reuse factor of *28* achievable over the 65 deg^2 coverage area

Table 2 Summary Comparison of L and S-Band.

Item	S-Band	L-Band
Tx Band (MHz)	2170-2200	1525-1559
Rx Band (MHz)	1980-2010	1626.5-1660.5
Target Capacity (channels)	12,000	12,000
Channel Bandwidth (kHz)	10	10
Tx-Rx Separation (MHz)	190	101.5
Spectrum Segmented	No	Yes
Available Spectrum (MHz)	6	5
Spectrum Spread (MHz)	6	34
Coverage Area (deg^2)	65.0	65.0
Area per Tx HPBW (deg^2)	0.51	1.03
Maximum Freq Reuse (F)	≅ 56	≅ 28
Max Capacity (Spectrum)	33600	14000
Min Number of beams	127	63
Freq Reuse Factor Reqrd	20	24
Beams per Cluster (Nbc)	7	9
Number of Beams (Nb)	140	216
C/I performance	Highest	Lowest
Number of elem (Tx&Rx)	144	2 X 72
Number of SSPAs	144	72
Average gain (dB)	44	42
Total Tx power (norm)	1	0.8
SSPA power rating (norm)	1	1.6
PIM Order	63	5
Single Antenna Possible	Yes	Risky
Multipactor Risk	Lowest	Highest
OBP complexity	Lowest	Highest

CONCLUSION

The available spectrum at L-Band tends to be narrower than that which is available at S-Band, resulting in a higher frequency reuse factor requirement for the L-Band system to support the same capacity. Because of its wider beams, the maximum frequency reuse factor achievable at L-Band is nearly half of that which can be realized at S-Band, as highlighed in **Table 2**. These two constraints yield a significant spectrum limitation for the L-Band system.

High capacity systems also have higher spectrum requirements. The high spectrum requirements are incompatible with the scarce L-Band spectrum and the lower frequency reuse factor yielded by its wider beams.

An advantage of the L-Band system is that it only requires about half the elements needed at S-Band. This reduces the number of SSPAs required and yields some cost savings.

Because there are fewer elements in the L-Band feed array, the power handling of the L-Band Tx chains is higher. The higher power combined with a lower multipactor threshold for L-Band yields a much higher multipactor risk.

The significantly lower L-Band PIM order results in much higher PIM risk. This increased PIM risk warrants special precautions in the fabrication of the filter/radiating element and reflector. It also makes it extremely risky to implement a common Tx/Rx antenna at L-Band. The implementation of a second reflector antenna and feed array assembly result in a significant increase of the payload mass and cost.

As identified in **Table 2**, the only advantage of L-band over S-band is the lower number of SSPAs required. It is evident that S-band yields significant advantages in the capacity, coverage strategy and risk areas. For all these reasons, it is strongly believed that S-Band will be preferred to L-Band for most future mobile communications systems.

ACKNOWLEDGEMENT

The authors wish to thank Mr. Sylvain Richard, from Spar, for his contribution to the Multipactor section of this paper.

REFERENCES

[1] **A.B. Salmasi**, A parametric Analysis of Performance Characteristics of Satellite-Borne Multiple-Beam Antennas, JPL Publication 80-52, 1980.

[2] **D. Boulanger**, M-SAT PIM Control Plan, Spar Publication 814611, 1991.

ACeS Antenna Feed Arrays

M. Forest, S. Richard and C.A. McDonach

Spar Space Systems
21025 Trans-Canada Highway, Ste-Anne-de-Bellevue, Québec
CANADA, H9X 3R2 Tel.: 514-457-2150, Fax: 514-457-2724
e-mail: mforest@spar.ca

ABSTRACT

Today's mobile in-orbit communication systems require state-of-the-art antenna technology that meets stringent requirements for power handling, passive inter-modulation, gain and mass to provide a link between the satellite and the mobile users. This paper presents a description of the feed arrays of the ACeS L-band antennas which provide mobile communications over a southeast Asian coverage. Emphasis is on the feed array and filter design trade-offs that were made to meet the stringent mass and RF performance requirements. Also addressed are the power handling design features that ensure PIM free and multipactor free operation. Low and high power test results are also discussed.

INTRODUCTION

Spar Space Systems has considerable experience in the design, fabrication and tests of high power antenna sub-systems. Building on the previous heritage [1,2,3], new and significantly improved L-band feed components were designed which are used for a new generation of geostationary high power payloads, namely the ACeS program.

Spar is involved in the ACeS program (Asia Cellular Satellite) as the supplier of the transmit and the receive L-band feed array sub-systems to the prime contractor, i.e. Lockheed Martin Missiles and Space of East-Windsor, New-Jersey, who in turn supplies the satellite to the Asian Cellular Satellite Consortium.

The transmit antenna feed array assembly (TAFAA) is a planar array consisting of 88 L-band cup-dipole radiators. Such large number of elements / components is typical of the current state-of-the-art antenna systems requiring advanced manufacturability in a high volume production environment. A bandpass filter is used behind each cup-dipole to provide the required out-of-band rejection. The TAFAA also consists of a feed support structure which interfaces with the spacecraft. The function of the transmit feed array is to receive a pre-determined distribution of RF signals from the matrix power amplifiers and to illuminate a mesh reflector to form 140 spot beams for transmission from the spacecraft to the earth. The receive antenna feed array assembly (RAFAA) is similar to the transmit feed array except that its function is to collect signals from the reflector and to distribute them through the LNA/BFN assembly. Figure 1 illustrates the geometry of the feed arrays.

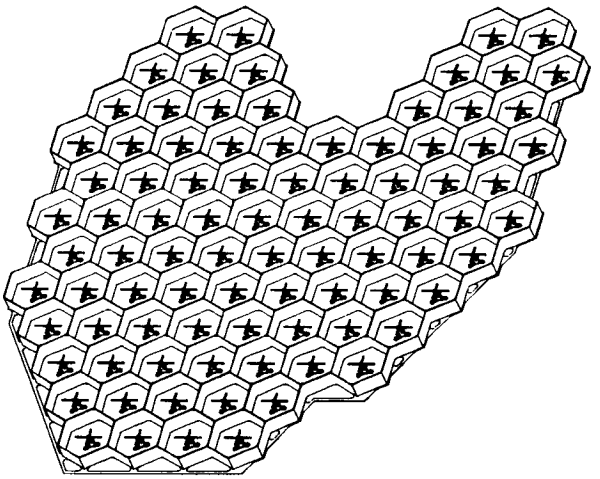


Figure 1 Transmit Feed Array Model

REQUIREMENTS

The operating bands for the two feed arrays are the standard L-band mobile frequency allocation, namely:

- 1525.0 to 1559.0 MHZ for Tx
- 1626.5 to 1660.5 MHZ for Rx

The minimum gain of the feed/filter element over these bands is required to be 10.2 dB and 10.5 dB for the transmit and the receive elements respectively. In addition, each element is required to meet the maximum axial ratio of 1.2 dB on boresight.

The maximum insertion loss of the bandpass filter is 0.5

dB and 0.6 dB for the transmit and receive bands respectively. There is also an 80 dB rejection specification between the transmit and the receive bands.

Being on a geostationary orbit, the ACeS L-Band antenna subsystem must offer high gain to communicate with small mobile earth handsets which have limited output power and low gain. The operating power of the transmit antenna is therefore very high. These considerations result in stringent power handling and thermal requirements. In addition, very low PIM levels must be achieved as a result of the high transponder gain. This translates in a maximum PIM level requirement of -128 dBm (with two 80 watt CW carriers). In terms of power handling, each feed/filter element is required to handle an average and peak RF power of 170 watts and 732 watts respectively without multipactor breakdown.

The above electrical requirements have to be achieved within a very stringent mass budget. The mass requirements of the transmit and receive feed arrays are 53 kg and 40 kg respectively.

DESIGN CONSIDERATIONS

The radiating elements are made of hexagonal cups in which cross-dipoles are mounted, providing a left-hand circularly polarized (LHCP) pattern. The cup height and the geometry of the dipole were optimized for gain and axial-ratio performance.

It was decided early in the program to proceed with an integral cup-array panel to reduce the risk of passive intermodulation (PIM). The 88 cup panel was therefore machined integrally from a single piece of aluminum. For the same reason, each dipole and each bandpass filter were also machined as integral parts. A series of anti-buckling rib patterns were designed at the back of the cup/array panel for optimum strength and mass performance.

The bandpass filters are mounted directly on the back of the cup/array panel and coupling to the dipole is provided by means of a coupling probe at the end of the dipole transmission line. The bandpass filters produce a quasi-elliptical transfer function to maximize the isolation over the receive or transmit band. The filter is implemented as a six-pole coaxial cavity design with coupling irises between the cavities to realize the desired transfer function. Metallic tuning screws are used in each cavity of the receive filters. However, for the transmit filters, an innovative tuning post design was implemented. This new technology yields an integral all-metal filter assembly with no metal to metal interface. This is a major advantage for power handling considerations (PIM and multipactor).

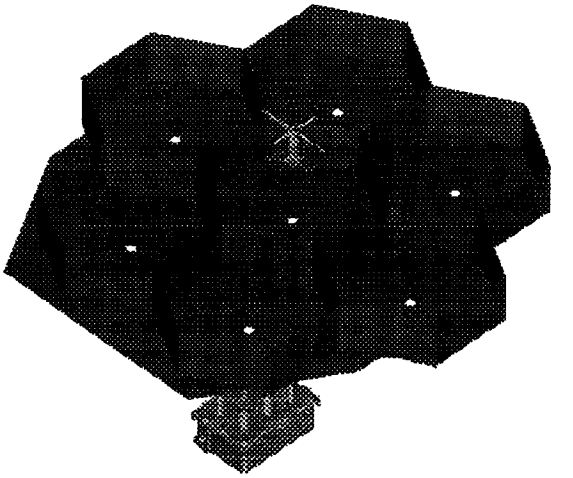


Figure 2 Seven Element Array Concept

Once all filters are assembled to the cup/array panel, a close-out panel is bonded to the rear resulting in a structural assembly. This assembly is then integrated to a third structure for attachment to the spacecraft.

Figure 2 shows a view of a seven element array illustrating the design concept.

Model Philosophy

In order to minimize risk to the program, a conservative model philosophy was adopted to make sure that all performance targets were met and that the manufacturing and assembly processes were validated.

The first step in the development plan was the breadboard phase. During this phase, two bandpass filter units (one Tx and one Rx) were built to determine internal dimensions required to achieve the RF response. Also, an 88-element adjustable cup-dipole breadboard array was built to determine the individual element patterns in the array environment. This array was tested in Spar's compact antenna test range (CATR) to optimize the dimensions of the radiating elements for optimum gain and axial-ratio performance in the array environment. In addition, it was used to validate a new state-of-the-art automated test chamber for field aperture probe test. This facility is based on a six degree of freedom gantry robot for probe positioning. Automated measurement is done through accurate positioning of the probe in front of each element of the feed array. Data acquisition of on-axis frequency response is also done automatically with a computer controlled network analyser. Using an auto-calibration procedure for DUT

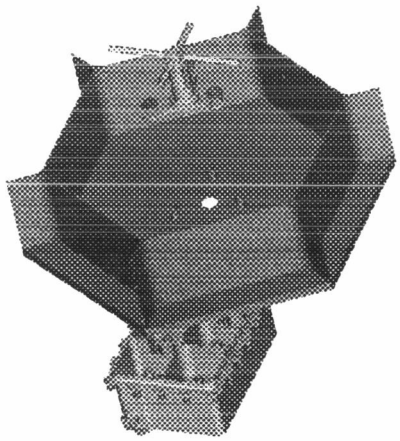


Figure 3 BPF / Cup-Dipole Test Assembly

localisation, the entire 88-element array frequency response (including filters) can be measured within 8 hours.

The next step in the development plan was the elegant breadboard phase (EBB phase). Both transmit and receive bandpass filter units were built as integral one piece items to validate the RF performance (Figure 3). These units were also used to verify the high power performance (PIM and multipactor). At this point, final modifications to the design were made to further improve performance margins.

In order to prove the manufacturing processes and methods, a seven element cup-dipole array was built during the manufacturing validation phase (MVM phase). The MVM is fully representative of the flight hardware, thus validating the entire manufacturing and assembly processes prior to flight hardware manufacturing. This phase was particularly useful to refine the innovative machining techniques that were developed to achieve error free machining in such high volume production quantities. PIM tests were also conducted successfully on this representative feed array.

Once all design aspects were finalized, two transmit qualification cup-dipole / filter assemblies (EQM) were manufactured per flight specifications. These two units were submitted to a rigorous qualification process which involved low power as well as high power PIM and multipactor tests and environmental testing.

Flight hardware production could then proceed with validated manufacturing and measurement processes.

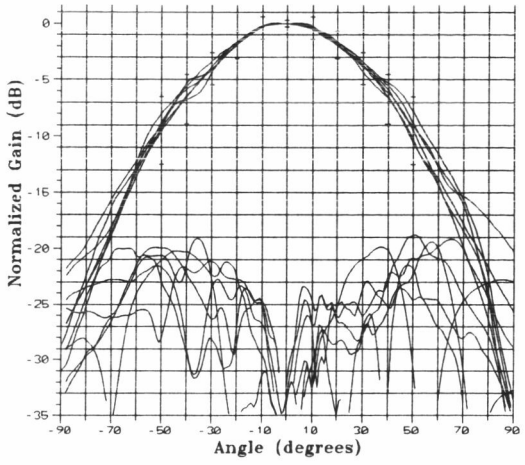


Figure 4 Typical Primary Pattern of TAFAA

PERFORMANCE

Cup-Dipole Elements

The main performance requirements of the cup-dipole element is a high gain and a low axial-ratio over the operational band. Both parameters were measured in Spar's compact antenna test range (CATR) with the full size 88-element breadboard array and were found to be within the specifications. Typical performance for the normalized radiation pattern of the cup-dipole is shown in Figure 4. Requirement for gain and axial-ratio were met with margin.

Band-Pass Filters

The bandpass filters are designed for minimum insertion loss and maximum isolation. The measured insertion losses of the transmit and receive BPF over the operating temperature range are 0.50 and 0.42 dB respectively, meeting the specifications. The isolation over the transmit or receive band is 85 dB minimum. Figure 5 shows a typical frequency response of a transmit bandpass filter.

The filters are also required to provide substantial rejection at low frequencies (between 1 and 1.45 GHz) as well as at C-band and X-band. All BPF isolation specifications were met and generally exceeded with significant margin.

PASSIVE INTERMODULATION

By design, the ACeS transmit antenna feed array is free of any measurable passive intermodulation (PIM). This results from the following design features:

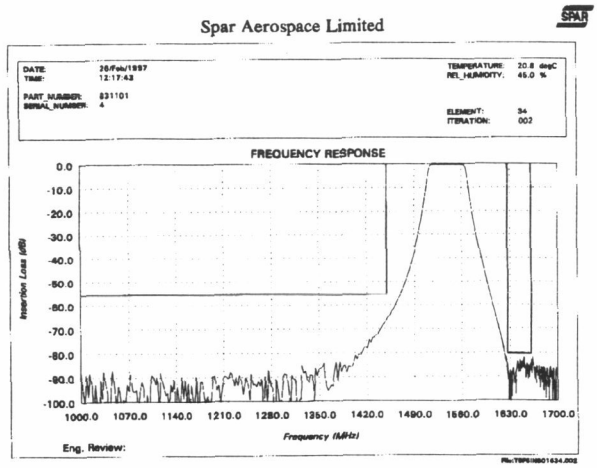


Figure 5 Transmit BPF Typical Response

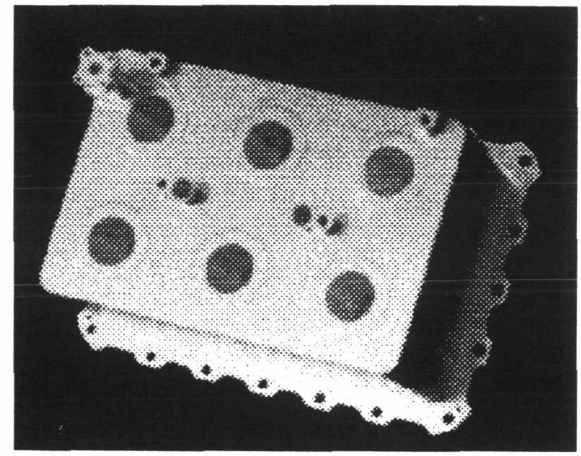


Figure 6 Receive Bandpass Filter

- integral cup panel assembly
- one piece dipole and feeding transmission line
- bandpass filter also made in one piece and directly mounted at the back of the panel

In order to produce PIM free hardware, a comprehensive PIM control plan has been implemented. Based on past experience, this control plan was put in place to ensure high quality of workmanship during the flight assembly process. All personnel involved in the design, fabrication, assembly, tests, inspection or quality assurance of PIM critical hardware were trained and provided with continuous technical support. Special care was given to the handling, the protection and the storage of critical parts. A portion of the integration area was dedicated to the assembly of the feed array and was restricted to PIM trained personnel only. Finally, rigorous inspection processes (called "PIM gates") were implemented to detect any anomalies prior to the final assembly of the feed array.

PIM Test Plan

PIM tests were done on the EBB, the MVM and finally the EQM units. During the EBB phase, the objective is to verify and refine if necessary the component design through experimentation (i.e. dipoles / filters). The MVM PIM test provides a verification of the overall assembly performance. Finally, the EQM phase fully qualifies the flight design.

For PIM testing of these units, two 80 watts carriers (in the transmit band) are routed to the input port of the DUT which is placed into an RF transparent thermal enclosure inside an anechoic chamber. The intermodulation signals are monitored using a receive probe in front of the radiating assembly. The measured signal is filtered and amplified using high rejection bandpass filters (Rx filter) and low noise amplifiers in cascade. Two different PIM orders can be measured simultaneously, namely the 5th and the 7th orders. The PIM signals are detected using highly stable spectrum analysers. The thermal enclosure allows testing between -50°C to +120°C

PIM Test Results

The first PIM tests were done during the elegant breadboard phase (EBB phase) where two EBB bandpass filter / cup-dipole elements were PIM tested over more than 23 thermal cycles. No PIM was measured above the test set-up noise floor (-140 dBm) during all these thermal cycles.

During EQM and MVM phases, similar results were obtained, thus confirming the PIM free performance of the design.

MULTIPACTOR

The following design features were implemented in the ACeS transmit antenna feed in order to ensure multipactor free operation:

- no small gaps used in high field regions
- gap size in high power/high field areas sufficient to keep the operating voltage well below the multipactor threshold voltage
- all metal components, i.e. no dielectric material exposed to high field levels, therefore eliminating the risk of outgassing and corona discharge

The risk of multipactor is inherently further reduced by the comprehensive PIM control plan where special care

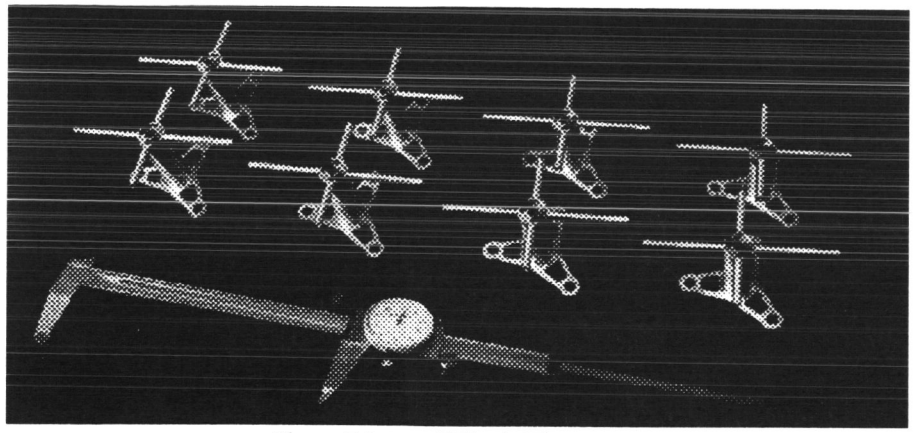


Figure 7 Integral Dipole Elements

is given to the handling, the protection and the storage of critical parts in order to eliminate the risk of contamination.

Multipactor Test Plan

Multipactor tests were done on the EBB and the EQM units. The EBB phase is a development phase where the objective is to determine the multipactor threshold of the components and modify their design if necessary. The EQM test phase provides a final qualification of the flight design.

A high power pulse source is used for multipactor testing. The tests are performed at hot with the full operating average power (170 watts).

Multipactor Test Results

The first multipactor tests were done during the elegant breadboard phase (EBB phase), where two EBB bandpass filter / cup-dipole elements were multipactor tested. Multipactor threshold was observed to be 900 watts which meets the 732 watt specification.

During the EQM phase, similar results were obtained, thus confirming the multipactor performance of the design.

CONCLUSION

The ACeS feed is considered to be the first of a new generation of high power mobile communication satellite antenna feeds. Two state-of-the-art L-band antenna feed arrays have been designed and qualified as part of the ACeS program. Full compliance with the specifications was achieved. Innovative component design and manufacturing techniques have been utilized resulting in significant improvements over past programs.

Difficult technical issues such as PIM and multipactor were successfully addressed through innovative design concepts and manufacturing techniques, as well as an extensive qualification program. This new feed technology yields considerable benefits in a high volume production environment. The flight feed arrays will be delivered in the fall of 1997.

REFERENCES

[1] S. Richard, K.N. Patel, Y. Patenaude, J.P. Langevin, J. Wang, *M-SAT L-Band Antenna Feed Arrays,* 15th International Communication Satellite Systems Conference, San Diego, February 1994.

[2] Y. Patenaude, E. Amyotte, P. Illot, F. Ménard, S. Gupta, *M-SAT L-Band Antenna Subsystem,* 15th International Communication Satellite Systems Conference, San Diego, February 1994.

[3] Y. Patenaude, S. Richard, F. Ménard, E. Amyotte, C.K. Mok, *M-SAT PIM and Multipactor Test Program,* International Mobile Satellite Conference, Ottawa, 1995.

The AstroMesh Deployable Reflector

Mark W. Thomson
Astro Aerospace Corporation
6384 Via Real, Carpinteria, CA 93013-2920
Phone: (805) 684-6641 FAX: (805) 684-3372

ABSTRACT

Astro Aerospace Corporation has developed the AstroMesh deployable reflector (notice of patent allowance granted) to address the demanding requirements of mobile satellite communication systems. The technology is mature, with models of approximately 6 and 12 meters offset circular aperture fully developed. Both models have been successfully validated with a full range of qualification tests to assure performance during launch, deployment and extended service in geosynchronous orbit. Each reflector passed rigorous passive inter-modulation (PIM) tests in which significant margins on the most demanding PIM specifications were achieved.

The AstroMesh reflector can be directly scaled to apertures of approximately 40 meters yet be integrated with communications systems that are launched in medium launch vehicles. Existing large launch vehicles could accommodate systems with apertures up to approximately 150 meters without fundamental design changes. The design is capable of unusually low deployed mass, stowage volume and cost, yet is very robust during and after deployment. It is compatible with L-band communications antenna systems to all feasible aperture diameters and with Ka band systems to approximately 10 meters.

ASTROMESH CONCEPT

Astro Aerospace Corporation has developed the AstroMesh reflector to make a revolutionary advance in deployable mesh reflector technology. Every stride in the development of the reflector, from structural concept to the use of highly specialized materials, has been driven by the unique requirements of offset-fed, extremely PIM-sensitive geostationary mobile communication satellite systems. The elimination of PIM products and electrostatic discharge (ESD) events has been achieved in every component of the deployed reflector, regardless of location. This elegant and readily scalable design supports low-cost, minimum schedule programs with rapid hardware delivery for commercial competitiveness. The 6-meter reflector hardware, designed and built according to MSAT main antenna specifications, is shown in Figure 1. The 6-meter reflector shown in Figure 1 provides an RMS surface accuracy of less than 0.6 mm from all sources and environments of which 0.4 mm is manufacturing error. All RMS values are calculated using the Jet Propulsion Laboratory's half-path length (HPL) error method.

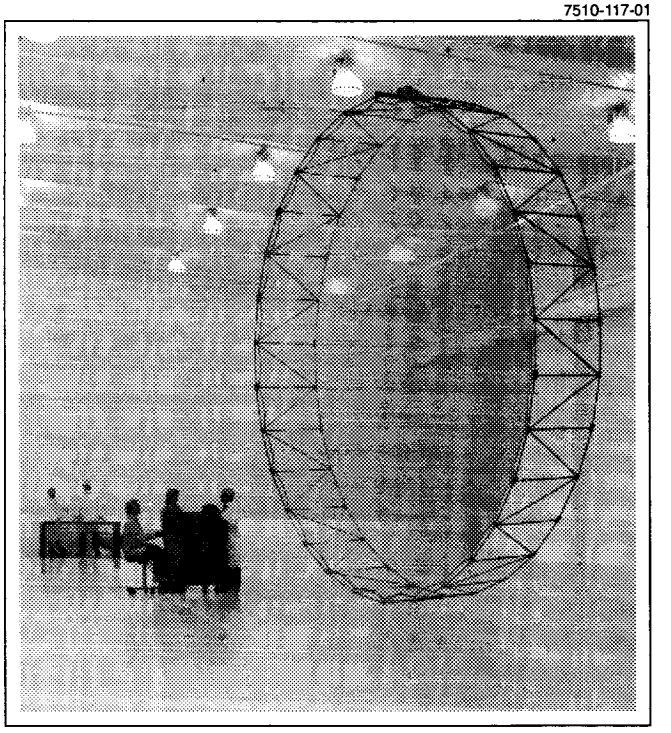


Figure 1. 6-Meter MSAT-Compatible Reflector.

The AstroMesh reflector is a lightweight and inherently stiff structure that precisely and repeatably deploys the reflective mesh surface regardless of environments. As shown in Figure 2, the reflector consists of stiff, doubly curved composite membrane structures called nets, placed back-to-back across a deployable ring truss. Tension tie assemblies evenly apply approximately normal forces between nets to permanently preload the stiff nets in tension. This forms a rigid drum-like structure that possesses exceptional structural efficiency, thermal dimensional stability and stiffness-to-weight ratios. Figure 3 provides additional reflector component nomenclature.

Highly RF reflective gold-plated molybdenum mesh is stretched across the convex side of the front net structure. This forms numerous flat facets in the mesh to approximate the required offset parabolic shape with the required precision. Higher surface accuracy is achieved by increasing the density of the net structures and decreasing the web thicknesses. Similar mesh continues service in orbit after over 20 years

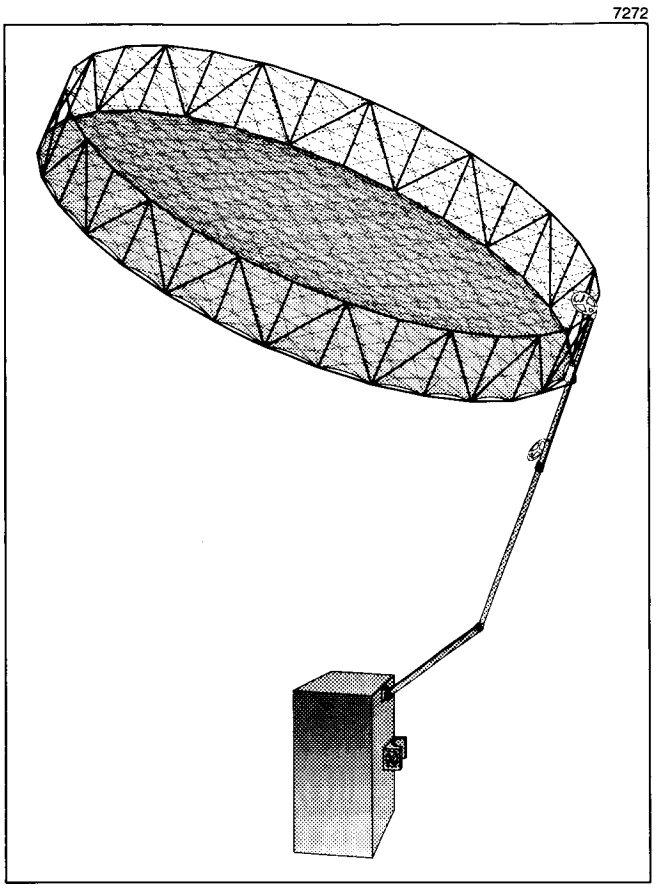


Figure 2. Antenna Subsystem with AstroMesh Reflector.

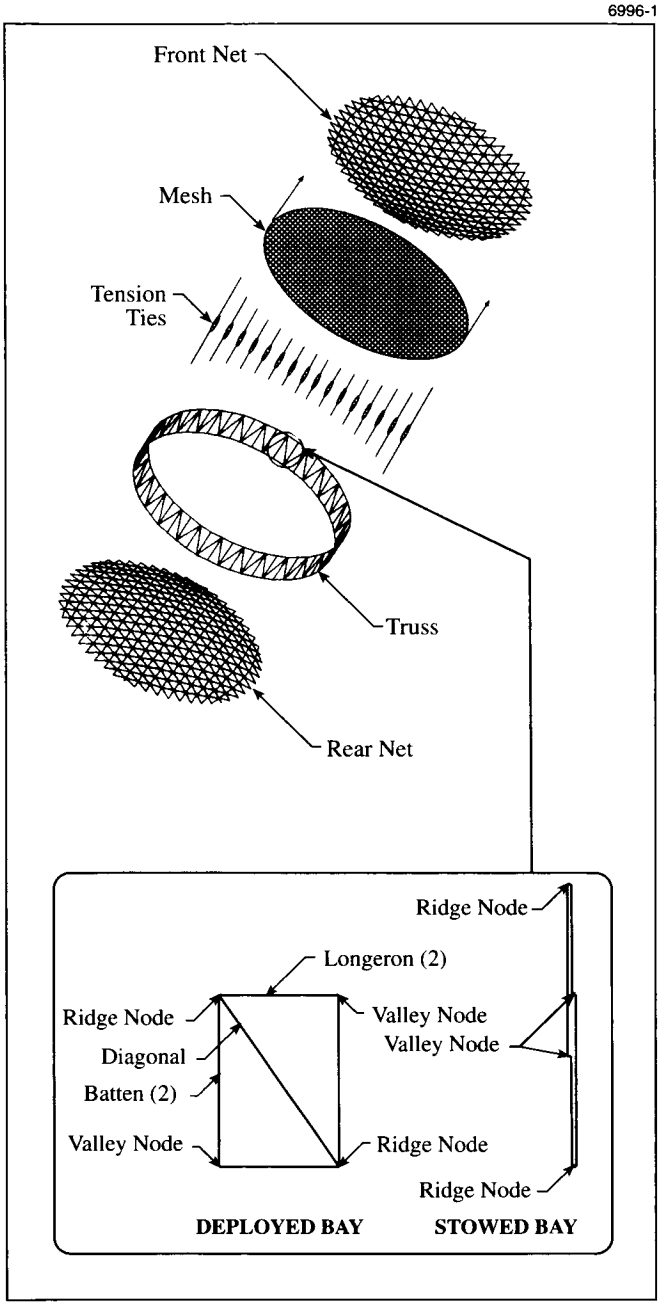


Figure 3. AstroMesh Reflector Component Nomenclature.

on NASA's TDRS system and in other flight systems. Recent testing of full-scale AstroMesh reflectors and numerous samples proves that mesh is viable and robust in PIM-sensitive antenna applications with appropriate design and fabrication measures. Using the most open mesh knit that is practical, insertion losses are less than .01 dB in L-band applications and transmissivity is typically below -35 dB. Overall reflector transparency (optical or aerodynamic) exceeds 80 percent of the frontal area.

DEVELOPMENT STATUS

The AstroMesh technology is the culmination of five generations of AstroMesh reflector hardware development over a period of 6 years. Reflectors from 2.5 to 12 meters have been built and tested to varying extents with a record of over 160 deploy and stow cycles.

The development of a PIM-free, large aperture AstroMesh reflector for space flight was recently demonstrated on a short schedule. A 12-meter reflector assembly was designed and built in 12 months, then qualified for a flight program. Astro believes that the attributes of the AstroMesh make it ideal for the main reflectors of any communications, radar or other large aperture antenna system operating up to and including Ka band.

The reflector concept evolved from a 25-meter aperture class reflector design study in 1990. The design progressed to hardware as a nondeployable 2-meter structural demonstrator in 1991, then a deployable 2.5-meter structural model with a metallic net and no mesh.

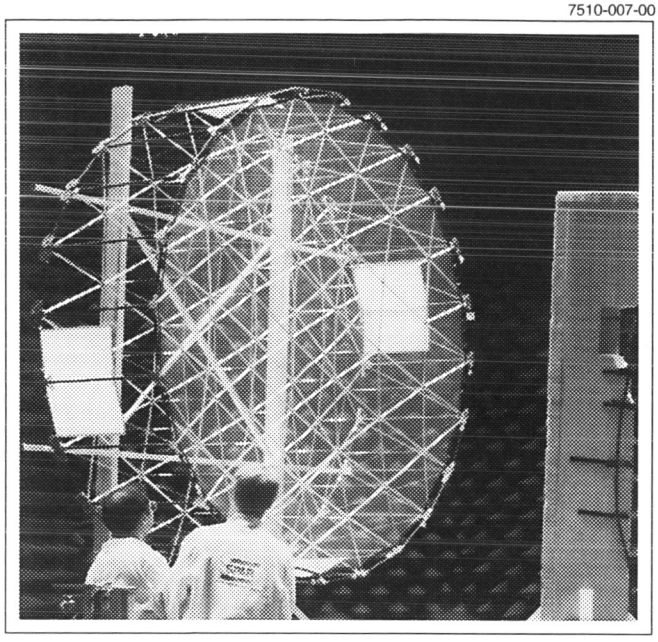


Figure 4. 2.5-Meter AstroMesh Reflector Range Testing at Spar Aerospace Limited in Montreal, Canada.

2.5-Meter Reflector

The 2.5-meter structure was then completely modified to include kinematic coupling in the deployable ring truss, flight-like mesh and RF transparent thermoplastic composite webs in the net membrane structures. This model was pivotal in validating deployment behavior including mesh management potential with 115 total deployments. No degradation of the hardware or mesh occurred. Actual RF patterns were measured in a compact range with this hardware up to and including X-band, as shown in Figure 4. A high degree of correlation to the predicted patterns was achieved.

6 and 12-Meter Reflectors

Two large-scale versions of the reflector, the 6 and 12-meter models, have been built with materials appropriate for space flight and successfully qualified by a suite of electrical and environmental tests. Table 1 gives an overview of the type of tests the large aperture models have undergone. Performance values that were achieved by the 6-meter model are included.

The 6-meter design given in Figure 1 was built with Astro funds using the MSAT main reflector specification as a model. It was built and tested during 1994 and 1995. The approximately 12-meter reflector was designed and built after completion of the 6-meter project. The 12-meter test performance met or exceeded levels that are typical of geostationary mobile satellite systems including the reflector, boom assembly and launch restraint devices. Confidentiality agreements preclude the publishing of any specific 12-meter design details or test results.

Several notable qualification tests were performed on the 6-meter reflector. During bulk thermal testing, a worst-case contour distortion of 0.25 mm RMS was predicted between test temperature extremes of -160°C to +120°C. The actual distortions at temperature could not be discerned by a video photogrammetry system with a resolution within 0.05-mm RMS. The same system could not discern any differences in RMS between cup-up and cup-down contour measurements. Based on test results and 1-g static analysis, cup-down contour measurements were deemed unnecessary.

Table 1. Large Scale AstroMesh Reflector Test History.

Program	Total Deployments	Photogrammetry Cup Up	Photogrammetry Cup Down	Contour Repeatability	Thermal Distortion with 250° C Gradient	Thermal Cycling	Sine & Acoustic Vibration	PIM	Thermal PIM for 6 & 12-Meter	PIM with Vibration	Deployment with Spacecraft ACS Loads	Deployed Frequency	Failure Modes	Transportation	Launch Acoustic
6-Meter Reflector Level	36	✔ 0.41-mm RMS	✔ (No measureable RMS differences)	✔ <0.1 mm RMS	✔ Not measureable	✔ -140° C to +135° C	✔ 3 axis sweeps to 50-g peak response	✔	✔ -165 dBM	✔	✔	✔ 2.0 Hz cantilever fixed boom interface	✔	✔	✔ Typical
Component Level								✔ samples		✔ samples					
12-Meter Assembly Level	14	✔		✔		✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Component Level						Boom only	Boom & all bonded assys with proof test		✔ samples		✔				

GENERAL TECHNICAL DETAILS

PIM and ESD Design

Astro has been successfully developing PIM-free deployable structures, mechanisms and reflectors for the very demanding PIM environments of geosynchronous mobile satellite telephony since 1992. Every aspect of an AstroMesh reflector has been designed, manufactured and validated to be PIM-free and noncharging in the presence of space plasma. PIM performance margins are well below typical communications systems noise floor levels of -150 dBM. Additional margin below the noise floor can range from an additional 15 dBM minimum for a 6-meter MSAT-type system to 35 dBM or more in 15-meter class designs, depending upon surface power levels and geometry. Experience gained developing two flightworthy designs has established consistent and highly effective PIM and ESD guidelines for every stage in a program. Formal PIM design and manufacturing approaches, processes and procedures are in place at Astro.

Mesh PIM Design

The primary method of AstroMesh reflector PIM control is inherent to the design. The AstroMesh reflector's structural insensitivity to mesh-induced forces makes it possible to install a monolithic piece of mesh which contains 100 percent continuous wire filaments within the mesh perimeter. No wires are cut or broken within the aperture area during manufacture of the reflector. In addition, overall deployed mesh tension levels can be high. The average tension value in typical gored mesh reflectors is approximately 2.6 N/m. The AstroMesh reflector has demonstrated an average tension field of over 11 N/m without deleterious effects on surface contour performance. Controlled PIM tests on mesh samples have confirmed 5 dBM or more additional margin with high tension mesh.

The 6-meter flight reflector was assembled with a single piece of mesh. Apertures of about 7.5 meters to 15 meters are assembled with a single high-performance seam along the major axis (due to current knitting machine size limitations). As in the remainder of the aperture, there are no cut or broken wires in this proprietary seam, because it joins the knitted selvage. Apertures of over 15 meters to 22.5 meters would require two seams parallel to the major axis and so on. Sample tests have proven that the Astro seaming technique yields uniquely robust PIM performance. Compared to the types of seams necessary in gored reflector designs, where the mesh must be cut, the AstroMesh provides an additional 5 dBM minimum PIM margin. Aperture edge treatment necessarily incorporates cut wires, but the cut wires are sandwiched in a proprietary PIM-free, flexible composite sandwich.

Structure and Mechanism PIM Design

In the remainder of the AstroMesh reflector structure, every component and assembly is designed and manufactured to be PIM-free. This is particularly important in the front-facing portion of the rim truss, where RF spillover might otherwise cause the generation of PIM products from the bonded graphite composite and aluminum deployable truss structure. Although the mesh is an excellent shield, components in the reflector RF "shadow" are treated the same, with a zero-PIM design philosophy throughout. This area is where Astro has spent the bulk of its PIM development activities. The primary reason is because of the need to eliminate ESD.

ESD Design

Astro's historical goal is to provide reliable, high conductivity bleed paths everywhere in a deployable structure when ESD is a concern. If the structure must also be PIM-free, that goal is difficult to achieve. The primary source of PIM products is nonlinear electrical conductivity within certain materials and between conductive components that are in contact. In a conductive structure, the larger a component is relative to the RF wavelength, the better "antenna" will it make in order to generate more energetic radiated PIM products. Since the RF induced electrical currents from these "antennas" are able to flow through ESD or other conductive interface paths, all interfaces must behave in a linear, stable manner. Unfortunately, conventional design details, materials and workmanship methods do not reliably meet this requirement. Paradoxically, one of the strengths of gold-plated mesh in a PIM environment is the uniform conductivity that is provided from a multitude of high-contact stress electrical contacts between closely spaced wires.

For the AstroMesh, Astro adopted a different approach to ESD control: The structure is designed to provide sufficient conductive coupling between any two conductive components to reliably bleed electrostatic charge potential while avoiding conductive paths that are potentially nonlinear or unstable enough to generate PIM. In order to implement this philosophy, a large repertoire of PIM-free high-performance space-grade materials has been developed and/or tested by Astro for use in a wide variety of applications within the reflector assembly. For instance, vacuum corona chamber testing of actual material charging voltages was performed on many low-conductivity thermoplastic and thermoset materials. Charging potentials of 100 volts or more in any environment are criteria for rejection.

Numerous other structural design and material usage methodologies have been developed and validated to eliminate PIM and ESD concurrently. A process of cycling each design feature through coupon, component and subassembly PIM sample build-up tests has been used with consistent success at Astro for developing electrically dissipative yet PIM-free hardware. First-time success in subsequent full-scale testing is typical.

Mass

The primary load-bearing support structure in the deployed AstroMesh reflector is the rim truss. Longeron members in the rim are in compression from the opposed, well-tensed nets and relatively large deployed static loads are developed. These loads are distributed in relatively short composite tubes

with small diameters and thin walls for high weight efficiency. The filamentary net structure and mesh are the only elements within the rim truss diameter. They represent about 10 percent of total reflector mass. The efficient truss and net design gives the AstroMesh unusually low mass and dramatic downward trends in mass-to-aperture area ratios as diameter increases.

All of the fixed mobile communications satellite systems made public at the time of writing have combinations of aperture and operating frequency that are readily accommodated by the AstroMesh. Rim truss and web members that would be required in each case need not display near-zero coefficients of thermal expansion to meet surface contour requirements in thermal extremes. Composite member lay-up design can thus be optimized for the required strength or stiffness with minimum weight.

The 6-meter reflector-only mass is 14.5 kg, and the present state of the art could lower that to approximately 11 kg. With favorable offset geometry (F/D > 0.75), 12 to 14-meter reflector mass is about 40 kg to 50 kg. Figure 5 shows reflector mass as a function of aperture size and F/D.

A complete AstroMesh antenna subsystem would include the deployable boom and thermal PIM blankets, all tie-downs and release mechanisms. The complete subsystem mass is approximately 1.7 to 2.3 times reflector mass for an F/D of 0.5 to 0.9 respectively, primarily due to boom length. The actual mass multiplier could vary significantly depending upon spacecraft (S/C) bus and launch vehicle accommodations, required deployed natural frequency and other system-specific parameters.

Stowed Volume

The AstroMesh reflector is stowed in a very compact package. Figure 6 shows an example of a stowed and deployed AstroMesh reflector. Truss members are stowed around the circumference of the stowed mesh and nets. The truss fittings or nodes are preloaded against each other or lightweight hoop structures to form a stiff barrel-like structure that is strapped into composite honeycomb cradles for launch. Aperture size to stowed length ratios vary from 3.4 to 4.2 as F/D goes from roughly 0.5 to 0.9, respectively. The deployed to stowed diameter ratio varies from roughly 15 to 25 as aperture varies from 6 to 25 meters. As with mass, actual stowed dimensions will vary depending on geometry and other system parameters. Various booms are available; folding or telescoping designs would be selected based upon length and/or S/C bus interface configuration.

Stiffness and Precision

The stiff, filamentary Aramid thermoplastic composite net webs are located to precisely conform to the required parabolic geometry with a built-in surface metric. The intersections of these uniaxial pultruded webs are located to within one part in 10^6 with proprietary manufacturing techniques. Surface figure adjustment after fabrication is not possible nor is it

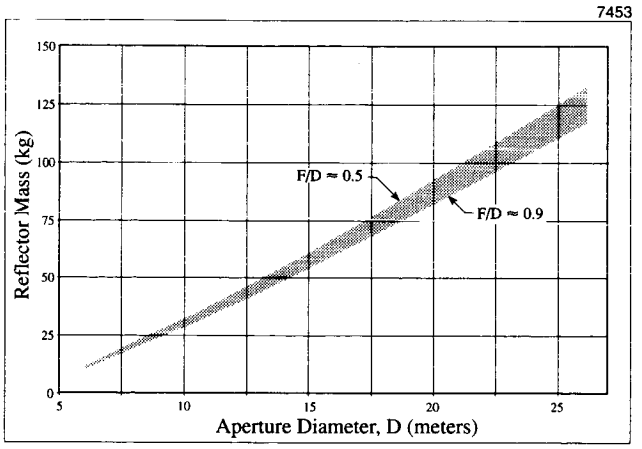


Figure 5. Reflector Mass as a Function of Aperture Size and F/D.

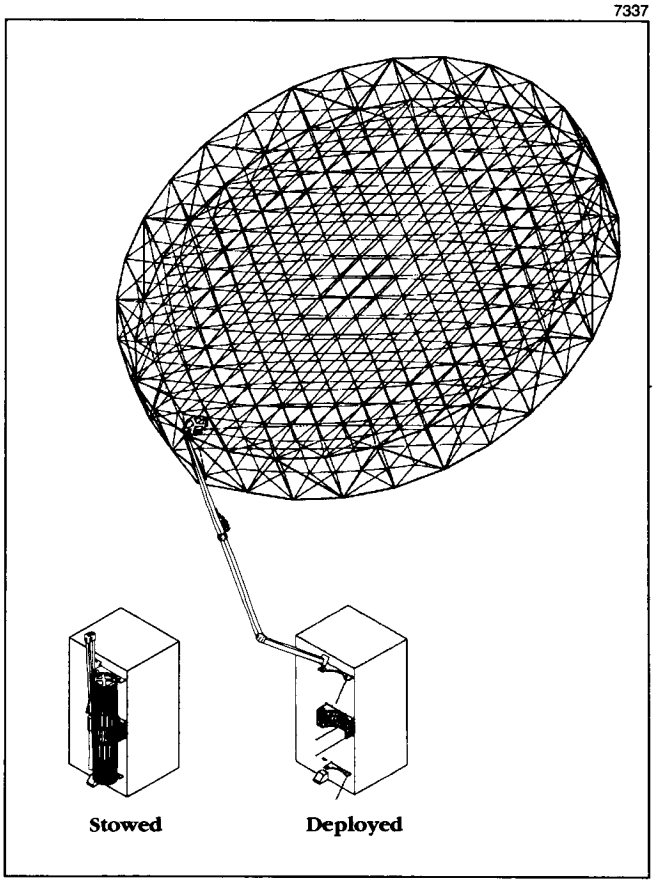


Figure 6. Stowed and Deployed AstroMesh Reflector.

required. The maximum D/W_{RMS} value practical with the AstroMesh technology is approximately 5.0×10^4 where D is aperture diameter and W_{RMS} is the HPL RMS error.

The mesh is held in tension behind the front net at about one-tenth of the total net loading value or less, about 8 to 12 N/m average lineal tension (three to four times the industry standard

for antenna mesh tension). This causes the mesh to be gently shaped by the net with corresponding accuracy. The highly tensed net structure is not significantly distorted by the relatively light mesh loading.

The combination of the tensed nets and the rim truss leads to the potential for very high deployed natural frequencies. The first cantilever mode of the reflector if fixed to an infinitely stiff boom varies inversely from approximately 2.0 Hz to 0.5 Hz for apertures of 6 to 25 meters, respectively. Such stiffnesses would not typically be approached in practice because of the additional boom mass that would be incurred. The greatest benefit of high global reflector stiffness is the elimination of significant gravity sag. Contour in gravity will be within 0.1 mm RMS of the zero-g value up to a 15-meter class design. Only one cup-down measurement is recommended on a given reflector project to validate the gravity sag behavior. This significantly reduces test cost and schedule for the remainder of the program.

Stowed reflector stiffnesses are such that such it is possible to achieve first major mass modes in vibration of 20 to 60 Hz inversely proportional to the deployed aperture sizes of 25 meters down to 6 meters. An optimum interface arrangement is shown in Figure 6. If it is necessary to allow the upper end of the stowed reflector to extend beyond the S/C structure, the free end is readily stabilized by the combination of the stowed boom and an additional deployable bipod (not shown).

Thermal Stability

The "deep" structural quality of the AstroMesh truss makes it quite insensitive to on-orbit thermal gradients, as previously mentioned. This makes it possible to design the composite truss tubes and web material in the nets without excessive emphasis on zero CTE materials. No thermal insulation is required for any reflector component. The deployable support boom would necessarily have thermal PIM blankets.

Deployment and Mesh Management

The reflector and mesh both have very robust deployment characteristics. Although this paper will not describe details of deployment, the deployment drive mechanism is given powerful geometric authority over reflector truss kinematics. The mesh and nets have been developed to be snag-free during deployment, with approximately 160 successful ground deployments completed. Deployment kinematics and dynamics are such that net/mesh and truss assemblies are decoupled during deployment. The truss acts as the launch containment structure for reflector net and mesh elements. The mesh is gently hung within the stowed ring truss, which acts as a hollow cylinder or barrel with debris-shielding lids on both ends. Truss compression when deployed eliminates all compliance from clearances in the pin joints that are employed. The joints can thus be designed with ample clearances for free-running diametric fits at all thermal extremes of deployment.

Cost and Schedule

The structural morphology of the AstroMesh reflector is very simple from a design and manufacturing perspective. Much of the design process is automated with customized software. Components in the truss and nets are multiple repeats of identical or slightly varying designs. They can thus be economically designed and manufactured with automated equipment. As an example, all the webs for a 15-meter class reflector can be manufactured and assembled into net structures on the truss in 3 weeks by several technicians. Truss assembly from the bonded strut assemblies takes a matter of hours. Most of the effort in an antenna subsystem development program is devoted to interfaces with the S/C such as the boom and tie-downs which are unique for each new S/C and launch vehicle combination. A completely new large aperture reflector can be designed and built to protoflight or qualification standards in as little as 12 months, depending upon diameter.

Reflector assembly can be completed very rapidly with 2 to 4-month spacing on delivery of completed units if needed by the customer, depending upon diameter. The high stiffness and low weight of the AstroMesh make it highly compatible with ground testing. The reflector can be stowed in 2 to 8 hours depending upon size, which eases tight test schedules. The mesh does not require special attention or handling during a stow process. Surface figure adjustments and cup-up versus cup-down comparisons for zero-g figure are not required. Deployment is tightly coupled to a plane by the truss kinematics and is very robust and reliable. Since the reflector structure is relatively insensitive to mesh loading, mesh knitting can take place without extensive stiffness evaluations that are very time consuming.

SUMMARY

The AstroMesh deployable reflector represents a revolutionary advance in the technology of mesh reflectors. The novel and efficient structural concept it employs leads to numerous strengths in the following areas:

- PIM and ESD design
- Mass and stowed volume
- Stiffness and precision
- Thermal stability
- Deployment and mesh management
- Cost and schedule

Every detail of the reflector has been developed specifically for geosynchronous satellite mobile communications systems. The AstroMesh reflector's capability to perform according to the challenging requirements typical of high-gain fixed mobile systems has been proven with flightworthy models having offset parabolic apertures of 6 and 12 meters.

A Study on On-Board Processor for Advanced Mobile Satellite Communications

Osamu Takeda, Shinichi Taira, Yoichi Kawakami, Norikazu Yamasaki, Yuichi Otsu
†Hideki Katagiri, †Shinji Senba
Advanced Space Communications Research Laboratory (ASC)
7F Hayakawa-Tonakai Bldg., 2-12-5, Iwamoto-cho, Chiyoda-ku, Tokyo 101 Japan
Phone: +81-3-5821-5088 FAX: +81-3-5821-5096
Email: takeda@asc.co.jp

† NEC Corporation 4035 Ikebe-cho, Tsuzuki-ku, Yokohama 224, Japan

ABSTRACT

This paper describes the design concept of an On-Board Processor (OBP) for a mobile communication satellite for use early in the next century. This OBP can exchange over 1,000 channels. To realize high flexibility for controlling the channel connection, base-band switching is applied. Most of the functions in the OBP are performed using digital signal processing techniques which have the advantage of reducing the size and power consumption of the OBP by adopting Application Specific Integrated Circuit (ASIC) technology.

INTRODUCTION

Many studies and development activities have been conducted worldwide in regard to the use of satellite communication systems for mobile communications. In Japan, the N-STAR has already been developed, and is now in operation. In addition, the Communications and Broadcasting Engineering Test Satellite (COMETS), which is a research and development satellite aimed at developing new technologies in communications and broadcasting fields, is scheduled to be launched in 1997. Under these circumstances, it seems that the demand for multimedia communications or high-speed communications is growing gradually. Accordingly, an increase of channel capacity is expected to respond to various types of transmission rates and to realize low-cost communications, even in mobile satellite communication systems. Consequently, future OBPs are expected to have large channel capacities, high-speed processing, high reliability, and low power consumption through the development of ASIC technology.

ASC has been conducting research and development in key technology areas such as large deployable antennas, high-power transponders, large-capacity OBP, and miniature mobile Earth terminals[1][2].

This paper presents the current study results of an OBP which is suitable for a mobile satellite communications system using a geostationary satellite.

SURVEY OF OBP

In general, multibeam technology is required to make mobile terminals compact and realize highly efficient frequency utilization. In order to realize a mobile satellite communications system, a satellite antenna using for multibeam; a high efficiency, and high output power amplifier; and an OBP are required. Furthermore, satellite switching between several beams is necessary in the OBP. Satellite switching is divided into radio and intermediate frequency (RF/IF) band switching and base-band switching. Base-band switching is superior to RF/IF band switching for flexible use of the channel exchange. However, the RF/IF band switching is more effective from the viewpoint of hardware scale. Therefore further development of device technology for base-band switching is highly desirable and will resolve the problem of hardware scale.

Table 1 Comparison of OBP functions (GEO System)

Function	ACeS[1]	APMT[2]	ASC[3]	MSAT[4]	ACTS[5]	Japanese projects or organizations: ETS-VI	COMETS	N-STAR	(SCR)[6]	ASC[7]
Base-band switching	11,000ch	16,000ch	16,000ch	2,000ch			16ch		384ch	1,000ch
·Exchange with regeneration	(✓)	(✓)	(✓)	(✓)	✓		✓		✓	✓
·Transparent exchange	(✓)	(✓)	(✓)	(✓)	✓					✓
RF/IF band switching					✓	✓	✓	✓	✓	✓
Digital filtering	✓	✓	✓	✓			✓		✓	✓

1)ACeS: Asia Cellular Satellite System
2)APMT: Asia Pacific Mobile Telecommunication
3)ASC: Afro-Asian Satellite Communication
4)MSAT: Mobile Satellite Communication system
5)ACTS: Advanced Communications Technology Satellite
6)SCR: Space Communications Research Corporation
7)ASC: Advanced Space Communications Research Laboratory
(✓): Not certain

Comparison of OBP functions in Japan and other country's projects about GEO system are shown in Table 1. N-STAR and Engineering Test Satellite- VI (ETS- VI), which was launched in 1995, utilize IF band switching at the satellites; base-band switchings are performed at the terrestrial base station. In contrast, the maximum channel capacity in the Japanese systems has been limited to 384 [3][4] so far for base-band switching systems. This channel capacity is fairly small compared with other systems. There still seems to be a stage of experimentation, and greater channel capacity is expected for practical use.

SYSTEM TARGET

Considering hand-held phone services expected at the beginning of the next century and technology trends for on-board hardware and hand-held terminals, ASC has been studying the performance of mobile satellite communications system using 2-ton class geostationary satellites.

Table 2 shows the major system parameters in the current study[3]. S-band (2.6/2.5 GHz) was chosen for the mobile link and Ka-band (30/20 GHz) for the feeder link in accordance with frequency assignments specified in WARC-92.

The information rate was selected to satisfy an application for multimedia communications. Multicarrier Time division multiple access (MC-TDMA) was selected to realize high performance, low-cost miniature hand-held mobile Earth terminals usable for multimedia communications[1]. To maintain the time delay, including signal processing time, below 0.5 seconds, channel calls from one mobile terminal to another must be exchanged at the satellite[1].

Table 2 System parameters

Number of beams	2 to 4
Frequency band	
· Mobile link	2.5/2.6 GHz
· Feeder link	30/20 GHz
Information type	voice, data
Information rate	voice: 5.6 kbps
	data: 32.0 kbps
Transmission rate	70.0 kbps
Modulation	π /4 shift QPSK
Multiple Access	MC-TDMA (Multicarrier-TDMA)
Method	Both up and down link
Forward Error	· convolutional coding
Correction	· Viterbi decoding
	3-bit soft decision
	- constraint length: 7
	- coding rate: 1/2
Time Delay	Less than 0.5 seconds

OBP REQUIREMENTS

ASC is planning to construct a mobile satellite communications system using a hand-held terminal and 2-ton geostationary satellite as the next century communications system. To realize the system, channel exchange should be performed by one hopping, and the OBP should have functions such as base-band switching, regeneration, and greater channel capacity.

Base-band switching: In a multibeam system, on-board switching is useful in terms of flexible channel exchange. Although there are two switching schemes in general, RF/IF-band and base-band switching, base-band switching is preferable for relatively low speed data handling because of the flexibility and efficiency of channel control. Consequently, base-band switching is indispensable for mobile satellite communication.

Regeneration: By employing forward error correction (FEC) both in the up and down links, the link margin can be increased in power restricted system.

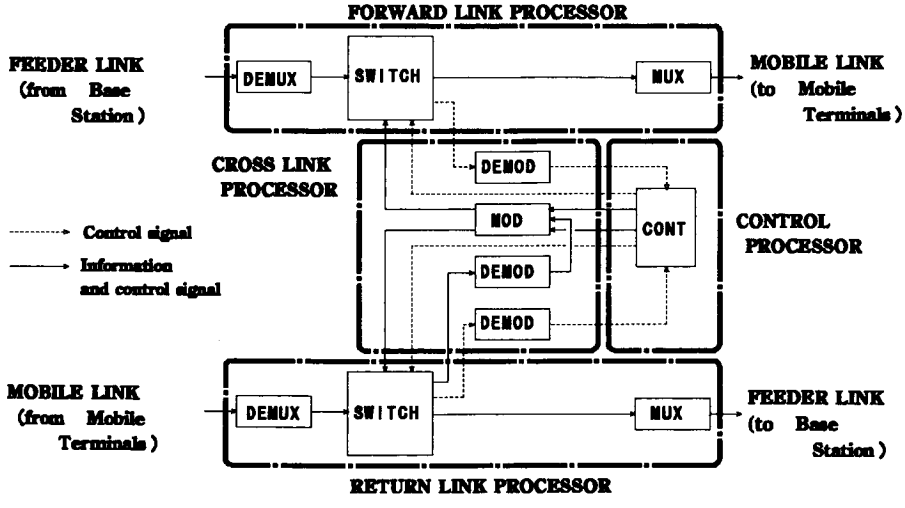


Fig. 1 OBP Block Diagram

Large channel capacity to reduce communication costs: To realize a large channel capacity, it is necessary to minimize the circuit size since satellite mass and power are strictly limited. We propose using digital signal processing technology because the benefit of the power and weight reductions required in the near future through ASICs qualified for the space use. Satellite channel capacity can also be improved through efficient demand assignment by base-band switching.

Figure 1 shows the block diagram of an OBP incorporating the requirements above mentioned. The OBP is composed of the Forward Link Processor, Return Link Processor, Cross Link Processor, and Control Processor. Digital signal processing technology which integrate a complex functions into the ASIC is adopted in these four Processors.

FEATURES OF ASC's OBP

In ASC's OBP, two kinds of communication schemes are utilized as describe below.

Communication between mobile terminals: The OBP performs channel exchange with regeneration in this communications system. This will contribute to reducing the size of mobile terminals because of the coding gain obtained by FEC. In addition, this communication scheme can endure a disaster because the almost of channel exchange is performed by the OBP and is independent of terrestrial equipment.

Communication between a mobile terminal and a subscriber in a public network through the gateway: The OBP also performs channel exchange without regeneration of information signals. In this case, the channel exchange function is shared by the OBP and the base station connecting to the Public Communication Network. This scheme contributes to the large channel capacity, adopting fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT) technologies.

The control processor efficiently exchanges channels by the demand assignment method. The communications software in the control processor will be downloaded through the feeder link, in order to perform some desired experiments.

UNITS COMPOSING ASC'S OBP

ASC's OBP is composed of demultiplexers, in which more than 100 FDM-multiplexed carriers per beam are demultiplexed; switches which exchange more than 100 carriers per beam; demodulators and decoders; modulators and encoders; and controllers. Some of them are described below.

Demultiplexer Unit: The forward and return link processor is composed of the demultiplexer, the switch, and the multiplexer units shown in Fig.1. Figure 2 shows the functional diagram of the forward and return link processor. There are several methods applicable to the demultiplexer processing digital signal such as per-channel method, multi-stage method, bulk-FFT method, and polyphase FFT method. In case of adopting polyphase FFT method, the number of multiplication per sampling is the smallest among them. It is important to minimize the number of multiplication considering hardware size. Consequently, ASC's OBP adopts the polyphase FFT method which is suitable as a demultiplexer when using many carriers[4].

The polyphase FFT is mainly composed of a polyphase filter obtained by distributing the impulse response coefficients of the desired basic filter to the subfilter elements, and an N-point (N=128) FFT corresponding to the number of multiplexed carriers in the unit circuit. The demultiplexed signals are supplied to the switch units at 50-KHz sampling rate.

Polyphase Filter: The Polyphase Filter consists of the delay elements, the subfilter elements, and the phase rotation elements shown in Fig.2.

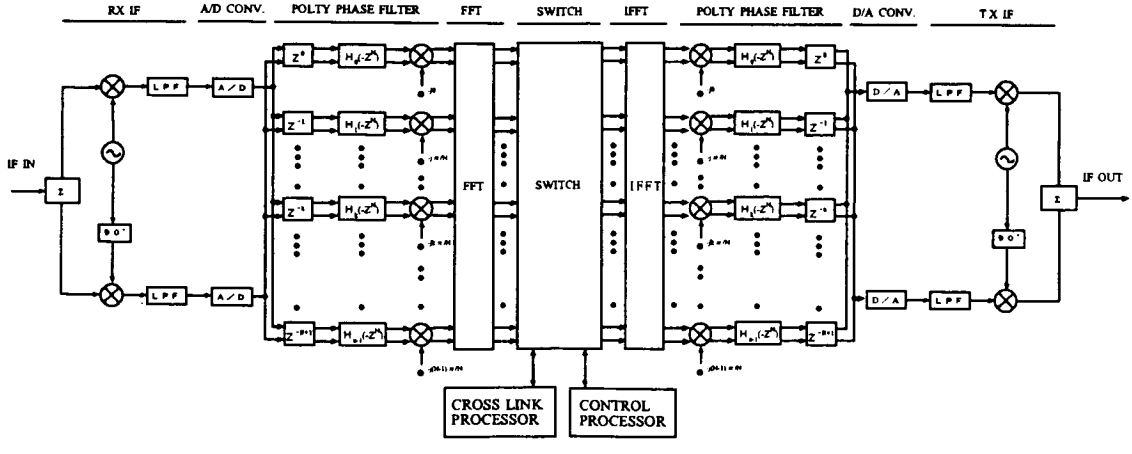


Fig. 2 Forward and Return link Processor

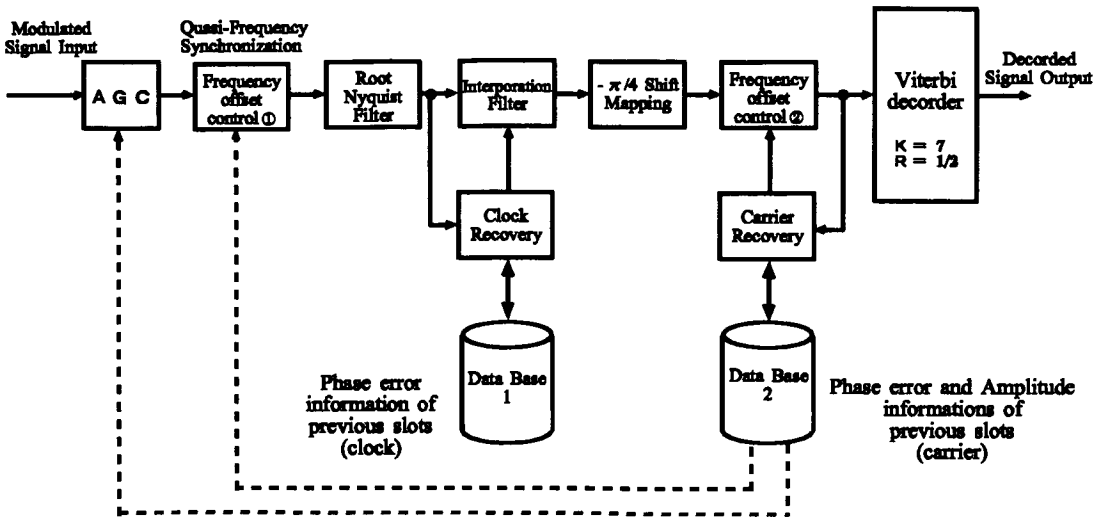


Fig. 3 Demodulator Schematic Diagram

The basic filter is a finite impulse response (FIR) roll-off type filter. Attenuation exceedingly 35 dB in the adjacent channel can be obtained for the Doppler shifts, the frequency stability of the system, and the dynamic range of each carrier.

FFT: N (N=128), which stands for the number of carriers demultiplexed simultaneously, affects the size of circuit. In order to minimize this size, the pipeline junction method is adopted in the FFT algorithm considering signal control simplicity. The number of multiplier circuits is also minimized by accessing the same multiplier in the divided times.

Multiplexer Unit: The polyphase IFFT method is selected in the multiplexer unit because the design can be simplified by applying a design similar to that of the demultiplexer to the multiplexer as shown in Fig.2.

Demodulator Unit: Figure 3 shows the Demodulator schematic diagram. We propose using the digital demodulation method because the power and weight reductions can be achieved by using ASIC technology. The signals provided to the demodulator unit have already been distributed to each carrier but are still multiplexed in the frame format composed of five time slots. Each demodulator circuit processes the several carriers by accessing in the divided time. The sampling rate of these inputs with both real and imaginary forms is 50-KHz, which is equal to the carrier spacing between the carriers. The interpolation filter transforms the signal from a 50-KHz sampling rate to a 35-KHz symbol rate. The demodulators have another filter which limits the noise bandwidth. The frequency response of this filter is root Nyquist, so that the Nyquist characteristic can be obtained in the communication path. If there is any frequency shift in the signals supplied to the root Nyquist filter, however, the influence of intersymbol interference will increase. To minimize the interference, a quasi-frequency synchronization circuit is installed in the front of the root Nyquist filter.

Mobile communications requires a wide capture range and short lock up time. To satisfy these requirements, the reverse modulation method of estimating the frequency offset is adopted as the carrier recovery circuit, and a phase-error-estimating method utilizing a Fourier transformation is adopted as the clock recovery circuit, considering both the performance and the circuit size. The phase information of the previous time slot is utilized to reduce the length of the preamble .

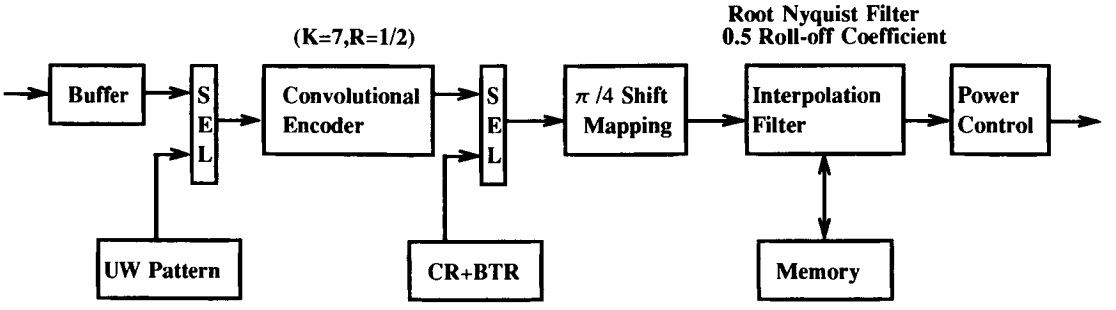


Fig. 4 Modulator Schematic Diagram

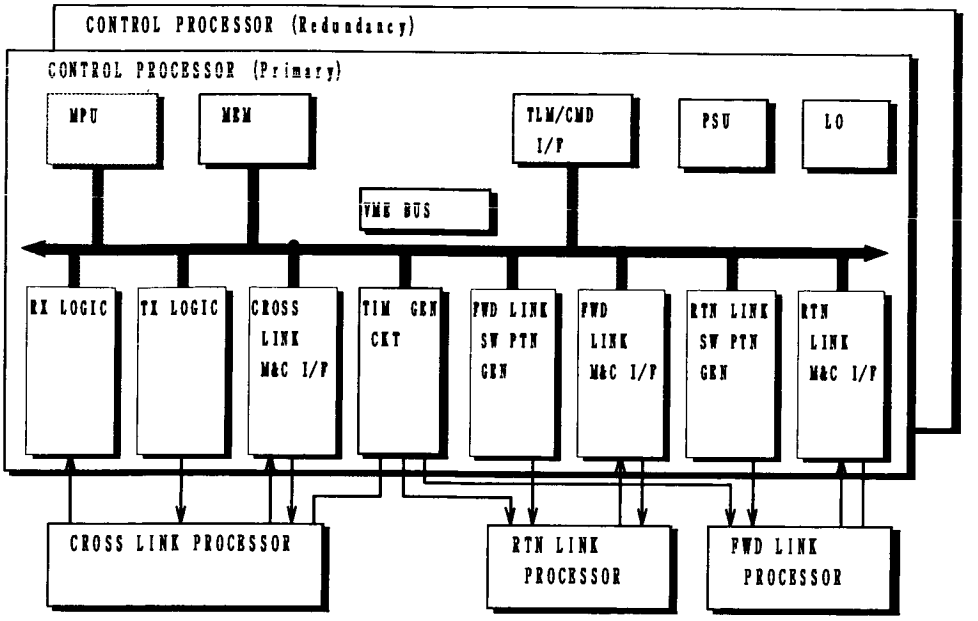


Fig. 5 Controller Schematic Diagram

Modulator Unit: Figure 4 shows the schematic diagram of the Modulator. Digital processing technology is also used in this Modulator. The decoded signal outputs in Fig. 3 are convolutionally encoded, after adding the Unique Word whose pattern length is 16 bit. They are then added to the preamble and rotated in accordance with the π /4 shift mapping. An interpolation filter with a 0.5 rolloff coefficient transforms the signal from a 35-KHz symbol rate to a 50-KHz sampling rate, and shapes the wave form by limiting the frequency bandwidth. To reduce the size and power consumption of the Modulator, the convolutional calculation is performed using the memory as explained below. In general, the interpolation filter convolves the input data with the impulse response coefficients using a multiplier.

The multiplier circuit is bigger than other circuits. It is beneficial to avoid using a multiplier in order to reduce the circuit size. Consequently, we store the results of multiplying the input data and the impulse response coefficients in memory beforehand, and the interpolation filter brings the desired data corresponding to the input data from the memory instead of the multiplier.

Controller Unit: Figure 5 shows the schematic diagram of the Controller. The Controller receives the control signals from the Cross Link Processor through the Rx Logic circuit, and transmits the control signals to the Cross Link Processor through the Tx Logic circuit. The switch pattern assigned by the controller in response to the mobile terminal's requirements is supplied to the switches in the Forward Link Processor and the Return Link Processor. The Controller has the major functions below to perform the flexible channel control for demand assignment.

- Data analysis of the control packet and processing.
- Channel exchange function with regeneration
- Transparent channel exchange function
- Mobility Management (Registration of mobile terminal's positions)

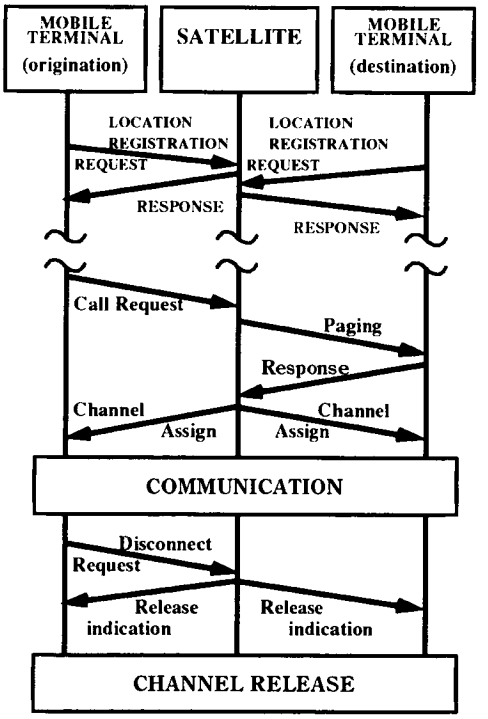


Fig. 6 Basic Procedure

- Timing adjustment corresponding to the mobile terminals which have various time delays
- Call Control (Channel search, Channel assignment, Monitoring of the carrier, Channel release)
- Radio Frequency Transmission Management

As additional functions, the Controller generates timing signals which are used for communications system standards and distributes them to the other units. Statistical information such as the status of channel connection and the number of control packets which is received normally are obtained periodically. Communications software in the control processor can be downloaded through the feeder link, in order to perform some desired experiments.

The Controller is most important for assigning communications channels to the mobile terminals. In order to increase the Controller reliability, the perfectly redundant structure is adopted.

The basic procedure for the communications between the mobile terminals is shown in Fig.6[4]. As soon as the Earth Stations are turned on, they verify the beam they belong to using the information broadcast from the satellite. They then request location registration whenever there is a difference between the beam data memorized before turning on and that obtained after turning on. Timing is simultaneously adjusted in this location registration process. After that, a call procedure is performed as shown in Fig.6. ASC is now studying suitable configurations of control for over 1,000 channels including the communications between a mobile terminal and a subscriber in a public network through the gateway.

Conclusion

This paper has described the design concept of an OBP for a mobile communications satellite for use early in the next century. This OBP has the capacity of over 1,000 channels and realizes high flexibility for controlling channel connection using the base-band switching method. It is still under development at this stage. After the performance evaluation tests, it will be installed on a Japanese Engineering Test Satellite which is planned to be launched around 2001.

Acknowledgement

The author would like to thank the members of ASC for their valuable guidance and suggestions during discussions about this study.

References

[1] K. Yoneyama, Y. Otsu, T. Miyoshi, Y. Kawakami, H. Hara, H. Hoshino, "R&D on S-band Mobile Communications and Sound Broadcasting System by Geostationary Satellite for the Next Decade," AIAA International Astronautical Federation-96-M.3.04. (1996)

[2] K. Yoneyama, Y. Otsu, T. Miyoshi, and H. Hara, "Research and Development Plan of Advanced S-band Satellite Mobile Communications and Broadcasting system," Proc. 1994 IEICE Fall Conference, No.B-186, Sept.1994.

[3] M. Endo, T. Kumagai, T. Yamamoto, Y. Otsu, T. Kikuti, and N. Komiyama, "Future Satellite Communications System using Millimeter-wave and On-board processor," ICDSC9-A1-5 (1992)

[4] Y. Otsu, T. Yamamoto, H. Maruyama, and M. Itoh, "Future Personal Communications through Space Platforms with Advanced On-bord Switching Systems," Asia Pacific Microwave Conference WS 6-5 (1994)

Session 12
Advanced System Concepts I

Session Chairperson—*Edward W. Ashford*, ESA (ESTEC), The Netherlands
Session Organizer—*Polly Estabrook*, Jet Propulsion Laboratory, USA

Next Generation Mobile Satellite Communication Architectures, Networks and Systems

Dr. James R. Stuart and Dr. Janet G. Stuart
Consultants
1082 West Alder Street, Louisville, CO 80027-1046, USA
tel: +1 (303) 666-0662, fax: +1 (303) 666-0388

ABSTRACT

This paper describes possible revolutionary next generation mobile communications satellite missions and architectures. The satellite communications evolutionary background, current GSO's (geostationary orbit satellites), 'Little', 'Big' and 'Broadband' NGSO's (non-GSO's), and enabling satellite technologies are briefly discussed. Converging technological and economic forces will drive us inevitably to the next decade's communication satellite systems and network architectures. The increasing pace of new capabilities, Moore's law, and the convergence of key technologies over the next few years provide the next generation satellite revolution 'push', while the global and local commercial communications demand and economics, orbit slot and spectrum scarcity, rapidly dropping communications costs, privatization, multiple market entries and competitive shakeouts provide the 'pull'. The result of these forces is a rapid evolution of communications satellites and network architectures that will depart significantly from today's GSO and NGSO systems. The revolutionary next generation architectural paradigms for distributed space systems is described. New system architectures are predicted for GSO's (virtual satellite clusters), NGSO's (multi-mission, hybrid payload constellations, clusters), and the likely next big waves: interlinked hybrid networks, station-keeping, stratospheric platforms (aerostats), micro-nanosatellites and trends towards clusters of large area satellites.

TABLE OF CONTENTS

1. BACKGROUND

In 1945, Arthur C. Clarke first predicted that satellites in orbit approximately 36,000 kilometers above the equator, with a period of 24 hours, could maintain a fixed location as seen from the ground. In this geostationary orbit (GSO), a satellite could receive signals from the ground and transmit them over roughly a third of the Earth's surface. For more than three decades now, GSO satellites have been virtually the exclusive means of providing space-based communications (e.g., TV broadcast, long distance telephony, etc.). The communications revolution is rapidly changing space-based communications services, systems and networks.

The 'Negroponte Flip' was the vision of the communications revolution in the 1970's. The 'Negroponte Flip' described the trend that narrowband (voice) access would move from telephones to wireless (i.e., cellular), and broadband (i.e., TV) access would move from rabbit-ear and roof-top antennas to wire (cable).

The 1980's vision of the communications revolution evolved to the 'Pelton Merge', which said that all communications traffic would be digital packets of mixed, multimedia data carried on hybrid networks of interconnected wire (cable, fiber) and wireless (terrestrial, satellite) transmission media. In the 1990's, this has evolved to a much higher level than just the interconnected wired and wireless transmission media of the 'Information Superhighway' (in part through government studies such as 'Global Grid').

The communications revolution is now viewed as the emergence of the 'Global Information Infrastructure' (GII): a universally accessible web of multiple interconnected networks, permitting access to widely distributed private and public data bases, and providing ready transmission of information in any format, to any one, in any place, at any time. The current view of the emerging GII revolution includes an entire emerging system of:

1) human users (and developers),

2) user information appliances (computing and consumer electronics),
3) accessed information, data bases and computing resources, and
4) networks (interconnected)

In this GII view of the communications revolution, space communications is subsumed into a critical niche of the emerging GII network composed of many intricately tangled, overlaid, interconnected networks:
1) wired and wireless
2) terrestrial, high altitude and space
3) physical and virtual
4) commercial, private and government

The emerging GII elements will be driven by the large commercial market, and will migrate to the web elements that are 'commercially most efficient' (i.e., reliable, ubiquitous, seamless, flexible, capable and cost-effective). History shows that the surviving, successful elements will be interoperable with open interfaces and accepted standards, allowing a wide array of competing information appliances and software tools and value-added features. Similarly, through the inescapable competition with the next generation of evolutionary and revolutionary innovations, the successful space communications elements of the GII will be driven to provide the 'commercially most efficient' solutions for commercial service applications, for example:
- significant new capabilities and features
- significant improvements in the price/ performance ratio, etc.

In looking out into the future at various possible evolutionary paths over the next decade, the impact on the commercial communications marketplace is the figure of merit for choosing likely successful next generation space communications architectures, technologies and designs. The dominant evolutionary selection rule for the next generation space technology developments and innovations will be their ability to enable successful commercial market-driven applications.

Commercial satellite communications is such a big business because it drives a positive economic feedback spiral. Access to communications is addictive and self-reinforcing. E. F. Tuck, Teledesic's founder and co-chairman, has often said that 'No one has ever lost money overestimating the demand for access or bandwidth'. Demand is growing exponentially in the USA and globally for:
1) access and connectivity (the highest value-added network element)
2) all service types: [e.g., mobile, monitoring, personal communications services (PCS), bandwidth-on-demand, broadcast, down-load, interactivity, etc.]
3) service feature differentiation (e.g., availability, quality, bundled services, inter-operability, transparent interconnections, etc.)
4) affordable terminal and service costs

2. NEW SATELLITE SERVICES AND SYSTEMS

In recent years, many new GSO projects continue to expand the range of new communications services (e.g., direct to home, direct audio broadcast, mobile voice, broadband, etc.) to more regions of the world, and several commercial remote sensing projects have begun. Most strikingly, a number of major non-GSO (NGSO) constellation projects have been started, costing in the billions, to provide a range of mobile and fixed communications services globally. These are called 'Little' NGSO's, 'Big' NGSO's and 'Broadband' NGSO's.

These NGSO systems are not an aberrational phenomenon. Rather, they reflect a fundamental evolution in satellite-based communications networks. Just as networks on the ground have evolved from centralized systems, built around a single powerful mainframe computer, to decentralized networks of interconnected personal computers, so too are satellite networks evolving from centralized systems, consisting of a single powerful GSO satellite, to decentralized networks of interconnected NGSO satellites. Many of the same technological developments underlie both trends. While GSO will continue to play an important role in space-based communications, the NGSO's are forging brand new directions (e.g., low-cost high volume satellite component production, global market penetration, etc.)

The 'Little' NGSO's (e.g., OSC's Orbcomm, GE's Starsys, Vita, etc.) are a class of UHF/VHF, low data rate, store-and-forward constellations (from 3 to 48 satellites in low earth orbit with non-contiguous global coverage) which will provide non-real time and near-real time digital data communications to low-cost mobile and fixed terminals. They will provide a wide range of new services, such as monitoring, alarm detection, short digital messages, facsimiles, electronic mail, paging, etc. These projects each represent commercial financing commitments ranging from 10's to 100's of millions of dollars.

The 'Big' NGSO's (e.g., Motorola's Iridium, Loral's Globalstar, TRW's Odyssey, ICO-P, etc.) are a class of L-band low data rate constellations (from 12 to 66 satellites with continuous global coverage) which will provide mobile voice services, extending the range of terrestrial cellular phone networks for users anywhere on the globe. As with the 'Little' NGSO's, the 'Big' NGSO's use low earth orbit (LEO) and medium earth

orbit (MEO) satellites to enable their low-power, small terminals. The 'Big' NGSO' projects each represent commercial financing commitments over a billion dollars.

The largest licensed constellation, proposed by Teledesic, is a 'Broadband' NGSO, operating at Ka-band, which will use 840 satellites in NGSO to globally provide broadband channels supporting video-conferencing, interactive multimedia and real-time, digital network connections with a service cost comparable to urban wireline networks. In the case of Teledesic, LEO satellites enable the low-delay required to provide 'fiber-like' interconnection with the broadband terrestrial networks, as well as small antennas and compact electronics. The largest of the commercial space communications ventures to date, the Teledesic project received its FCC license March 1997 and will cost a projected nine billion dollars.

Other 'Broadband' NGSO's are continuing to be proposed to meet this potentially explosive market demand, such as:

1) Alcatel's Sativod 20/30 GHz constellation (64 satellites), and

2) Motorola's recently filed (September 1996) M-Star 40/50 GHz constellation (72 satellites) costing some 6.1 billion dollars.

These new NGSO commercial communications projects have been enabled by the convergence of global market demand for communications and new space technologies (e.g., power, data processing electronics, low thrust propulsion, multi-functional structures, multi-beam antennas, etc.) that can most effectively meet that market need.

3. NEXT DECADE'S COMMUNICATION SATELLITE SYSTEM DRIVERS

For the next decade, successful commercial market-driven applications will drive successful space technology developments and innovations. Several important market-driven forces are converging to 'pull' the successful next generation space systems from the pool of possibilities ('push' technologies). Commercial satellite communication market demands which are 'pulling' successful technologies, new designs and new architectures include:

- Larger bandwidths
- Higher frequencies
- Higher satellite powers
- Smaller, lower cost terminals
- Lower bit error rates (BER)
- Higher frequency reuse
- Higher energy storage demands
- Smaller, agile beams (larger antennas)
- Tighter pointing control and knowledge
- More precise stationkeeping
- Intersatellite links (ISL)
- Interconnectability (hybrid networks)
- Rapid network capacity reconfiguration/ adaptation to market demands

The emergence of vibrant new commercial space business opportunities is unleashing the powerful market-driven selection criteria for determination of space systems evolutionary paths. Government-specified technology developments, architectures and missions are yielding dominance to market-driven 'commercially most efficient' technologies and system architecture solutions. Satellite communications currently feeds the entire commercial space industry.

Private funding for large GSO space ventures is now relatively easy to obtain. The significant private funding raised by some of the 'Big' NGSO's to date,

- Iridium with ≥1.9 B$
- Globalstar with ≥1.8 B$
- ICO-P with ≥1.5 B$

is an example of financial market determination of the pool of 'commercially most efficient' competitors. Recent FCC (U.S. Federal Communications Commission) spectrum auctions show that the financial and commercial markets are bullish on the satellite communications future

Market-driven, commercially funded space technology developments enabling successful commercial applications will determine successful space systems evolutionary paths. For example, Moore's' law (which has been correct for the past two decades and is expected to apply for at least the next decade) states that the performance/price ratio of digital electronics (e.g., processors, random access memory, etc.) doubles every 18 months.

This has an enormous effect on commercial competitiveness of later-to-market entries and systems designed for more frequent upgrade replacements. As satellite mass/performance and launch costs/mass continue to decline, the commercial marketplace may reverse the historic trend for longer life satellites with fixed and dated capabilities by favoring upgraded, replaceable systems with shorter lifetimes (as in other technologically dynamic products such as consumer electronics and personal computers).

As another example, market-driven demand for access to spectrum to launch new communications ventures is at an all time high. Spectrum scarcity is now the rule. Licensed spectrum is the critical path now for most satellite communications development projects. Scarce spectrum is the reason for the clamor for more 'Little' and 'Big' NGSO' spectrum, and the clamor for GSO slots at all frequencies.

Applications for C-band (4/6 GHz) GSO's over East Asia have piled up to where they now represent more than seven times the available orbital slots. Thirteen USA Ka-band (20/30 GHz) GSO applicants propose to launch 74 GSO satellites to provide wireless broadband communications. Six of these included 60 Ghz intersatellite links and five proposed to provide voice and messaging services. All plan to launch within three to five years if they receive FCC licenses. The world of commercial space communications applications changed forever January 1996, during the FCC's first ever auction of U.S. satellite spectrum, when MCI acquired the DBS (direct broadcast service) frequencies at the 110°W orbit slot with a winning bid of $682.5 million (FY96$).

Scarce, precious spectrum drives technologies that get more frequency reuse, primarily with smaller spots, and more MHz per km^2. Note that since the only commercially successful space based businesses to date are based on, or support, photon processing, for example:

- solar photons to electrons
- solar to RF (radio frequency) photons
- RF to RF photons
- visible photons to RF, etc.

a dominant trend in emerging systems is 'Aperture Rules!'. Smaller spots means more, larger antennas. More frequency reuse also means more powerful satellites with larger solar arrays. Larger antennas and sloar arrays lead to more flexible satellites that require lower thrust propulsion for stationkeeping to minimize flexible structure vibrations. The design of successful next generation GSO satellites will require large antenna farms in each licensed slot.

Similarly, there are numerous significant satellite technology advances 'pushing' the new system designs and architectures and vying for commercial evolutionary selection:

- High power generation and storage
- Low-thrust electric propulsion
- Deployment mechanisms, e.g.: shape memory metal, HOP (high output paraffin), inflatables, etc.
- Multi-functional composite structures
- Precision pointing
- Precision guidance and navigation
- Efficient HPA's (high power amplifiers)
- Multi-beam antennas, arrays and lenses
- Large aperture deployable antennas
- Active control of flexible body dynamics
- Digital data processing electronics
- Optical LAN's (local area networks)
- Optical intersatellite links
- Modulation, coding and multiple access schemes
- Autonomous satellite operations and operating system software

For example, the demand for bandwidth and clear spectrum will move satellites to increasingly higher frequencies, jumping into the 90 Ghz range and even to the point where optical communications may become commercially attractive. Optical spectrum is virtually untouched and unregulated. A laser payload for space-ground communications can use smaller transmitters with higher energy concentrations on the target receiver. With site diversity, laser communications can even have over 90% availability for space to ground laser communications. Laser communications improvements (e.g., production costs, optical detectors capable of detecting modulatable light with first order roll-off of >750 Mhz, etc.) may spark new commercial market-driven applications.

Wireless communications will be naturally differentiated by the market selection of competitive service providers. The market criteria, such as:

- Services offered
- Service costs
- Service quality
- Availability (outage)
- Terminal costs

are determined by the system architecture features, such as:

- Coverage area (global, regional, etc.)
- POP (serviceable population)
- Bandwidth
- MHz/km^2
- Cell size,
- Latency/delay
- ROI (return on investment)
- Initial cash requied and payback period

The wireless market and services are naturally differentiated by platform altitude (e.g., GSO's, MEO's, LEO's, aerostats and terrestrial towers). Aerostats are a new class of powerful commercial communications and remote sensing platforms stationkeeping above most of the atmosphere at around an altitude of 20 km with a 150 km footprint. Aerostats fill an important market-valued niche in the spectrum of locations from GSO to MEO to LEO to aerostats to terrestrial towers. Each has a natural coverage area (global, regional, local) and range. As an example, one can provide remote mobile cellular voice (or imaging) from a regional GSO, but for most markets not as efficiently or competitively when compared to services from a global LEO (50 times closer) or a local aerostat (2000 times closer). Attendant terminal and service costs provide market differentiation. Each platform altitude will evolve under competitoion to offer those services in those markets for which it is the best suited.

Candidate future evolutionary paths for the next generation GSO's and NGSO's are described below. As stated before, the dominant evolutionary selection rule for space technology developments and innovations will be their ability to enable successful commercial market-driven applications.

4. FUTURE GSO COMMUNICATION SATELLITES AND NETWORKS

Satellite communications has always been dominated by GSO's in hierarchical communication architectures. There are approximately 3,200 transponders in orbit now with 2,000 licensed transponders currently on order. Over the last 30 years, GSO satellite communications has enjoyed an extraordinary and successful evolution:

1) Bigger, more powerful, more efficient, longer life GSO satellites
   - 1000 times more cost effective
   - 100 times higher power
   - 50 times more freq. reuse efficiency
   - 10 times longer lifetimes
2) Increasingly smaller user ground terminals
3) Migration to ever larger bandwidths and higher frequencies
4) New capabilities (e.g., satellite switching, intersatellite links, precision navigation for collocated satellites, high-power payloads, large deployable antennas with small agile beams, etc.)
5) New markets and services (e.g., broadband data, mobile voice, direct-to-home, etc.)

Emerging technologies and commercial market competition will enable even further successful evolution of GSO's. The following are likely evolutionary paths for:

1) The next generation GSO communication satellite designs (virtual satellite comprised of a close formation-keeping cluster of specialized comsat elements),
2) Several new next generation GSO communication hybrid network architectures which provide new communications network capabilities and new, unique, value-added communications services, not currently available using today's GSO's in their current hierarchical communication network architectures.

GSO Distributed (Cluster) Satellite

The value of spectrum drives the next generation GSO's to get more frequency reuse with smaller spots. Smaller spots means more and larger antennas on each platform. Previously successful next generation communications satellites have always needed more solar power and larger antennas. The design of successful next generation GSO satellites will likely require even larger antenna farms in each licensed slot to be economically competitive.

A next generation architecture that seems to be 'pulled' to the market is a distributed, *virtual* satellite. This distributed GSO would be composed of:

1) a central switching 'shepherd' satellite which station-keeps in the GSO slot and
2) an interconnected cluster of close formation-keeping 'sub-satellites', each with a large antenna and a broadband intersatellite link (ISL) to the central switching satellite.

The sub-satellites receive and transmit at licensed frequencies to ground terminals; received digital signals are shipped over the ISL to the switching satellite for routing over the ISL to the appropriate sub-satellite's spot. Since all traffic is processed digitally, the sub-satellites can each operate at different bands allowing complete cross-strapping of links. An ideal architecture for this cluster GSO would allow the central switching satellite to accept additional sub-satellites into the local star network as market demands increase or new spectrum becomes available. Such a GSO virtual distributed satellite would allow:

- incremental deployment (track new markets, minimize intitial cash, etc.)
- graceful introduction of new spectrum technologies, capacity and features
- replaceable elements (shorter lifetimes)
- graceful degradation modes
- continuous occupation of the orbital slot.

GSO's in New Hybrid Networks

GSO's will inevitably be linked into hybrid networks with other GSO's, NGSO's (LEO's, MEO's), new aerostat platforms and new terrestrial networks. Next generation GSO's will be interconnected to other networks forming larger more capable hybrid networks providing brand new, synergistic network capabilities and new, unique value-added services for all individual network elements to sell. Using interoperable broadband optical intersatellite links, GSO's will interconnect seamlessly with other operator's (and other vendor's) GSO's. Using broadband optical intersatellite links, GSO's will interconnect to LEO and MEO constellations, as well as to stratospheric aerostat communications platforms. GSO's will be interconnected through interlinked microwave DTH (direct-to-home) receive-only and NGSO terminals for new, asymmetric interactive network capabilities allowing narrowband data request through the NGSO terminal and broadband downloading through the DTH receive-only VSAT's (very small aperture terminals).

5. FUTURE NGSO COMMUNICATION SATELLITES AND NETWORKS

Emerging technologies, commercial market competition and competitive alliance efficiencies will drive the evolution and further service differentiation of the LEO's, MEO's and other NGSO's. The following are the likely evolutionary paths for the next generation 'Little', 'Big and 'Broadband' NGSO's

'Little' NGSO's Evolution

Driven by the diverse market demands, competitive service differentiation and market shakeouts, 'Little' NGSO's could evolve in several likely directions to different successful solution strategies.

1) Some current and new 'Little' NGSO's will offer robust, global, near real-time coverage through a network of distributed ground relay stations which will route packets optimally in hops through the ground relay/satellite network.

2) Some 'Little' NGSO's will move to larger constellations of smaller low power, low cost satellites in order to minimize holes, increase capacity, increase frequency reuse and provide real time services within more 4000 km footprints. The emergence of miniaturized, mass produced satellite components which can be assembled automatically into simple satellites will inevitably drive this path. The evolutionary trend of this 'commercially most efficient' path is to very large constellations of large-aperture, distributed nano-satellites.

3) 'Little' NGSO's will develop high performance, low delay (fiber-like: roundtrip < 150 ms) inter-connected networks by adding low cost, low data rate intersatellite links and satellite switching to all (or some) satellites and using distributed ground relay stations to interconnect the rest.

4) Some successful 'Little' NGSO's will move to interoperable terminals which can access other competitive 'Little' NGSO's, providing each alliance competitor with large market benefits. Then 'Little' NGSO's will move to strategic business operations alliances with standardized intersatellite links and protocols to interlink directly with GSO's, with other 'Little/Big/Broadband' NGSO systems and with aerostat platforms, providing each alliance partner with larger new markets and new value-added service offering benefits.

5) Finally 'Little' NGSO's will move to multi-license shared platform strategic business alliances and joint ventures with:

- Different 'Little' NGSO licensed payloads on shared facility constellation, or
- 'Little' NGSO licensed payloads piggybacking on a licensed 'Big' NGSO constellation (complete global coverage on much more reliable constellation), providing new combined services and feeding data into the 'Big' NGSO network.

The marginal costs to piggyback a typical 'Little' NGSO licensed payload on a typical licensed 'Big' NGSO constellation are a compelling ≤ 30% of the current stand-alone 'Little' NGSO project costs. Such NGSO strategic partnering (piggyback payloads, interoperable terminals, shared satellites, etc.) will likely cause a major aftershock of consolidations and alliances as others seek to replicate the competitive advantages.

'Big' and 'Broadband' NGSO's Evolution

Driven by the demand for bandwidth, large diverse market, and competitive alliance efficiencies, 'Big' and 'Broadband' NGSO's will similarly evolve to more spectrum efficient and capable systems, and interoperable terminals which can access other competitive NGSO's, and interconnected into transparent hybrid networks (with GSO's, other NGSO's and aerostats), then perhaps to an ultimate 'commercially most efficient' solution of multi-license multi-frequency shared constellation

1) 'Big' NGSO's will likely move to higher capacity, smaller beam satellites with much higher frequency reuse, agile, multiple frequency (multiple license) payloads, inter-satellite links and on board satellite switching and new value-added processing

2) Some successful 'Big' NGSO's will likely move to strategic alliances with standardized intersatellite links and protocols to interlink directly with other terrestrial and space-based alliance partners, providing each with benefits of larger capacity, new markets and new value-added service offerings.

3) Finally 'Big' NGSO's will likely move to multi-license shared platform strategic business alliances and joint ventures, such as different 'Big' NGSO licensed communications (and remote sensing) payloads on a shared facility constellation, or on a licensed 'Broadband' NGSO constellation (complete global coverage on much more reliable constellation), providing new combined services such as inter-network paging and trunking data through the 'Broadband' NGSO network

6. NEXT BIG WAVES

Several major trends and innovations are on the horizon and are likely to change the future of satellite communications.

The first wave is the emergence of the new disciplined fields of design and process architecting, for example:

- space infrastructure design architects (technical, business, programmatics)
- space production and manufacturing process design architects, etc.
- distributed network and hybrid network design architects
- space infrastructure/network/business meta-architects (higher level)

Coupled with digital electronics and micro-nano technologies, these trained architects will undoubtedly lead the commercial evolution of space in new revolutionary directions.

The second wave is the move to hybrid communications networks linking GSO's, NGSO's, aerostats and terrestrial networks to form larger, more capable communications services. The effect on competitive alliances and standardization will be dramatic. Similarly, NGSO strategic partnering (interoperable terminals, piggyback payloads, shared satellites, etc.) will cause a major aftershock of consolidations and alliances as others seek to replicate the competitive advantages.

The third wave is the introduction of aerostats. There are a number of commercial projects being proposed using either airplanes or lighter-than-air vehicles (most are unmanned). Aerostats offer the ability to carry collections of multipurpose payloads over high population density areas and provide:

1) unique complementary services through strategic business alliances to existing terminals by extending regional coverage (e.g., cellular, PCS, etc.), or providing brand new services (e.g, high user density connections for NGSO satellite networks, local TV channels for DTH, etc.), and/or

2) directly competitive services using new terminals specifically for aerostat services (e.g., PCS, tracking, location, wireless broadband last mile connections, remote monitoring, etc.). The introduction of aerostats will provide unprecedented new competition and cooperative opportunities for our next generation space systems.

And finally, the fourth major wave coming is the emergence of new technologies likely enabling a revolutionary new class of *ultra-low cost*, *ultra-lightweight*, large area satellite designs and missions. In one communications constellation embodiment, each satellite would be able to radiate only 1/4 watt RF with open loop electronic steering (or to a beacon using conjugate phase steering.) These independent satellites could be physically connected into long lengths (e.g., 100's in a row) providing a total 250 watt RF. The satellites could communicate along the row to provide coordinated beam pointing. Several new technologies have converged to allow such satellites to efficiently maintain attitude control, power storage, maneuvering capabilities, and to allow each satellites to communicate to other neighboring rows through innovative intersatellite links. Such schools of rows could form enormous distributed apertures capable of unprecedentedly small communications spots with large powers at many orders of magnitude reductions in cost.

7. CONCLUSIONS

Converging technological and economic forces drive us inevitably to the next decade's communication satellite systems and network architectures. The increasing pace of new capabilities, Moore's law, and convergence of key technologies over the next few years provide the next generation satellite revolution 'push', while the global and local economics of communications demand, orbit slot and spectrum scarcity, rapidly dropping communications costs, privatization, multiple market entries, competitive shakeouts and competitive alliance efficiencies provide the 'pull'. The dominant evolutionary selection rule for space technology developments and innovations will be their ability to enable successful commercial market-driven applications. The result is a rapid evolution of communications satellites and network architectures that will depart si gnificantly from today's systems (e.g., GSO clusters of specialized satellites, multi-license/frequency shared NGSO constellations, NGSO clusters, interlinked hybrid networks, aerostat 'stratospheric GSO' platforms, and the move to swarms and clusters of large, low-cost satellites).

Author Biography-- Dr. James R. Stuart is an independent, international consultant specializing in advanced commercial space systems. He is also currently the Chief Technical Advisor for two 'Little NGSO' programs (a licensed Mexican constellation, and a planned U.S. constellation), staff consultant for Teledesic (a 'Broadband NGSO'), and principal or board member in several entrepreneurial technology and space companies involved with communications, satellites and launchers. Dr. Stuart has held positions as Vice President and Chief Architect at Teledesic Corp. in Kirkland, WA. and Chief Scientist and Chief Engineer at Ball Space Systems Division in Boulder, CO. He was previously the founding Chief Engineer of Orbital Sciences Corporation, the Assistant Laboratory Director of the Laboratory for Atmospheric and Space Physics, the creator and first Project Manager of the Mars Observer Project at NASA/Jet Propulsion Laboratory, where he was also Manager of Advanced Planetary Programs. Dr. Stuart has been a memberof the graduate faculty of the University of Colorado at Boulder for the past 16 years: in the Electrical Engineering Department, in the Center for Space Construction, and currently in both the Telecommunications and Aerospace Engineering Sciences Departments. He received his Ph.D. in Systems Engineering (1979), M.S. in Operations Research (1977), and M.S. in Electrical Engineering (1974) from the University of Southern California, and his B.S. in Physics (1968) from the University of Washington. Dr. Stuart has received numerous professional awards, including NASA's Exceptional Service Medal for his project management of the Solar Mesosphere Explorer Project, JPL's highly successful, first modern small satellite project. He is also listed in Via Satellite's "Top 100 Executives in the Satellite Communications Industry". Dr. Stuart holds 8 patents and has published over 90 professional papers on the topics of small satellite systems, space technologies, communications and satellite economics.

Co-author Biography-- Dr. Janet Gleave Stuart received her B.S. in Mechanical Engineering from Utah State University in 1987, an M.S. in Aerospace Engineering Sciences from the University of Colorado at Boulder in 1990, and her Ph.D. in Aerospace Engineering Sciences, also from the University of Colorado at Boulder, in 1995 and also attended the International Space University in Japan in 1994. Dr. Stuart has worked as an engineer at Honeywell Satellite Systems, in Phoenix, and at the Lawrence Livermore National Laboratory, and has consulted for various satellite telecommunications companies, including Leo One Panamericana, of Mexico City, Mexico, and Teledesic, Corp., of Kirkland, WA. Dr. Stuart performed postdoctoral research at the National Renewable Energy Laboratory's National Wind Technology Center in Golden, CO, from 1995-1997, where she worked on wind turbine control systems, including smart-rotor/adaptive-rotor active control systems. She received the National Research Council's Research Associateship Award and is currently conducting her NRC research in large flexible satellite smart structure controls at the Naval Postgraduate School in Monterey, CA.

Optimum Space-Ground Tradeoffs for Millimeter Wave Communication

Paul Christopher
Stanford Telecommunications, Inc.
1761 Business Center Drive, Reston, VA 20190, USA
Phone: 703-438-8017 FAX: 703-438-8112
email: pchristopher@sed.stel.com

ABSTRACT

Optimum satellite frequencies and altitudes are found for downlink limited systems.

1. BACKGROUND FOR AN OPTIMIZATION ANALYSIS

Thomas Teichmann (1) had key insights into satellite communication economics in the early '60s. He recognized that stringent link equation requirements could be maintained by either of two extreme methods: Huge ground stations could be combined with small satellites, or conveniently small ground stations could be combined with a prohibitively large space investment. Figure 1-1 outlines the background of the analysis. A downlink limited communication link may be satisfied with an adequate combination of satellite transmitter (Pt), satellite antenna (At), and ground receiver (Ar). The downlink requires a critical product of the three variables, as shown by a simple equation which can be generated with the aid of unusual units:

$$P_R = P_t \frac{A_R A_t f^2}{r^2}$$

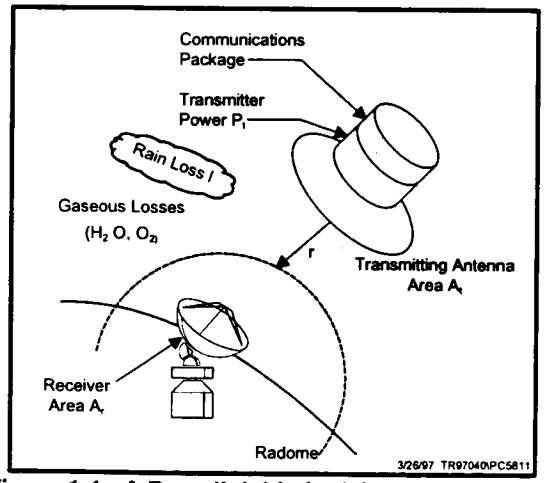


Figure 1-1. A Downlink Limited Communications Link which can be Optimized for Cost/Bit and Frequency

where: f = carrier frequency (GHz)
r = separation distance (ft)
A_R = receiving antenna area (ft^2)
A_t = transmitting antenna area (ft^2)

However, Teichmann recognized that system economics varies in an entirely different way. He postulated that ground system cost was proportional to antenna size, and this was later confirmed by Pope (2) across a large range of frequencies which extended into the millimeter wave region. He also recognized that space costs were strongly dependent on satellite mass and the ability of launch vehicles to deliver large satellite mass at low cost. Teichmann wrote system cost equations similar to those developed in the Appendix. The Appendix also adds Wilson's suntracker (3) atmospheric attenuation to the analysis so that a wide range of frequencies from 1 - 90 GHz may be considered.

Optimum ground and space segment sizes may be found which minimize total system cost. When the ground station and satellite are varied against each other while maintaining the necessary link equation, the optimum ground station antenna system cost turns out to be the same as the optimum satellite antenna system cost. Figure 1-2 illustrates the system cost, relative to an optimum tradeoff, for varying ground antenna size. Note that a quadrupling of the ground antenna size implies a doubling of total system cost. Launch costs are left totally general for Figure 1-1, but system costs for different launch costs are compared in Figure 1-3 for 10^4\$/lb and 10^3\$/lb. Only a factor of $10^{(2/3)}$ exists for the ratio of system costs because ground and satellite system costs are asked to compensate for each other's weaknesses during the optimization.

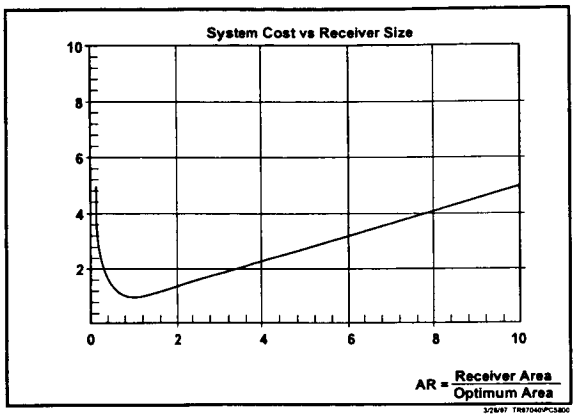


Figure 1-2. System Cost as a Function of Receiver Size

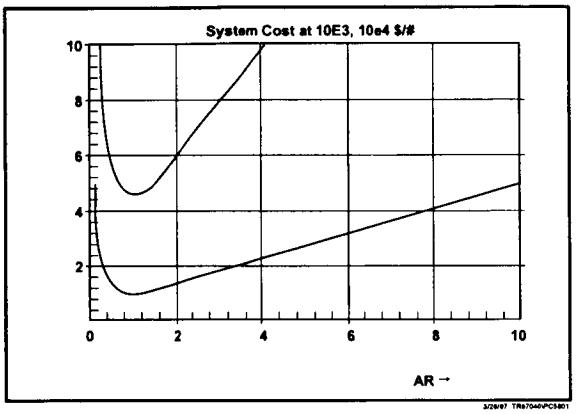


Figure 1-3. System Costs for Different Launch Costs 10^3 $/lb, 10^4 $/lb

2. ATMOSPHERIC ATTENUATION

The Bell Labs suntracker experiments of the late '60s offered a rare opportunity to observe a wide range of frequencies simultaneously across a full range of weather conditions. The attenuation statistics can be used, at a given frequency such as 16 GHz, to construct or verify rain attenuation models. Wilson went further, however, and established correlation between attenuation over a significant part of the millimeter wave region between 30 - 90 GHz. He found that rain attenuation (dB) was typically proportional to $f^{(1.81)}$. The rain attenuation statistics may be represented as Figure 2-1.

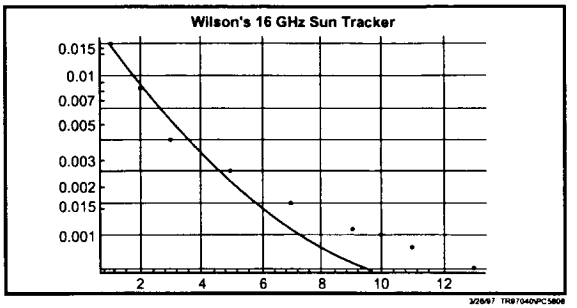


Figure 2-1. Wilson's 16 GHz Sun Tracker Attenuation Statistics (May-June BSTJ 1969)

Gaseous attenuation is also important in the millimeter region. Convenient forms for gaseous attenuation were found (4, 5) for both oxygen and water vapor attenuation which allow representations as Figure 2-2. The small water vapor resonance at 22 GHz and an overwhelming oxygen resonance at 60 GHz are clearly identifiable.

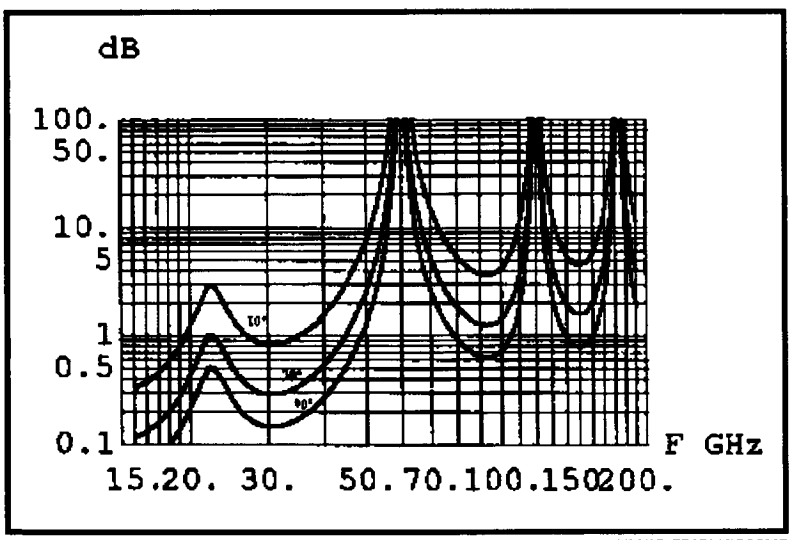


Figure 2-2: Integrated Liebe-Van Vleck Gaseous Attenuation on Ground-Satellite Links (7.75 g/m^3 water vapor) 10° elevation, 30° elevation, and 90° elevation

Total atmospheric attenuation is used here as the sum of gaseous and rain attenuations. A Teichmann analysis would compensate the atmospheric loss optimally by sizing the key ground station and satellite sizes as proportional to (power loss)$^{(1/3)}$. This would imply that expected cost per information bit (Appendix) may be described for a Teichmann analysis over a wide range of frequencies. Figure 2-3 indicates relative expected cost/bit vs. frequency. Three attractive regions appear at 17, 30-40, and 90 GHz. Of course, communication systems typically require reasonable reliability and objections to the higher bands may be justified on the basis of reliability. Figure 2-4 includes a second ground station for independent reception, and the system costs shift upward. The higher bands are seen to remain attractive.

The millimeter wave band may be especially attractive for aircraft. Attenuation drops, and reliability increases for aircraft at 10,000 feet as in Figure 2-5.

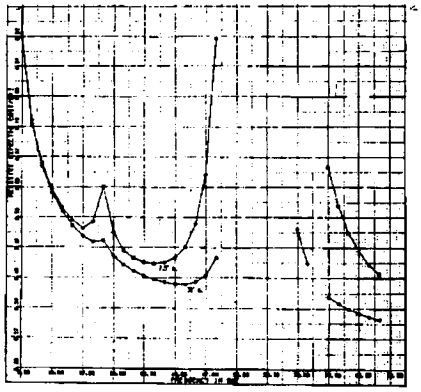


Figure 2-3. Expected Cost/Bit vs. Frequency (no diversity, altitude = 0)

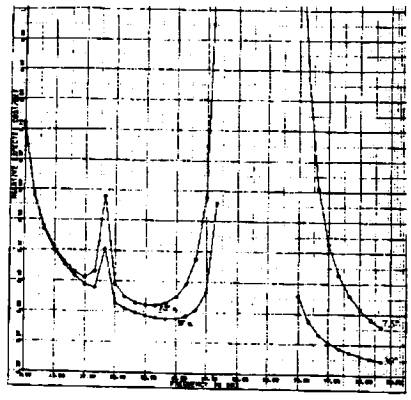


Figure 2-4. Expected Cost/Bit vs. Frequency Dual Diversity, Altitude = 0)

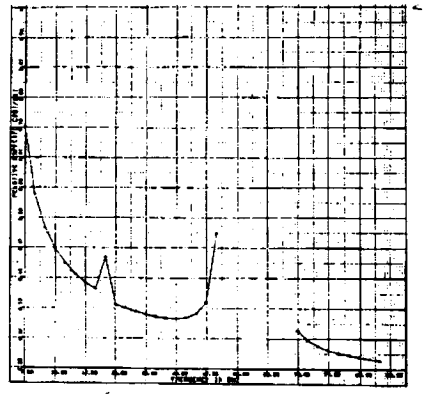


Figure 2-5. Expected Cost/Bit vs. Frequency (No Diversity, Altitude = 10,000 Feet)

System Cost may also be presented as a 3D plot, as Figure 2-6. The optimum receiver size in the plot is at AR = 1, where AR is relative to optimum size.

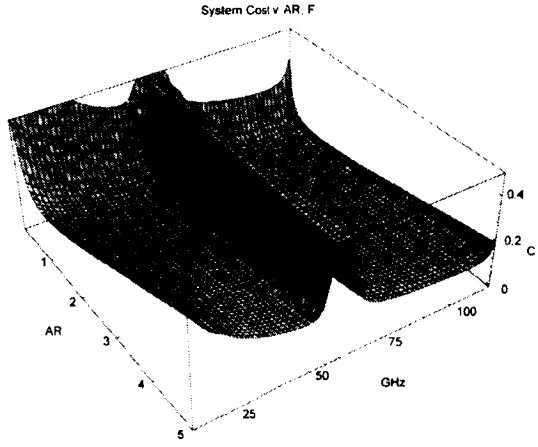


Figure 2-6. System Cost vs. Fraction of Optimum Receiver Size and Frequency

3. OPTIMUM ALTITUDE FOR MILLIMETER WAVE SATELLITES

System cost was seen to depend on the launch cost per unit in the Teichmann analysis. Cost per unit mass increases with satellite altitude, and this suggests that a tradeoff for optimum altitude may be done.

If

Satellite value = (present value of lifetime) x

$$\frac{P_{vis}^{N}}{(cost / lb) \bullet (r^2)}$$

where

P_{ris} = Earth visibility area
N = 1, if crosslinks available
2, if single hop required

plots may be done for satellite value.

Lifetime vs. altitude may be considered as in Figure 3-1. Figure 3-2 addresses satellites with crosslinks. It shows satellite value with and without lifetime considerations. The top curve, with lifetime, peaks sharply near 1000 km. Figure 3-3 shows single hop satellites with cost effectiveness peaking in the 2000-10,000 km range.

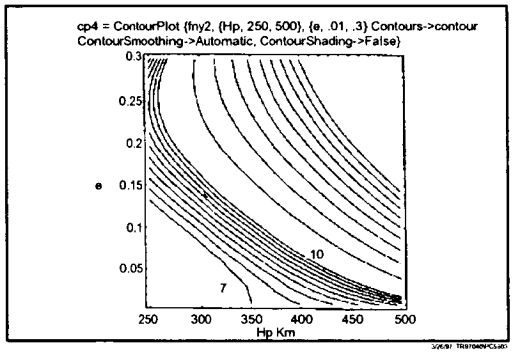


Figure 3-1. Lifetime Contour Plot

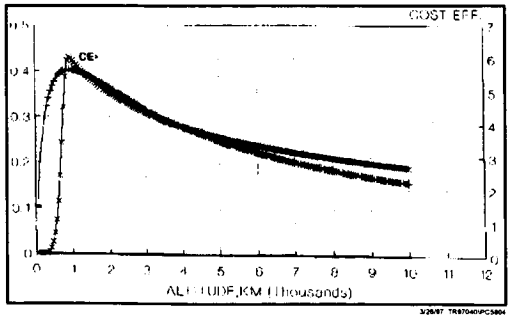


Figure 3-2. Cost Effectiveness vs. Altitude Includes Lifetime, Present Value, ISLs

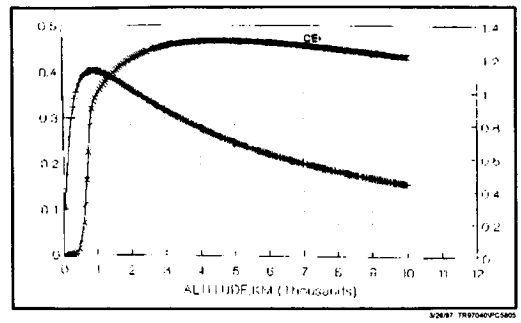


Figure 3-3. Cost Effectiveness vs. Altitude (No ISLs) (Nearly Optimum Form 2000-10,000 km)

CONCLUSION

Optimal system tradeoffs allowed excess atmospheric losses for millimeter wave systems to be compensated, so that system costs increased slowly as $(loss)^{(1/3)}$. The 30-47 GHz and 90 GHz regions were attractive. Optimum altitude relations were also suggested by the Teichmann analysis. If crosslinks were easily available for convenient satellite configurations, altitudes near 1000 km were especially attractive. Satellites without crosslinks would favor 2000-10,000 km altitude.

REFERENCES

1. T. Teichmann, "Some Elementary Considerations of Satellite Earth Communication Systems", IRE Trans. Space Electronics and Telemetry, Vol. SET-7, June 1961.
2. D. L. Pope, "Parametric Representations of Ground Antennas for Communication System Studies", Bell System Technical Journal, Vol. 47, No. 4, Dec. 1968.
3. R. W. Wilson, "Sun Tracker Measurements of Attenuation by Rain at 16 and 30 GHz", BSTJ, Vol. 48, May-June 1969.
4. A. H. Jackson, P. Christopher, "A LEO Concept for Millimeter Wave Communication", Proc. IMSC, Ottawa, June 1995.
5. A. H. Jackson, P. Christopher, "Optimum Millimeter Wave Regions for Satellite Communication", URSI Radio Science Meeting, Baltimore, MD., July 1996.

ACKNOWLEDGMENT

Vickie Granberg of Stanford Telecom interspersed the paper with a pressing schedule of STel documents.

APPENDIX

A Teichmann analysis begins with a form of the ideal information rate (Shannon rate) as:

$$I = W \log_2 (1 + S) \text{ bits/sec} \quad (1)$$

where:

I = information rate in bits/sec
W = information bandwidth (Hz)
S = required signal-to-noise ratio.

The required power, P_r, as measured at the input to the first receiver of the ground terminal, can be expressed as:

$$P_r = nSW \quad (2)$$

where:

n = received noise power plus internal noise density

Satellite instrumentation power requirements were:

$$P_{ic} = wJ$$

where:

w = power requirement per bit/sec (watt-sec/bit)
J = actual information transmission rate (bits/sec)

The ratio of received to transmitted power, P_t, is (within a few percent):

$$\frac{P_r}{P_t} = \frac{A_r A_t f^2}{r^2} \quad (3)$$

where:

f = carrier frequency (GHz)
r = separation distance (ft)
A_r = receiving antenna area (ft^2)
A_t = transmitting antenna area (ft^2)

The units appear inconsistent in Equation (3). A squared velocity of light term originally appeared in the denominator, but it was numerically canceled because of the odd units chosen. These inconsistent units will later (optimum A_r, A_t relations, equations (12) and (15)) cause an ambiguity between time and length. These inconsistent units are retained here only because we do not wish to change useful existing development until it is necessary (e.g., to introduce real atmospheric losses).

If the mission cost can be amortized over a fixed number of bits, Q, the useful mission life is Q/J.

Also, if the actual system operates at a fixed fraction, β, of the Shannon rate,

$$J = \beta I \text{ bits/sec},$$

the cost equation can now be written in terms of cost (dollars) per bit of information. Cost of receiver operation per bit of information is:

$$\frac{a}{J} = \frac{\alpha}{\beta W} \quad (4)$$

where:

a = cost of operating receiver system/sec of transmission

$$\alpha = \frac{a}{\log_2(1+S)}$$

Cost of receiver maintenance is:

$$\frac{bA_r}{J} = \frac{bA_r}{\beta I} = \frac{bA_r\alpha}{\beta W a} \quad (5)$$

where:

b = cost of maintaining receiver system/sec of mission life/ft^2 of receiving antenna.

The next important ingredient in the Teichmann cost model was the cost of the space system. He made the startling proposal in 1961 that cost of the space segment should be a linear function of the weight of the space segment. This is one reason why the Teichmann model did not gain wide acceptance at the time, but is also one reason why we are now recalling the Teichmann model.

Satellite weight = antenna weight + communications system weight + transmitter system weight

$$= kA_t + gwJ + \frac{hr^2 P_r}{f^2 A_r A_t} \quad (6)$$

$$= kA_t + gw\beta W \frac{a}{\alpha} + \frac{hW}{\gamma A_r A_t}$$

where:

k = transmitter antenna weight (lb/ft^2)
g = communications system weight (lb/watt)
h = transmitter system weight (lb/watt)

$$\gamma = \frac{f^2}{r^2 nS}$$

The transmitter weight of Equation (6) is that required by free space loss.

Satellite cost per bit

$$= \frac{c}{Q}\left[kA_t + gw\beta\frac{aW}{\alpha} + \frac{hW}{\gamma A_r A_t}\right] \quad (7)$$

where:

c = costs per pound on orbit.

Total system cost per bit of information, C, is:

$$C = \frac{cq}{Q}w\beta\frac{a}{\alpha} + \frac{bA_r\alpha}{\beta Wa} + \frac{ckA_t}{Q} + \frac{\alpha}{\beta W} + \frac{chW}{Q\gamma A_r A_t} \quad (8)$$

Three major variables may be optimized simultaneously. The partial derivatives of cost, C, with respect to W, A_r, and A_t, may be found, and optimum solutions derived for these variables.

(A designer with fixed transmission rate requirements does not wish to study the change of cost with respect to bandwidth. The variation of system cost/bit with respect to bandwidth does not essentially change the interesting features of the final result of Equation (18), however.) The equation for $\partial C/\partial W$ is included here for completeness.

$$\frac{\partial C}{\partial W} = \frac{cgw\beta a}{Q\alpha} - W^{-2}\left(\frac{bA_r\alpha}{\beta a} + \frac{\alpha}{\beta}\right) + \frac{ch}{Q\gamma A_r A_t} \equiv 0 \quad (9)$$

$$\frac{\partial C}{\partial A_r} = \frac{b\alpha}{\beta Wa} - A_r^{-2}\frac{chW}{Q\gamma A_t} \equiv 0 \quad (10)$$

$$\frac{\partial C}{\partial A_t} = \frac{ck}{Q} - A_t^{-2}\frac{chW}{Q\gamma A_r} \equiv 0 \quad (11)$$

The optimum A_t can be solved immediately from Equations (10) and (11):

$$A_{t_{opt}} = \left(\frac{hb\alpha Q}{\gamma\beta ac}\right)^{1/3}\left(\frac{1}{k}\right)^{2/3} \quad (12)$$

One form of A_r is:

$$A_r = W\left(\frac{hk}{\gamma}\right)^{1/3}\left(\frac{\beta ac}{b\alpha Q}\right)^{2/3} \quad (13)$$

This form for ground receiving antenna area is useful when the information bandwidth is not optimum. The optimum information bandwidth may be found from Equation (9):

$$W_{opt} = \left(\frac{Q}{cgwa}\right)^{1/2}\left(\frac{\alpha}{\beta}\right) \quad (14)$$

which leads to:

$$A_{r_{opt}} = \left(\frac{ca}{Q}\right)^{1/6}\left(\frac{hk\alpha}{\gamma\beta}\right)^{1/3}\left(\frac{1}{gw}\right)^{1/2}\left(\frac{1}{b}\right)^{2/3} \quad (15)$$

When Equations (12), (14), and (15), are substituted into the satellite weight, Equation (6), optimum satellite transmitter weight is seen to be equal to satellite antenna weight. Substitution of these optimum parameters back into the cost expression, Equation (8), yields:

$$\text{ground cost} = \left(\frac{cgaw}{Q}\right)^{1/2} + \left(\frac{bhk\alpha}{a\gamma\beta}\right)^{1/3}\left(\frac{c}{Q}\right)^{2/3} \quad (16)$$

$$\text{satellite cost} = \left(\frac{cgaw}{Q}\right)^{1/2} + 2\left(\frac{bhk\alpha}{a\gamma\beta}\right)^{1/3}\left(\frac{c}{Q}\right)^{2/3} \quad (17)$$

$$C_{opt} = 2\left(\frac{cgaw}{Q}\right)^{1/2} + 3\left(\frac{bhk\alpha}{a\gamma\beta}\right)^{1/3}\left(\frac{c}{Q}\right)^{2/3} \quad (18)$$

(When transmission bandwidth is fixed, Equations (12) and (13) would be used for substitution into the cost expression, Equation (8). The result for C_{opt} would then have been:

$$C_{opt} = \frac{cq}{Q}w\beta W\frac{a}{\alpha} + \frac{\alpha}{\beta W} + 3\left(\frac{bhk\alpha}{a\gamma\beta}\right)^{1/3}\left(\frac{c}{Q}\right)^{2/3}$$

The final term is identical with that in Equation (18) and it is the only frequency dependent term considered).

Global Multimedia Mobile Satellite Communications System (GMMSS)

Shingo Ohmori,* Masao Matsumoto,** and GMMSS Study Group
*Communications Research Laboratory
**Ministry of Posts and Telecommunications
Telephone: +81-423-27-7501, Fax: +81-423-27-6698, E-mail: shingo@crl.go.jp

Abstract

The Ministry of Posts and Telecommunications (MPT) has begun feasibility studies on a global multimedia mobile satellite system (GMMSS), which can provide global personal communications service by means of a group of low-earth-orbiting (LEO) satellites with a user data rate of 2 Mbps to handy terminals. The target for commercial implementation is the year 2010. The GMMSS concept is considered the second phase of the future land mobile telecommunications system (FPLMTS), and the second generation of the big-LEO systems. This paper gives an outline and the main results of the GMMSS study, which will be published in a report by the MPT. The technical feasibility study concludes that user data rates of 2 Mbps could be achieved in the S and C bands, but that, it would be very difficult to provide the service in the Ka and millimeter-wave bands.

1. Introduction

In recent years, innovative new concepts for mobile (personal) satellite communication systems using a group of low-earth-orbiting (LEO) satellites have been proposed primarily in the United Sates. These systems, such as Iridium, Globalstar, Odyssey and ICO, have aimed to provide worldwide communication services with user data rate around 4.8 kbps, [1] primarily for voice communications with handy terminals. Also, the standardization of the future land mobile telecommunication system (FPLMTS, IMT-2000) is under way at the International telecommunication Union (ITU) in order to implement worldwide personal communications services at a user bit rate over 2 Mbps in the year 2000. With the IMT-2000, data rates of 2 Mbps can be achieved for indoor offices, but are as slow as 9.6 kbps for satellite links. [2]

The Ministry of Posts and Telecommunications (MPT) has begun the feasibility studies on a global multimedia mobile satellite system (GMMSS), to provide global personal communications service at a 2 Mbps user data rate to handy terminals by using a group of LEO satellites. The system is targeted for commercial implementation in 2010. This concept is considered the second phase of the FPLMTS and the second generation of the big-LEO systems.

We here present an outline and the principal results of the study report, which was drafted by the study group listed at the end of this paper.

2. Future Prospects for Mobile Satellite Communications

2.1 Services

In the year 2010, when the Global Multimedia Mobile Satellite System is expected to begin providing commercial service, the communications market will be characterized by the following factors.

(1) Multimedia communications

High-speed networks consisting of optical fiber networks using the ATM transmission mode will have already been established all over the world. Through these high-speed backbones, the present voice communications, computer communications, and broadcasting will be merged into multimedia communications.

(2) Global mobile communications

In 200 years, satellite systems such as Iridium, Globalstar, Odyssey and ICO, using big-LEO and/or goestationary earth orbiting (GEO) satellites, will have already been put into commercial service. These systems are going to provide mainly voice communications (4.8 kbps), through handy terminals. They are considered the first generation of the personal satellite communication systems. The ITU is now drafting the standards for the FPLMTS (IMT-2000), which aims to achieve user data links of 2 Mbps for indoor office, 512 kbps for outdoor and pedestrian, 144 kbps for vehicular, and 9.6 kbps for satellite links. Second-generation big-LEO systems or second generation FPLMTS are anticipated by 2010. Expectations are high that the second generation of big-LEO satellite systems with provide multimedia services for personal use with not only voice but also video transmissions at a rate of around 2 Mbps.

2.2 Estimation of market for the GMMSS

(1) Potential users

The following figures are assumed in estimating the market for the GMMSS. The total population of the world will be 7 billion in the year 2010. Of the total population, 21% (1.5 billion), will use mobile communications, and 1% of the mobile communication users (15 million) will use GMMSS systems. Based on the number of users estimated users for the year 2010, the market for mobile satellite communications would be 76 billion US dollars per year. If the share of the GMMSS were 10% of the total market, $7.6 billion a year could be expected for the GMMSS.

(2) Cost of the GMMSS

System cost = (Total income/year) x (ratio of company share)
x (ratio of depreciation amount) x (satellite life)
=$8.2 Billion x 70% x 20% x 10 years
= about $11.5 Billion

3. Technical feasibility studies on the GMMSS

Technical feasibility studies on the GMMSS, focusing on provision of 2-Mbps user links, were carried out for the S-, C-, Ka- and millimeter-wave bands. Satellite link parameters, especially in LEO systems, are extremely sophisticated to set, because of their there large number; they include frequency bands, orbital altitudes, constellation of satellites, number of satellites, and so on. Hence, we chose the link parameters shown in Table 1

to facilitate comparison of the feasibility studies of different frequency bands.

Table 1. Satellite link parameters in the GMMSS

Antenna efficiency	40%
Feeder loss of satellite	2 dB
Feeder loss of mobile terminal	2 dB
Antenna gain degradation at edge of service coverage area	3 dB
Pointing error of satellite antenna	1 dB
Receiver noise temperatures of satellite and mobile terminal	400 K
Antenna noise temperatures	300 K (satellite) 150 K (mobile)
Rain attenuation margin	0 dB (S band, C band)
Fading margin	5 dB (S band, C band)
(Rain attenuation + fading) margin	7 dB (Ka, MM wave)
	7 dB
Modulation	QPSK
Required Eb/No (at BER=10-5, 2 dB hardware degradation)	7.7 dB

In the feasibility studies on the GMMSS, the key parameters of satellites shown in Table 2 were assumed to be technically fabricated in the year 2010.

Table 2. Key parameters of satellites feasible in the year 2010

Number of satellites	Under one thousand
Satellite antenna diameter	About 10 m (S band)
	Under several m (Ka-band)
Antenna beamwidth	Over 0.2 degrees
Number of beams	Under one thousand
Number of antenna elements (direct radiating active phased array)	Under one thousand
Maximum transmission power	About 1 kW

4. Feasibility studies on different frequency bands

Feasibility studies were done on S-, C-, Ka- and millimeter-wave bands on two types of terminals. The first terminal was a handy terminal having a non-tracking (omni-directional) antenna with a gain of 3 dBi. The second one was a portable terminal larger than the handy terminal because it used a tracking antenna with a gain of 11 dBi. Both terminals had a transmission power of 1 W. In each frequency band, not only 2 Mbps, but also 384 and 64 kbps rates were studied. Examples of link budgets in different frequencies are listed in Table 3. More detailed results of the studies are described in the following sections.

Establishing detailed system configurations for the GMMSS is very

intricate, because of large number of parameters in system and hardware designs. In any event, it can be presumed that satellite antennas with a large number of radiating elements are key technologies in achieving 2-Mbps service for handy terminals.

Table 3. Link budgets of the GMMSS

Mobile terminal	Non-tracking (1W, 3dBi)	Tracking (1W, 11dBi)	Non-tracking (1W, 3dBi)	Tracking (1W, 11dBi)	Tracking (1W, 11 dBi)	Tracking (1W, 32 dBi)	Tracking (40mW,33dBi)	Tracking (40mW,35dBi)
Frequency	S-band		C-band		Ka-band		Millimeter-wave	
User link data rate					(One-way)	(Tow--way)		
Up-link	2 Mbps		2 Mbps		4.2 - 32 kbps	2 Mbps	64 kbps	
Down-link	2 Mbps		2 Mbps		2 Mbpsx50ch	2 - 100 Mbps	2 Mbps	
Satellite								
Altitude	700 km		700 km		700 km		1,000 km	10,000 km
Mini. elevation	20 deg.		20 deg.		20 deg.	40 deg.	40 deg.	
Numbers	10 planes X 18		10 planes X 18		180	50-800	903	66
Antenna size	11 m Φ	4.3 m Φ	7.8 m Φ	3.0 m Φ	33 cm Φ (Rx) 50 cm Φ (Tx)	0.2 m Φ	0.1 m Φ (Rx) 1 m Φ (Tx)	1 m Φ
Beam footprint	40 x 100 km	90 x 200 km	22 x 56 km	51 x 113 km		50 km Φ	9 km Φ	90 km Φ
Elements number	4,000	600	12,500	1,900	500		40,000	
Beam number/satellite								
Up-link	100 (with beam scanning)		1,000 (with beam scanning)		100	450	128 cell	1,800
Down-link	100 (with beam scanning)		20 (with beam scanning)		100	200	128 supercell	1,800
Tx power	1 W/ch		1 W/ch (50 W/ Beam)		10-20 W/ch	1 W/ch	30W/Trapon	100W/Trapon
Channel capacity	1,000 ch		1,000 ch		50-100 ch	1,000 ch	800 ch	100 ch
Terminal								
Antenna size	Patch/Helical	27 cm Φ	Patch/Helical	11 cm Φ	3.8 cm Φ	20 cm Φ	10-15 cm Φ	10-20cm Φ
Number of antennas	1	2	1	2	2	2	2	2
Antenna gain	3 dBi	11 dBi	3 dBi	11 dBi	11 dBi	32 dBi	33 dBi	35 dBi
Tx power	1 W	1 W	1 W	1 W	1 W	1 W	400 mW	
Multiple access								
Up-link	SCPC or CDMA, TDD		SCPC or CDMA, TDD		Multicarrier		FDMA	
Down-link	Multicarrier TDMA, TDD		Multicarrier TDMA, TDD		Multicarrier, TDMA		TDM	
Inter-satellite link	Ka-band, MM-wave, Laser		Ka-band, MM-wave, Laser		Ka-band, MM-wave, Laser		Laser	

4.1 S band

The S-band system has the advantage that a large antenna can be constructed having a relatively small number of antenna elements. For a minimum elevation angle of 20 degrees and an orbital altitude of 700 km, the satellites would number about 180. In the case of a handy terminal, the size of the satellite antenna would be about 11 m. A direct-radiating active phased array antenna , if required, would be very difficult to produce. In the case of a portable terminal, the size of the satellite antenna would be about 4 m. This could be achieved even by present technologies, but would have the disadvantage of being too large for personal use. Considering the complexity of beam forming networks, we assume a total of 100 beams. These beams would scan the area required by the satellite.

4.2 C band

In an up link, a single beam is directed to a single user. Because the beamwidth becomes narrower than the S band at higher frequencies. A satellite must have the same number of beams as there are users in the service area. In a down link, on the other hand, the number of beams will be reduced to about 20 (100 ch/50 ch) by employing time division multiplexing (TDM) of 50 channels per beam. In this case, users will be able to transmit high-speed data at 100 Mbps by means of high-speed transmission terminals. In the case of handy terminals, the number of satellite antenna elements

required has grown to about 12,500, which is three times larger than that for the S band. This point is a disadvantage, compared to the case of the S band. Nonetheless, the C band system will be favored candidate for the GMMSS.

4.3 Ka band

Providing 2-Mbps transmission services in Ka-band systems would be extremely difficult, especially in up-link. This is because the large number of beams in up link and the extremely large number of antenna elements are too difficult to produce even by 2010. Use of a portable terminal with a tracking antenna might make two-way 2-Mbps service possible. If around 100 kbps transmissions are assumed only in up link (one-way), 2-Mbps transmission would be possible in down link for a portable terminal with a non-directional antenna. Such systems would closely resemble the Teledesic system. [3]

4.4 Millimeter-wave band

In the millimeter-wave system, it would be extremely hard to provide 2-Mbps service for personal handy terminals for both up- and downlinks. However, if portable terminals with tracking antennas were used, 2-Mbps service could be provided for downlinks. Each satellite has its own service coverage, which is a fixed footprint on the earth's surface. Inter-satellite communications will be carried out by laser links at a high speed of 1.2 Gbps. Each satellite coverage area comprises of 128 super cells, and each super cell comprises of 100 cells. A broad beam, which covers a super cell, provides 64 kbps services for both up and down links. In downlinks, a scanning beam scans 100 cells within a super cell according to the demand from each cell.

5. Conclusion

Table 4 summarizes the feasibility studies on the GMMSS.

Table 4. Summary of feasibility studies on S-, C-, Ka- and MM-wave systems

Frequency	S-band		C-band		Ka-band		MM-wave	
Data rate	Up	Down	Up	Down	Up	Down	Up	Down
2 Mbps	○	○	○	○	△	○	△	△
384 kbps	◎	◎	◎	◎	○	◎	△	○
64 kbps	◎	◎	◎	◎	◎	◎	○	○
9.6 kbps	FPLMTS (satellite)							
4.8 kbps	Iridium, ICO							

◎ : Feasible by advancing present technologies

○ : Feasible with technical breakthrough

△ : May not be possible to achieve by 2010 .

According our studies, development of the Ka band and millimeter-wave systems by 2010 would not be possible, because the gap is too wide between the present technologies and those required. On the other hand, it would be possible to achieve bi-directional 2-Mbps user links for handy terminals for the S- and C-band systems. Also, it would be possible to provide bi-directional 2-Mbps service in the Ka-band if portable terminals were used as fixed stations. These systems would correspond the Teledesic system, which has been proposed for providing 1.5-Mbps service for portable terminals by means of 870 LEO satellites. In the case of millimeter-wave systems, almost the same as with the Ka-band systems, 2-Mbps service would be extremely difficult to achieve, especially in up link. However, if transmissions between up- and down-link are assumed to be asymmetrical, it might be possible to provide a 2-Mbps rate for down links only.

According to our studies, as 2-Mbps transmission in user links is achieved, antenna diameters will become about 10 m for the S band, and about 5 m for the Ka band. Such antennas will have several thousand radiating elements, each element with a very small module consisting of filters, amplifiers, and digital phase shifters. Development of a direct-radiating, active phased array antenna would be one of the most important breakthroughs.

The study described in this paper was performed by the study group for global multimedia mobile satellite systems (GMMSS) for next-generation LEO systems. Organized by Ministry of Posts and Telecommunications, the group consists of Communications Research Laboratory (CRL), National Space Development Agency of Japan (NASDA), Nippon Telegrapgh and Telephone Corporation (NTT), Kokusai Denshin Denwa Co. Ltd. (KDD), NEC Corporation, Mitsubishi Electric Corporation, Hitachi, Ltd. Fujitu Ltd., Toshiba Corporation, Japan Telecom, Space Communications Corporation, IDC, Shikoku Electric Power Company and Telecommunications Advancement Organization (TAO). The TAO will start a research and development program on the GMMSS from April 1997 by organizing an R&D team form the private-sector comprises that took part.

References

[1] Proceedings of the Asia-Pacific Satellite Communications Conference '96, Korea, November 1996.

[2]11th Meeting of ITU-R Task Group 8/1, Geneva, October 1996.

[3]Proceedings of Second Ka Band Utilization Conference and International Workshop on SCGII, Florence, Italy, September 1996.

A Geostationary Satellite System for Mobile Multimedia Applications Using Portable, Aeronautical and Mobile Terminals

G. Losquadro
Alenia Aerospazio
Divisione Spazio
Via Saccomuro 24
00131 Roma
ph. +39 6 4151 2621
fax +39 6 4151 2102
Losquadr@alespazio.it

M. Luglio
Dipartimento di Ingegneria Elettronica
Università di Roma "Tor Vergata"
Via della Ricerca Scientifica
00133 Roma
ph. +39 6 7259 4453
Luglio@utovrm.it

F. Vatalaro
ph. +39 6 7259 4464
fax +39 6 2020519
Vatalaro@utovrm.it

ABSTRACT

Satellite systems shall play a very significant role in providing wideband personal services, complementing terrestrial systems. Satellite multimedia applications will require wide bandwidths, high degree of flexibility on-board and intelligence at terminal level. The combined use of Ka band (20-30 GHz) and EHF band (40-50 GHz) can extend interactive multimedia and personal services to ubiquitous and mobile users.
A satellite system named SECOMS/ABATE (Satellite EHF Communications for Multimedia-mobile Services/ACTS Broadband Aeronautical Terminal Experiment) is under study by a team of 16 European and U.S. Partners[1], in the frame of the ACTS (Advanced Communication Technology and Services) program of the European Community.
The system will be composed by:

- Ka band system component to answer business and mobile users needs for large bandwidths, interactivity and access mobility at a reasonable cost;
- EHF band system component to be added and integrated with the Ka band system to increase capacity, expand set of services and enhance mobility over the same Ka band component service area.

The paper will mainly focus the attention to the system architecture. The major characteristics of the space segment, the ground control station and the portable, aeronautical and mobile user terminals will be outlined too, providing the updated results of the study.

OBJECTIVES OF THE PROJECT

The main scopes of the project are the system architecture identification, the technological development and a demonstrator for the definition of future advanced mobile multimedia service operating in the Ka (20-30 GHz) and EHF bands (40-50 GHz), capable to provide the widest class of telecommunication services [1, 2].

The project foresees two different tasks characterised by different environments in which the provision of the broad-band services shall be experimented:

- SECOMS sub-project, dedicated to the general Mobile MultiMedia Satellite Communications;
- ABATE sub-project, dedicated specifically to the Aeronautical Mobile Communication Environment.

In the frame of the SECOMS sub-project the following objectives shall be reached:

1. a broadband system architecture corresponding to identified service requirements and terminals;
2. a model for the EHF channel that matches with the measures to be performed during field trials [3];
3. a model of the overall EHF satellite link including on-board, EHF channel and ground equipment;
4. a land-terminal prototype, including receiving antenna, RF section, baseband, user interface, controller with a special developed hardware for 20-30 GHz and with some off the shelf components for the 40-50 GHz band;
5. the validation of the overall concept through field trials performed at 20-30 GHz with ITALSAT satellite and at 40-50 GHz with airplane support.

In the frame of the ABATE sub-project the definition of the future advanced Aeronautical Mobile Satellite System (AMSS) operating at Ka band capable to provide broadband telecommunication services as well as the classical services associated to low bit rate, shall be reached by the provision of:

1. a broadband aeronautical mobile communication system architecture corresponding to the service and terminal requirements;
2. a satellite link design optimised for the aeronautical service, based on measured channel characteristics;
3. a novel broadband aeronautical terminal design as well as the relevant satellite payload optimised for maximum system throughput as well as for the aeronautical services to be provided;
4. an aeronautical terminal prototype, utilising available equipment, developed in the frame of an American experimental program, Advanced Communications Technology Satellite (ACTS) [4], and other (mo-dem, antenna ponting sub-system) to be developed;
5. the corresponding mo-dem section prototype to be used at the gateway station;

[1] Alenia Aerospazio (I), Alcatel Espacio (E), Dassault Electronique (F), DLR (D), EAC (F), INMARSAT, Intracom (Gr), JPL (USA), Nuova Telespazio (I), Politecnico di Milano (I), Siemens (D), Space Engineering (I), University of Bradford (UK), University of Surrey (UK), Università di Roma Tor Vergata (I), Università di Roma La Sapienza (I).

6. the validation of the developed Aeronautical Terminal (AT) prototype and corresponding earth station equipments, through airplane flight trials, performed at 20-30 GHz, by the use of the operating ITALSAT satellite.

The ABATE sub-project allows to link the aeronautical mobile experimental R&D / trial activity, which is on going in the frame of the American ACTS Program.
By building a collaboration between the two projects, large mutual advantages are expected in the frame of joint demonstration, joint experimentation, key issues addressing and knowledge base accumulation and advancements for the implementation of a broadband multimedia onboard service.

OVERALL SYSTEM ARCHITECTURE

The objective is to demonstrate feasibility of Ka and EHF bands exploitation for implementation of mobile, including aeronautical, multimedia services with a system having a star topology architecture.
A constellation of at least two geostationary satellites (one primary and one back-up satellite) is foreseen for each sub-project component. However, depending on needed capacity, a constellation of more than two satellites may be necessary. The user equipment segment includes terminals of different capabilities, for different applications. The gateways, system and network control centres perform the functions of interface with the space segment, and of control and supervision of the system, respectively. Interfaces are also provided with fixed and mobile public networks, to whom the system is intended to be integrated, as well as with private networks.
The Aeronautical Mobile Communication system based on the concepts proposed in the frame of the ABATE sub-project shall operate in the environment depicted in figure 1. The ground segment will be characterised by some GESs (Ground Earth Stations) that will be interfaced with different ground subnetworks, having the requirements to provide different type of services. In the figure it is possible to see system architecture in which all the aeronautical ground subnetworks are foreseen to be connected with the satellite mobile subnetwork.
The use of high frequencies in a hostile environment such as the aeronautical one, must cope with:

- challenging channel conditions (relatively high values of Doppler frequency shifts and Doppler rates, shadowing, Doppler spread, multipath etc.);
- use of low-cost,low-power consumption components and subsystems for the Aeronautical Terminal).

In the frame of the ABATE sub-project several activities are being carried on. The first activity is devoted to the definition of the requirements for the Aeronautical Services (AS) and for the Aeronautical Terminal (AT).
The second activity is devoted to system feasibility studies, aimed at defining an advanced system to achieve very high performance in terms of efficient and flexible use of the communication resources (bandwidth and power) for aeronautical service. In the frame of this activity a link design is foreseen based on a characterisation of the aeronautical channel, in order to quantify the impairments signal due to the aeronautical environment (such as the Doppler frequency shift effect, the shadowing/diffraction and consequent fadings due to the aircraft structure, etc.).

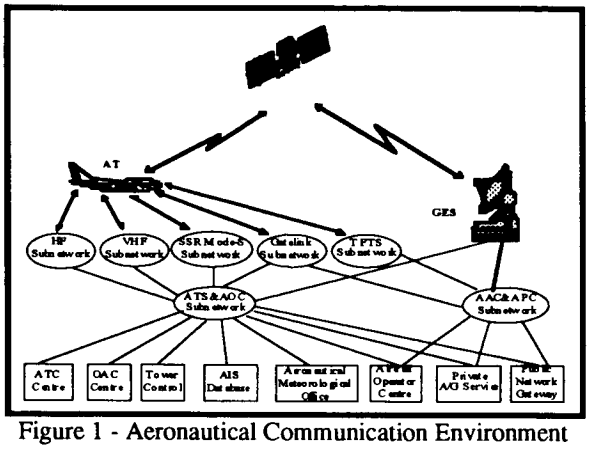


Figure 1 - Aeronautical Communication Environment

The last activity is devoted to definition and construction of a demonstrator for the inter-trial experiment, that will make use of the sub-system prototypes available from the American ACTS Programme. The demonstrator will be aimed at verifying that the main outcomes of the basic system studies are fulfilled. The adopted scheme for the demonstrator is the same as for the Ka land mobile broadband communication demonstrator (SECOMS sub-project), based on maximum use of existing earth stations facilities and Italsat satellite for experiments and demonstrations.
This project will consider mainly the GEO scenario: however activities are in progress to verify and compare the performances of non GEO systems having characteristics similar to SECOMS/ABATE.

SECOMS SUB-PROJECT: ARCHITECTURE

Several analyses and trade-offs have been carried on to identify the SECOMS system architecture which best optimises the performances of both the space and earth segments. The following concepts have been adopted:

1. use of multiple co-located satellites, connected through Inter-Satellite Links (ISLs), to best adapt the SECOMS system to the market demand and evolution;
2. implementation of a "dual-band" integrated system to support a wide range of services with different data rates in different bands (Ka and EHF);
3. minimum system configuration at service start-up: one Ka band + one EHF band satellite (plus spares);
4. use of High-Gain Spot Beams on-board antennas ;
5. regenerative payloads implementing On-board Fast Digital Processing (OFDP) allowing direct interconnection among users within the service area;
6. use of the most appropriate Multiple Access/Distribution (MA/D) technique;
7. use of terminals with limited EIRP and G/T;

Furtherly, the following aspects have been addressed to define the SECOMS network architecture and the associated system elements:

- throughput requirements of the supported multimedia services,
- cost effective design for mobile satellite terminals,
- interfacing with external public and private terrestrial/satellite networks allowing evolution towards Global Information Infrastructure (GII),
- mobility management requirements.

A great variety of data traffic capabilities are envisaged, ranging from few kbit/s up to 2 Mbit/s. Therefore, several types of satellite terminals characterised by different user interface(s), uplink data traffic capacity, and operational frequency band are proposed.

The interworking aspects with the existing Public Terrestrial Networks (PTN) have been duly considered in defining the network architecture. The presence of large fixed earth stations or Gateways (GTW) will assure the distribution of the traffic originated from the PTN towards the large population of mobile users directly connected to the "SECOMS satellite network" and vice-versa. Two other fixed satellite earth stations have been also identified: Service Provider Stations (SPS) and Inter-Link Stations (InLS). The former for providing access to/from external Services Providers aiming at direct diffusion of locally generated data/information; the latter allowing interconnection with terrestrial broadband backbones and different satellite systems covering other parts of the globe (i.e. USA, Asia,..) and providing the same kind of multimedia services.

Point-to-point and point-to-multipoint interconnections among mobile users and mobile users-fixed stations everywhere located within the coverage area are requested by the supported multimedia services which are also characterised by different real time constraints and variable transmission rate demand.

The solution based on a centralised ground station, for system and mobility management, has been selected as the most effective for the SECOMS system. Thus the following network key elements have been identified:

- Ka-band payload,
- EHF-band payload,
- Mobile Sat-Terminals,
- Fixed Stations (GTW, SPS, InLS),
- Master Control Stations (MCSs).

These elements shall interface the following external elements:

- Satellite Control Centres (SCC) for monitoring and commanding of the payload functions;
- SECOMS Network Operator for system administration and billing;
- Service Providers managing "virtual" private networks;
- Public Terrestrial Networks (PTNs);
- User equipments.

A system architecture based on two subnetworks interconnected via ISL, including the Ka-band and EHF band elements respectively, has been selected.

Figure 2 shows the Ka-band network architecture. It includes the satellite terminals, the Ka-band payload, the Gateways, the Service Provider Stations, the Inter-link Stations and the Master Control Stations. Three types of Sat-Terminals have been identified:

- Sat-Terminal A (SaT-A), for low-medium data rate users (U/L: up to 160 kbit/s; D/L: 2.048 Mbit/s);

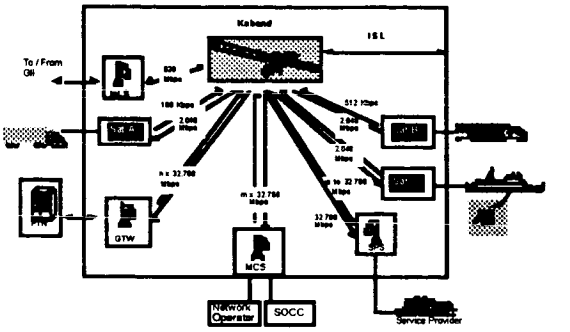


Figure 2: Ka-band SECOMS Network Architecture

- Sat-Terminal B (SaT-B), for medium-high data rate users (U/L: up to 512 kbit/s; D/L: 2.048 Mbit/s);
- Sat-Terminal C (SaT-C), for high data rate users, (up to 2.048 Mbit/s both in uplink and downlink).

Full connectivity is provided among different SaT types for many services. Uplink transmissions from fixed earth stations are performed at n x 32.768 Mbit/s. Ka-band payload downlink data rate is 32.768 Mbit/s (4 Mbit/s for SaT-A). Minimum switching capability provided on board the payload is 16 kbit/s.

Figure 3 describes the EHF-band SECOMS network architecture which is very similar to Ka band one.

One type of Sat-Terminal is foreseen, the Sat-Terminal Type D (SaT-D), providing connections among low data rate users, requiring up to 64 kbit/s both in uplink and in downlink. The uplink and downlink transmissions from fixed earth stations are performed at n x 1.024 Mbit/s, where n depends on the expected uplink signalling and data traffic. The minimum switching capability provided on board the payload is 4 kbit/s.

The coverage for Ka band and for EHF subsystems is shown in figure 4 and in figure 5 respectively.

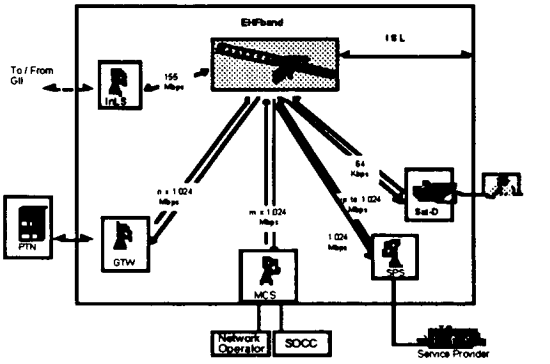


Figure 3: EHF-band SECOMS Network Architecture

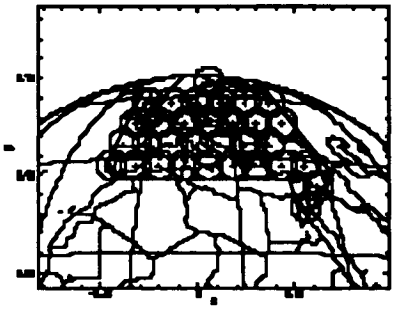


Figure 4: Ka band Coverage

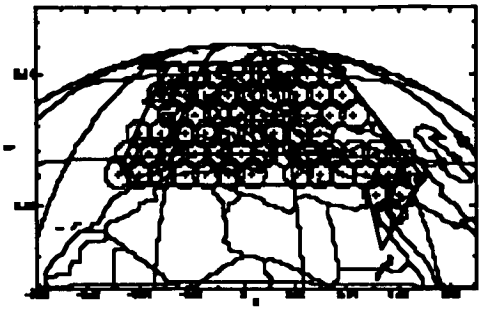


Figure 5: EHF band Coverage

SECOMS SUB-PROJECT: LINK DESIGN

The activities carried on in this frame concern MF-TDMA/TDM access technique[2].
Studies have been conducted for coding, modulation and interference analysis. The link budgets have been calculated considering the 32 beams coverage pattern for the Ka band subsystem and 64 beams coverage pattern for the EHF subsystem. Un-homogeneous traffic distribution between different spotbeams has been assumed for both Ka and EHF components. Due to the different traffic capacity and atmospheric attenuations among spot beams a non-uniform power distribution has been required to the payload to provide, within the coverage area, a "quasi" uniform power flux density on ground. Figures 6 and 7 show power distribution for Ka band and for EHF band subsystems respectively.

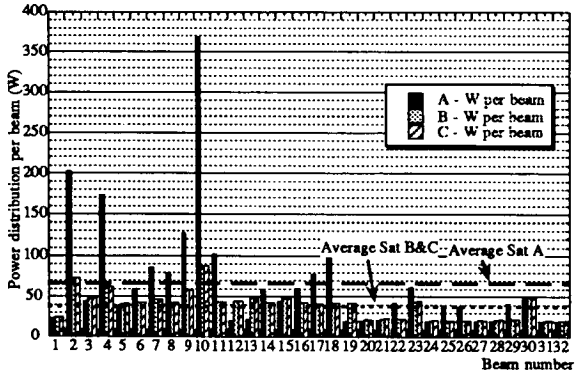


Figure 6: Power distribution from Ka band payload

The modulation is QPSK. The adopted coding scheme is concatenated RS (204,188) + convolutional (1/2, 7) for both links in the EHF component while in the Ka band subsystem, in order to alleviate payload size, it is only RS in the uplink and concatenated in the down link.
The BER is equal to 10^{-10} while the link availabilities are: 99% for mobile terminals (lower in beams n. 2 and 3 in Ka band and in beams 2, 3 and 4 in EHF band); 99.5% for SPS and InLS; 99.9% for GTW and MCS.
The available strategies for carrier and bit timing synchronisation have been identified, and relevant documentation has been collected. Since symbol synchronous MF-TDMA has been adopted, Viterbi&Viterbi algorithm [5] seems to be the most appropriate to perform carrier recovery in up link.

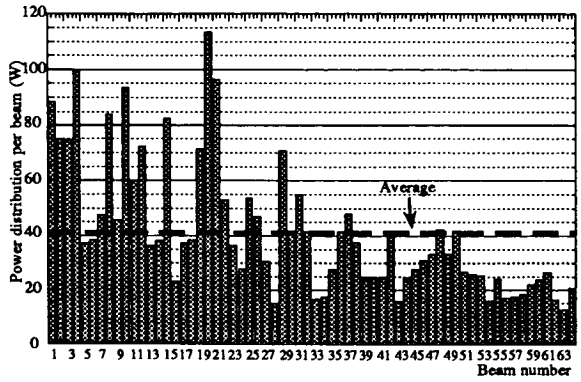


Figure 7: Power distribution from EHF band payload

Analyses are in progress, with particular attention devoted to problems induced by the mobile environment. The system parameters up to now defined are summarised in table 1 (referred to MF-TDMA case). Main results concerning the dimensioning of the user terminals are summarised in table 2.

Parameter	Ka-band	EHF-band
Satellite Position	12° E	12° E
Max. Satellite Traffic Cap.	3.6 Gbit/s	300 Mbit/s
Nominal Sat. Traffic Cap.	3 Gbit/s	250 Mbit/s
Numb. of Spot Beams	32	64
Type of Sat. Terminals	SaT-A,SaT-B,SaT-C	SaT-D
Type of FES	GTW, SPS, InLS	GTW, SPS, InLS
SaT-A uplink inform. rate	16 to 160 kbit/s	--
SaT-B uplink inform. rate	16 to 512 kbit/s	--
SaT-C uplink inform. rate	16 to 2048 kbit/s	--
SaT-A,B,C D/L infor. rate	2.048 Mbit/s	--
SaT-D uplink inf. rate	--	4 to 64 kbit/s
SaT-D downlink inf. rate	--	64 kbit/s
Fixed Earth Stat inf. rate	32.768 Mbit/s	1.024 Mbit/s
Frequency: Uplink Downlink	28 - 30 GHz 18.2 - 20.2 GHz	43.5 - 44.5 GHz 39.5 - 40.5 GHz
Coverage	extended Europe	extended Europe
Spot Beamwidth	0.7° (at -3 dB)	0.49° (at -3 dB)
Polarisation	Circular	Circular
Sat. G/T at tr cross p (tcp)	16.3 dB/K	18.5 dB/K
Satellite EIRP at tcp	55.8 dBW (Mob) 42.0 dBW (Fixed)	54.0 dBW (Mob) 41.7 dBW (Fixed)
Fixed Earth Stations G/T	23 dB/K (SPS) 27.3 dB/K (GTW) 31.9 dB/K (InLS)	26.8 dB/K (SPS) 30.5 dB/K (GTW) 37.3 dB/K (InLS)
Fixed Earth Stations EIRP	61.5 dBW (SPS) 70.5 dBW (GTW) 83.5 dBW (InLS)	53.7 dBW (SPS) 57 dBW (GTW) 75.5 dBW (InLS)
Link Quality	BER < 10^{-10}	BER < 10^{-10}

Table 1: SECOMS Main System Parameters

Mobile Terminals	SaT-A	SaT-B	SaT-C	SaT-D
Uplink Data Rate (kbit/s)	160	512	2048	64
EIRP (dBW)	35.7	41.7	47.7	32.5
G/T (dB/K)	4.3	11.7	11.7	5.5
RF Output pow (ave) (W)	3.2	1.4	6	1.1
RF Output pow (w.c.) (W)	3.7	1.7	7	1.9
Receiving Ant. Gain (dB)	34.04	36.23	36.23	31.27
Antenna Dim. (cm x cm)	27.5	35.4	35.4	10.0

Table 2: MF-TDMA Link Budgets Results

[2] Another subteam is devoted in analysing CDMA/CDM access.

ABATE SUB-PROJECT: ARCHITECTURE

Aiming at the support of the broadband UMTS passenger communication services, the ABATE system network architecture directly derives from the one defined in the frame of the SECOMS sub-project (see figure 8).

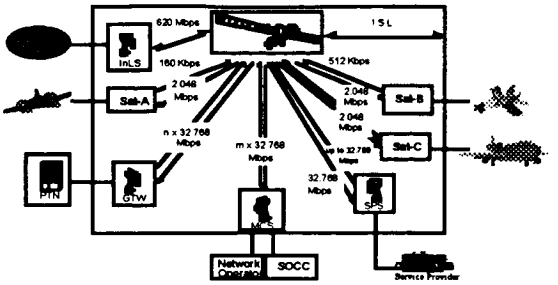


Figure 8: ABATE Network Architecture

It consists of only one segment (the Ka-band one) and foresees the use of a constellation of co-located satellites interconnected through dedicated high rate inter-satellite links (ISLs). Three types of mobile terminals (SaT-A, SaT-B and SaT-C) are baselined, dedicated to aircraft having respectively few, medium and large number of passengers. A preliminary link dimensioning for these three type of terminals has been performed considering two different beam aperture angles (0.7° and 1.4°) corresponding respectively to the conditions of aircraft on ground and in flight (different atmospheric attenuations) and characterised by different values of link availability (99% and 99.9%). GTW, SPS and InLS are also present in the ABATE network to provide system interconnectivity with public and private terrestrial/space networks. The ABATE system service area is shown in figure 9 including also the trans-Atlantic routes.

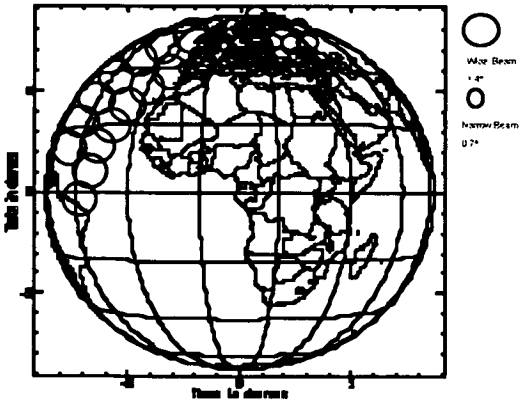


Figure 9: ABATE Coverage Area

ABATE SUB-PROJECT: LINK DESIGN

Work is currently in progress to in deep analyse for the MF-TDMA/TDM satellite access scheme, the aeronautical link performances and BER degradations due to the imperfection of the channel. Analyses on the synchronisation issue and power control procedures have been started. Particular attention will be devoted to problems induced by the aeronautical environment. Table 3 summarizes the presently obtained results.

Parameter	Value	
Satellite Position	12° E	
Max Sat. Traffic Capacity	4 Gbit/s	
Coverage	Extended European + Atlantic Ocean	
Numb. of Spot Beams	32	14 (TBC)
Spot Beamwidth	0.7° (at -3 dB)	1.4° (at -3 dB)
Polarisation	Circular	Circular
Satellite G/T	16.4 dB/K	10.4 dB/K
Satellite EIRP	55.8 dBW 47.0 dBW	60.5 dBW Mob Us 47.0 dBW Fix Stat
Link Availability	99% (on ground)	99.9 % (in flight)
Operational Frequency Band	Uplink:30-31GHz Downlink:20-21GHz	
Link Quality	BER < 10^{-10}	
Type of Satellite Terminals	SaT-A, SaT-B, SaT-C	
Type of Fixed Earth Stations	GTW, SPS, InLS	
Satellite Terminals Tx Information Rate	SaT-A: from 16 to 160 kbit/s SaT-B: from 16 to 512 kbit/s SaT-C: from 16 to 2048 kbit/s	
Satellite Terminals Rx Information Rate	SaT-A: 4096 kbit/s SaT-B and C: 32.768 Mbit/s	
Satellite Terminals G/T	SaT-A: 4.4 dB/K (1 dB of margin) SaT-B: 13.1 dB/K SaT-C: 13.1 dB/K	
Satellite Terminals EIRP	SaT-A: 38.7 dBW (1 dB of margin) SaT-B: 44.5 dBW SaT-C: 50.5 dBW	
Fixed Earth Stations Tx and Rx Information Rate	GTW & SPS: m x 32.768 Mbit/s InLS: 620 Mbit/s	
Fixed Station G/T	see table 1 (SECOMS)	
Fixed Station EIRP	see table 1 (SECOMS)	

Table 3: ABATE Main System Parameters

The analyses have been implemented considering two different type of spot beamwidths: 0.7°and 1.4°.
Concerning the distortion budgets both analytical calculations and dedicated software simulations using TOPSIM are underway. The degrading effect due to power amplifier non linearity (OBO and modulation losses), in band filter distortions (amplitude and group delay) and modulation distortion (rise time, I v.s. Q mis-alignment, amplitude and phase errors) are under investigation.

Aeronautical Channel Characterisation

The aeronautical channel for level altitude and time scale smaller than 10 min., has been identified as a Rice channel with a large Ricean parameter (> 25 dB). For aircraft ascent/descent further research is needed, because JPL experienced a slowly varying amplitude variation (< ±2.5 dB) in their experiments with a C-141 Starlifter. This could be related to shadowing/scattering from the tail structure, but other effects could also be the cause of this variation (e.g. temperature induced changes). No comparable variability has been observed by JPL's experiments with their Rockwell Saberliner, which is much smaller and has a different tail structure.
Technical data for different types of aeroplanes has been gathered and scenarios for typical flight phases will be developed. Geometric analyses, especially for the tail structure, are in progress. The Doppler effect has been analysed and a formula is being developed to predict the Doppler shift in dependency of the aircraft known parameters (latitude, longitude, over ground velocity, heading and pitch). Also the differential Doppler shift has been analysed and the U-turn of a plane has been identified as worst case for differential Doppler.

USER TERMINALS

The user terminals have been grouped within the main categories: individual, portable and mobile terminals. Tables 4 and 5 summarise the main characteristics of identified terminals.
Uplink power control technique for both the fixed stations and the mobile users terminals have been investigated to reduce the co-channel interference among spot beams reusing the same transmitted frequency band. The effectiveness of open loop and closed-loop power control for various link factors (rain attenuation, link distance,...) were analysed for both MF-TDMA and CDMA access schemes.

Terminal	SaT-D	SaT-A	SaT-B, C
Case	Palm-top	Lap-top	Briefcase
Use	Individual		
Mobility in operat.	Personal	No	
User pos. w.r.t. ant.	Near	Medium	
Pointing Requirem.	Manual (coarse); Auto (fine)		
Beam Steering	Electronical	Mechanical	
Atmospheric Losses	Yes		
Shadowing Losses	Yes	No	
Doppler effect	Low	No	
D/L reception rate	2.2222 Mbit/s	4.4444 Mbit/s	35.55 Mbit/s
U/L max tx data rate	147.6 kbit/s	184.5 kbit/s	590.3 kbit/s (B), 2361.1 kbit/s (C)
Antenna type	Flat	Flat, Remoteable	
Ant. Electr. Design	Act. array ant.	Passive printed antenna	

Table 4: Portable Terminals Characteristics

Terminal	SaT-A	SaT-B	SaT-C
Mobile Type	Car	Aeroplane,Ship,Train Cruise,Bus,Truck	
Use	Individual	Ind/Group	Group
Mobility during operation	Yes	Yes	
User position w.r.t. ant.	Far/Shielding	Far/Shielding	
Pointing Requirements	Auto	Auto	
Beam Steering	Electronical	Mechanical	
Atmospheric Losses	Yes	Yes	
Shadowing Losses	Yes	Rare/Environment	
Doppler effect	Yes	Yes	
Downlink max rate	4.444 Mbit/s	35.55 Mbit/s	
Uplink max data rate	184.5 kbit/s	590.3 kbit/s	2361.1 kbit/s
Antenna type	Not protuberant	Protuber. acceptable	
Antenna Electrical Design	Act. array antenna	Pass. printed antenna	

Table 5: Mobile Terminals Characteristics

EXPERIMENTS

Two different network architectures will be used for the SECOMS/ABATE experiments. The first one, used in the frame of SECOMS sub-project for the satellite trials of the land mobile broadband communication system, will be based on:

- a fixed Gateway Station (GS), made available for the selected geostationary system and appropriately modified in the baseband section for its utilisation in the particular application;
- a transparent transponder (Italian Global Coverage) used for both the forward (GS-to-mobile) and the return link (mobile-to-GS);
- a land-mobile terminal.

The second one, used in the frame of ABATE sub-project for the satellite trials of the aeronautical mobile broadband communication system, will be based on:

- a fixed Gateway Station (GS), made available for the selected geostationary system and appropriately modified in the baseband section for its utilisation in this particular experiment, taking into account the need to measure in the demodulation section the transmitted data clock error (w.r.t. the local GES master clock) and to insert this information in the forward data stream, to the AT to allow the correct TX data clock adjustment
- a transparent transponder (Italian Global Coverage) used for both the forward (GS-to-mobile) and the return link (mobile-to-GS);
- an Aeronautical Terminal based on the already partially developed subsystems and on sections developed in the frame of the ABATE project.

CONCLUSIONS

The demand of multimedia services together with personal mobility is expected to grow in the next years. In this scenario satellite systems have to be considered an essential element to meet these requirements.
A system study on the feasibility of a geostationary satellite system for mobile (including aeronautical) multimedia services is ongoing. The most important parameters have been presented being almost finalised. In order to validate the feasibility, several test-beds and experiments are planned both for mobile and for aeronautical terminals.

REFERENCES

[1] G. Losquadro, F. Vatalaro, *A Satellite System for Multimedia Personal Communications at Ka-band and beyond*, Ka band Utilization Conference, October 10-12, 1995, Rome, Italy, pp. 273-281.

[2] G. Losquadro, *Multimedia Services Using Portable and Mobile Satellite Terminals*, ACTS Mobile Communications Summit, November 27-29, 1996, Granada, Spain, pp.411-418.

[3] A. Jahn and E. Lutz, *LMS channel measurements at EHF-band*, IMSC '97, Pasadena, CA, June 17-19, 1997.

[4] W. F. Cashman, *ACTS multibeam communication package: Description and performance characterization*, 14th International Communication Satellite Systems Conference and Exhibit, AIAA, Washington DC, USA, March 22-26, 1992, pp. 1151-1161.

[5] A. J. Viterbi, A. M. Viterbi, *Nonlinear Estimation of PSK-Modulated Carrier Phase with Application to Burst Digital Trasmission*, IEEE Trans. on Information Theory, vol. IT-29, No. 4, July 1983, pp.543-551.

Space Segment Integration in Future Mobile Systems

Jonathan P. Castro, Amre. El-Hoiydi

Centre Suisse d'Electronique et de Microtechnique SA.
Jacquet-droz 1 CH-2007 Neuchatel, Switzerland
Tel. +41 327 205 588, Fax +41 327 205 720
Email: jonathan.castro@csemne.ch, amre.el-hoiydi@csemne.ch

INTRODUCTION

The integration between the terrestrial and the space segment will depend on the satellite operating scenarios. In a stand-alone satellite network the spacecraft system will be a self-contained entity, where the type of connection with the Terrestrial Universal Mobile Telecommunication System (T-UMTS) for example, will be contingent upon the *degree of integration* allowed by a Core Network (CN). This paper, focuses on defining a high level integration using a Broadband Integrated Services Digital Network (B-ISDN) as the core system.

INTEGRATION LEVELS

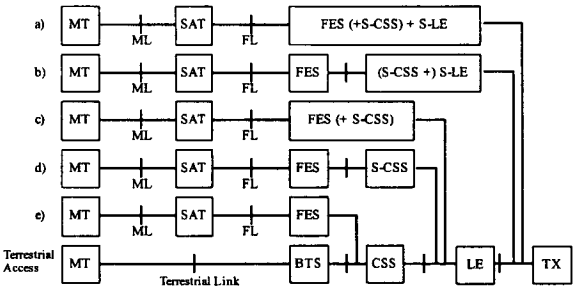
When considering the five levels of integration identified, i.e., *Geographical*, *Services*, *Network*, *Equipment*, and *System* integration, the latter one portraits an advanced connection between the satellite and the terrestrial networks. In this last case, UMTS as a 3rd generation mobile system, regards the space segment as one more cell, i.e. the Satellite Cell in its cellular hierarchy.

Since *Geographical* integration implies coexistence of the satellite with the terrestrial system, and *Service* integration the provision of similar services without any common elements in both systems; there is little interest in these two types of integrations for 3rd generation mobile systems that foresee a transparent transition between the segments, as well as conjoint service offering.

On the other hand, *Network* and *Equipment* integration which imply that the two segments are interconnected through an Internetworking Unit (IWU) may be of some interest. The difference between *Network* and *Equipment* integration is the number of common components and parameters used at the IWU. However, these integrations still keep the segments totally independent. Thereby, seamless intersegement handover is not practical. Therefore, *System* integration offers a more structural advantage, because it would allow common elements throughout the two segments whenever possible. The *System* integration option will be detailed in the following. However, as a first step on the network integration design, the CN access point options will be introduced.

Core Network Access Point: For the Satellite-UMTS (S-UMTS), the Network Entities (NE) comprised in the *user plane* are: Mobile Terminal (MT), Satellite (SAT), Fixed Earth Station (FES), Cell Site Switch (CSS), Local (LE) and Transit Exchange (TX). The *control plane* encompasses, in addition to the previous NEs, also the Mobile Service Control Points (MSCPs) and the Mobile Service Data Points (MSDPs). Management entities are not covered here. Since in many cases the *user plane* transfer also related *control plane* information, we use a '*transport plane*' that includes typical *user plane* and all *control plane* functions directly bounded to the *user plane*.

Figure 1 illustrates the CN access points assuming the *transport plane*. The highest access point occurs at the TX, options (a) and (b); the next one is at the LE, options (c) and (d); and finally, the last point is at the CSS, option (e). Thus, Figure 1 illustrates the optimised transport access point alternatives.



MT	Mobile Terminal	CSS	Cell Site Switch
SAT	Satellite	LE	Local Exchange
FES	Fixed Earth Station	TX	Transit Exchange
ML	Mobile Link	+	Reference Point
FL	Feeder Link	S-CSS	Satellite Specific CSS
S-LE	Satellite Specific LE		

Figure 1: S-UMTS Core Network Access Point alternatives in the transport plane [3]

System Integration: The System Integration for practical purposes sub-divides itself further in: *Signalling*, *Data*, *Control* and *Core* integration. The last one represents the maximum integration at the highest access point of the CN.

In the *Signalling* integration exchanges between the satellite and the terrestrial segments occur through a

signalling network. This integration considered as the lowest level of the System integration does not afford higher improvements than the Network and Equipment integration levels. Since a UMTS user may need space and terrestrial services, the data support part of both Terrestrial and Satellite segments would not be fully integrated; thereby it will introduce a major limitations to UMTS service providers.

Data integration implies having a combined database between the terrestrial and the satellite networks. A common MSDP would ease service consistency, flexibility of access and quality of service. However, to afford Global roaming while keeping signalling load to a minimum, a common MSCP will also be necessary.

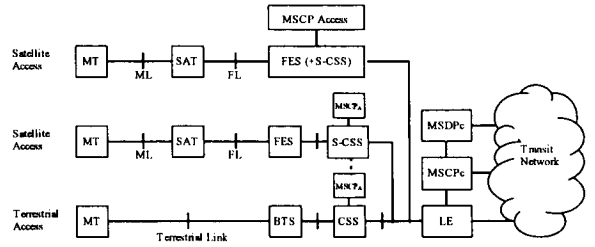
Control integration thus improve further the *Data* integration by facilitating inter-segment handover. In this case, the two segments can share certain core network functions located in the MSCP of the local exchanges. The overall resource allocation and control, as well as routing strategies would then be optimised taking advantage of the *unified* view of both segments. This will increase access flexibility, and permit uniform service provision.

Core integration means full integration of the Terrestrial and Satellite networks at the TX point. This design will required a satellite specific LE. The integration with the terrestrial segment may be limited to Intelligent Network (IN) capabilities available through the core network (inter-LE handover, for example). The Satellite-LE (S-LE) would contain then all the mobility functions including the call control. However, such integration would impose large demands on the CN and require the service providers to implement the service logic in the satellites. Therefore, this option is not recommended.

Control integration having the access point at the LE level seems to be the best option. This approach will maximise the commonalties between the MSCPs and the MSDPs without demanding big changes in the fixed networks. For example, services can be controlled by the service providers at the level of the LE. In addition, by connecting the satellite segment to a standard terrestrial LE, the terrestrial service providers can also deploy services over the satellite segment.

There are two options for connecting the satellite segment at the LE level. The first option is to directly connect the FES to the LE (alternative (c) of Figure 1 and first case of Figure 2), in which case the FES includes the CSS functions. The 2nd option is to connect the FES to a satellite specific CSS which is in turn connected to the LE (case (d) of Figure 1 and 2nd case Figure 2). The FES comprises then primarily the satellite transmission system.

An advantage of designing a satellite specific CSS equipment, with a standard FES-CSS interface, is that it can be reused for many different satellite constellations wanting to be UMTS compatible. The CSS, in the terrestrial segment concentrates many BTSs and performs intra-CSS handovers. In the satellite segment, a S-CSS will be connected to one FES only, because of the geographical distribution of the FESs. Its role is to handle intra-FES handovers and interface the FES with the B-ISDN LE. The name Cell Site Switch has been kept for the satellite segment to clarify the architecture commonalties between the segments, although it could have been called the "FES Site Switch".



MT	Mobile Terminal	CSS	Cell Site Switch
SAT	Satellite	LE	Local Exchange
FES	Fixed Earth Station	S-CSS	Satellite Specific CSS
ML	Mobile Link	+	Reference Point
FL	Feeder Link	MSDP	Mobile Service Data Point
S-LE	Satellite Specific LE	MSCP	Mobile Service Control Point

Figure 2: Integration at the LE point[3]

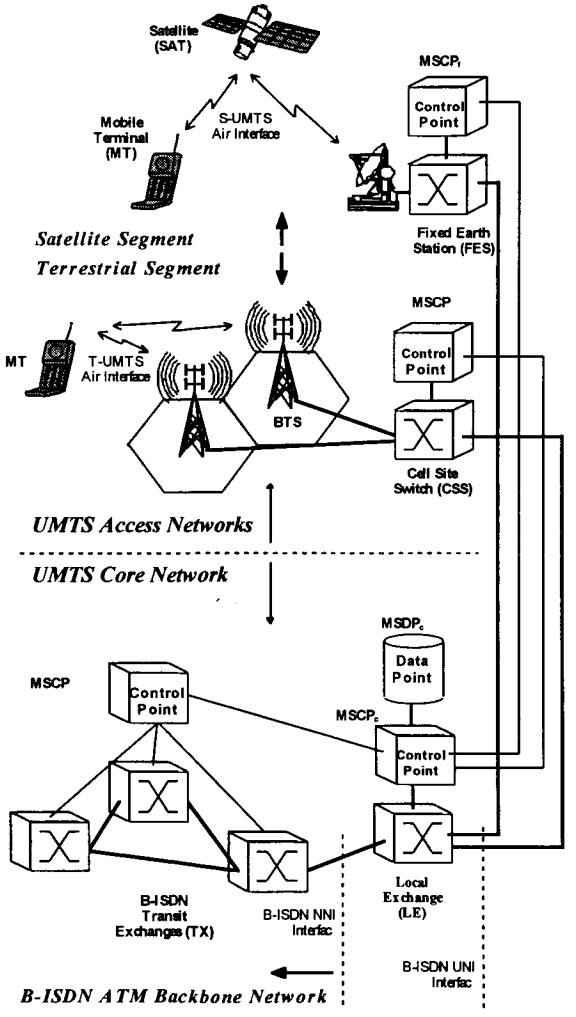
Based in the preceding integration concepts, Figure 3 illustrates an integrated UMTS reference configuration. It assumes that the UMTS elements (i.e., FES, CSS, LE, TX) will be based on B-ISDN using ATM for the transport of user signalling information.

For the terrestrial case, in the *user plane* the Mobile Terminal (MT) communicates with the Base Transceiver Station (BTS) in the Radio Access System (RAS) via the terrestrial UMTS radio interface. The Cell Site Switch (CSS), which routes information between the fixed network and a concerned BTS, is located in the RAS. In the other hand, a BTS itself may be directly connected to the B-ISDN fixed network. In addition, other B-ISDN terminals could also be connected to the same local exchange.

Thus, the integration of fixed and mobile systems using B-ISDN as the CN, maximises the commonality of the required functionalities. Once paging and handover are accounted for in mobile networks, functions such as call handling, mobility management for personal and service mobility, transport of user data, registration, location update, and switching need to be available in both fixed and mobile networks. Since the space segment in UMTS

assumes global coverage access to fixed of mobile networks will be transparent. Thus, by integrating mobile and fixed communications into one network, duplication of common functionalities will be prevented.

In the following, before looking at the options in the Satellite Network Architectures, we will first briefly define the main mobility functions. This will provide the background to see the impacts due to the different architecture configurations. The type of Satellite architecture will influence the procedures or execution flow of the Mobility Functions (MFs).



BTS	Base Transceiver Station
B-ISDN	Broadband Integrated Services Digital Network
MSCP	Mobility and Service Control Point
MSDP	Mobility and Services Data Point
NNI	Network Network Interface
UNI	User Network Interface

Figure 3 UMTS Satellite and Terrestrial Reference Configuration Example, (adapted from [3])

THE S-UMTS MOBILITY FUNCTIONS

Session Set-up/ Session Release: The collection of call preparations events is defined as the *Session Set-up* procedure (e.g. getting access to the network, exchanging keys, user authentication, terminal authentication and collecting part of the user profile). In UMTS the two types of Sessions are User Session and Terminal Session.

User Session: User session implies the user choosing a particular service and the service provider accepting or refusing the request. After a trusted association between the user and the service, encryption keys are produced. The user can then initiate or receive a number of successive or simultaneous calls within the session.

Terminal Session: Terminal session is possible only when the network knows the terminal. Since, there are no permanent terminal identifiers within a network, a procedure to make the terminal known is initiated by either the terminal or the UMTS network operator/service provider. Then, terminal authentication occurs during the terminal session set-up. Thus, a terminal session is used for the delivery of services, when the personal delivery is not requested. Only the identity of the terminal is verified by the network. The smart card authentifying the user must not be in the terminal.

The difference between the user and terminal session is the security level. The latter has less strict security requirements.

Call Set-up/ Call Release: The main satellite specific issues to consider during the call set up are:

1. reduced capacity,
2. larger delays,
3. frequent handovers (e.g. in LEO systems),
4. low number of FESs (i.e. far apart from each other),
5. Class of satellite (e.g. Bent Pipe, Cross Connect, and Switching), and
6. type of integration with the core network (particularly B-ISDN units, e.g. CSS, LE, etc.).

In the case of the last item (6), e.g. when the integration between S-UMTS and B-ISDN happens at transit exchange level (while for the terrestrial segment this integration occurs below the local exchange level), then a completely independent satellite call handling model becomes an option to be considered. However, the goal is to maximise the functional commonalties between the two segments by using as much as possible common building blocks.

User Registration/User Deregistration: The registration procedure a user tells his service provider where he wants to receive calls. However, before a user can register on a terminal, he should first have an active user session running. Once he has a trusted association with his service

provider, he can request the service provider to register him for a particular service on his terminal. Then, the terminal capability to support the requested service and the authorisation. If these are satisfactory, the service provider will store information in his database for the registration, i.e. he will store the location and the terminal where the user wants to receive the indicated service.

Attach/Detach: Since in the space segment radio resources are generally more scarce and more expensive than in the terrestrial segment, to limit the signalling on the radio link through attach/detach procedures is more critical. In addition, due to the large coverage of the satellite spot beams, a different trade-off can be expected in the S-UMTS. The main benefits of using the attach/detach procedures in the space segment are:

1. Reduced signalling on part of the fixed network path,
2. and less paging resources utilisation on the radio link.

A drawbacks can be the extra procedures in the UMTS network architecture, which in some case may increase system complexity, e.g. due to the extra check necessary when a mobile terminates a call.

Handover: In general, handover procedures are used to maintain calls when the mobile terminal is moving with respect to the coverage entities, and also for traffic management reasons or special user demands. On the S-UMTS side two main handover types are Inter-segment HO and Intra-segment HO.

Inter-Segment Handover: Assuming a satellite component is integrated at system level with the terrestrial component, the handovers from a satellite footprint to the cellular cells or vice versa can happen:

1. when the terminal leaves or enters into the cellular coverage,
2. when traffic congestion arises, and
3. when the preference of the subscriber is in question.

Intra-Segment Handover: Intra-segment handover occurs, e.g. when non-GEO satellite footprints move at high speeds with respect to the Earth's surface, and the number handovers due to the movement of the terminals vs. the cell areas are negligible with respect to the number of satellite handovers. The two types of HO are:

1. Intra-satellite between spot beams of same satellite, and
2. inter-satellite between spot beams of adjacent satellites.

Message Service: The difference between the T-UMTS and the S-UMTS (transmit/receive) message service lies in the geographic scale of the delivery of a broadcast message. Since a spot beam is very large, the geographic area where the message is broadcast covers at least the diameter (e.g. 1000 km) of one spot beam.. This is one of the main advantages of the S-UMTS segment, particularly when large areas are wanted to broadcast effectively. On the other hand, this can also be a disadvantage when smaller areas than a spot beam coverage should be broadcast precisely.

Location Update/ Domain Update: Location management includes the non-call related location procedures, i.e., location updates and domain updates. These two procedures are invoked automatically by the terminal, on behalf of the users registered on the terminal [2]. The location update function in a cellular mobile system allows the network to page a mobile terminal for an incoming call [4]. In satellite environment, where an antenna's radio coverage is much wider than in terrestrial networks, the optimum trade-off between precision location update and wide-area paging is different from that in the terrestrial cellular environment due to the scarce and costly available bandwidth.

Security and Data base issues: In contrast to the terrestrial segment, the satellite air interface has a double hop (MT-Satellite and Satellite-FES). Thus, in terms of security there are two possibilities, 1st, when the satellite segment is used to provide services within a session, the security level can be as in the terrestrial UMTS; and 2nd when the satellite link can is used out of a session, satellite feeder links encryption could be applied. In regards to the database aspects, the main concern is the usage of common data for both the terrestrial and satellite segments of UMTS. This implies that the UMTS service provider should be capable of providing services using both satellite and terrestrial network operators.

S-UMTS NETWORK ARCHITECTURES

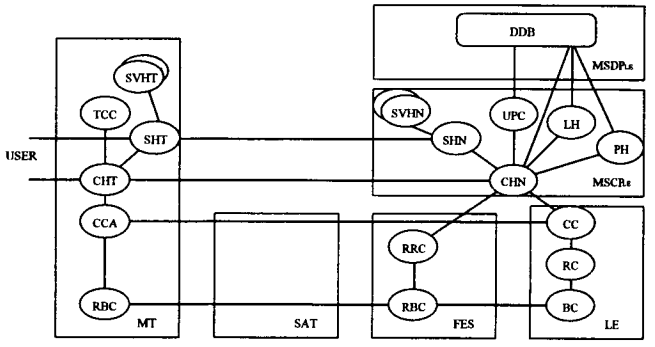
Following the integrated model we can envisage two types of S-UMTS architectures, i.e. the *Evolutionary* and the *Revolutionary* ones. The impact of these architecture configurations will be illustrated with the description of the 'calling handling function. This function offers the best representation of all the satellite architecture cases.

The Evolutionary Architectures: The T-UMTS/S-UMTS integration based in the evolutionary approach would assume that S-PCN's such as Globalstar, ICO and Iridium presently under implementation, may at a given time be capable to meet UMTS requirements by taking into account the highest level of integration possible in the context of mobility procedures, and satellite network feasibility in terms of the B-ISDN Core Network interface. The evolutionary architectures, which incorporate the *Bent Pipe* and *Cross Connect* satellites, would also imply that their air-interface have evolved to handle the transmission rates expected for the satellite networks. This includes rates up to 64 or 128 kbps.

Bent-Pipe (BP) Satellite Networks: Transparent (or bent pipe) satellites would correspond to the Globalstar and ICO systems. The following reasons justifies the first scenario for integration of S-UMTS and T-UMTS:

1. Using transparent constellations would keep the satellites as simple as possible within the state-of-the-art technology. Thus allowing rapid UMTS integration with a low cost space segment.

2. In addition the *Control Integration* facilitates a maximum commonality between the terrestrial and satellite access networks.



BC: Bearer Control
CC: Call Control
CCA: Call Control Agent
CHN: Call Handler Network
CHT: Call Handler Terminal
LH: Location Handler
SVHN: Service Handler Network
SVHT: Service Handler Terminal
PH: Paging Handler
RBC: Radio Bearer Control
RC: Resource Control
RRC: Radio Resource Control
SHN: Session Handler Network
SHT: Session Handler Terminal
TCC: Terminal Capability Check
UPC: User Profile Check

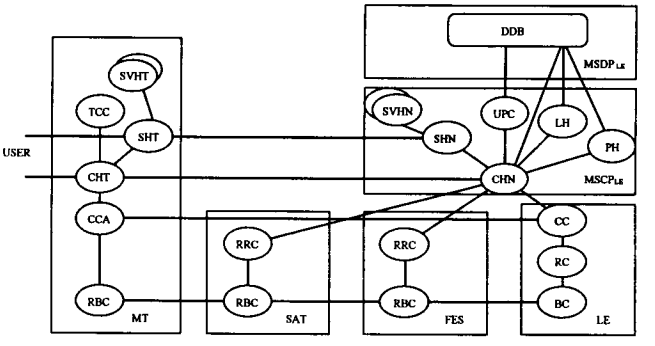
Figure 4 Call Handling in BP Satellite Architecture [3]

In Figure 4, with a transparent satellite architecture, the FES-SAT-MT double hop performs the function of the BTS-MT single hop. The FES is equivalent to a BTS, in what concerns the call handling [3]. The mapping between feeder and mobile links do not afford dynamic configuration (on a per call basis). Because the call handling model in the bent pipe satellite is not seen.

All Functional Entities (FEs) on the user side are of course in the mobile terminal. The question is how to allocate the functions at the network side. The Call Control (CC) is in the LE in order to reuse the B-ISDN protocol and ease the service integration between both segments. The Resource Control (RC) and Bearer Control (BC) are concerned with fixed resources and bearers and must hence be in the LE as well.

The Radio Bearer Control (RBC) function in the FES talks with the BC to open a B-ISDN link between the FES and the LE. It also talks with the RBC in the MT to open the radio link. It controls the mapping between the two links. The Radio Resource Control (RRC) is allocated to the FES because it must closely collaborate with the RBC. All other control entities are allocated to the MSCP of the LE. The allocation of these entities to a MSCP at the level of the FES would not bring a big reduction in the signalling delay, since the FES and the LE are likely to be close to one another. The disadvantage would be not to share the same functional entities with the terrestrial segment, i.e. to duplicated the control software and probably equipment.

Cross-Connect Satellite Networks: Cross-connect (XC) satellites as the ones used in the Iridium system, are a natural technical step forward when trying to add some complexity on board with respect to a fully transparent satellite [3]. Thus, the ability of the cross-connect satellites to support ISL efficiently when compared to transparent satellites would allow more efficient Radio Resource management and inter-satellite handovers. The FES in XC satellite architecture would gather a set of transceivers to handle feeder link communications with the satellites in view at a given time. Functionally, the FES should be considered to be the equivalent of the terrestrial BTS and CSS grouped together, except that part of the Radio Bearer and Resource Control functions would be handled by the cross-connect satellite.



BC: Bearer Control
CC: Call Control
CCA: Call Control Agent
CHN: Call Handler Network
CHT: Call Handler Terminal
LH: Location Handler
SVHN: Service Handler Network
SVHT: Service Handler Terminal
PH: Paging Handler
RBC: Radio Bearer Control
RC: Resource Control
RRC: Radio Resource Control
SHN: Session Handler Network
SHT: Session Handler Terminal
TCC: Terminal Capability Check
UPC: User Profile Check

Figure 5 Call Handling in XC Satellite Architecture [3]

For example in Figure 5, the cross-connect satellite has some Radio Bearer Control capabilities (RBC). This is made possible because the broadband control protocols are adopted only up to the CSS, so that radio bearer control is separated from call control. In this way, a call request will arrive to the appropriate FES and the Radio Call Control will initiate an end-to-end Radio specific call (i.e., a call from the FES to the MT). The FES in XC satellite architecture would gather a set of transceivers to handle feeder link communications with the satellites in view at a

given time. Functionally, the FES should be considered to be the equivalent of the terrestrial BTS and CSS grouped together, except that part of the Radio Bearer Control functions would be handled by the cross-connect satellite.

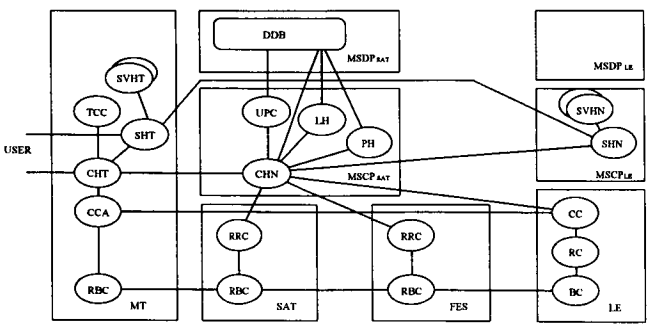
The difference with the architecture for transparent satellite is that the Bearer Control (BC) functions and the intra-satellite resource control are on board. This resource control can execute intra-satellite handovers independently of the Earth control. The link of the RRC with CHN is used for other Resource Control tasks, in which the overview of the ground control is necessary.

The Revolutionary Architecture: The idea behind the "revolutionary" approach is to have full access to network functionalities on board the space segment, so that the UMTS space segment support of the ground segment is reduced to its minimum. Of course, interface between the Core Network and the Satellite Access network would be realised at the LE level through the FES. Thus, an associated ground segment would still be necessary for satellite control.

The Switching (SW) Satellite Network: The next technical and natural step would be to increase the functionality of the payloads by adding switching capabilities.

From the proposed S-PCNs only the Teledesic system may fall in this category. However, because it is meant to operate in the Ka-band frequency, it does not correspond to the UMTS integrated network.

In the case of switching satellite we will add on board the satellite the functions of Connection Management and Mobility Management which were handled in the terrestrial access network for transparent and cross-connect satellites. One can think at the satellite as supporting the full set of functions of the MSCP and CSS. The FES should then be limited in its functions to a set of transceivers supporting feeder link traffic with the satellites, and to a minimum set of low level CSS functions to handle connection to the core network through an associated Terrestrial CSS.



BC: Bearer Control
PH: Paging Handler
CC: Call Control
CCA: Call Control Agent
CHN: Call Handler Network
CHT: Call Handler Terminal
LH: Location Handler
SVHN: Service Handler Network
SVHT: Service Handler Terminal
RBC: Radio Bearer Control
RC: Resource Control
RRC: Radio Resource Control
SHN: Session Handler Network
SHT: Session Handler Terminal
TCC: Terminal Capability Check
UPC: User Profile Check

Figure 6 Call Handling in SW Satellite Architecture [3]

As illustrated in Figure 6, the call handling of the system can be enhanced by placing part of the Radio Call Control (RCC) on-board, since a minimum RCC functionality needs to be in the FES. Thus, part of radio specific call controls are performed prior to establishing the Call Handler Mobile (CHM) to Mobile Call Handler Originating side (MCHO) relationship. This option gives full flexibility for the implementation of landing strategies. The satellite is able to select the most appropriate landing point for a call, thereby reducing the call-set-up time from eliminating the FES re-selection.

However, adding full switching and IN capability on board the satellite could be very demanding in term of the overall processing requirement, besides increasing the overall signalling load. While the implied processing and storage requirements seem to be achievable in a very near future, the overall power requirements to support a fully functional B-ISDN switch on-board and ISL's along with a large number of user channels need further investigation.

CONCLUSIONS

This contribution has presented the integration options of the Satellite-UMTS segment to its terrestrial counterpart. It has also described the influence of the different S-UMTS architectures in the call control function for 3rd generation mobile networks. At present the inputs reflect the preliminary specification of the S-UMTS architectures, in later activities the suggested configurations will be optimised after verification of the performance analysis through network simulations.

Acknowledgements: Part of this work has been performed under the sponsorship of the European Space Agency (ESA), through a consortium composed by CSEM, Logica, Matra Marconi Space, and the Swiss Telecom PTT. Therefore, the authors appreciate and acknowledge the collaboration and inputs of all the participants.

REFERENCES

[1] J.P. Castro Ed. "UMTS Conflicts with S-PCN approaches," WP 200 report on the "Network Techniques for Dynamic Satellite Constellation" project. ESA contract 11825/96/NL/US. Final Version Nov. 25th 1996.

[2] RACE 2066 MoNet, UMTS System Structure Document, final document, December 1995.

[3] J.P. Castro Ed., *Configuration of Satellite-UMTS Network Architectures*, WP 300 report, "Network Techniques for Dynamic Satellite Constellation" project. ESA contract 11825/96/NL/US. Draft Version 2.0, Feb. 28th 1997.

[4] A. El-Hoiydi, R.J. Finean, *Location Management for the Satellite - Universal Mobile Telecommunication System*, IEEE - ICUPC, Cambridge, USA, Sept. 29th - Oct. 2nd, 1996, pp. 739-744, vol. 2.

A Mobile Satellite Ground Segment Architecture Interoperable with GSM

Gary A. Johanson
Westinghouse Wireless Solutions Company
930 International Drive
Linthicum, MD 21093 USA
Phone: +1 410-765-9045 FAX: +1 410-765-9745
email: johanson.g.a@wireless.westinghouse.com

Eric Copros
Matra Marconi Space
31, rue des Cosmonautes - Z.I. du Palays
31077 Toulouse Cedex - France
Phone: +33 562 19 73 70 FAX: +33 562 19 76 52
email: eric.copros@tls.mms.fr

ABSTRACT

This paper describes a ground segment system that is designed to support advanced multi-beam Mobile Satellite Systems (MSS) that offer interoperability with ground-based Global System for Mobile communications (GSM) cellular networks and support an advanced satellite design featuring on-board switching capability. The features of the overall system model are described highlighting the advancements that are possible with near term technology including satellite digital processing for flexible signal routing and sufficient power and sensitivity to enable communications to hand-held terminals from Geosynchronous (GEO) based satellites. These technological advances in space and terminal technologies impose new requirements on the ground network infrastructure. Coupled with these new requirements is the desire to provide dual mode handsets that communicate within the current GSM network or the advanced GSM based Personal Communications Networks as well as with the satellite network. Since subscribers may maintain a home registration in either the satellite or the GSM network, interoperability must be maintained between the two networks. This interoperability also imposes requirements and constraints on the ground infrastructure which are described in the paper. System parameters are introduced that define the top-level requirements of the ground segment for these advanced systems given the satellite and terminal technology. These parameters are used to define an approach to the ground network that can meet the challenge of new satellite and terminal technologies. The paper concludes by offering a suggested architectural approach to a GSM interoperable MSS system and provides a summary of the differences between the satellite system and the GSM network that must be considered by infrastructure suppliers.

SYSTEM MODEL

The overall system model is illustrated in Figure 1. The figure shows a GEO based mobile satellite system with major architectural elements being: the Satellite, the Network Control Center (NCC), the Gateways, and the user terminals (UT). This GEO satellite based system provides wireless access services over a large portion of the earth's surface using a Time Division Multiple Access (TDMA) scheme similar in most respects to the terrestrial GSM network except for the physical geometry between the elements. Only recently has satellite technology advanced to a degree which permits access to/from hand held User Terminals (UT) from the GEO orbit.

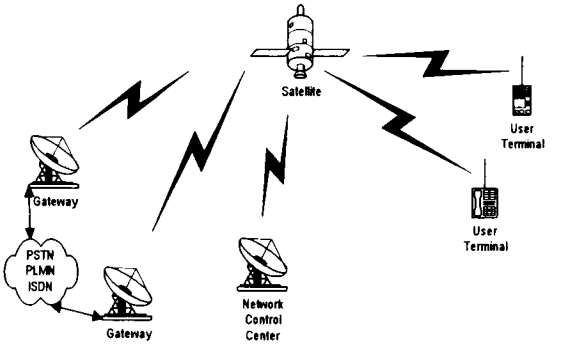


Figure 1 - System Model

Spacecraft Technology

Spacecraft technology has improved in recent years to permit communications with, and more importantly between, hand held UTs. The basic enabling technologies on the spacecraft include: large antenna apertures, multi-beam feed technology, and Digital Signal Processing (DSP). Large antenna apertures (>10 meters) combined with multi-beam (>100 beam) feed technologies provide the gain and frequency reuse necessary to close the link with a disadvantaged UT and to reuse the user link

frequencies as many times as possible. Having a hand held user terminal is considered a market penetration necessity. Minimizing the use of the scarce user link frequency band (assumed to be in the L-band) is an operational necessity in order to provide enough channels to make an economically viable service offering. Digital signal processing has been proposed to enable a "switching" function in the spacecraft. This switch allows direct communications user terminals without the necessity of passing the signals through a gateway which would introduce additional irritating delay in conversations. This processing may be pass-through or regenerative. This paper assumes a pass through technology which provides most of the DSP benefits at a significant reduction in complexity and development risk.

Terminal Technology

Terminal technology has also had advancements similar to satellite technology. To match up with the new satellite payload capabilities, terminal manufacturers are ready to provide advanced handsets capable of direct communications via GEO satellites. The hand held variety of these UTs will most probably be dual mode to provide either terrestrial cellular/personal communications or satellite communications. Users of these terminals will expect the same features as they have now such as:

- One or more days of standby time
- Several hours of talk time
- Easy/convenient operation
- Reduced size with the same feature set
- A high level of security
- Privacy

Due to its rapid world-wide expansion, the leading contender for the cellular mode is GSM or the GSM based Personal Communications Network (PCN) which is being deployed now.

Ground Network Impacts

The combination of advanced satellite communications payloads and the emergence of convenient hand held user terminal technology places new demands on the ground network.

The ground network supporting advanced satellite payloads and advanced terminals has several challenges in key development areas:

- Network capacity and call processing load
- Cellular interoperability
- Call routing and system features

Network Capacity: Advanced satellite payloads are capable of carrying on the order of 10,000 to 30,000 Erlangs of user traffic. This traffic figure assumes a full rate voice coder/decoder (CODEC). This amount of carried traffic is much larger than is anticipated in even the largest switching platforms today. This implies a distributed ground switching network. Even though a single ground entry point could be envisioned for a GEO based satellite system, a distributed network would most probably be desired by the service providers in order to allow independent operations to occur in geographically dispersed regions. If these regions are a good market for mobile satellite technology, then they probably have limited terrestrial infrastructure in place now, making a central network connection even less desirable since the calling patterns from remote sites tend to be geographically clustered near the originating point. As shown in Figure 1, the system model allows for dispersed network connections called "Gateways". These gateways perform the real time call management within the network and provide the mobility management required for subscribers.

There are two severe resource constraints facing the satellite service provider: Radio Frequency (RF) spectrum and orbital slots. These resources are very scarce and require delicate negotiations to obtain them. Of prime concern in this paper, is RF spectrum. The mobile satellite bands have either been already in use for many years for mobile services or have significant interference problems in many areas of the world. Current operators holding spectrum resources will want to retain as much bandwidth as possible. New service providers will want to obtain as much bandwidth as possible. Frequency coordination meetings are held periodically with representatives from all interested parties. Attendees of these meetings agree that there is not enough spectrum to satisfy everyone's desires. Therefore, significant compromise is necessary to obtain any spectrum at all. This imposes a severe constraint on system designers to maximize the efficiency of spectrum utilization so as to balance system capacity with spectrum availability.

Call Processing Load: Satellite air interface definitions usually allow for a half rate voice CODEC to be implemented. This could lead to a doubling of initial capacity for voice calls. If a call duration of 60 seconds (for all call attempts) is assumed, the call loading on the network for half rate voice calls could run as high as 1,000 call attempts per second. This is at least an order of

magnitude higher than the capability of currently operating voice capable land mobile satellite systems such as Optus (Australia), MSAT (USA and Canada), Movisat (Mexico), and INMARSAT (within a single ocean region). This peak call processing load will, in practice, become smaller since some percentage of the calls will be data or fax calls and not all users having half rate voice CODEC capability will be active simultaneously. Still, distribution of the call processing load to gateways will reduce to a minimum the processing requirements that may still be centralized.

Cellular Interoperability: If terminal technology exists for dual mode handsets, then interoperability between the two networks is implied. If the subscriber is "homed" in the Public Land Mobile Network (PLMN), then he will become a visitor in the satellite when he "roams" there after losing contact with the PLMN. The converse is also true. The subscriber wishes to give out only one directory number for his handset and expects to be reached wherever he is currently registered. NGWs with PLMN compliant Mobile Application Part (MAP) interfaces will solve this issue, and an individual subscriber may be homed in either service. Another feature of cellular systems is the ability to hand over a call in progress to avoid its being dropped. This is certainly a requirement in PLMNs due to the relatively small coverage area of any one cell site. A GEO satellite, even with 100 to 200 user coverage beams, has "cells" of significantly larger size. This reduces or eliminates the necessity to hand over calls in progress between satellite beams. Hand over of calls in progress between the satellite network and the PLMN are likewise considered unneeded. This simplifies the interface requirements between the satellite network and the PLMNs, requiring only simple mobility management (the transfer of current location data) outside of calls and the routing of new calls to the appropriate gateway.

Call Routing: Flexible call routing is expected of modern telecommunications networks. When a GEO satellite system is envisioned, this capability becomes especially desirable. A UT located at one edge of coverage may be connected to another or a ground termination point many thousands of kilometers away. Add the DSP capability of a new generation of satellites and this user to user connection can be accomplished in a single "hop" through the satellite. This is an exceptionally powerful technique to reduce backhaul (or tail) costs. Another call routing technique, also very powerful for least cost routing, is the ability to accept a terrestrially originated call, to a satellite terminal subscriber, at any gateway in the network. Since the gateway can have access to the subscriber data, the most cost effective routing can be specified on a call by call basis. This may mean, of course, that the terrestrial networks to which the satellite system is connected must be able to accomplish this routing. One way to accomplish this would be to assign a unique country code to all the subscriber directory numbers within the system.

System Features: Many potential satellite system subscribers will examine the service features available from a service supplier before making a commitment to purchase the handset and the service. This will be especially true when there is competition in the marketplace. The system supplier must offer a competitive range of services similar to those currently available in a terrestrial network. System service providers may also compare the offerings of satellite system suppliers before making a choice as to which service to offer. The satellite system designer, therefore, is constrained to provide all of the typical services offered in a GSM terrestrial network to the extent possible.

GROUND SYSTEM DESIGN

A ground system architecture that takes advantage of the opportunities offered by advanced satellite design and hand held user terminals is illustrated in Figure 2. This architecture embodies solutions for the development constraints or challenges identified above.

System Architecture: Start with a very powerful satellite that includes a payload capable of digital processing. Satellite providers today have a range of options to choose from when selecting such items as a large reflector. Many of these items shave not been commercially available until rather recently. The payload should have a capacity for several thousand simultaneous full duplex voice circuits and data rates for bearer services up to 9,600 bps. When improved voice CODECS are implemented, this will almost double the total satellite capacity to perhaps a few 10s of thousands of half rate voice circuits. The feeder link would typically operate in the Ku-band or C-band providing direct connectivity to the gateways for the majority of the traffic.

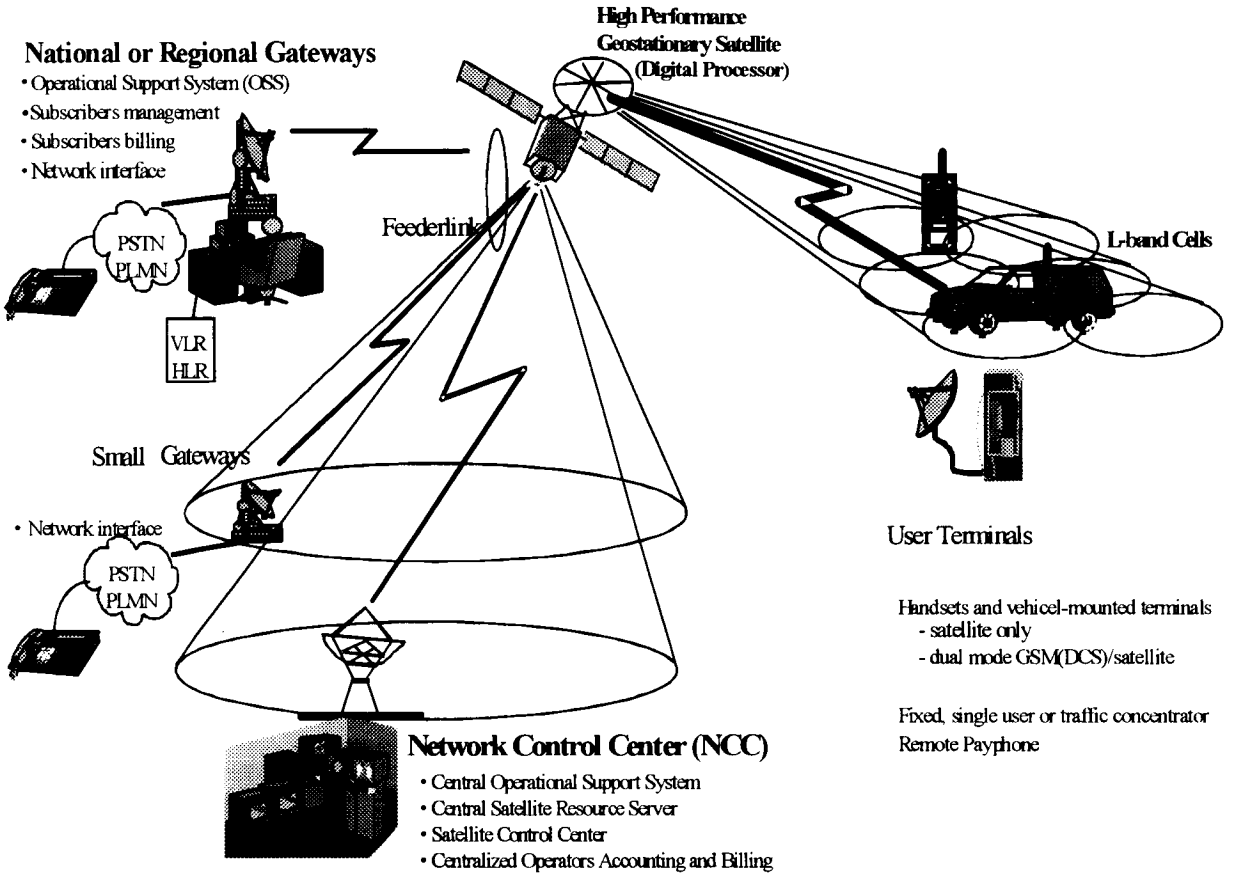


Figure 2 - Ground System Architecture

A significant percentage of the capacity would be routed on board for flexible user terminal to user terminal communications in a single hop. The frequency band allocated for user terminal communications would typically operate in the L-band. Both L-band and Ku/C-band would support multiple feeder link or user link beams. Standards for mobile satellite communications air interfaces and terminals are emerging and being promoted by such notable suppliers as Ericsson. With flexible and powerful space and terminal components available, the model must include the ground network components that take advantage of these capabilities.

The Network Control Center (NCC) provides satellite and payload control, broadcasts system wide information in all beams, provides high level management of all system resources and allocates resources to regional operators having gateways. It provides administrative services and billing of resource utilization to the regional operators. The NCC provides network synchronization, and the architecture supports a backup (redundant) NCC facility which can take over operations should the primary NCC experience a failure or for other operational reasons. A diagram of the NCC facility is shown in Figure 3. The key sub-elements include the Central Satellite Resource Server (CSRS), the Network Signaling and Synchronization Access (NSSA), the Central Operational Support System (COSS) including the Operations and Maintenance (OAM) facility and the Regional Operator Accounting Center (ROAC), the Satellite Control Center (SCC) which includes the In Orbit Test (IOT) facility. The Satellite Radio Subsystem, the Link Monitoring System (LMS), the Signaling Transfer Point (STP), and interfaces to external terrestrial networks are also part of the NCC.

The CSRS maintains the pool of satellite resources available within the system. Some resources may be dedicated to regional operators for long or mid term periods which is handled through administrative

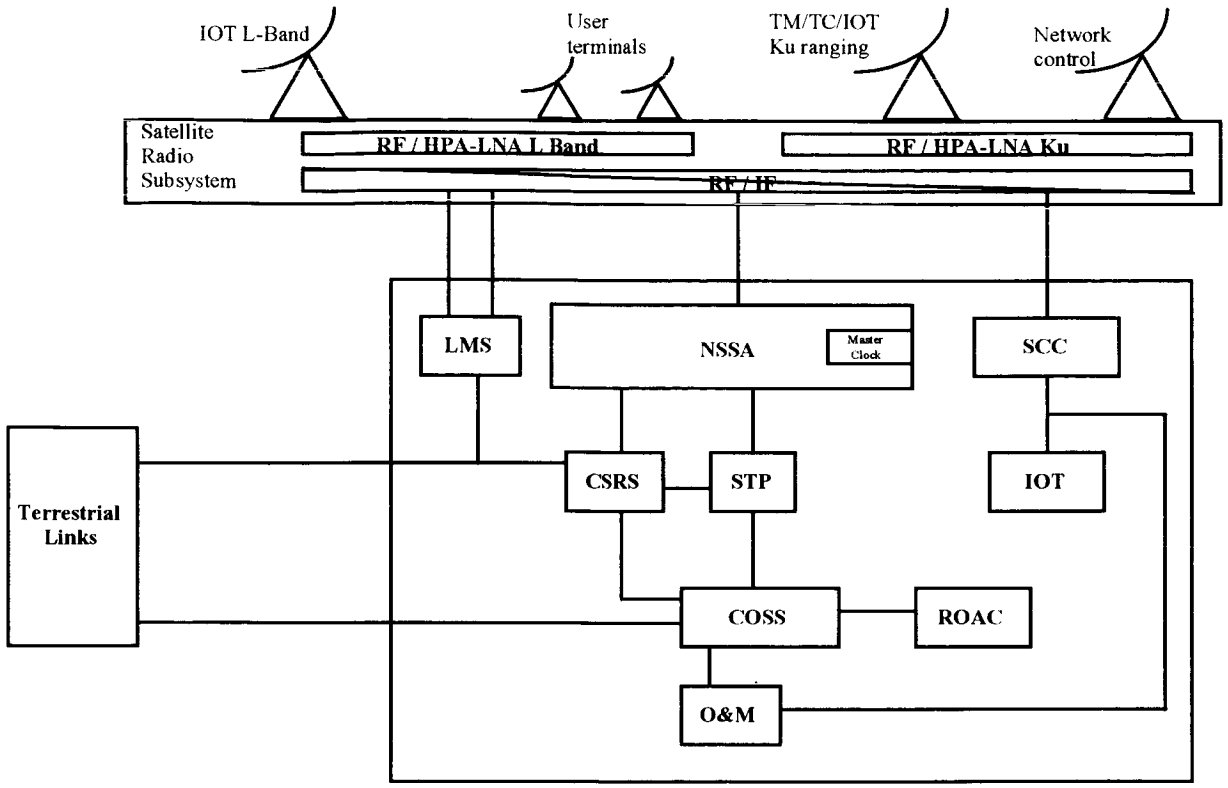


Figure 3 - NCC Diagram

procedures. Regional gateways may also request additional resources on a call by call basis which is handled via the CSRS. Release of these resources may involve a payload switching reconfiguration which is initiated by the CSRS, taking into account the current satellite connectivity, resource availability, interference conditions, and power limitations.

Payload reconfigurations are relayed to the satellite via the SCC which is in charge of satellite monitoring and control. The SCC works with the IOT which is responsible for on orbit testing, satellite subsystem calibration, and real time monitoring. Satellite performance monitoring and calibration functions are assisted by the LMS which may have several physical locations to monitor signals in many beams. These local or remote LMSs are usually located at the NCC, backup NCC, and at gateways.

Communications to the NCC from remote locations is via the signaling channels which use the STP for message routing. The STP is also used for routing messages between gateways and the NCC or other gateways and for exchanges between the network management functions at the NCC and at the gateways. Some of this routing, especially to the backup NCC, may be accomplished via external terrestrial data links.

The NSSA broadcasts general information about the network in all of the beams using the Broadcast Common Control Channel and the Synchronization Channel which is used to synchronize the entire network to a master clock reference. The NSSA accesses the satellite via the Satellite Radio Subsystem which handles all satellite communications requirements at a given site from intermediate frequencies through RF (Ku-band or L-band).

The COSS, ROAC and OAM systems form the central network management function to provide configuration, accounting, performance, security, and fault management of the entire network. The ROAC handles the special requirements associated with providing customer care and billing services between the satellite carrier operator and the regional operators. While overall network management is supervised at the NCC, the regional operators have some autonomy in operating and maintaining their portion of the network - the Gateways.

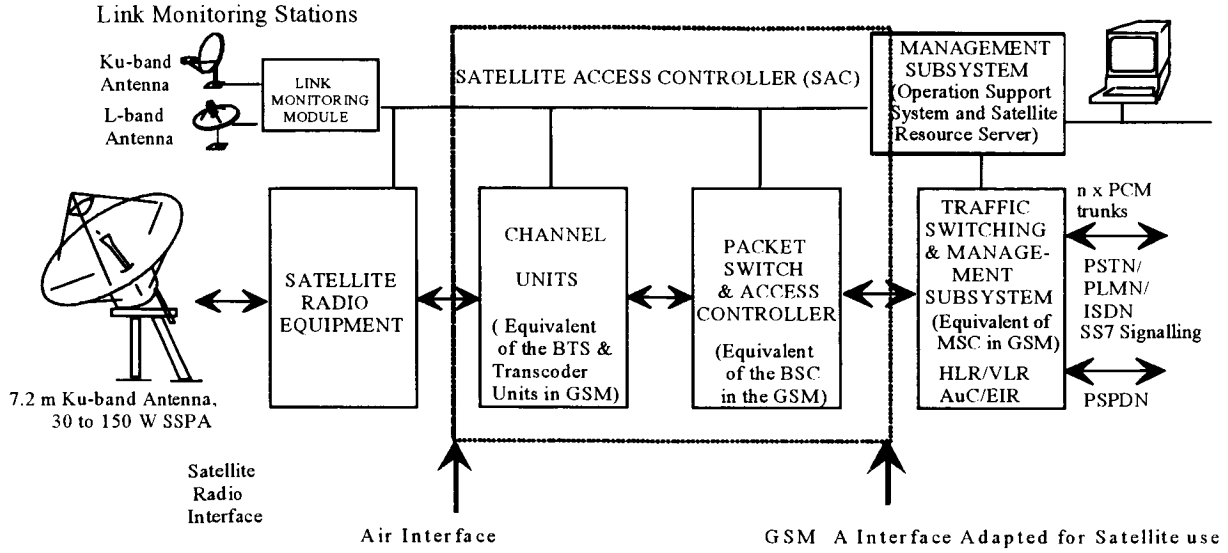


Figure 4 - NGW Diagram

Gateways may be one of three types: a National Gateway (NGW), a Local Gateway (LGW), or a Micro Gateway (μGW). Figure 4 shows the diagram of a typical National Gateway.

The NGW is associated with a service provider or regional operator and contains a full complement of mobility management facilities including: Home Location Register (HLR), Visitors Location Register (VLR), Authentication Center (AuC), Equipment Identity Register (EIR), and Gateway Mobile Switching Center (GMSC). These should be standard GSM products with the exception of the A interface which requires changes to system timing requirements due to the satellite system geometry. Replacing the Base Station Subsystem as specified in GSM is a satellite base station subsystem shown as the Satellite Access Controller and the Satellite Radio System. Changes to the GSM standards are minimized by concentrating these changes to the "left" of the A Interface (as shown in Figure 4) and leaving the MSC, AuC, VLR, EIR, and network interfaces basically unchanged. As discussed earlier, a LMS may also be located at a gateway. A NGW has some operational autonomy from the NCC for many of the local network management functions and for individual subscriber management. The NGW is responsible for least cost call routing, database management, real time switching, subscriber management, and equipment management.

LGWs or μGWs are extensions of the NGW but with lesser capabilities. In principle, they may be viewed as remote satellite base station subsystems except that they are expected to terminate calls directly into the Public Switched Telephone Network (PSTN) or allow for the origination of calls locally. This extension requires additional functionality for call routing within the network. This call routing is accomplished in cooperation with the NGW.

SUMMARY

A model for aground network for use in a GEO based satellite system has been described which promotes interoperability with GSM networks. The network reuses the GSM functionality, with extensions for use in a satellite network. The GSM functionality, to be implemented with standard components, already satisfies many of the user or operator requirements such as privacy and security as well as convenient operation if the subscriber or operator is already familiar with a GSM network. The distributed nature of the architecture ensures that a very large call processing load can be handled by the network over a more centralized architecture. Individual gateways perform the call processing functions, including: paging, access grant, and terminal random access, as well as data base management functions associated with subscriber profiles, equipment registration, and authentication. Allowing individual operators complete autonomy in the management of their subscribers is another benefit that accrues to a distributed architecture. The recommended architecture is designed to satisfy the users and operators of advanced systems offering mobile satellite services.

Handover Performance in an Integrated GSM and Satellite Mobile Communication System

W.Zhao, R.Tafazolli, B.G.Evans
Mobile Communication Research Group
Centre for Communication Systems Research (CCSR)
University of Surrey, Guildford, Surrey, GU2 5XH U.K.
Tel: +44-1483-259810 Fax: +44-1483-259504 Email: W.Zhao@ee.surrey.ac.uk

ABSTRACT

This paper aims at evaluating inter-segment handover initiation performances in an integrated GSM and satellite mobile communication system. Performance evaluation is based on an analytical model developed for inter-segment handover. Concerned handover performances are handover position, unnecessary handover probability and forced termination probability. In GSM to satellite direction, handover is initiated based on signal strength and MT-BS distance measurement. In satellite to GSM direction, a MT position assisted handover initiation criterion is proposed. Results are compared with that of GSM handover performances. The performance analysis indicate that most of the inter-segment handover performances are comparable with those of GSM handover for the proposed handover initiation criteria.

1. INTRODUCTION

It has been well known that handover position plays an important role for most of the handover performances [1-2]. Pre-mature handover generally results in unnecessary handover and delayed handover results in increased probability of forced call termination. To initiate accurate handover in GSM network, the MT-BS distance has been used as one criterion for handover initiation [3]. In GSM network, handover position can be well under control by using MT-BS distance in the handover initiation. The same rule also applies to handovers between different segments in a GSM and satellite integrated system. The introduction of MT-BS distance information results in more accurate inter-segment handover position, therefore decreased probability of unnecessary handover and probability of forced call termination.

The way to estimate MT-BS distance during a GSM to satellite handover is similar to that in GSM handover [4]. During a GSM call, the serving GSM BS is able to measure round trip delay from BS to MT. Based on this delay, MT-BS distance can be evaluated with the same uncertainty. This distance then can be used as a criterion to trigger handover from GSM to satellite. On the other hand, the estimation of MT to target BS distance during a call using satellite link is not so easy. Without distance information, the handover in satellite to GSM direction has to be distributed in a much wider range compared with GSM to satellite handover.

Solution to this problem is found in this paper by introducing a MT positioning technique during a call through satellite radio link. In a dynamic mobile satellite system, an active MT's position can be estimated through satellite LES by measuring delay and Doppler shift parameters. With appropriate filtering scheme, positioning uncertainty can be as low as 2-3km. Detailed positioning technique has been presented in [5] and it will not be addressed in this paper. Assuming location of GSM target BS can be available in the LES, then distance from MT to target BS can be calculated and a handover from satellite to GSM can be triggered by the distance.

The remainder of this paper will be organised as follows. Section 2 introduces the proposed inter-segment handover initiation criteria. Based on the criteria, handover initiation analytical model is described in section 3. In section 4, performances of inter-segment handover are evaluated and results are compared with GSM handover. Finally, section 5 gives conclusions.

2. HANDOVER INITIATION CRITERIA

Similar to GSM, parameters available to initiate inter-segment handover are still signal level, signal quality and MT-BS distance. In the GSM to satellite handover direction, the way to perform handover measurements is similar to GSM one. In the satellite

to GSM handover direction, MT positioning technique has to be used in order to evaluate its distance to target GSM BS.

In this paper, it is proposed that inter-segment handover in both handover directions are based on signal level and distance measurements. In GSM to satellite handover direction, GSM signal level and MT-BS distance are used in the handover detection phase. A handover can be triggered if any one of these is out of range. In the handover decision phase, the optimised target satellite is selected by looking for a spotbeam which has the highest signal level. In satellite to GSM handover direction, the two phases have been combined into one. This handover is triggered by two simultaneous events: GSM signal is over pre-defined threshold and MT-BS distance is within pre-defined range.

Based on above analysis, inter-segment handover initiation conditions can be given by

$$\begin{aligned} &\left\{\left[\overline{r_g(x)} < P_{ho\cdot g}\right] \ or \ \left[\hat{d}_{mb} \geq d_{\max}\right]\right\} \ and \\ &\left\{\overline{r_s(x)} \geq P_{ho\cdot s} + H\right\} \\ &HO \ \ from \ \ cell \ \ to \ \ beam \\ &\left\{\overline{r_g(x)} \geq P_{ho\cdot g} + H\right\} \ and \ \left\{\hat{d}_{mb} \geq d_{\max}\right\} \\ &HO \ \ from \ \ beam \ \ to \ \ cell \end{aligned} \tag{1}$$

In which $\overline{r_g(x)}$ is averaged signal level on GSM link, $P_{ho\cdot g}$ is the GSM link handover threshold, $\hat{d}_{mb}$ is the measurement distance from MT to GSM BS, $d_{\max}$ is the handover distance threshold, $\overline{r_s(x)}$ is the averaged signal level received from satellite radio link, $P_{ho\cdot s}$ is the satellite link minimum required signal level, H is handover hysteresis margin.
In particular, if $d_{\max} \to \infty$ (for GSM to satellite handover) or $d_{\max} \to 0$ (for satellite to GSM handover), the signal level and distance based handover becomes a signal level based handover.

3. ANALYTICAL MODEL

In this paper, analytical model introduced in [6] has been used to evaluate the received signal level under a particular shadowing environment. On the GSM radio link, Okumura model is applied for path-loss evaluation. On the satellite radio link, path-loss variation has been assumed to be pre-compensated by the satellite, therefore signal level received by the mobile has a constant mean.

First handover position

It has been identified that the first handover position has a great impact on the overall handover performances. Performance objective for inter-segment handover is to make its handover position comparable with GSM handover position.

To initiate a handover from one segment to another, handover initiation conditions expressed by (1) is applied. If the MT is in position x, probability to initiate a handover can be given by

$$P_H(x) = P_{hf}(x) \cdot P_c(x) \tag{2}$$

in which $P_{hf}(x)$ is the probability that handover condition expressed by (1) is satisfied, it is given by

$$P_{hf}(x) = P_{\det ect}(x) \cdot P_{dec}(x) \tag{3}$$

and $P_c(x)$ is the probability that the MT still remains on the original link up to position x, we have

$$P_c(x) = \prod_{x'=0}^{x}\left[1 - P_{hf}(x')\right] \tag{4}$$

$P_{\det ect}(x)$ is handover detection probability, $P_{dec}(x)$ is handover decision probability. Based on (1), for GSM to satellite handover direction, we have

$$\begin{aligned} P_{\det ect}(x) &= P\left\{\overline{r_1(x)} < P_{ho\cdot g} \ \ or \ \ \hat{d}_1 \geq d_{\max}\right\} \\ &= P\left\{\overline{r_1(x)} < P_{ho\cdot g}\right\} + P\left\{\hat{d}_1 \geq d_{\max}\right\} \\ &\quad - P\left\{\overline{r_1(x)} < P_{ho\cdot g}\right\}P\left\{\hat{d}_1 \geq d_{\max}\right\} \end{aligned} \tag{5}$$

$$P_{dec}(x) = P\left\{\overline{r_s(x)} \geq P_{ho\cdot s} + H\right\} \tag{6}$$

For satellite to GSM handover direction, we have

$$P_{\det ect}(x) = P\left\{\overline{r_1(x)} \geq P_{ho} + H\right\} \tag{7}$$

$$P_{dec}(x) = P\left\{\left\{\overline{r_g(x)} \geq P_{ho\cdot g} + H\right\} and \left\{\hat{d}_1 \leq d_{\max}\right\}\right\}$$
$$= P\left\{\overline{r_g(x)} \geq P_{ho\cdot g} + H\right\} \cdot P\left\{\hat{d}_1 \leq d_{\max}\right\}$$

(8)

In particular, if $d_{\max} \to \infty$, the signal level and distance based handover becomes a signal level based handover.

Probability of forced termination

In GSM, the forced call termination criterion is based on a radio link counter S. If the MT is unable to decode a SACCH message, S is decreased by 1. In the case of a successful reception of a SACCH message, S is increased by 2. In any case S shall not exceed the value of RADIO_LINK_TIMEOUT. If S reaches 0, a call dropping is declared [7]. To establish an analytical model, this criterion is simplified to be the follows. If the current radio link is heavily shadowed for a given duration ΔT (or equivalently ΔL in distance), at the same time, conditions of handing over to another link is not satisfied, the call is forced to be terminated. In this section, the same idea is also applied for inter-segment handover. ΔT can be calculated by

$$\Delta T = RADIO\_LINK\_TIMEOUT \cdot T_{SACCH} \quad (9)$$

For GSM to satellite handover, T_{SACCH} is the SACCH frame duration on GSM radio link. For satellite to GSM handover, T_{SACCH} is the SACCH frame duration on satellite radio link. It should be noticed that the simplified forced termination criterion only considers call dropping for signal level reason, it does not take the co-channel interference into account.

Based on above definition, the forced termination probability in both handover directions can be given by

$$P_D(x) = P_c(x) \prod_{x-v\cdot\Delta T}^{x} \left[P\left\{\overline{r(x)} < P_{\min}\right\}\right] \quad (10)$$

$\overline{r(x)}$ is the average received signal level from GSM radio link (GSM to satellite handover) or from satellite radio link (satellite to GSM handover), $P_{\min}$ is the minimum acceptable signal level on GSM radio link (GSM to satellite handover) or on satellite radio link (satellite to GSM handover).

Probability of unnecessary handover

The probability of unnecessary handover is defined to be the probability that a MT handovers from one link to another and then initiates another handover back to the first one within a pre-defined duration. The main reason of the unnecessary handover is the signal level profile unevenness because of various local environment. An unnecessary inter-network handover produces additional signalling load on both radio and terrestrial links, much longer handover interruption and lower call quality.

If $P_H(x)$ is the probability for the MT to handover from one segment to another and $P_{hb}(x)$ is the probability that handover condition in backward direction described by (1) is satisfied, then $P_U(x)$, the probability to initiate an unnecessary handover in position x is

$$P_U(x) = P_H(x) \sum_{x'=x}^{x+v\cdot\Delta t} P_{hb}(x') \quad (11)$$

$P_H(x)$ is still calculated from (2), $P_{hb}(x)$ is given by

$$P_{hb}(x) = P_{det\,ect}(x) \cdot P_{dec}(x) \quad (12)$$

Depending on the handover direction, $P_{det\,ect}(x)$ and $P_{dec}(x)$ can be either calculated from (7-8) (GSM to satellite handover) or from (5-6) (satellite to GSM handover).

4. NUMERICAL RESULTS

Numerical results will be presented in this section based on previous introduction. The following assumptions will be applied throughout the numerical analysis.

In GSM network, cell radius R=30km. With a hexagonal cell layout, distance between neighbouring BSs is 52km, therefore the ideal GSM handover position is $d_{\max} = 26km$. In each BS, EIRP=30dBW. BS antenna height is 100m, MT antenna height is 3m. Standard deviation of shadowing on each GSM radio link is 8dB. With 30km/s MT velocity and 0.5s sampling interval, shadowing correlation between two sampling position is 0.9. During a call, the signal is averaged

over 32 samples which corresponds to 15s average window.

In satellite network, the transmission EIRP from each satellite is 33.2dBW. The variation of satellite path attenuation due to satellite elevation angle's change is assumed to be pre-compensated on the satellite. Therefore during the selected analysis window (it takes one hour for the mobile travelling through the considered range 15km to 45km), the attenuation change is purely caused by satellite antenna's radiation pattern, this has been assumed to be 2dB. The MT antenna has 2dB gain. On satellite radio link, standard deviation of shadowing is 8dB with 0.8 shadowing correlation between two sampling positions.

Handover position

Based on (2) - (6), GSM to satellite handover probabilities are calculated under various conditions, results are plotted in Fig. 1 against the distance from the MT to its serving BS. The same graph also shows the handover probabilities between two GSM cells. From this graph, with the selected handover initiation criteria, inter-segment handover and inter-cell handover achieves similar performance. For H=2dB, most of the handovers will happen around ideal handover position (26km to serving BS).

The satellite to GSM handover probabilities are calculated based on (7) and (8) and results are plotted in Fig. 2. Since MT positioning uncertainty is high, MT-BS distance estimation error is also high, therefore the satellite to GSM handover position will not be as accurate as that of GSM to satellite one.

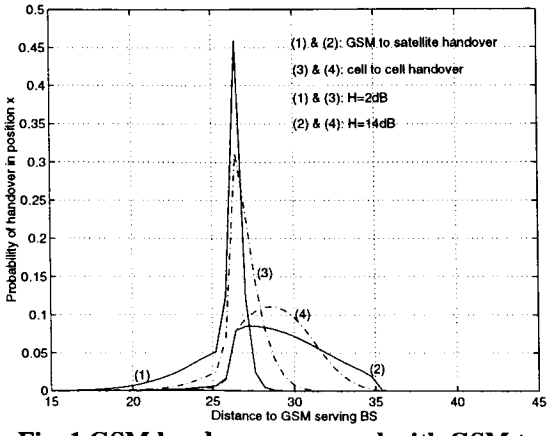


Fig. 1 GSM handover compared with GSM to satellite handover

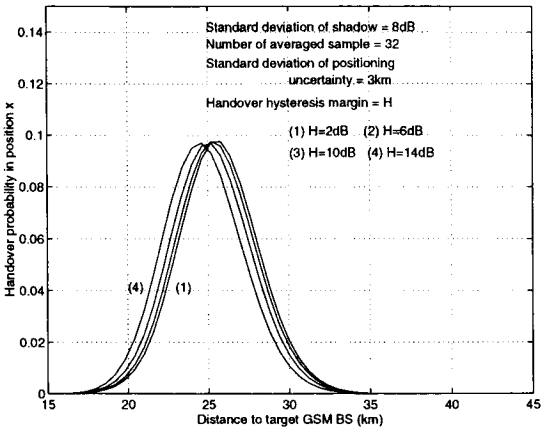


Fig. 2 $P_H(x)$: beam to cell handover (based on signal level and position)

Probability of forced call termination

Based on (10), probabilities of forced termination for GSM to satellite handover are calculated and results are plotted in Fig. 3. Again, the same graph also shows the $P_D(x)$ during a GSM handover. From this graph, the GSM to satellite handover has a lower probability of forced termination. This improved $P_D(x)$ comes from the fact that GSM to satellite handover has more accurate handover position for the considered initiation criterion.

$P_D(x)$ results for satellite to GSM handover are given in Fig. 4. In this graph, the first curve is a signal level based handover, the second one is based on signal level and MT-BS position. Since a signal level and position based handover moves the average handover position closer to GSM BS, it generally results in increased $P_D(x)$. Meanwhile, the $P_D(x)$ in this handover direction is 2-3 orders lower compared with handover in the other direction, it is considered that this increased $P_D(x)$ will not influence system performance significantly.

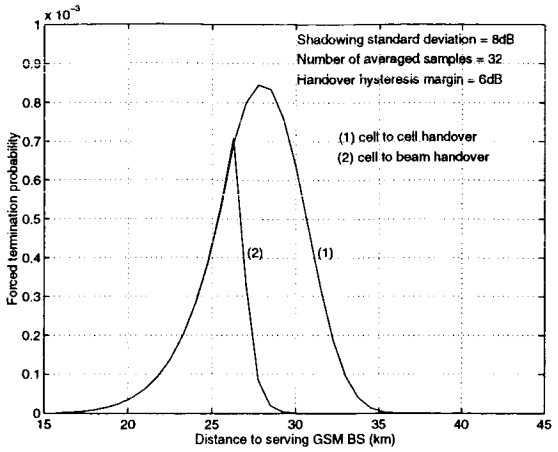


Fig. 3 $P_D(x)$: cell to beam handover compared with GSM handover

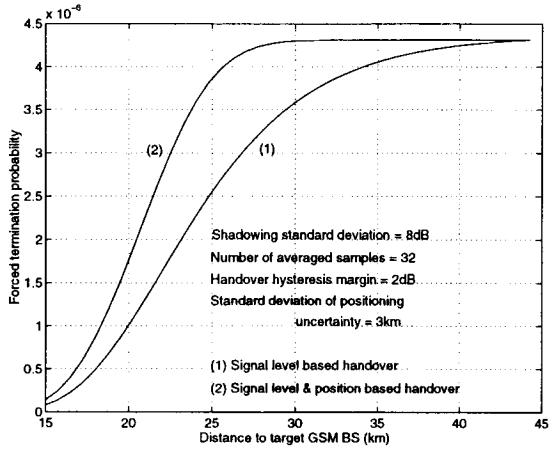


Fig. 4 $P_D(x)$: beam to cell handover

Probability of unnecessary handover

The analysis of $P_U(x)$ is only carried out in GSM to satellite handover direction, since handovers in the other direction have similar performance. $P_U(x)$ is evaluated from (11) and results are plotted in Fig. 5. Again, $P_U(x)$ for inter-segment handover is compared with $P_U(x)$ during GSM handover in the same graph. In order to reduce the $P_U(x)$, handover distance thresholds has been staggered by *2km*, e.g. $d_{\max} = 27km$ for GSM to satellite handover and $d_{\max} = 25km$ for satellite to GSM handover. In this way, the average GSM to satellite handover position is well behind the average satellite to GSM handover position, therefore most of the unnecessary handovers can be avoided. From Fig. 5, with a staggered handover position in different handover directions, $P_U(x)$ for inter-segment handover can be similar to that of GSM handover.

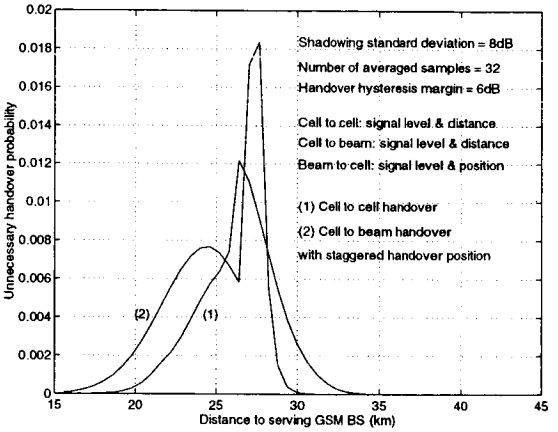


Fig. 5 $P_U(x)$: cell to beam handover compared with GSM handover

5. CONCLUSIONS

Inter-segment handover initiation criteria have been proposed in a GSM and satellite integrated mobile communication system. Handover performances are evaluated based on analytical model and results have been compared with GSM handover performances. In GSM to satellite handover direction, handover initiation is based on signal level and MT-BS distance. With the selected handover hysteresis margin and signal level threshold on satellite radio link, numerical results show that the GSM to satellite handover achieves more accurate handover position, lower call termination probability and similar unnecessary handover probability compared with GSM handover. In satellite to GSM handover direction, the handover is initiated by GSM signal level measurement and MT's positioning estimation. With the positioning assistance, handover position can be distributed in a much smaller range compared with signal level based handover, at the same time, the achieved call termination probability has similar value to GSM handover. The analysis has been based on ICO satellite constellation and link power assumptions, but the same conclusion can be also applicable to other dynamic satellite constellations.

REFERENCES

1. **Rajiv Vijayan,** "A model for analysing handoff algorithms", IEEE Transactions on Vehicular Technology, Vol 42, No. 3, Aug, 1993.

2. **R.Beck and F.W.Ho,** "Evaluation of performance of field strength related handover strategies for micro-cellular systems", Proc. 3rd Nordic Seminar on Digital Land Mobile Radio Communication, Copenhagen, Denmarj, 1988.
3. **ETSI** - European Telecommunications Standardization Institute: GSM Recommendation 03.09. Handover procedures.
4. **ETSI** - European Telecommunications Standardization Institute: GSM Recommendation 05.10. Radio subsystem synchronisation.
5. **W.Zhao, R.Tafazolli, B.G.Evans,** "A UT positioning approach for dynamic satellite constellations". Proceedings of the Fourth International Mobile Satellite Conference (IMSC'95), Ottawa, Canada, June 1995. pp35-41.
6. **W.Zhao, R.Tafazolli, B.G.Evans,** "A combined handover algorithm for dynamic constellations", Electronic Letters, 28th Mar., 1996, Vol. 32, No. 7, pp622-624.
7. **ETSI** - European Telecommunications Standardization Institute: GSM Recommendation 03.05. Technical performance objectives.

An FEC Booster for UDP Applications over Terrestrial and Satellite Wireless Networks

D. S. Bakin, W. S. Marcus, A. J. McAuley, T. M. Raleigh
Bellcore
445 South Street, Morristown, NJ 07960-4698
e-mail: {dbakin, wsm, mcauley, tom}@bellcore.com

Abstract

We describe an experimental packet Forward Error Correction (FEC) booster with fast erasure decoding, called the FZC Booster, implemented in the Unix kernel. We show that, as expected, an FZC Booster pair decreases the effective packet loss of UDP applications with both random and bursty packet errors (from simulated terrestrial and satellite wireless networks). Most importantly, it does so without reducing robustness since it does not modify or significantly delay data packets. Moreover, FZC Boosters can be dynamically placed anywhere in the network or end systems, acting only on selected streams.

1 Introduction

A global information infrastructure must capitalize on the unique capabilities of terrestrial and satellite wireless networks. However, the existing protocols (e.g., TCP and UDP) perform poorly with the noise inherent in most wireless networks [1] and the high latency of satellite links [2]. The problem is especially apparent in the transport protocols. Finding the best solution to poor protocol performance over terrestrial and satellite wireless networks is thus critical to the success of a global information infrastructure.

An important example of poor protocol performance is the high packet error rate seen by real time [3] and multicast [4] applications using UDP (often with RTP). The UDP protocol has error detection but no error correction; thus a noisy wireless environment increases packet loss rates.

The primary goal of this paper is to describe a robust method of reducing effective packet error rate using a protocol booster [5] with a packet Forward Error Correction (FEC) code and erasure decoding. To distinguish it's erasure decoding, we call it an "**FZC Booster**" rather than an FEC Booster. In addition, we provide some initial experiment results using the FZC Booster with UDP.

Section 2 describes packet FEC; Section 3 describes Protocol Boosters; and Section 4 provides some initial results.

2 Packet Forward Error Correction

2•1 Packet FEC with Erasure Decoding

An FEC code allows recovery of lost information without retransmission. The FEC encoder adds redundant information to a data steam. The FEC decoder uses the redundancy to recover the original data, despite corrupted or lost information.

Erasure decoding is a subset of FEC decoding, where the FEC decoder knows the positions of the errors. Compared with an error decoder, an erasure decoder uses its knowledge of the error positions to double error correction capabilities and reduce computation needs. It assumes that packets either arrive correctly or are erased. For the case when packets could arrive in error (e.g., using UDP without the checksum option), then error detection must be done before erasure decoding (to turn packet errors into packet erasures).

With any packet FEC code, the transmitter adds h parity packets to k data packets. An erasure decoder at the receiver can reconstruct the k data packets from

*This work was supported by the Defense Advanced Research Projects Agency under contract #DABT63-95-C-0073.

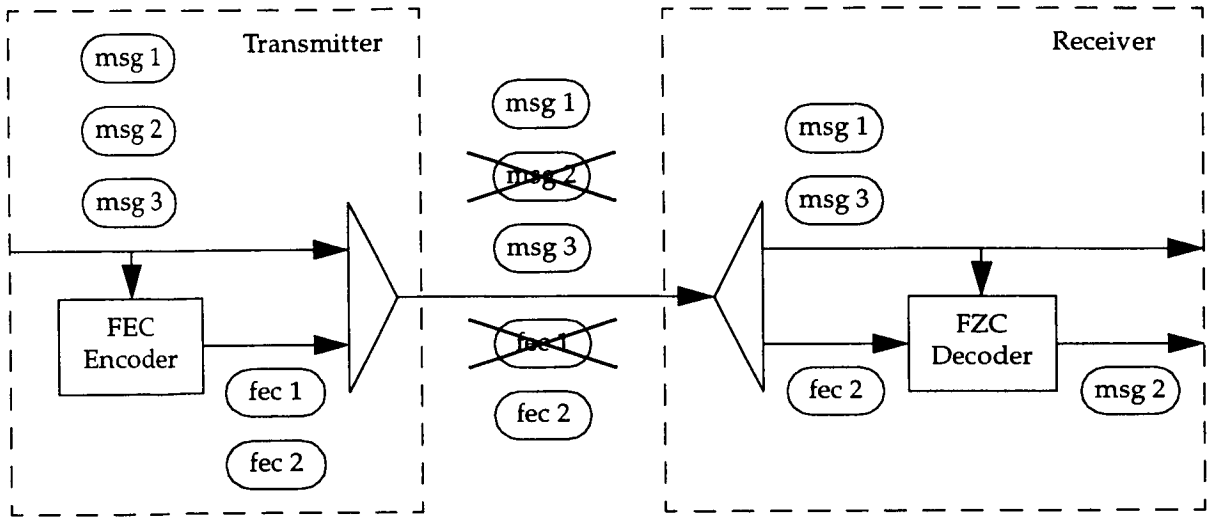


Fig 1. Packet Forward Error Correction with Erasure Decoding

any k of the h+k packets. For example, Figure 1 shows a transmitter sending k=3 data packets (msg 1, msg 2, msg 3) with the FEC encoder adding h=2 parity packets (fec 1, fec 2). Despite two lost packets (msg 2 and fec 1), the receiver gets the three data packets. If packets are forwarded as they are received (as they are in our FZC Booster), then the packets may be misordered (msg 1, msg 3, then reconstructed msg 2).

2•2 *FEC Code suited to Booster Application*

We use the FEC code and erasure decoding algorithm described in [6], because of several desirable properties:

- It allows a fast software implementation (1 to 20Mb/s depending on the value of h).
- The receiver need not know the number of parities (h) sent (simplifies dynamic control algorithms).
- The receiver need not wait for, nor calculate, any parity packets beyond the number of missing data packets.

Both the encoder and decoder generate weight matrix B. This is computationally intensive, but need only be performed once, at start-up. The h parity packets are the matrix product of the first k columns and h rows of B with the k data packets. The decoder waits until k packets have arrived. If the k packets are the k data packets we are finished; otherwise, if we are missing g data packets, then we perform a matrix inversion and a couple of matrix multiplications to recover the missing g data packets. A complete description of the FEC code and erasure decoding algorithm can be found in [6].

3 FEC booster

This section describes the concept of Protocol Boosters, as well as the application and implementation of packet FEC with erasure decoding in the FZC Booster.

3•1 *Improving Protocol Performance*

Broadly speaking, three classes of solutions have been proposed to improve protocol performance on noisy wireless networks:

- Add powerful link error control protocols to improve the performance of existing protocols.
- Create new (or modify existing) transport protocols that work more efficiently over terrestrial and satellite wireless networks.
- Adapt the protocol to the specific environment, using different protocols depending on the network conditions or application requirements.

Link error control is needed to reduce the bit error rate on noisy wireless links to allow most packets to get through. Adding more powerful link error control can further reduce the effective packet error rate. The extra error control, however, also reduces the available bandwidth and increases latency; while providing functionality that may not only be unnecessary for many applications, but may also be inefficient [5].

New transport protocols allow an application to choose only the functionality that is needed [7]. Unfortunately, new protocols take years to

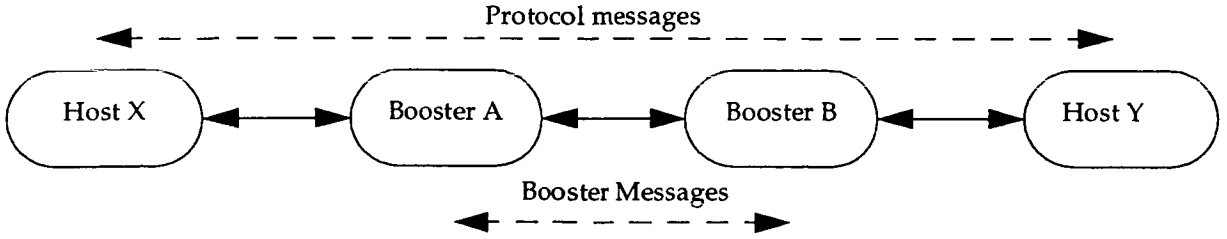


Figure 2 Two-Element Booster

standardize and even longer to become widely deployed. Also, any transport protocol must work in the heterogeneous network environment and with diverse applications, forcing some compromise in efficiency. For example, TCP's congestion avoidance mechanism is necessary to prevent network overload, but also limits effective throughput when there is loss, even with large windows [8], improved congestion avoidance mechanisms [9] and selective acknowledgments [10].

Adapting a protocol to application requirements or network conditions can quickly provide efficiency improvement. Existing approaches, however, such as terminating one protocol and starting a new protocol, are not robust. We have developed a fast and robust way to adapt protocols using protocol boosters [5].

3•2 Protocol Booster

A booster is a parasitic module that transparently improves a protocol's performance. The booster can reside anywhere in the network or end systems, and provides a more robust architectural alternative to existing protocol enhancement techniques, such as protocol conversion and protocol termination.

A booster is parasitic because it may add, delete or delay protocol messages, but never originates, terminates, or converts that protocol. A multi-element booster may define new protocol messages to exchange among themselves, but these protocols are originated and terminated by booster elements and are not visible external to the booster. Figure 2 shows, for example, the information flow in a two-element booster.

A booster is transparent to the protocol being boosted. Thus, elimination of a booster does not prevent end-to-end communication, as would, for example, the removal of one end of a conversion protocol (such as a TCP/IP header compression unit).

Multiple protocol boosters can operate on the same protocol. Although a booster itself is not a protocol, a protocol and a booster combine to form a new protocol. With a set of boosters, it is natural to talk about the protocol family generated by the boosters. Our initial work has been the development of a family of error control protocol boosters for the popular TCP/IP communication stack. The booster family will allow users to selectively improve performance of existing applications using standard TCP/IP stacks, making it appear as if they are using an advanced version of TCP/IP over the parts of the network using the boosters.

3•3 FZC Booster

The "**FZC Booster**" uses a packet FEC code with erasure decoding. This booster is part of our TCP/IP booster family which also includes ARQ boosters, a reorder booster, and an error detection booster (it can turn corrupted packets into erasures for the FZC Booster). Each booster is designed to provide a specific function and work harmoniously with the other boosters in the family.

The FZC Booster reduces the effective packet loss rate on noisy links, such as terrestrial and satellite wireless networks. Although packet error correction is normally most efficiently and flexibly done by packet retransmission (ARQ), FEC is desirable for some latency constrained and multicast applications, or where the return channel is unavailable or slow and where the loss of a single packet causes other packets to need retransmission. The FZC Booster is not, however, well suited for dealing with packet loss due to congestion [11]. (This can be handled by other members of our TCP/IP error control booster family).

Figure 3 shows a typical application. The FZC Booster at the transmitter side of the wireless network adds parity packets. The FZC Booster at the receiver side of the wireless network removes the parity

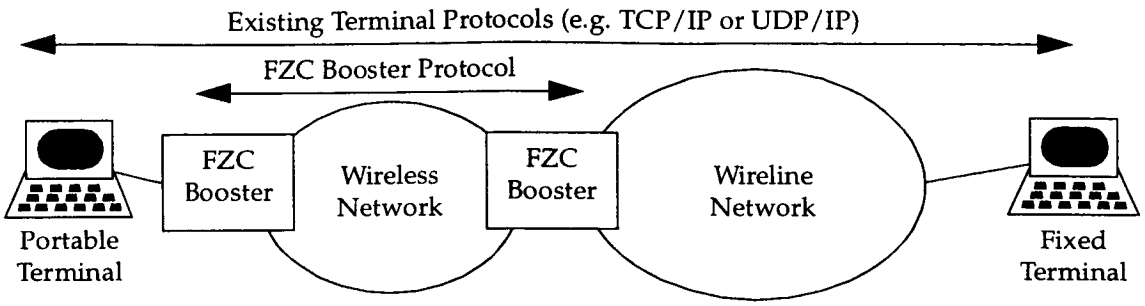


Figure 3 Communication over a wireless network with a FZC Booster

packets and regenerates missing data packets. This appears similar to link FEC; however link FEC only corrects bits (or words) not IP packets and cannot be applied between any two points in a network (including the end systems). Also the FZC Booster can be applied incrementally. In Figure 3, for example, we could add an extra FZC Booster at the Fixed Terminal. If this booster adds h1 parity packets and the second booster adds h2 parity packets, then the portable terminal can recover from up to h1+h2 packet erasures. In the reverse direction the second booster could reduce the number of redundant packets to reduce congestion in the wireline network.

3•4 Implementation

We added (at compile time) Protocol Boosters support to the FreeBSD 2.1.0 version of UNIX for the i386 (Intel) architecture. The support resides in the IP portion of the TCP/IP networking stack. Every IP packet processed is examined to detect if it is a member of a boosted communication channel. If boosting is specified, the packet is handed to the appropriate booster/debooster. Our implementation currently identifies boosted channels by source and destination IP addresses; however, we can add stronger filtering based on other information, such as port number.

For each FZC Booster channel, the FZC booster caches, then immediately forwards, each data packet it receives: whether the packet is from an upper layer protocol or the IP Forwarder. The only modification to each data packet is that the FZC booster overwrites the IP packet's 16-bit identification field with a sequence number, allowing the decoder to know the packet's position information. This does not change the end-to-end UDP datagrams or TCP segments.

After receiving k packets (k is defined per channel) on a given channel, the cached packets are zero-padded to the size of the largest packet in the cache. Also each packet's size and protocol type are appended to the packet's tail. The transmitter performs the matrix multiplication (see Section 2.2) over the payload, padding, and appended tail of the k packets. The h overcode packet payloads produced by the FEC encoder are then prepended with an IP header and a booster header. This IP header contains a prototype field identifying it as a protocol booster packet and a sequence number in the 16-bit identication field. The booster header contains the type of booster (FZC Booster), the value k, and the sequence number of the first of the k packets. The h packets are then transmitted toward the same destination as the data packets.

The FZC debooster at the receive side caches incoming data packets and immediately forwards them to either an upper layer protocol or towards their eventual destination. Overcode packets are also cached, but are not forwarded. Packets are released from the cache when either (1) an entire collection of k data packets are present, (2) the received data packets plus parity packets equals k, or (3) when the cache occupancy dictates cache content replacement. Only in situation (2) are the matrix computations performed to generate the missing data packets.

4 Experiment Results

This section describes the results of running UDP with different levels of FEC overcode over different emulated wireless networks.

4•1 Error Model and Loss Module

The amount and types of errors on wireless networks depend upon link conditions and link error control. In general, however, errors are not random but come in a burst of consecutive bits. Whether the

packet errors are also bursty depends on the size of packets and average burst error lengths. Based on the actual packet loss results obtained in [12] we constructed two basic packet error models: random and bursty. These error models were used in a loss module, placed in the IP receive path, that can delete packets based on the chosen error model.

4•2 Experimental Setup

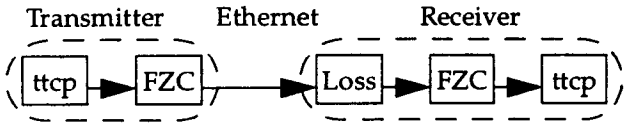


Fig 4 Experimental Setup

As our application we used public domain Test TCP program (**ttcp**), with the UDP option. We ran ttcp between two machines on a dedicated ethernet. Both ends ran the FZC Booster and the receiver also ran the loss module. The experiments used a packet size of 512 bytes, with 1000 packets per trial, and an FEC block size of k=20 packets.

4•3 Results

Figure 5 shows the percentage of packets successfully transmitted versus the amount of random packet loss. As we would expect, with no FEC the effective packet loss is roughly the same as the loss rate defined in the loss module; increasing redundancy decreases the effective packet loss.

Figure 6 depicts the percentage of packets successfully transmitted versus both random and bursty packet loss. Although burstiness further increases the network packet loss, FEC is still effective. In fact, for the same percentage of overcode, we can approach the effective packet loss on links with random errors by choosing a larger block size (k).

Conclusion

We show that, as expected, an FZC Booster pair decreases the effective packet loss of UDP applications with both random and bursty packet loss (from simulated terrestrial and satellite wireless networks). Most importantly, it does so without reducing robustness since it does not modify or delay data packets. Moreover, FZC Boosters can be dynamically placed anywhere in the network or end systems, acting only on selected streams.

References

[1] R. Caceres & L. Iftode, "The Effects of Mobility on Reliable Transport Protocols," Proc of the 14'th ICDCS, June, 1994.

[2] M. Allman, C. Hayes, H. Kruse, S. Ostermann, "TCP Performance over Satellite Links," 5'th Int'l Conference on Telecommunication Systems, 1997.

[3] A. Albanese, J. Blomer, J. Edmunds, M. Luby, M. Sudan, "Priority Encoding Transmission," Proc. of 35th Symposium on Foundations of Computer Sciences, Santa Fe, pp. 604-612, Nov 1994.

[4] N. Shacham, J. S. Meditch, "An Algorithm for Optimal Multicast of Multimedia Streams," Proceeding of IEEE INFOCOM '94, Toronto, Canada, 1994.

[5] D. Feldmeier, A. McAuley and J. Smith, "Protocol Boosters," Submitted to IEEE JSAC.

[6] A. J. McAuley and D. C. Feldmeier, "A Powerful Forward Erasure Correction Code Suited to High Speed Software Implementation,"submitted to IEEE Transactions on Computers.

[7] R. Durst, G. Miller, & E. Travis, "TCP Extensions for Space Communications," Proceedings of ACM Mobicom 96, November 1996.

[8] V. Jacobson, R. Braden, D. Borman, "TCP Extensions for High Performance," RFC 1323, May 1992.

[9] W. Stevens, "TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms, RFC 2001, January 1997.

[10] M. Mathis, J. Mahdavi, S. Floyd, A. Romanow, "TCP Selective Acknowledgment Options, RFC 2018, October 1996.

[11] E. Biersack, "Performance Evaluation of Forward Error Correction in an ATM environment," IEEE Journal on Selected Areas in Communication, Vol. 11, No. 5, pp. 631-640, May 1993.

[12] A. J. McAuley, D. S. Pinck, T. Kanai, M. Kramer, G. Ramirez, H. Tohme, L. Tong, "Experimental Results From Internetworking Data Applications Over Various Wireless Networks Using a Single Flexible Error Control Protocol," Fifth WINLAB Workshop, East Brunswick, New Jersey, April 26-27, 1995.

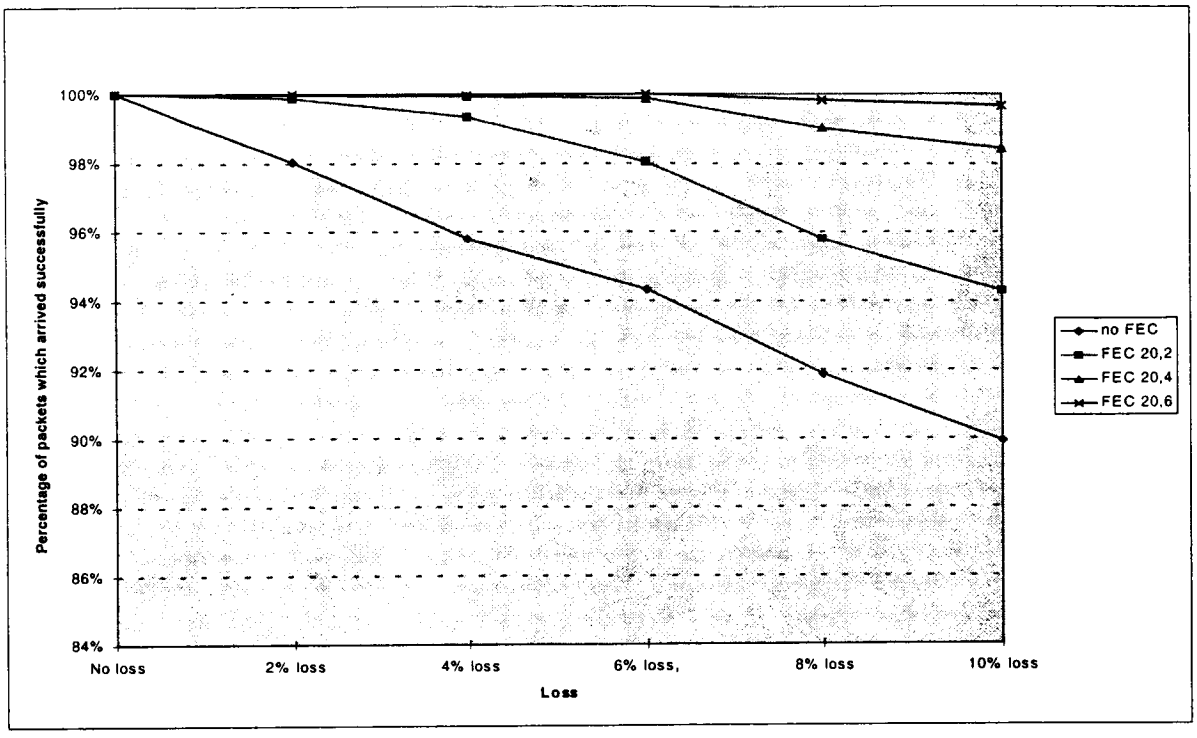


Figure 5 FZC Booster with Random Errors

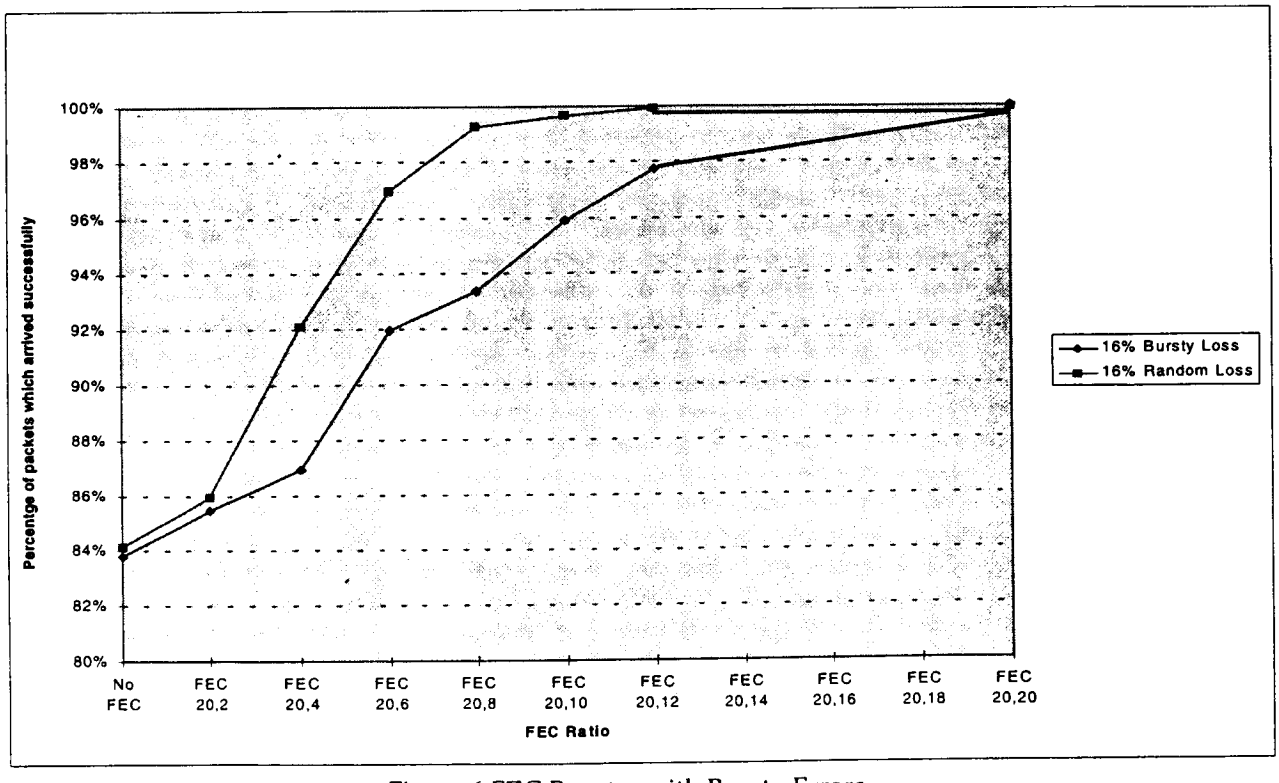


Figure 6 FZC Booster with Bursty Errors

Panel 2

A METRIC TO EVALUATE MOBILE SATELLITE SYSTEMS

Elizabeth L. Young
Consultant to JPL
162 Wilton Creek Road
Hartfield, VA 23071
Phone: 804-776-9436 FAX: 804-776-7220
RRLOVELL@postoffice.worldnet.att.net

ABSTRACT

The concept of a "cost per billable minute" methodology to analyze mobile satellite systems is reviewed. Certain assumptions, notably those about the marketplace and regulatory policies, may need to be revisited. Fading and power control assumptions need to be tested. Overall, the metric would seem to have value in the design phase of a system and for comparisons between and among alternative systems.

INTRODUCTION

In early 1995, a class of advanced graduate students at MIT, under the instruction of the Jerome Clarke Hunsacker Professor of Aeronautical Engineering, Robert R. Lovell, tackled the problem of how systems engineers could establish a useful method of evaluating different systems architectures during the design process for satellites intended to support mobile communications. They were aided by guest lecturers from several leading aerospace companies and mobile system providers. Their work led them to postulate that: "The primary metric used to evaluate the design is the system cost per useful virtual circuit minute." [1] The short-hand for this metric became "cost per billable minute" to produce a defined level of profitability or internal rate of return (IRR), which the study targeted as 30%. [2] One half of this equation was the number of billable minutes any system could address; the other half represented the matching satellite system life cycle costs, which included development, deployment, ground infrastructure, insurance and operations. [3] A "virtual circuit" was defined as a full duplex voice connection, of predetermined quality, between two users. "Useful" in the primary metric included consideration of the network capacity, individual spacecraft capacity, and customer or user geographical distribution.

The students used the "cost per billable minute" metric to compare and contrast certain proposed and theoretical mobile satellite systems and to design a hypothetical system. (While they examined the probable dimensions, performance and costs/prices of handheld units, they did not include any assumptions about variations in handhelds that would influence the "cost per billable minute" metric.) Their conclusion was that several different types of mobile systems could be successful, provided that assumptions about the worldwide market were correct. The class study, which was published, also became the basis for two masters' theses and an AIAA paper. The purpose of the present article is to revisit certain of the assumptions that were imbedded in the MIT work and to speculate on variables that were not within the scope of the study but which could affect system choices and ultimate profitability of mobile satellite systems.

MARKET ASSUMPTIONS REVISITED

A key finding of the study was that the single most important variable in assessing likely success is market share, rather than, e.g., selection of orbit, access scheme, or complexity of the space or ground systems. [4] This assumption itself could, of course, be challenged, but even if it is correct, the underlying assumptions about the nature of the

present and future market for mobile satellite services become critical.

One such assumption was that by the year 2013, the addressable market would be 6.5 billion minutes. [5] Another was that growth would be capped in areas of the world where cellular was likely to be dominant; therefore, the greatest growth for satellite services was predicted to be in China, the CIS and Russia, other areas of the Far East and Latin America. (It was further assumed that only 10% of the users would be in an urban environment, while at least 40% would be in suburban and 50 % in rural areas.) Some consideration was taken of business users in the maritime and aeronautical environments.

One could challenge any one or more of these assumptions, but a few key questions emerge. First, will the absolute appetite for mobile communications continue to grow worldwide? We have some evidence that it will. Inmarsat, the first and oldest mobile satellite system, continues to increase its minutes and revenues even as other new mobile systems, e.g., MSAT and OPTUS, have come on line. On the cellular front, the Strategis Group predicts that "by 2002 . . . there will be over 500 million cellular and PCS subscribers worldwide, more than tripling the end-year 1996 cellular and PCS subscriber base of nearly 140 million". [6]

However, the fact that cellular and PCS systems ARE growing, leaves a very real question about whether the size of the addressable market for satellites may have been overestimated. Obviously, issues such as quality, ease of use and prices of satellite services will dictate to a great extent how well they will compete with non-satellite services. Interestingly, the MIT study concluded that even with as many as five mobile satellite systems addressing the worldwide market, all can be cost effective, but that as any system falls below 31% of the market toward 10%, its prices will likely approach $10.00 or more a minute retail, which may not be competitive with other satellite systems or with cellular and PCS. Another conclusion derived from the study was that "room for multiple [mobile satellite] systems exists since the initial systems will be unable to meet the full market demand". [7]

As a corollary question, have markets been underestimated, e.g., by not including the full complement of possible aeronautical, maritime and industrial users, and by not assuming satellite technology will be deemed superior to terrestrially based systems in some settings? Also, should projected mobile satellite user minutes be capped at 125 million in more populated markets (as was assumed in the MIT study) or should correlative cellular/PCS and satellite minute growth be assumed? These are a few of the questions amenable to further research. With new, non-geo mobile satellite systems coming on line in the near future, we also have no absolute way of knowing whether their availability will release a pent-up demand.

The MIT study assumed that mobile satellite users' behavior would mirror that of fixed communications users, i.e., with peak use occurring between 9 a.m. to noon and 2 p.m. to 5 p.m. on weekdays. However, if mobile users are content with voice mail reception for a large percentage of their calls and if data transmissions begin to substitute for voice, it is at least possible that peaks may be redefined, allowing more efficient use of circuits -- and lower prices to users.

Additional assumptions about critical user requirements included: ubiquitous coverage; 90% availability of any system (meaning a call could be placed and completed about 90% of the time); 95% reliability (the call will not be dropped more than 5% of the time); call duration (an average of 5 minutes for a voice call or 20 minutes for a data transmission); voice quality on a par with cellular; and price (customers would expect competitive pricing but would also expect to pay more for satellite-based connections than for PSTN or local cellular calls). [8] As more customers have experience with both cellular and satellite services, these assumptions need to be retested, especially those concerning quality -- is, for example, mere voice recognition enough in many cases? Is call duration as predicted? And most important, what is the price elasticity?

TECHNICAL ASSUMPTIONS REVISITED

It is beyond the scope of this paper to examine in detail all the technical assumptions that led the MIT students to conclude that most planned (or hypothetical) mobile satellite systems could be profitable given sufficient market share. However, a few of the key assumptions that entered the calculations of "life cycle" cost would need to be reviewed before making any rigorous application of the MIT methodology to real systems. These assumptions are discussed below.

Satellite construction and replacement costs dominated cost estimates, followed by launch service and insurance costs. Gateway costs were assumed not to be a significant variable between and among systems, even for LEO systems. A spacecraft recurring dry mass cost of $77,000 per kilogram was assumed. Varying ratios of non-recurring to recurring costs were assumed, based on differences between and among possible systems, e.g., GEO, LEO, MEO. None of these assumptions appears to violate industry experience, although what is learned in building the new generation of satellites may provide a new dry mass cost average number.

Greater elevation angles were assumed to provide a higher quality of service (absent any other considerations) since signal fading and blockage are especially dependent on the elevation angle. [9] Actual practice will reveal how significant this assumption is. Higher link margins and multiple satellites may serve to overcome problems of potential signal fading and blocking. Indeed, Violet notes, in commenting on the study, "the major assumption in the system capacity simulations that could most significantly change the cost per billable minutes results includes the fading and power control assumptions." [10] The power control assumption made was that mobile satellite systems of any type could provide power control within 2 dBW to any user. Atmospheric variations, multi-path effects and uncertainty of antenna gain in an omni-directional hand-held make accurate power control difficult at best.

ISSUES BEYOND THE MIT STUDY

The "cost per billable minute" metric contains implicitly embedded assumptions that may or may not prove accurate in the real world. One of the most potentially troubling of these is the nature of regulatory constraints. It is assumed that regulation will not significantly hinder deployment and operation of systems, thus leaving all market predictions dependent only on user behavior and the technical capabilities of the system. In fact, regulatory policies (or the lack thereof) can operate in several ways. As we saw in North America before the ultimate licensing of MSAT, the regulators handled competition by forcing applicants to undergo a long series of negotiations before any system could ultimately be designed, built and brought to market. If a satellite system comes on line many years after the initial system builders estimate costs and profitability, it is likely the early estimates will be wrong.

Today, the LEO, MEO and HEO systems are, by definition global (or quasi-global). Therefore, they need permission to operate wherever they plan to serve users. Individual countries may deny licenses for unspecified periods of time, having a devastating effect on the "billable minutes" part of the equation. As spectrum becomes scarce, there may be policy decisions restricting or denying assignment of certain bands to mobile satellites, with these decisions coming either at the global (ITU), regional or national levels. Perhaps the most significant restriction may be the forced sharing and coordination of scarce spectrum, since sharing may take considerable time to negotiate and result in service restrictions for some systems.

Another variable that could force different considerations in building the "cost" part of the model is the availability and cost of launch vehicles. If we encounter a period of failures of several different launchers, significant delays and price increases may result. The MIT study assumed competition in the launch vehicle field. If and when competition decreases (although, today, there seem to be more countries trying to

enter this business), then these costs and associated risks may also rise.

While the MIT evaluation of the mobile satellite marketplace did take into consideration competition from cellular and PCS, no dramatic changes in any technology -- whether ground or space based -- were assumed. If we should see a breakthrough in how mobile users can be served, resulting in dramatically lower prices in equipment and service but with no loss of quality and reliability, market assumptions may need to be revised further.

A FINAL LOOK AT THE METRIC

As we have seen, while matching system costs against probable billable minutes seems to be a valid and intrinsically sound methodology, certain key variables require continuous review. Most notable among these are the size and probable behavior of the market. As a corollary, one could add the likely behavior of regulators and the results of regulatory policies. Technical assumptions will doubtless become modified with experience, and such assumptions will further need to be modified if user behavior differs from what is predicted (e.g., if users are willing to put up with certain types of gaps in service or service quality). Also, if dramatic changes ensue in the next few years -- whether in mobile technology, the launch vehicle market, or the world economy -- all of what seem sound assumptions in estimating costs and profitability may be in need of drastic revision.

The "cost per billable minute" metric has much to commend it as a tool to evaluate different system designs. Even with the assistance of such a tool, we cannot help but note that the creation of mobile satellite systems is a difficult and risky business! It is extremely challenging to design complex systems that may take anywhere from seven to ten years from conception to deployment and that have a design life-time of 12 or more years (which period may encompass two complete successive generations of a given LEO system, for example). Nonetheless, we are clearly entering an era when the demand for mobile communications is growing, and with that growth is coming a proliferation of systems to serve the demand. In the last one hundred years, it has been more common to underestimate telecommunications demand than to overestimate it. While this does not mean that everyone entering the market will be profitable, it does at least paint a scenario with room for competing systems, of differing designs, entering the market at different times -- which is exactly what the MIT study predicted.

ACKNOWLEDGMENT: This work was performed for the Jet Propulsion Laboratory, California Institute of Technology, spnsored by the National Aeronautics and Space Administration.

REFERENCES

[1] Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, MIT Mobile - Project Iris, Cambridge , MA, June, 1995, p. B3.

[2] Ibid., pp. 20-21.

[3] Ibid.

[4] Michael D. Violet, "The Development and Application of a Cost per Minute Metric for the Evaluation of Mobile Satellite Systems in a Limited-Growth Voice Communications Market", M.A. Thesis submitted to MIT, Cambridge, MA, September, 1995, p. 220.

[5] MIT Mobile - Project Iris, op. cit., p. 13.

[6] The Strategis Group, World Cellular and PCS Markets: 1997, Washington, DC, 1997.

[7] C.C. Gumbert, M.D. Violet, D.E. Hastings, W.M. Hollister, and R.R. Lovell, "Assessing Mobile Satellite Systems Using a Cost Per Billable Minute Metric," a paper presented at the AIAA conference on Communications Satellite Systems, Washington, DC, February, 1996, p. 12.

[8] MIT Mobile - Project Iris, op. cit., p. 15.

[9] Ibid., p. 23.

[10] Violet, op. cit., p. 235.

Session 13
Modulation, Coding and Multiple Access II

Session Chairperson—Dariush Divsalar, Jet Propulsion Laboratory, USA
Session Organizer—Marvin Simon, Jet Propulsion Laboratory, USA

Design and Performance of Precompensated Frequency Modulation (PFM) for use with a Quasi-Synchronous CDMA Return Link

S. Crozier[1], P. Guinand[1], B. Mazur[2]

[1]Communications Research Centre, 3701 Carling Ave., P.O. Box 11490, Station H, Ottawa, Ontario Canada, K2H 8S2; ph: 613-998-9262; fax: 613-998-1686; email: stewart.crozier@crc.doc.ca
[2]Square Peg Communications Inc., 3701 Carling Avenue, Building T3, P.O. Box 11490, Station H Ottawa, Ontario, Canada, K2H 8S2; ph: 613-820-7817; fax: 613-993-3122; email: http://www.crc.doc.ca/innovation/SquarePeg.html

Abstract

This paper presents the design and performance of a constant envelope modulation scheme, known as precompensated frequency modulation (PFM), appropriate for use with a quasi-synchronous code division multiple access (QS-CDMA) return link for future satellite based mobile and personal communications systems. The design is based on the results of a recent study performed for the European Space Agency (ESA), on the performance of candidate options for a third generation mobile satellite system [1]. This study involved detailed simulation and numerical analysis of a number of candidate forward and return link technologies. Only the design and performance of PFM, as it applies to a QS-CDMA return link, is presented in this paper. The QS-CDMA/TDMA frame structure and return link synchronization approach are described in more detail in a companion paper [2].

Introduction

Precompensated frequency modulation (PFM) is a new constant envelope modulation technique developed at the Communications Research Centre (CRC). As with any constant envelope modulation scheme, the transmit spectrum always falls off more gradually than a corresponding linear modulation scheme with root-Nyquist filtering. This implies increased adjacent channel interference (ACI). A tradeoff exists between ACI and bit error rate (BER) performance in additive white Gaussian noise (AWGN). The potential gain with constant envelope signals is in the use of a more power-efficient non-linear amplifier (such as a switching amplifier or a class-C amplifier operating in saturation) which can lead to a higher peak output power capability, longer battery life, smaller heat sinks, and lower cost. Satellite applications are particularly well suited to this type of approach due to the limited dynamic range of adjacent channel interferers. That is, the near-far problem is much more benign that that usually associated with many terrestrial applications. A strong line-of-sight (LOS) component and accurate power control also favor constant envelope approaches.

The PFM approach is similar to other continuous phase modulation (CPM) schemes in that multi-level data pulses are used to excite a pulse shaping filter followed by an FM modulator, or equivalent I-Q modulator [3-8]. There is an important difference with PFM, however. Non-Nyquist-3 pulse shape filtering and/or tight receive filtering usually results in phase errors in the signaling constellation observed at the output of the receive filter. These phase errors can be partially compensated for at the transmitter by adjusting the levels of the data pulses. The precompensation can be applied using a simple table look-up approach. The table is created using a novel non-linear algorithm [6], which takes into account the characteristics of both the transmit pulse shaping filter and the receive filter.

PFM with 4-level symbols and a modulation index of h=0.25 is denoted as PFM4 and is intended for use with a conventional $\pi/4$-QPSK receiver [7]. PFM with 2-level symbols and a modulation index of h=0.5 can be made to work with either a $\pi/2$-BPSK receiver or an OQPSK receiver [8]. The precompensation is different for the two cases. With a $\pi/2$-BPSK receiver the PFM signal is denoted as PFM2. With an OQPSK receiver the PFM signal is denoted as PFMO. As one might expect, PFM2 is the least bandwidth efficient of the above three schemes. PFMO is generally more bandwidth efficient than PFM4, but it can not be used in a (QS-) CDMA application, using (quasi-) orthogonal spreading codes, without requiring phase coherency between users. Thus, our primary interest here is in the performance of PFM4 since it is more appropriate for the QS-CDMA return link of the ESA system [1, 2].

As with conventional $\pi/4$-QPSK modulators, differential encoding is optional. The differential encoding inherent in any CPM or PFM scheme is easily eliminated by simply precoding the data. Coherent detection (or coherent despreading in the case of CDMA signals) is often the

preferred choice for PFM4. For the purposes of analysis, ideal coherent detection is assumed for the BER results presented herein.

This paper presents the bit error rate and spectral performance results obtained for PFM4 in the context of the QS-CDMA return link of the ESA system. It is shown that PFM4 can offer both good BER performance and bandwidth efficiency for this application.

The degradation when compared to the performance of a linear $\pi/4$-QPSK modem is typically no more than 0.7 dB, and less on average, depending on the loading in the adjacent and same frequency channels. Compensating for this degradation by increasing transmit power is more than offset by the increased power efficiency of a saturated amplifier, as compared to a linear amplifier with back-off. Overall power efficiency gains on the order of 2 to 6 dB can be expected, depending on the amplifier technology.

System Model

PFM4 is a constant envelope modulation scheme designed to work with a conventional $\pi/4$-QPSK receiver. Figure 1 shows the signaling constellation for a $\pi/4$-QPSK system. Assuming ideal root-Nyquist transmit and receive filters, perfect symbol timing, perfect carrier frequency and phase recovery, and no noise, this is the signaling constellation that would be observed after the receive filter. There are 8 possible phase points. Only 1 of 4 phases is possible per symbol period. For example, if the square points are valid for the current symbol then only the circular points are valid for the next symbol. The only difference between $\pi/4$-QPSK and QPSK is that each complex symbol is rotated by an additional $\pi/4$ radians relative to the last symbol. Rotating the symbols in this manner eliminates the transitions through the origin and reduces the amplitude fluctuations of the filtered signal.

Figure 2 shows the complex baseband equivalent system model under consideration. The transmit chip bits[1] are passed to the PFM4 modulator, which generates a constant envelope signal. The non-linear amplifier does not introduce any further distortion because the PFM4 signal is constant envelope. The signal is then corrupted by AWGN, and ACI from 2 adjacent channels. The same modulation format is assumed for all three channels. The channel spacing and ACI power levels are parameters that influence the design of the PFM4 modulator. The signal is then passed through a receive filter and detected in an ideal coherent $\pi/4$-QPSK detector.

[1] The term "chip bit" is used to refer to the channel bits of the spread CDMA signal. Note that the chip bit rate is twice that of the chip symbol rate when a PFM4 or $\pi/4$-QPSK type of modulation is being used.

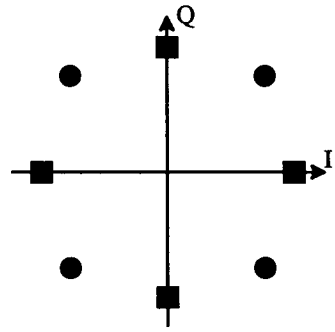


Fig. 1: Signaling constellation for $\pi/4$-QPSK and PFM4.

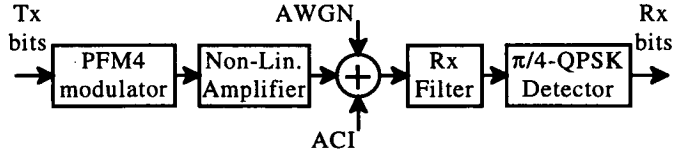


Fig. 2: System model used for PFM4 design.

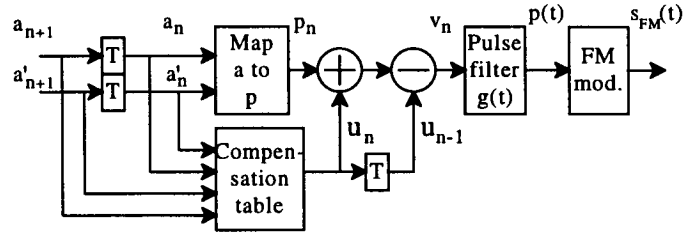


Fig. 3: PFM4 modulator with a 4-bit compensation table.

Figure 3 shows the PFM4 modulator in more detail. Except for the compensation, the model represents a standard 4-level FM modulator with modulation index h=0.25. The I and Q transmit bits, a_n and a'_n, are mapped in pairs to 4-level pulses, p_n, with amplitudes from the set {-3, -1, +1, +3}. These four amplitude levels result in corresponding phase changes of $\{-3\pi/4, -\pi/4, +\pi/4, +3\pi/4\}$ at the output of the FM modulator. The signaling constellation is the same as that for $\pi/4$-QPSK, as shown in Figure 1. Differential encoding is optional. The differential encoding inherent in the PFM4 approach is easily eliminated by simply precoding the data.

The pulse shaping filter, $g(t)$, is used at the output of the pulse amplitude modulator (PAM) to help determine the overall spectrum of the FM modulated signal. There is no need for this pulse response to satisfy the Nyquist-3 property [3]. The type and length of this filter, used for the results presented in this paper, are specified later. The resulting FM signal is constant envelope.

The receiver used for PFM4 can be identical to that used for a $\pi/4$-QPSK modulator, with root-Nyquist receive filtering. Such tight receive filtering normally causes severe distortion to the FM signaling constellation. The phase errors can be partially compensated for at the transmitter by adjusting the levels of the data pulses. This precompensation can be applied using a simple table look-up approach. The table is created using a novel non-linear algorithm which takes into account the characteristics of both the transmit pulse shaping filter and the receive filter. The details of this algorithm can be found in [6,7,8]. The amount of precompensation can be controlled to optimize the BER performance with ACI for specific channel spacings and adjacent channel power levels.

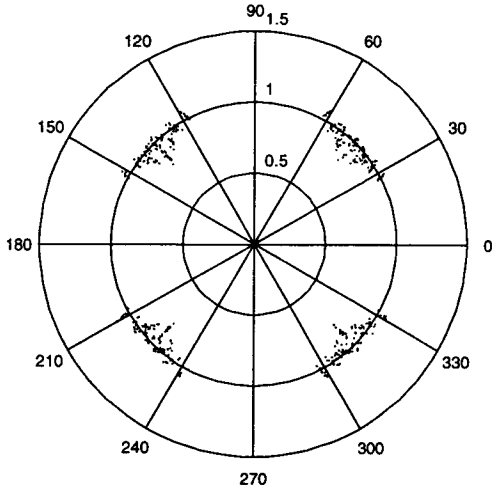


Fig. 4: Scatter diagram for PFM4 case A.

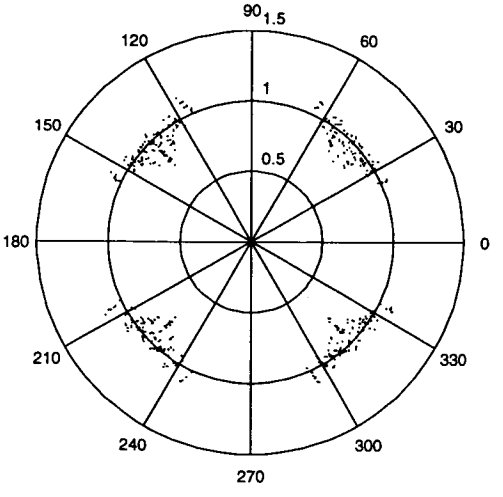


Fig. 5: Scatter diagram for PFM4 case B.

Table 1 lists the parameter settings for 2 different cases, denoted as A and B, corresponding to channel spacings of 0.75 and 0.70 chip bit rates, respectively. The theoretical minimum signaling bandwidth for $\pi/4$-QPSK, with 0% roll-off, is 0.50 chip bit rates. Case A corresponds to an excess bandwidth of 50%. Case B corresponds to an excess bandwidth of 40%. The pulse shaping filter used in the FM modulator was a time-domain raised-cosine-squared pulse of duration nT, where T is the chip bit period and n is the number of chip bit periods, as given in Table 1.

Table 1: Parameter settings for PFM4 cases A and B.

PFM4 Case	Channel Spacing (chip bit rates)	Pulse Shaping Filter Length (chip bit periods)	Rx filter roll-off (0.0 to 1.0)
A	0.75	4.0	0.50
B	0.70	4.5	0.40

An ideal coherent $\pi/4$-QPSK receiver with frequency-domain root-raised-cosine (RRC) receive filtering was assumed. The 2-sided 3 dB bandwidth was set to the 0.50 chip bit rate value in both cases. Thus, the receive filter was root-Nyquist. The receive filter roll-offs are as listed in Table 1 for each case. The precompensation table was optimized in each case to give the best BER performance at a chip bit E_b/N_0 ratio of -13 dB and an ACI level of +18 dB, relative to the desired channel, with the specified channel spacing. This is based on the assumption of an information E_b/N_0 of 5 dB, with rate 1/2 coding, and a spreading factor of 31 for the QS-CDMA return link [1,2]. The ACI level is based on the assumption of 64 equally powerful adjacent channel interferers, corresponding to 2 fully loaded adjacent QS-CDMA channels. An 8-bit precompensation table was used for all the results presented here.

Performance Results

Scatter Diagrams

Figures 4 and 5 show the scatter diagrams for PFM4 cases A and B respectively, as listed in Table 1. The scatter diagrams were obtained at the output of the receive filter using the decision samples for the even symbols only. The scatter diagram for the odd symbols is simply rotated by $\pi/4$ radians. Ideally, each scatter diagram should look like the signaling constellation shown in Figure 1, for the 4 circular points only. It can be seen that the distortion increases as the channel spacing is decreased. This is expected, and is partly due to the precompensation algorithm trading off ISI for spectral efficiency. Note that the ISI shown in these scatter diagrams is typically very small compared to the other sources of noise and interference in a CDMA application.

Power Spectrums and ACI

Figure 6 shows the transmit power spectrums for the two PFM4 cases listed in Table 1. These power spectrums were obtained using simulation and FFT-based techniques. The spectrums of hard-limited (HL) and linear 50% roll-off

π/4-QPSK are also included for comparison purposes. The frequency scale is normalized to the chip bit rate for convenience. Note that the power spectrum for case B is narrower than that for case A. This is to be expected since the close channel spacing for case B makes it more sensitive to ACI.

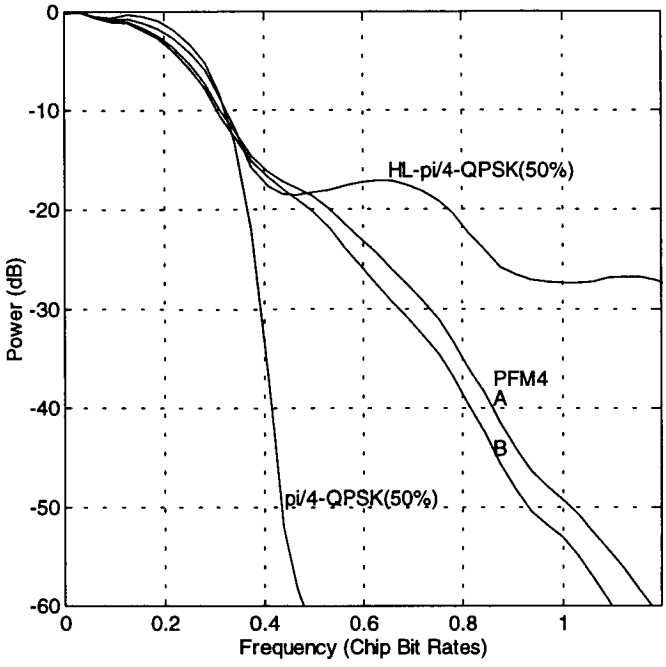


Fig. 6: Power spectrums for PFM4 cases A and B, and hard-limited and linear 50% roll-off π/4-QPSK.

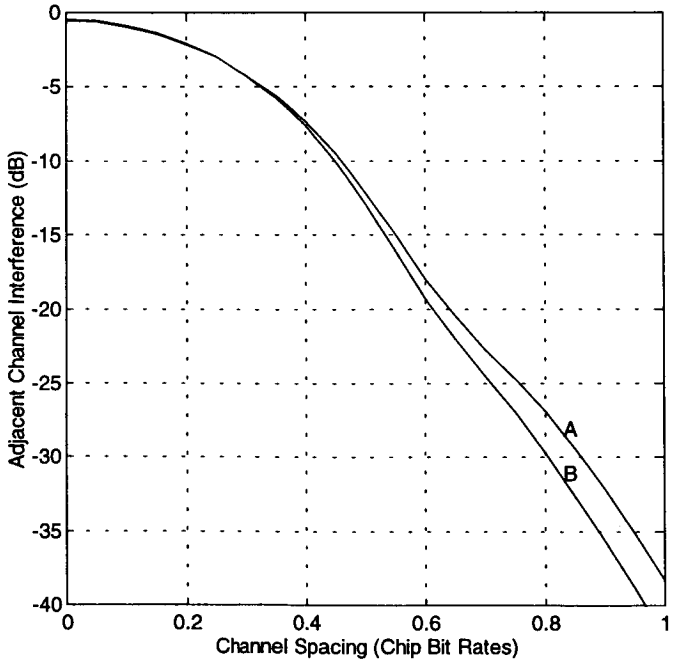


Fig. 7: ACI for PFM4 cases A and B.

Figure 7 shows the corresponding ACI results for PFM4 cases A and B. Here the ACI is defined as the ratio of the total power passed by the unit energy receive filter when offset by the channel spacing specified on the x-axis, and the total power of the PFM4 transmit signal. In other words, these ACI results indicate the relative interference power caused by one equal power adjacent channel interferer. Note that the ACI for a channel spacing of 0.70 chip bit rates (40% excess bandwidth) varies from -22.7 dB to -24.5 dB for cases A and B, respectively.

BER Results Without ACI

The BER performance results for cases A and B were computed assuming an ideal coherent π/4-QPSK receiver, without ACI. The theoretical BER performance for an ideal π/4-QPSK transmitter with a linear amplifier (i.e. ideal antipodal signaling) was used for comparison purposes [5]. Figure 8 shows the amount of degradation relative to ideal. At a chip bit E_b/N_0 ratio of -13 dB, which is the assumed nominal operating point for the QS-CDMA return link, the degradation from ideal varies from 0.37 dB for case A to 0.42 dB for case B. It can also be seen that the degradation in dB is only a mild function of the chip bit E_b/N_0 ratio, around this operating point.

BER Results With ACI

Figure 9 shows the amount of degradation in dB obtained with an ACI level of +18 dB, corresponding to 2 fully loaded adjacent QS-CDMA channels. The channel spacing for each case is as listed in Table 1.

As can be seen, at a chip bit E_b/N_0 ratio of -13 dB, the degradation from ideal varies from 0.47 dB for case A to 0.53 dB for case B. These are excellent results for constant envelope signaling, considering that the channel spacing for case B is only 0.70 chip bit rates.

Example Performance for a QS-CDMA Return Link

The previous PFM4 results are directly applicable to systems using asynchronous CDMA. The performance with quasi-synchronous CDMA (QS-CDMA) is slightly different. The design of a QS-CDMA return link, using Preferentially-Phased Gold Codes (PPGCs), is described in [1,2]. A critical issue is the performance with chip timing errors between the users. Selected performance results are presented here for π/4-QPSK and PFM4 modulation.

There are three ways in which chip timing error causes degradation to the BER performance of QS-CDMA. They are: a) energy loss due to signal sampling not being performed at the optimum time; b) increase in intersymbol interference (ISI) of the desired signal because the sampling is not being performed at the optimum time; and c) increase in the interference from other users because of the increase

in the cross-correlations between codes when they are time offset. The third source of degradation is worse for PFM4, compared to ideal π/4-QPSK, due to the inherent ISI designed into the modulation, even with perfect timing.

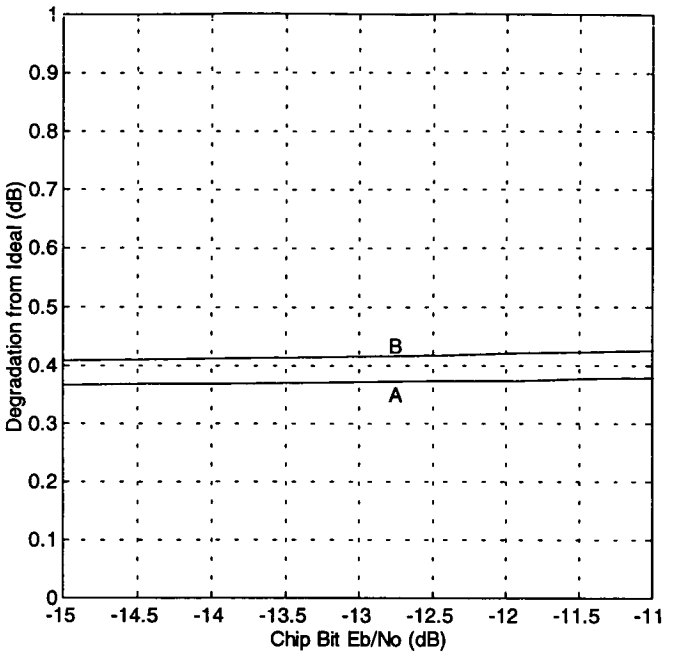


Fig. 8: Degradation from ideal for PFM4 cases A and B, without ACI.

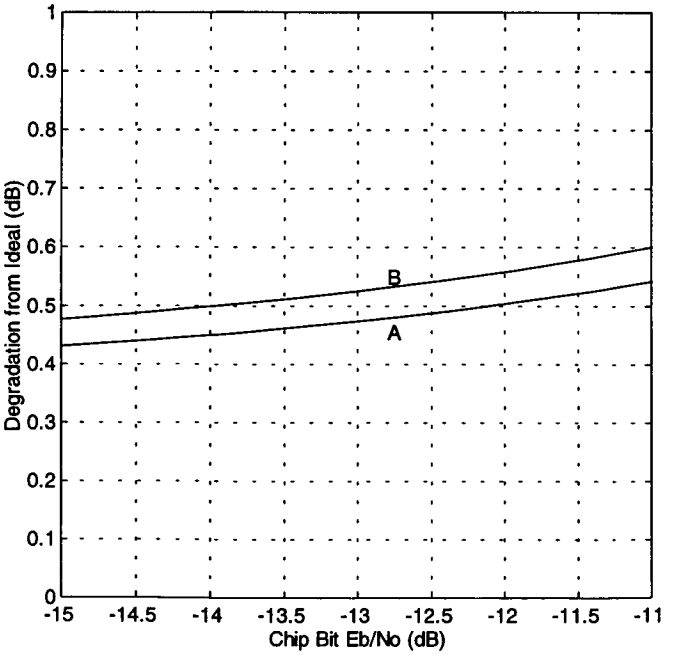


Fig. 9: Degradation from ideal for PFM4 cases A and B, with ACI.

The approach used to evaluate the sensitivity to timing errors was to conduct simulations to evaluate for fixed timing errors the three sources of degradation mentioned above. These were then combined to obtain estimates of the degradation in the effective E_b/N_0 experienced by a desired user, where E_b now refers to the energy in a coded channel bit. An actual channel E_b/N_0 of 2 dB was used, which corresponds approximately to a BER of 10^{-3} for a constraint length 9, rate ½, code; which was one of the coding strategies of interest in [1]. The results are based on using Gold codes of length 31 with 32 users, i.e. a fully loaded system. Root-raised-cosine 50% roll-off filtering was used for the π/4-QPSK results. The same receiver was used for the PFM4 case, as indicated by option A in Table 1.

Figure 10 shows the results for ideal π/4-QPSK. The desired user was assumed to have a fixed timing offset and the other users were assumed to have Gaussian distributed timing errors with zero mean and standard deviation σ, measured in chip symbol periods. With no timing errors for all the users, the degradation from E_b/N_0=2 dB is about 0.2 dB. This is mainly due to the non-zero cross-correlations between Gold codes. With a fixed timing error of 0.1 chip symbol periods for the desired user and σ=1/16 for the other users, the degradation is about 0.5 dB.

Figure 11 shows the corresponding results for PFM4. With no timing errors for all the users, the degradation from E_b/N_0=2 dB is about 0.75 dB. This is about 0.55 dB worse than the corresponding ideal π/4-QPSK case. With a fixed timing error of 0.1 chip symbol periods for the desired user and σ=1/16 for the other users, the degradation is observed to be about 1.1 dB. This is about 0.6 dB worse than the corresponding π/4-QPSK case.

The above QS-CDMA results do not include the effects of ACI. The additional degradation for PFM4 with ACI is easily derived from the previous results, and is about 0.1 dB for two fully loaded adjacent frequency channels with 50% channel spacing. Thus, the total degradation with ACI is no more than 0.7 dB. Note that, while the results for PFM4 may look slightly worse than the π/4-QPSK alternative, PFM4 modulation allows saturated operation of terminal amplifiers, which provides significant compensatory advantages. The degradation with partial loading is also significantly less.

Conclusions

PFM4 is a constant envelope modulation scheme designed to work with a conventional π/4-QPSK receiver. It was shown that PFM4 offers both good BER performance and bandwidth efficiency. Thus, PFM4 is an attractive option for the QS-CDMA return link of the ESA system.

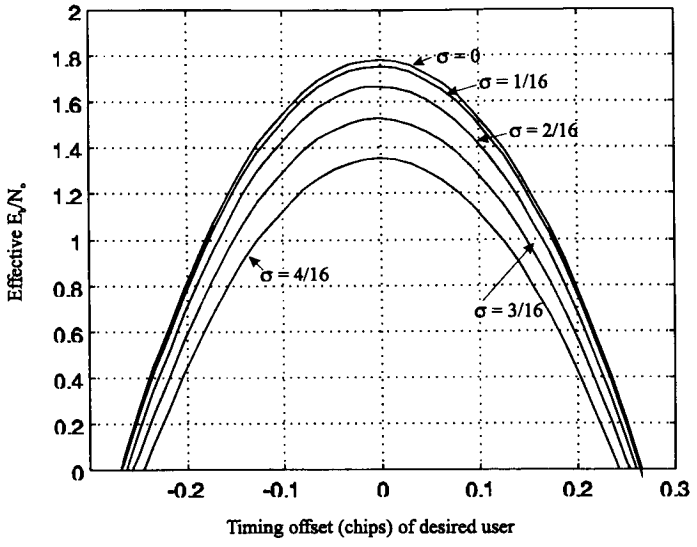


Fig. 10: Effective despread channel E_b/N_0 for fixed timing offsets of the desired user and random timing offsets for the other users; ideal $\pi/4$-QPSK, fully loaded with 32 users, actual channel E_b/N_0 of 2 dB.

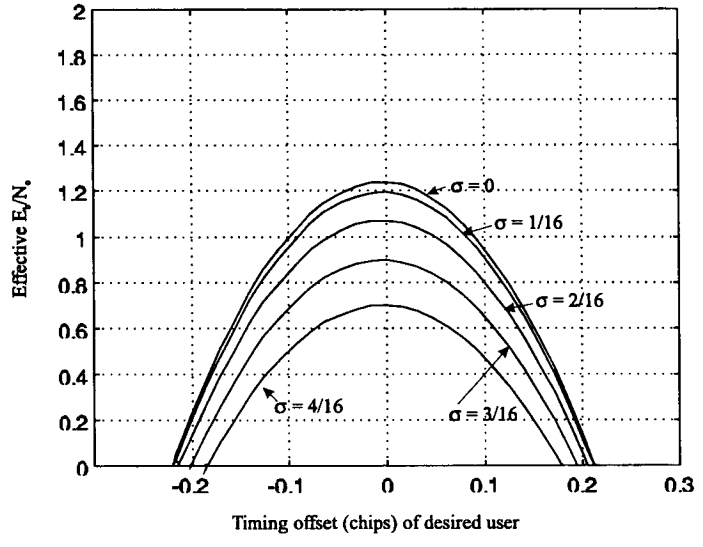


Fig. 11: Effective despread channel E_b/N_0 for fixed timing offsets of the desired user and random timing offsets for the other users; PFM4 case A, fully loaded with 32 users, actual channel E_b/N_0 of 2 dB.

Based on the above results, the BER degradation of PFM4, compared to linear $\pi/4$-QPSK, is typically no more 0.5 dB for asynchronous CDMA, and 0.7 dB for QS-CDMA. The degradation is less on average, depending on the loading in the adjacent and same frequency channels. This degradation is more than offset by the increased power efficiency of a saturated amplifier, as compared to a linear amplifier with back-off. Overall gains on the order of 2 to 6 dB can be expected, depending on the amplifier technology.

It should also be mentioned that linear amplifiers are never perfectly linear. They all introduce some AM-to-AM and AM-to-PM distortion, the severity of which depends on the backoff level. This distortion will degrade the performance of linear modulation schemes, but will not degrade further the performance of constant envelope modulation schemes, such as PFM4. Thus, when compared to the performance of linear modulation schemes with practical linear amplifiers, the degradation associated with PFM4 is even less than that stated above.

References

[1] B. Lyons et al., "A Transmission System for Future Satellite Mobile and Personal Communications", Final Report, ESTEC Contract No. 11515/95/NL/US, 1997.

[2] S. Crozier, B. Mazur, L. Erup, "Design of a Synchronous CDM Forward Link and a Quasi-Synchronous CDMA Return Link for Future Satellite Based Mobile and Personal Communication Systems", International Mobile Satellite Conference (IMSC'97), Pasadena, California, June 16-18, 1997.

[3] T. Aulin and C-E Sundberg, "Continuous Phase Modulation - Part I: Full Response Signaling - Part II: Partial Response Signaling", IEEE Transactions on Communications, Vol. COM-29, No. 3, pp. 196-225, March 1981.

[4] B. Sayar and S. Pasupathy, "Nyquist 3 Pulse Shaping in Continuous Phase Modulation", IEEE Transactions on Communications, Vol. COM-35, No. 1, pp. 57-67, January 1987.

[5] S. Crozier, "A Comparison of Three QPSK Type Modulation Schemes for Mobile Satellite SCPC Voice and Data Services", 16-th Biennial Symposium on Communications, Queen's University, Kingston, Ontario, Canada, paper A.2.7, May 27-29, 1992.

[6] U.S. Patent Application 08/300,189, S. Crozier, "Precompensated Frequency Modulation (PFM)", Filed September 22, 1994. Continuation-In-Part Filed March 22, 1996.

[7] S. Crozier, "Precompensated Frequency Modulation (PFM) designed for $\pi/4$-QPSK Receivers", 18-th Biennial Symposium on Communications, Queen's University, Kingston, Ontario, Canada, pp. 343-346, June 2-5, 1996.

[8] S. Crozier, "Precompensated Frequency Modulation (PFM) Designed for $\pi/2$-BPSK and Offset-QPSK Receivers", The 8-th International Conference on Wireless Communications (Wireless'96), Calgary, Alberta, Canada, pp. 155-170, July 8-10, 1996.

Design of a Synchronous CDM Forward Link and a Quasi-Synchronous CDMA Return Link for Future Satellite Based Mobile and Personal Communication Systems

S. Crozier[1], B. Mazur[2], L. Erup[3]

[1]Communications Research Centre, 3701 Carling Ave., P.O. Box 11490, Station H, Ottawa, Ontario, Canada, K2H 8S2
ph: 613-998-9262; fax: 613-998-1686; email: stewart.crozier@crc.doc.ca
[2]Square Peg Communications Inc., 3701 Carling Avenue, Building T3, P.O. Box 11490, Station H, Ottawa, Ontario, Canada
K2H 8S2; ph: 613-820-7817; fax: 613-993-3122; email: http://www.crc.doc.ca/innovation/SquarePeg.html
[3]Communications Services Division, Directorate of Telecommunications Programmes, European Space Agency,
European Space Research and Technology Centre, ESTEC/TST, P.O. Box 299, Noordwijk
NL 2200 AG The Netherlands; ph: +31 71 5653161; fax: +31 71 5654598; email: lars@t.estec.esa.nl

Abstract

This paper describes the high-level design of CDM/CDMA based forward and return links for future satellite based mobile and personal communications systems. The design is derived from the results of a recent study performed for the European Space Agency (ESA), on the performance of candidate options for a third generation mobile satellite system [1]. This study involved detailed simulation and numerical analysis of a number of candidate forward and return link technologies. Only the CDM/CDMA approaches are presented in this paper. The forward link uses a synchronous CDM intra-beam/asynchronous CDM inter-beam approach. The return link uses a quasi-synchronous CDMA intra-beam/asynchronous CDMA inter-beam approach. With the given assumptions, these approaches were determined to provide the highest link capacities for a multi-beam scenario, and appear feasible to implement.

Introduction

Satellite systems providing personal communications services to mobile users have been operating for some time (e.g. Inmarsat, MSAT). A second generation of these systems that extends services to hand-held terminals is currently being implemented (e.g. ICO, Iridium, Globalstar). Concurrently, standardization efforts for the next generation of mobile and personal communications systems is in progress. These future systems are intended to provide a wide range of telecommunications services and integrate a number of technologies that are currently separate (e.g. cordless phones, cellular and microcellular phones, and satellite communications). The third generation of mobile satellite communications systems is therefore envisioned to be an integrated component of this all-encompassing personal communications network (PCN).

This paper presents part of ESA's research into the missions, systems, and technology of this next-generation PCN satellite component. Through previous work, ESA identified highly elliptical orbit (HEO) satellite constellations as particularly well suited to service provision to intermediate-latitude regions, due to the high elevation angles which can be achieved, and a direct line-of-sight link with resulting low levels of shadowing and multipath most of the time. In keeping with the objectives of the next-generation PCN system, emphasis has been placed on providing a wide range of services, including voice and video telephony, high speed data, and one-way paging. The bit rates supported range from 1 to 64 kbps.

System Overview

The system architecture is as follows. One or several base stations (BSs) communicate with a potentially large population of user terminals (UTs) through a constellation of satellites in HEO orbit. Feeder links connect the BSs to the satellites, and mobile links connect the satellites to the UTs. There are no direct UT to UT connections. System performance is expected to be largely determined by the design of the mobile links and the associated transmitters and receivers; the feeder links are of secondary importance in this regard.

The mobile links are assumed to operate in the ~ 2 GHz radio band, with bandwidth allocated in ~ 2 MHz increments, and separate frequency bands for the forward and return directions. A fielded system might utilize several, typically contiguous, ~ 2 MHz bandwidth increments, depending on the capacity required and spectrum allocated.

Satellites with quasi-linear, transparent, transponders and antennas having several tens of spotbeams (e.g. 37 or 61) are assumed. Spotbeams from two satellites are considered distinct even if their footprints happen to coincide. The satellites might be in one of several candidate constellations, ranging from that typified by the elliptical portion of the Ellipso Borealis constellation (3 hour period, 7846 km apogee, 519 km perigee, 20° edge of coverage) to that typified by the MHEO(8) orbit (8 hour period, 26784 km apogee, 1000 km perigee, 40° edge of coverage).

The BSs are fixed ground stations with sophisticated processing capabilities, that communicate with the UTs through the satellite links, and that provide interconnections to terrestrial communications systems (and possibly other satellite communications systems). The system design generally places most of the processing burden on the BSs in order to minimize the cost and complexity of the UTs.

Provision has been made for several types of UTs:

- Vehicular terminals having a *G/T* of about -22.0 dB/K and access to all services. Vehicular terminals cannot easily relocate to avoid shadowing of the direct line-of-sight link(s) to the satellite(s).
- Hand-held terminals having a *G/T* of about -23.5 dB/K and access to toll quality telephony and lower speed services only. Time division duplex (TDD) operation is applied to obviate the need for a diplexer and its associated loss. Hand-held terminals will not operate in heavily shadowed locations.
- Transportable and fixed terminals having a *G/T* of about -20.5 dB/K and access to all services. These terminals can be situated to eliminate shadowing.

Forward link transmissions from each BS are modulated using quadrature phase shift keying (QPSK). Return link transmissions from a given UT are modulated using either: a) $\pi/4$-shifted QPSK ($\pi/4$-QPSK), or b) precompensated frequency modulation (PFM4). The latter can also be demodulated using a $\pi/4$-QPSK receiver and is intended for use with saturated amplifiers in low-cost UTs (e.g. hand-held terminals). Further details on the design and performance of PFM4 can be found in [2]. All forward and return link transmissions are coded using a rate ½, constraint length 9, convolutional code. Justification for these modulation and coding design choices is given in [1, 3].

Multiple Access Signal Structure

Forward Link

Transmissions from the BSs to the UTs within any given spotbeam will share the forward link using (quasi-) orthogonal code division multiplexing (O-CDM) with a time division multiplexing (TDM) overlay. This is conceptually illustrated in Fig. 1. All transmissions take place at the same nominal center frequency and use an allocated bandwidth of ~ 2 MHz. If multiple ~ 2 MHz bandwidth slices are available, then several O-CDM/TDM signals can be used.

The O-CDM/TDM signal time axis is divided into a number of equal-duration frames, and then subdivided into smaller time periods. The smallest allocated unit of time is called a slot, and the unit of signal transmitted during a slot is called a burst. Datagram traffic (e.g. asynchronous messages) is sent as part of the broadcast portion of the frame. Stream-mode traffic (e.g. telephony or continuous data) is transmitted in the remaining portion of the frame.

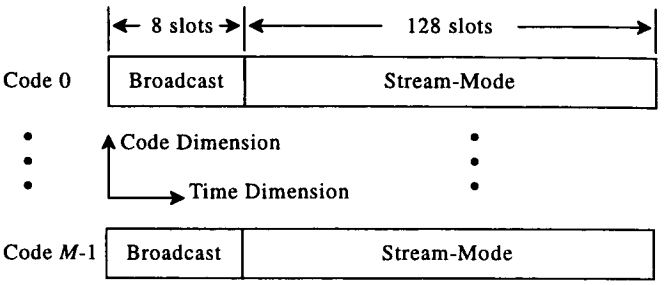


Fig. 1: Forward link frames for M different O-CDM codes.

In addition to datagram traffic, the broadcast time period is used to transmit a synchronization word (SW) and various system data, as illustrated in Fig. 2. The SW is used for system synchronization, as discussed later. The system data provides out-of-band signaling information.

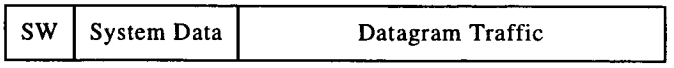


Fig. 2: Structure of forward link broadcast time period.

Preferentially-phased Gold codes (PPGCs) having a code period of 31 chips are used to perform the code division multiplexing [4]. These codes are generated by cyclic shifting and modulo-2 adding a pair of m-sequences. For any given pair of m-sequences there are a maximum of 32 such codes (implying $M \leq 32$ in Fig. 1), and only one m-sequence pair can be chosen for any given O-CDM multiplex. These codes have the property that their mutual cross-correlation is small (i.e. they are quasi-orthogonal) if they are aligned in time to within a small fraction of a chip period. This in turn results in a quasi-minimization of intra-beam system self-noise. With a maximum terminal data rate of 64 kbps, the code period of 31 chips was chosen to give the desired spread bandwidth.

Each BS/UT pair communicating through the same spotbeam will be assigned a distinct combination of time slot(s) and O-CDM code(s), i.e. a different channel assignment. In order to achieve quasi-orthogonality among the O-CDM codes, the BS(s) time and frequency synchronize all forward link transmissions within each spotbeam.

Forward link stream-mode data is formatted into N equal-duration blocks per frame, where N is a fixed value ranging from 1 to 64 (1 to 64 kbps data rate). As part of the channel coding and interleaving strategy, each coded and interleaved forward link stream-mode data block is split in half and transmitted as two bursts in the frame. These bursts are separated in time to improve performance in fading.

The same set of O-CDM codes are reused in all the spotbeams (of all the satellites), resulting in inter-beam interference. Randomization codes are overlaid over each frame transmitted to a given beam so that spill-over transmissions to other beams appear noise-like following

derandomization and despreading by the UTs in those beams. The randomization codes are applied over the entire frame with the exception of the SW which appears at the beginning of each frame. The phase of each randomization code is reset at the beginning of each frame. This permits a single pair of randomization codes (one for the in-phase channel and another for the quadrature channel) to be shared by all the spotbeams and satellites in the system. A UT differentiates among multiple spotbeams and satellites in view as follows:

- Same satellite transmissions to adjacent spotbeams are time offset from one another by 8, 16, or 24 chip durations. A 1-in-4 time offset reuse pattern is applied.
- Different satellite transmissions are time offset from one another by a large fraction of a frame duration (e.g. ½ or ¼ of a frame, depending on the constellation and the number of satellites that can simultaneously be in view).

Return Link

In like manner to the forward link, the return link time axis is divided into a number of equal duration frames, and then subdivided into smaller time periods. The smallest allocated unit of time is called a slot. Signals are transmitted as a series of bursts, where the duration of each burst is 1 slot.

Stream-mode traffic transmissions to the BSs from UTs in a spotbeam will share the return link using quasi-synchronous code division multiple access (QS-CDMA) with a time division multiple access (TDMA) overlay. In addition to this, random access transmissions, comprising datagram traffic and special synchronization bursts, share a separate random access time period in each frame.

The return link frame structure is conceptually illustrated in Fig. 3. All transmissions take place at the same nominal center frequency and use the ~ 2 MHz allocated bandwidth. Return link frames from a given spotbeam are given a fixed time offset (as measured by the UTs) relative to the forward link frames associated with that same spotbeam.

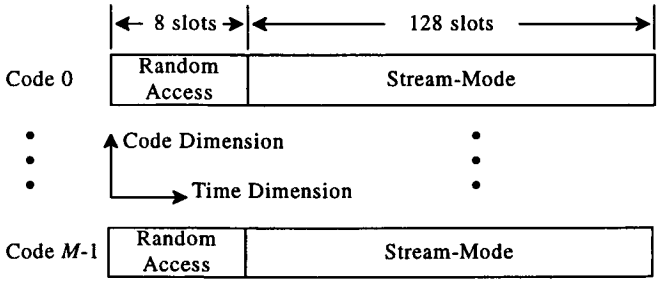


Fig. 3: Return link frames for M different QS-CDMA codes.

The return link uses the same set of period 31 PPGCs that are used on the forward link. There are a maximum of 32 such codes (implying $M \leq 32$ in Fig. 3). Again, with a maximum terminal data rate of 64 kbps, the code period of 31 chips was chosen to give the desired spread bandwidth.

Stream-Mode Traffic

Each UT/BS pair communicating through the same spotbeam will be assigned a distinct combination of time slots and QS-CDMA codes, i.e. a different channel assignment. In order to achieve quasi-orthogonality among the QS-CDMA codes, and thereby minimize system intra-beam self-noise, the UTs approximately time and frequency synchronize all return link transmissions within each spotbeam (quasi-synchronous operation).

Like the forward link, return link stream-mode data is formatted into a series of N equal-duration blocks per frame, where N is a fixed value ranging from 1 to 64 (1 to 64 kbps data rate). Each of these coded and interleaved blocks is split in half and transmitted as two separate stream-mode data bursts in the frame.

Like the forward link, the return link bandwidth and QS-CDMA codes are reused in all the spotbeams (of all the satellites), resulting in inter-beam interference. This interference is randomized, and thereby rendered equivalent to Gaussian noise having the same aggregate power, by overlaying randomization codes on each return link burst. System data encoded in the broadcast portion of the forward link identifies the randomization codes to be used on the corresponding return link, and these designated randomization codes are applied over each stream-mode data burst in the return link. The phase of each randomization code is reset at the beginning of each burst. Each satellite only requires 4 pairs (I and Q) of time offset randomization codes that are assigned using a 1-in-4 reuse pattern to its return link spotbeams (similar to the forward link). The randomization codes used by any satellite are different from those used by any other satellite. Transmissions from a given spotbeam will be derandomized by the BSs receiving that spotbeam, but spillover transmissions received from adjacent spotbeams or satellites will remain random.

Random Access Transmissions

Three types of bursts share the return link random access time period: asynchronous datagram traffic bursts, synchronization acquisition bursts, and synchronization reacquisition bursts (together designated random access bursts). The partitioning of this time period for the three different uses is illustrated in Fig. 4.

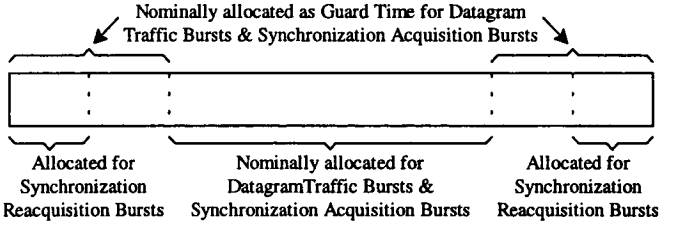


Fig. 4: Structure of return link random access time period.

Asynchronous datagram traffic bursts are used to transmit datagram data, while synchronization acquisition bursts are transmitted by the UTs to allow the BSs to accurately estimate their return link frequency and timing errors as part of a closed-loop control system. UTs acquire forward link frame timing from the SWs and use this to estimate return link frame timing. This estimate presumes that the UT is at the center of the selected spotbeam. To allow for the corresponding timing uncertainty, UTs nominally transmit their asynchronous datagrams and synchronization acquisition bursts into the center portion of what they estimate to be the random access time period. The first and last portions of the random access time period constitute adequate guard times in this context.

Asynchronous datagrams and synchronization acquisition bursts are transmitted using the unslotted Aloha protocol, because they cannot in general be scheduled to avoid mutual interference. Asynchronous datagram traffic bursts and synchronization acquisition bursts may overlap and cause mutual noise-like interference because their timing cannot in general be controlled. However, they do not collide in the Aloha sense (i.e. mutually cause high-level interference resulting in the loss of all overlapping bursts) unless they time align to within a chip period (which is very unlikely).

Synchronization reacquisition bursts are used by the UTs to reacquire return link frequency and timing if synchronization has been lost (unless an excessive amount of time has elapsed, in which case UTs must revert to the initial acquisition mode and use synchronization acquisition bursts to reacquire the return link and reestablish the call). On the expectation that UT timing is only slightly in error, synchronization reacquisition bursts are transmitted at the leading and trailing edges of the random access time period. While these transmission windows overlap the nominal guard times provided for asynchronous datagram traffic bursts and synchronization acquisition bursts, they will be lightly loaded by such bursts, thereby lowering the interference environment for the reacquisition bursts.

Random access bursts can be communicated over adjacent spotbeams of the same satellite, if they happen to be transmitted from the spotbeam overlap region. This redundancy can be exploited by the BS(s) to improve the reliability with which these bursts are received.

Synchronization reacquisition bursts within the same spotbeam do not interfere with one another because they are scheduled by the BS(s). Such bursts from adjacent spotbeams of the same satellite mutually interfere, but do not collide, because the return link frames of adjacent spotbeams are time offset by at least 8 chips, whereas reacquisition burst timing error will not exceed a few chips.

All asynchronous datagram traffic bursts, synchronization acquisition bursts, and synchronization reacquisition bursts in the system have the same structure, as illustrated in Fig. 5. The same SW as used for the forward link is transmitted at the beginning of each of these return link bursts. This is followed by an information field and a cyclic redundancy check (CRC) field. The CRC and information fields of each burst are error correction coded and interleaved as a single block. Each burst is then spread using the same pair of QS-CDMA codes in phase quadrature. Lastly, the portion of each burst following the SW is randomized using the same pair of randomization codes in phase quadrature. The same pair of randomization codes is used in all bursts in all spotbeams for all satellites. However, these randomization codes are different from any of those that are used for return link stream-mode data bursts.

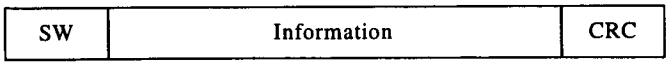


Fig. 5: Structure of a return link random access burst

Synchronization

The entire system is synchronized to a common time and frequency reference replicated in each BS. This reference corresponds to the nominal timing and frequency of forward link signals passing through the constellation of satellites.

A hierarchy of four synchronization processes are used to achieve and maintain this synchronization:

- BSs synchronize their local time and frequency references to the common system reference.
- BSs offset the transmit times, chip rates, and frequencies of their uplinks so that the signals arrive at the intended satellite(s) in synchronism with system time and frequency, aside from the fixed time offsets imposed between spotbeams and satellites as part of the forward link signal multiplex, and aside from frequency precompensation applied to zero satellite downlink Doppler at each beam center.
- UTs acquire forward link synchronization using the periodic SWs.
- UTs execute a closed loop control process in cooperation with the BSs to establish and maintain return link synchronization.

At the two highest levels of the synchronization hierarchy, a BS is first slaved to the common system reference, and then synthesizes a local system reference to be provided to its UTs. These steps can be executed in a variety of ways. The focus in the following discussion is on the two lowest levels of the synchronization hierarchy, where UTs first acquire the forward link using the periodic SWs, and then use the forward link reference in conjunction with feedback from the BSs to synchronize their return links.

Forward Link

UTs rely on the SW to acquire forward link time and frequency. The same 63-symbol SW is used throughout the system, and consists of one pair of unique words (UWs) transmitted in phase quadrature by the BSs using one pair of O-CDM codes. This SW is transmitted at the beginning of the forward link broadcast time period in each frame in each spotbeam. While the SW is being transmitted in a given spotbeam using the designated O-CDM codes, there is no transmission using the other 30 O-CDM codes.

The UTs use the SW for initial frequency estimation, frame synchronization, and symbol/chip timing estimation. The SW symbol/chip pattern has been designed with a hierarchical structure to enable the use of matched filter detection techniques by the UTs. To avoid the need for parallel matched filters (or significant detection loss), the BSs precompensate the frequency of each forward link signal for satellite Doppler, such that the residual satellite Doppler is zero at each spotbeam center. They also transmit the SW at a 6 dB higher power level than the data to any UT. All of the above permits acquisition by a low-cost UT.

As part of the initial acquisition process, each UT scans the received signal for occurrences of the SW. With a good SW detector architecture and a proper choice of detection thresholds, false and missed detections can be made quite unlikely. This is analyzed in detail in [1]. Depending on the location of the UT, a single cluster of 1, 2, or 3 candidate SWs might be detected, mutually offset in time by 8, 16, or 24 chip periods. These correspond to the forward link transmissions directed to different spotbeams from a given satellite. Alternatively, two or more clusters of SWs (each composed of 1, 2, or 3 candidate SWs) might be detected, mutually offset in time by fractions of a frame duration. These correspond to the forward link transmissions directed to different spotbeams from different satellites. By choosing the SW associated with the largest correlation peak, the UT can select the "closest" satellite and the "closest" spotbeam. (The relative strengths of the candidate SWs might also be used for position determination purposes.)

Once the strongest SW has been acquired, the UT can then more accurately estimate time and frequency for the forward link signal, and use these estimates to despread, demodulate, and decode the broadcast portion of the forward link frame. System data within the forward link broadcast informs the UT of the relationship between the forward link frequency and the forward link chip rate. Together with the frequency reference, this relationship enables the UT to accurately flywheel chip timing through the frame.

Return Link

Time and frequency references acquired from the forward link SW are adequate for return link datagram transmissions (using the random access time period). However, more precise return link synchronization is required if a UT needs to transmit stream-mode data to a BS. This more precise return link synchronization is achieved through a closed-loop control process executed in cooperation with the BS.

Recognizing that precise synchronization may occasionally be lost, a closed-loop reacquisition process, executed in cooperation with the BS, has also been defined to quickly restore synchronization.

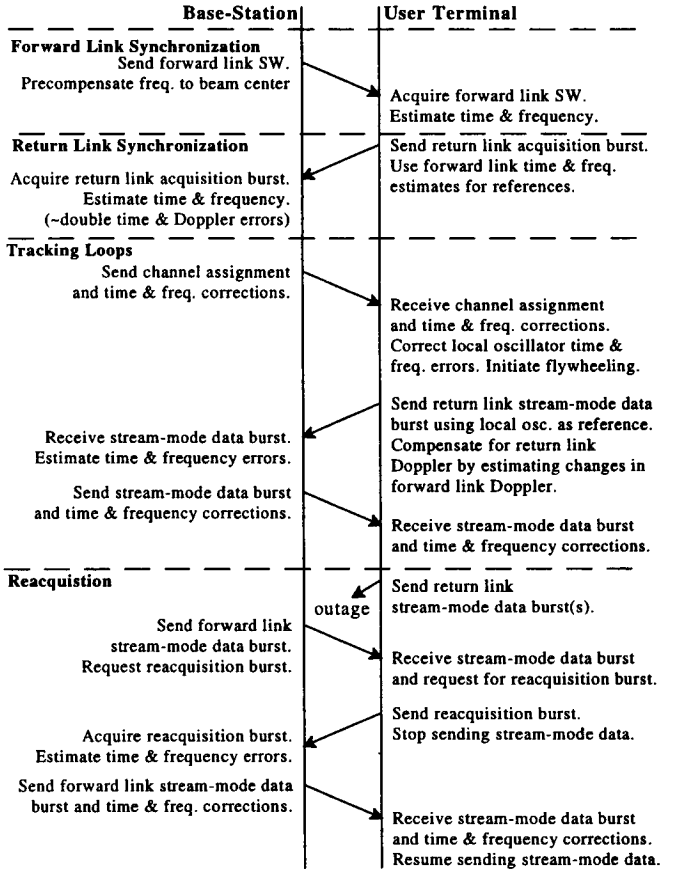


Fig. 6: Return link synchronization acquisition and reacquisition timeline.

The return link synchronization process is illustrated in timeline fashion in Fig. 6. First, there is the acquisition of the forward link SW by the UT as described previously. This establishes the time and frequency references for the forward link, as far as the UT is concerned, and allows the UT to recover the broadcast system data. The forward link time and frequency references establish the initial time and frequency references for the UT to use in the return link random access time period.

Second, there is the acquisition, at the BS, of a synchronization acquisition burst sent by the UT. This is either to initiate a call or to respond to a call initiation

received from the BS as part of the forward link broadcast system data. The acquisition burst allows the BS to measure the time and frequency errors associated with the UT transmission, and to send this information back to the UT via the forward link system data field.

Third, once the UT has been assigned a transmission time period(s) and QS-CDMA code(s), and stream-mode data bursts begin to be exchanged between the UT and the BS, time and frequency tracking loops can be set up by the UT to maintain the precise synchronization required for the channel. These loops rely on the signal being present and quasi-orthogonal to the other signals being simultaneously transmitted by other UTs. This implies the need for a fourth component, which is a strategy for fast reacquisition of time and frequency when synchronization is lost. A loss of synchronization may be caused by trying to flywheel through long periods (seconds) of shadowing or fading, or simply caused by noise within the time and frequency tracking loops. Extreme changes in UT motion might also cause a loss of synchronization, depending on the sophistication of the tracking loop design.

The system design is based on the use of a common reference oscillator at the BS for deriving both the forward link chip rate and the forward link transmit frequency. This leads to a known relationship between the chip rate and carrier frequency received at the UT, even when significant UT Doppler is present. If the UT uses the measured receive frequency of the SW as a reference from which to derive the chip rate, then the timing error will be dominated by the error in measuring the received frequency plus the error accumulated in flywheeling from the last received SW. The latter error is a function of the short-term stability of the UT oscillator and the interval between SWs. The Doppler frequency will only affect the timing drift if there are significant changes in the Doppler frequency between SWs.

The UT must know the frequency used by the BS, including any spotbeam-dependent frequency precompensation, in order to know the relationship between the transmit frequency and the chip rate. Thus, the transmit frequency, unique to each spotbeam, must be broadcast to the terminals using the forward link broadcast channel. The chip rate is fixed for all spotbeams so that they can remain time synchronized, with a fixed time offset.

With adequate short-term stability ($\sim 1 \times 10^{-9}$) for the UT oscillator, and using this oscillator as a reference, changes in the measured forward link frequency reflect changes in the Doppler frequency. Thus, the UT can use these measurements to anticipate and precorrect for frequency and timing errors caused by Doppler on the return link, without having to wait for frequency correction messages from the BS. Thus, except for very rapid changes in Doppler between each forward link SW, the ability to flywheel between correction messages is almost entirely dependent on the short-term stability of the UT local oscillator.

Through the use of decision-directed remodulation and correlation techniques, this synchronization strategy for the time and frequency tracking loops does not require additional synchronization overhead within the stream-mode data bursts. If synchronization degrades and the data cannot be detected with acceptable accuracy then synchronization will likely be lost. At this point the reacquisition bursts are used. Reacquisition bursts contain significant overhead (a known SW), but will only be required infrequently, and as requested by the BS.

Shortly after a loss of UT return link synchronization has been declared by the BS, or after an unexpected termination of transmission by the UT, the BS signals the UT to send a reacquisition burst at a scheduled time. Providing the UT has retained (or reacquired) forward link synchronization, and receives this message, it will comply, and return link synchronization can be quickly reestablished, thus permitting the return link stream-mode data flow to resume. Otherwise, after a timeout period, the BS will abandon the call, and the BS (or UT) will attempt to reacquire synchronization from scratch, using the forward link broadcast and/or return link acquisition bursts.

Concluding Remarks

With the given assumptions, the above design was found to provide the highest forward and return link capacities for a multi-beam scenario, and appears feasible to implement. Further details can be found in [1]. PFM4 modulation is described in [2]. Tools for evaluating the channel capacity of the proposed approach, and example capacity results, are presented in [3].

References

[1] B. Lyons et al., "A Transmission System for Future Satellite Mobile and Personal Communications", Final Report, ESTEC Contract No. 11515/95/NL/US, 1997.

[2] S. Crozier et al., "Design and Performance of Precompensated Frequency Modulation (PFM) for use with a Quasi-Synchronous CDMA Return Link", International Mobile Satellite Conference (IMSC'97), Pasadena, California, June 16-18, 1997.

[3] M. Moher et al., "Power and Bandwidth Tradeoffs for a Third Generation Satellite System", International Mobile Satellite Conference (IMSC'97), Pasadena, California, June 16-18, 1997.

[4] R. De Gaudenzi et al., "Bandlimited Quasi-Synchronous CDMA: A Novel Satellite Access Technique for Mobile and Personal Communication Systems", IEEE Journal of Selected Areas in Communications, Vol. 10, No. 2, pp. 328 - 343, February 1992.

A Mobile Satellite Modem for Helicopter Applications

W.G. Cowley, M.P. Lavenant,W. Zhang*

Defence Information Systems Centre
Institute For Telecommunications Research
University Of South Australia
The Levels SA 5095

***Communications Division**
Defence Science And Technology Organisation
PO BOX 1500, Salisbury SA 5108

Phone: +61 8 302 3316 FAX: +61 8 302 3873
e–mail: bill@spri.levels.unisa.edu.au

ABSTRACT

The paper describes a mobile satellite modem designed to provide flexible low–rate data and voice communications for helicopters via the L band transponder of the Australian domestic satellite . The project was constrained to use an existing low gain antenna on the helicopter.

The resulting communications channel exhibits severe periodic signal fading due to rotor blade obstructions, plus the usual mobile satellite channel characteristics of low signal to noise ratio and large doppler variations. The paper describes how these difficulties were overcome by optimising the modulation, channel coding and modem signal processing for this channel. The modem has been implemented with digital signal processors and designed to allow a variety of coding and processing algorithms to be tested at various data rates.

In particular this paper covers the following topics:

the helicopter channel magnitude and phase response, design of a pilot–symbol–assisted carrier phase recovery scheme and its performance relative to differential PSK, frequency offset estimation and interleaved coding, and performance results for the modem .

1 INTRODUCTION

The Institute for Telecommunications Research (ITR) has recently undertaken a project in conjunction with the Australian Defence Science and Technology Organisation (DSTO) to demonstrate helicopter communications via the L band transponder of the Australian domestic satellite. The project was constrained to use an existing low gain (GPS) antenna installed on the aircraft. This paper gives an overview of the resulting modulation, coding and signal processing employed to produce reliable voice and low–rate data (300 bit/s to 4800 bit/s) communications over a communications channel which is regularly obstructed by rotor blades.

The paper is organised as follows. In section 2 the channel characteristics of this particular mobilesat application are briefly described. Modem signal processing to implement the desired modulation and coding in real–time is presented in section 3. In particular a differential detection scheme is compared to a "pseudo–coherent" technique which employs pilot symbols to counteract the severe magnitude and phase fluctuations, plus an interleaving arrangement suited to this channel. Section 4 describes the resulting performance, both for coded and uncoded modes of operation.

2 CHANNEL CHARACTERISTICS

To measure the effect of rotor blade obstruction on the received L band signal, a series of channel recordings were undertaken. In order to provide a suitably high signal to noise ratio to obtain clean measurements of the channel, the satellite was simulated by transmitting an L band carrier from a lighting tower while a Black Hawk helicopter manoeuvred on the

tarmac. The GPS antenna on this helicopter is located about 2.8 m from the rotor hub, and slightly offset from the centre line of the aircraft. After down conversion to a low intermediate frequency a DAT recorder was employed to capture the signal for later analysis. Channel recordings made in this way allowed prototype modem algorithms to be tested with "channel playback".

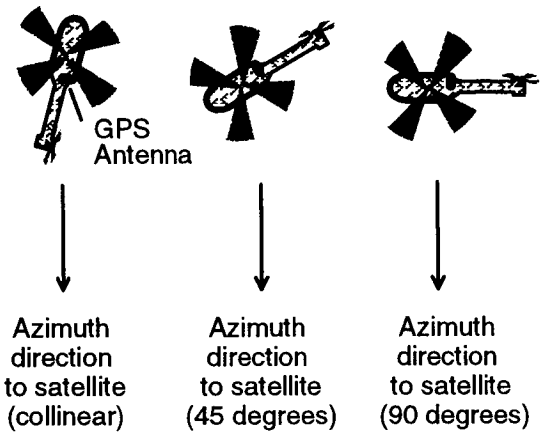


Figure 1 Helicopter orientations for channel recording.

Worst case signal fading occurs when the rotor hub, receiving antenna and transmitter are collinear (as shown in Figure 1), since it is only in this situation that the satellite signal is totally obstructed. Deep fades can be observed at the blade period, plus significant phase perturbations as shown in Figures 2 and 3. This helicopter has four blades and the rotor frequency is about 300 rpm. Magnitude and phase fluctuations for the other orientations are smaller (see [1] for details).

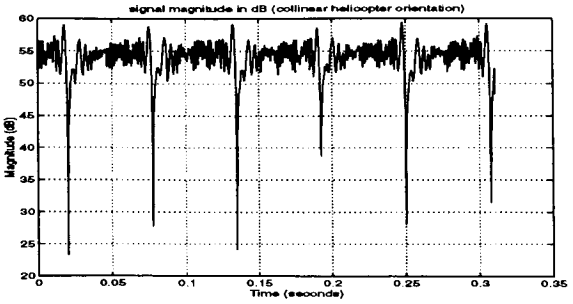


Figure 2 Signal magnitude (collinear).

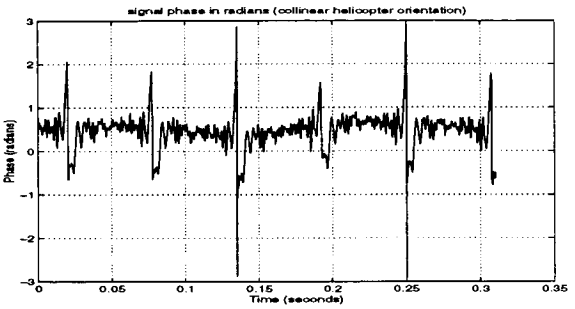


Figure 3 Signal phase (collinear).

It may be observed that the channel characteristics are very repeatable for each blade period, and that a chirping amplitude variation can be seen. Some insight into this interesting feature can be derived from the simple model shown in Figure 4. This model assumes two rays are incident on the antenna at A; one direct ray and one of relative amplitude a which is reflected from the rotor at R. Assume that the current rotor angle AHR is ϕ and that the hub to antenna distance HA is d. It is easy to show that the path difference between the rays is $2d\sin^2(\phi)$. By writing ϕ as a function of the rotor angular frequency ω_h, $\phi=\omega_h t$, and converting the path length difference into a phase difference at the carrier frequency (f_c) gives a received signal

$$r(t) = 1 + a\, e^{\{j\, 2\pi df_c \sin(\omega_h t)/c\}} \qquad (1)$$

It can be seen from (1) that the reflected component appears as a non-linear chirp in frequency. A plot of the amplitude of this function shows an approximate chirp of uniform amplitude. When a second exponential term is added to (1), to account for the next rotor blade at 90 degrees offset to the first, a more realistic model is obtained. With a small phase offset from 90 to simulate the lateral offset of the GPS antenna, this model gave the signal amplitude depicted in Figure 5. This illustrates the chirping amplitude characteristic and the interference between reflected rays causing the slow amplitude variations during the blade period. (It doesn't show the deep fade since the obstruction isn't modelled by (1).)

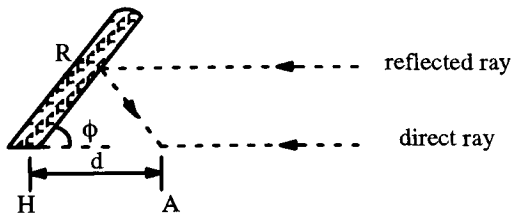


Figure 4 Simple model of reflection from a rotor blade

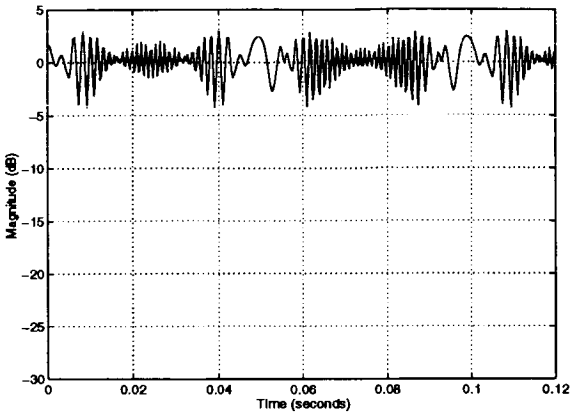


Figure 5 Plot of Predicted Signal Amplitude

Obviously the fade depths in the collinear case will be too large to allow any information to be received during the fade duration, and without suitable processing the resulting burst of errors would prevent useful communication. However with appropriate interleaving and error correction schemes, which are addressed in section 3.4, effects of the deep fades can be overcome.

3 MODEM SIGNAL PROCESSING

3.1 IF Conversion and Filtering

The modem signal processing is carried out on two TMS320C40 DSP chips. One DSP is used for the transmit and receive functions whilst the second DSP is used solely for the Viterbi decoding part of the signal processing.

Figure 6 shows the transmitter structure for the helicopter modem. The modem operates at a symbol rate of 2400 sps or 600 sps. The low transmit IF for the modem is centered at 4.8 kHz. The D/A operates at an f_s value of 19.2 kHz. Processing is performed at a rate of 8 samples/symbol (or 32 samples/symbols for data rates of 600 bits/s or lower). Root–Nyquist filtering with 40 % excess bandwidth is used.

The modem also has screen display features and voice communications capabilities. Voice can be transmitted and received using a 2400 baud rate LPC–10 based voice codec[4]. Screen display features include a scatter diagram display, a bit error distribution pattern and parameter viewing and setting[4].

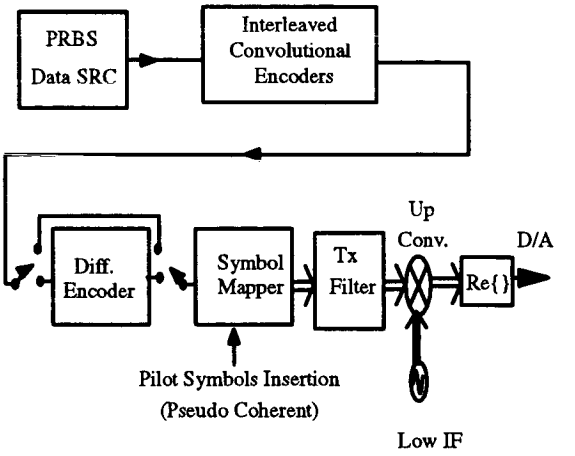


Figure 6 Modulator block diagram.

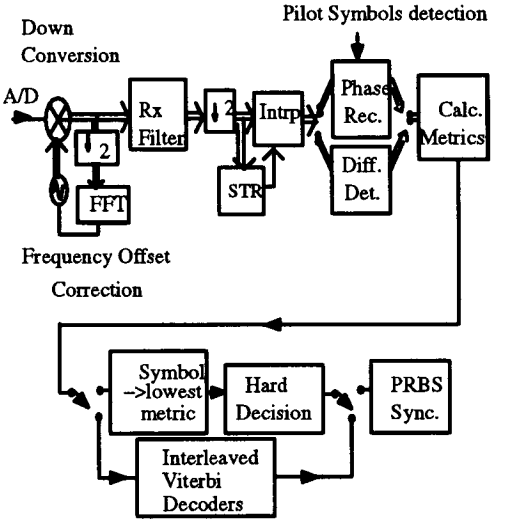


Figure 7 Demodulator block diagram.

Two detection schemes have been implemented and tested in this project: differential detection and a pseudo–coherent demodulator employing pilot symbols to assist in providing the phase reference. The latter scheme was used instead of multiple symbol differential detection since it can be shown to perform better [2], and provides soft decision outputs for the convolutional decoder.
Figure 6 shows a block diagram of the modulator using a PRBS generator as the data source. If differential coding is being used the symbols are differentially encoded, otherwise pilot symbols are inserted, for example, at a rate of one pilot symbol after every 9 information symbols. The symbol stream is then interpolated, say to 8 samples per symbol period, filtered and up converted prior to D/A conversion.

3.2 Frequency Offset Control

In the demodulator the bandpass signal is sampled, and frequency offset is corrected prior to receive filtering. Control of frequency offset is an important consideration in a low–rate modem where the signal is subject to rapid doppler changes, in this case some tens of Hz per second. When the demodulator starts, or when frame synchronisation has been lost, an estimate of gross frequency offset is first calculated. This uses the standard modulation stripping method of raising the signal to the M–th power, and using an FFT from which the largest spectral component is located. (Both BPSK and QPSK modulation are employed in this modem so M= 2 or 4.) This method multiplies the frequency offset (due to doppler and fixed frequency LO offsets) by M times, so a suitably high oversampling factor is required. In normal operation the frequency offset is tracked continuously by feedback of phase offset from the detector.

After removal of the frequency offset, the signal is receive filtered and decimated to four samples per T. Two symbol timing algorithms, one feedback and one feedforward, were tried. Since the channel perturbations do not effect symbol timing, both algorithms gave satisfactory performance (see [1] for details). A quadratic interpolator is employed to resample the signal according to the estimated timing offset.

3.3 Detection

In differential detection mode the single sample per symbol samples are used for differential detection. Differential metrics are calculated and are either used to derive the hard decision bit output or as input to the Viterbi decoder when convolutional coding is used.

During the pseudo–coherent mode of operation, pilot symbols are used to resolve the M fold channel phase ambiguity and assist in the estimation of phase offset. Consider a symbol stream :

...P A1 A2 A3 A4 P B1 B2 B3 B4 P C1 C2 C3 C4 P...

where P are pilots symbols and the As, Bs, and Cs represent information symbols (possibly with coding). In this case the pilot insertion rate is 20%. Phase recovery is performed on symbols B1 to B4 as follows. First an estimate of channel phase offset is calculated from the pilots. The information symbols and adjacent pilots are then rotated by this estimate and another phase estimate is calculated using the Viterbi and Viterbi [3] phase estimator with a amplitude nonlinearity of one.

Note that the minimum number of pilots that can be employed for each phase estimate is two and this gives the fastest response to rapid changes in channel phase. However it was observed when convolutional coding was employed (without interleaving in a AWGN channel) that more pilot symbols may be required in order to avoid $2\pi/M$ errors in channel phase estimate, which cause bursts of errors into the decoder and consequent unsatisfactory performance.

3.4 Error Control Coding

Forward error control (FEC) coding with appropriate interleaving is one of the key function blocks in this modem to combat the periodic signal fading. As shown in Figures 6 and 7, convolutional interleaved encoders and Viterbi decoders are employed. The advantage of this combination, compared with block interleaver and/or block code, is that it does not require frame synchronisation and provides good performance with moderate delay.

Let us examine the need for an interleaver. From section 2 it can be seen that the period of signal blockage is up to about 7 ms within every 57 ms period. For a 2400 bits/s modem, this corresponds to a loss of up to 17 consecutive symbols in one fading period. Without interleaving, the consecutive error symbols will fall into one constraint length of a convolutional code, or one codeword of a block code. Unless the constraint length (convolutional code) or codeword length (block code) is excessively and impractically large, the error patterns cannot be corrected.

We can then consider a block interleaver. Like a long block code, it requires frame synchronisation. This means not only additional effort in the demodulator, but also added overhead to accommodate the frame alignment word in the signal stream. This should be avoided if possible. On the other hand, the convolutional interleaver and multiple encoders/decoders are suitable for implementations on digital signal processors. The effective degree of interleaving for a given delay is also higher [5].

The convolutional interleaver and deinterleaver work as follows. The incoming information stream is distributed by a circular switch and fed into n convolutional encoders in a circular way. The encoder output symbols are then combined to form the

transmission stream. At the receiver, the demodulated baseband sequence is fed into n Viterbi decoders in a circular way. The output bits of the Viterbi decoders are then combined into one stream as the output. For a 2400 bits/s modem at least n=17 should be used. This means that there is one error due to fading in every 17 received symbols, which is correctable by a constraint length 7 code. Considering the decoding delay of a Viterbi decoder is 32 bit, then for a n=17 system has a decoding delay of 227 ms. If n=25 is used, to enable the modem to cope with longer fading durations, the delay is 333 ms. These additional delays are considered acceptable in our applications. For lower data rates the interleaver length n can be reduced accordingly.

Unlike a Rayleigh fading channel, in this periodic fading channel, there are no consecutive errors longer than a given length (namely 17), assuming ideal phase estimation. This means that all such errors, except those caused by the additive Gaussian noise, should be correctable by an appropriately designed decoder and interleaver. The performance should be that of a degraded Gaussian channel rather than a Rayleigh fading channel. It resembles the punctured convolutional code where only a minimal degradation (say within 2 dB) is involved. This is indeed the case as it is shown in our measurement results, which are described in Section 4.

4 PERFORMANCE MEASUREMENTS

This section of the paper presents a sub–set of performance results for the digital modem. Results are given for the differential and pseudo–coherent cases, both in the presence of AWGN only (ie with no helicopter channel interference) and with helicopter channel interference for the collinear case (see Figures 8 to 11).

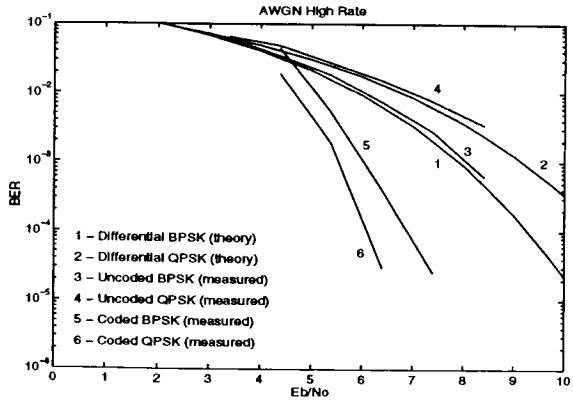


Figure 8 Differential performance (AWGN case).

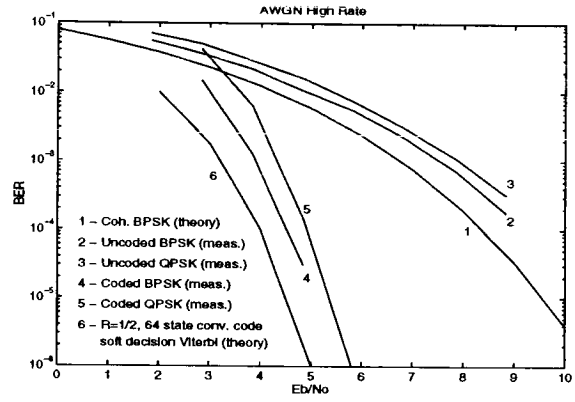


Figure 9 Pseudo–coherent performance (AWGN case).

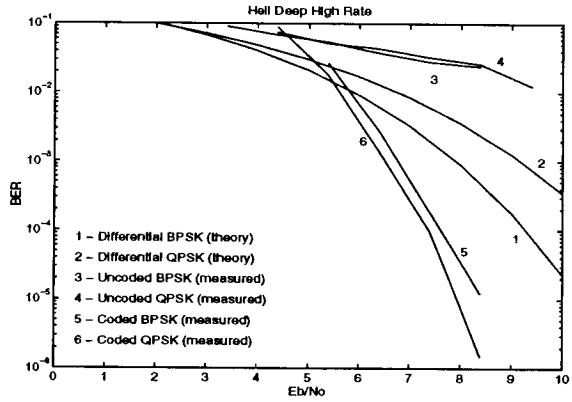


Figure 10 Differential performance (collinear case).

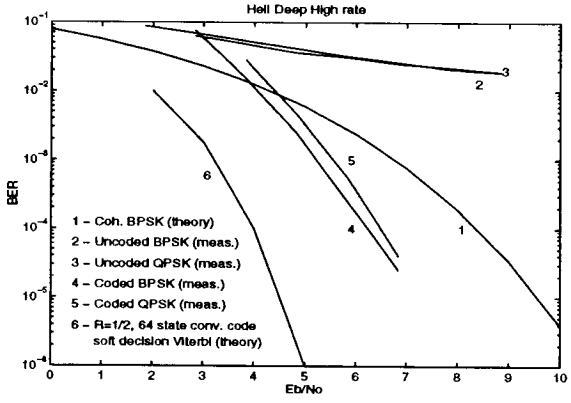


Figure 11 Pseudo–coherent performance (collinear case).

Results for the AWGN channel show a relatively small implementation loss and full coding gain. Uncoded results for the helicopter channel show a very high error rate as expected, due to blade obstruction.

It was found that the pseudo coherent BPSK modem configuration with interleaved convolutional coding of depth 25 offered a good measured performance against the recorded helicopter channel data. The performance loss relative to the AWGN channel case was measured to be 1.7 dB for the deep fade case, which represents the worst performance case in the presence of the helicopter channel environment.

5 CONCLUSIONS

The characteristics of a satellite to helicopter digital communications channel have been briefly described, and a simple model of the channel has been given which illustrates some of the channel characteristics. The modem signal processing required to implement the desired modulation and coding in real–time has been described. A differential detection scheme has been compared to a pseudo–coherent technique and performance results have been presented. These results show that when suitable algorithms are employed, there is only a slight degradation in BER performance incurred from the helicopter rotor blade signal perturbations. In flight testing of the modem will be required to verify the suitability of the scheme for the target platform.

ACKNOWLEDGMENT

The authors wish to thank Craig Burnet (ITR), Jeff McCarthy (DSTO) and Stephen Morris (DSTO) for their various contributions to the development of the mobile satellite modem. The work was performed under the sponsorship of DSTO, Contract No TO2059.

REFERENCES

[1] **Lavenant, M.P., Cowley, W.G., McCarthy, J., Burnet, C., Zhang, W., Morris, S. and Cook, S.,** "Digital Communications for the Satellite to Helicopter Channel", IAC–97, Sydney, Feb. 1997.

[2] **Cowley, W.G. and Crozier, S.N.,** "Comparisons Between Differential Methods and Pseudo–Coherent Techniques", DSP for Comms, Adelaide, Apr. 1994.

[3] **Viterbi, A.J and Viterbi, A.M.,** "Nonlinear Estimation of PSK–Modulated Carrier Phase with Application to Burst Digital Transmission", IEEE Transactions on Information Theory, VOL. IT–29, No4, July 1983, pp. 543–551.

[4] **University of South Australia,** Institute for Telecommunications Research, Defence Communications Research Centre, *FDRM Phase 2 Stage 3 Report Version 1.0*, November 1996.

[5] **Cook, S.C., Gill, M.C. and Giles, T.C.,** "A High–Speed HF Parallel–Tone modem", 6th International Conference of HF Radio Systems and Techniques, University of York, UK, 4–7 July 1994.

The Performance of Video and High-Quality Audio Telephony Services for Mobile Satellite Applications

Daniel Boudreau[1], John Lodge[1], Paul Guinand[1], Robert Lyons[2] and Lars Erup[3]

(1) Communications Research Centre, 3701 Carling Av., P.O. Box 11490, Station "H", Ottawa, Ontario, Canada, K2H 8S2, Tel.: (613) 990-6278, Fax: (613) 998-1686, email: Dan.Boudreau@crc.doc.ca
(2) Square Peg Communications Inc., 3701 Carling Av., Building T3, P.O. Box 11490, Station "H", Ottawa, Ontario, Canada, K2H 8S2, Tel.: (613) 820-7817, Fax: (613) 993-3122, email: http://www.crc.doc.ca/innovation/SquarePeg.html
(3) Communications Services Division, Directorate of Telecommunications Programmes, European Space Agency, European Space Research and Technology Centre, ESTEC/TST, P.O. Box 299, Noordwijk, NL 2200 AG The Netherlands, Tel.: +31 71 5653161, Fax: +31 71 5654598, email: lars@t.estec.esa.nl

ABSTRACT

This paper presents the performance of a joint source and channel coding design for a mobile satellite application involving video and high-quality audio telephony. For these services, source coding standards recently adopted by the International Telecommunication Union (ITU) are combined with matched channel coding techniques. Speech coding is performed at 8 kbits/s, using the G.729 standard, while video coding is accomplished with the H.263 standard at 51.2 kbits/sec. The two mobile telephony services are studied through computer simulations, in rural and urban mobile satellite communications conditions. The performance is expressed in terms of bit error rate (BER) before and after channel decoding, and in terms of subjective and objective quality measures. It is shown that the two telephony services, when operated in better conditions than a certain threshold, deliver good quality in most channel scenarios of interest.

1. INTRODUCTION

Many mobile satellite communication providers are in the process of investigating future services involving low-rate speech and video transmission. The European Space Agency (ESA) is researching source coding, channel coding/modulation and multiple access alternatives for the third generation European mobile satellite system. This paper presents the results of a performance study, for this European system, of joint source and channel coding designs for both the audio and the video telephony services. The digital speech coding is performed in both services by using the ITU-T G.729 standard at 8 kbits/s [1]. This standard produces toll quality speech, with an algorithmic delay of only 15 msec [2]. The video telephone uses the ITU-T H.223 [3] multiplexing concepts to combine the G.729 audio and the ITU-T H.263 video [4] at an overall rate of 64 kbits/s. The video telephone image format is QCIF (144 lines x 176 pixels), updated at 10 frames/sec.

The joint source and channel coding design, for the two services, was presented in [5]. This design involves the use of unequal error protection in the audio telephone, and the use of a powerful Reed-Solomon code in the video telephone. The channel coding design is performed by assuming that the transmission system provides the audio and video services with a baseline bit error rate (BER) of 10^{-3}. This error rate is provided by an inner constraint length 9 rate 1/2 convolutional code, coupled with QPSK (forward link) or $\pi/4$ QPSK (return link) modulation. Coherent demodulation (using reference symbols) and Viterbi decoding is accomplished in the receiver.

In this paper, the joint source and channel coding design is first reviewed. The key results of [5] are stated, and the coding methods retained in the final configuration are described. The channel scenarios implemented in the simulations are described in Section 3. The error statistics on the Viterbi decoder output bitstream, as obtained with the channel simulations, are studied in Section 4. These statistics are utilized in Section 5, to justify the design of the interleaver on the outer channel coder output. The results of the simulation of the overall audio and video telephony services are then presented and discussed.

2. JOINT SOURCE AND CHANNEL CODING FOR AUDIO AND VIDEO TELEPHONY

The results of a sensitivity analysis performed on the audio and video coders were presented in [5]. The G.729 speech coder operates on 10 msec frames, producing 80 bits for each of these time periods. The

analysis established that of these 80 bits, 12 were very sensitive to channel errors, 11 were moderately sensitive, while 57 were relatively insensitive. The threshold BER, for each of these classes, was established in order to produce a low subjective speech degradation. Because the BER thresholds for the two sensitive classes were very close, it was proposed to group the 12 most sensitive bits with 9 of the moderately sensitive bits, and work with two classes: the class of the 21 most sensitive bits, for which channel errors cause a significant degradation in the synthesized speech, and the class of bits that can be left protected by the inner channel code only. The sensitive bit numbers, using the numbering of the standard [1], are given in Table 1.

	Bit number
Sensitive class (21 bits)	2, 4, 5, 8, 19 to 25, 27, 46, 47, 51 to 54, 75, 76 ,80
Less sensitive class (59 bits)	others

TABLE 1: The ITU-T G.729 bits corresponding to the final sensitivity classes [5].

It was determined that in order to obtain a low subjective speech degradation, a maximum BER of $5x10^{-5}$ was required on the sensitive bits, along with a BER of $5x10^{-3}$ on the less sensitive bits. A (31,21) BCH code was proposed as an outer code for the sensitive bits. This results in a 90-bit coded frame, and a coded bit rate of 9 kbits/sec. On a binary symmetric channel, the BCH code results in a BER lower than 10^{-6} for an inner BER of 10^{-3} [5].

The results of the sensitivity analysis performed on the H.263 video standard have concluded that it is better to protect all the coded bits evenly, at an error rate of 10^{-5} or better [5]. An 8-bit (255,223) Reed-Solomon code was selected to provide this threshold, from the inner BER of 10^{-3}. On a binary symmetric channel, this Reed-Solomon code results in a BER lower than 10^{-10} for an inner BER of 10^{-3} [5]. It is also very efficient when long error bursts are encountered. An extra level of joint source and channel coding is also present in the video telephone design. The video coder is modified such that portions of the image are updated using only transform coding, without any reference to previous images (no prediction). This *forced updating*, performed at the image macroblock (16x16 pixels) level, 5 cyclically selected macroblocks per frame, allows the image to be completely refreshed every 2 seconds (20 frames), and prevents prediction error accumulation [6]. This allows image reception at a BER of 10^{-4}, although with a noticeable degradation.

3. CHANNEL CONFIGURATIONS

The two telephony services were tested on five simulated channel configurations: AWGN, Land Mobile Highway (LMH), Land Mobile Suburban (LMS), Land Mobile Urban (LMU) and Hand-Held Urban (HHU), on both the forward and the return links. The results presented in the paper are for the forward link only, since they do not differ significantly from those of the return link. The main characteristics of these channels are given in Table 2. The operating Eb/No represents the value of energy per bit to AWG noise power required to operate the system at the inner BER of 10^{-3}.

Channel		K Factor	Fading rate (speed)	Operating Eb/No at the 10^{-3} threshold
AWGN		--	--	3.2 dB
Land Mobile	Highway	15 dB	100 Hz (110 km/h)	4.0 dB
	Suburban	10 dB	65 Hz (70 km/h)	5.5 dB
	Urban	7.8 dB	35 Hz (40 km/h)	7.5 dB
Hand-Held Urban		7.8 dB	10 Hz (1.5 m/s)	6.5 dB

TABLE 2: Simulated channel configurations, in unshadowed conditions.

Simulations were also performed in shadowed conditions. The telephony services were designed for unshadowed conditions, and because of the large drop in SNR that shadowing produces, it is not expected that the services will be maintained at a good quality level when shadowing occurs. The simulation scenarios in shadowing conditions were limited to the land mobile urban and the land mobile suburban channels, both at BERs better than the threshold. These scenarios are indicated in Table 3.

Channel	K Factor (dB)	Eb/No (dB)
Land Mobile Suburban	10	6
Land Mobile Urban	7.8	8

TABLE 3: Simulated channel configurations, in shadowed conditions.

The shadowing profiles, for the conditions of Table 3, are given in Figures 1 and 2. They were obtained at the output of the shadowing model [7].

4. STATISTICS AT THE INNER VITERBI DECODER OUTPUT

The combination of a Viterbi decoder with a multipath fading channel is likely to produce error bursts at the input of the outer channel decoder. Some form of bit or symbol interleaving is therefore necessary to spread out the effects of these bursts, and to allow the BCH or the Reed-Solomon decoder to approach the performance predicted in Section 2, for a binary symmetric channel.

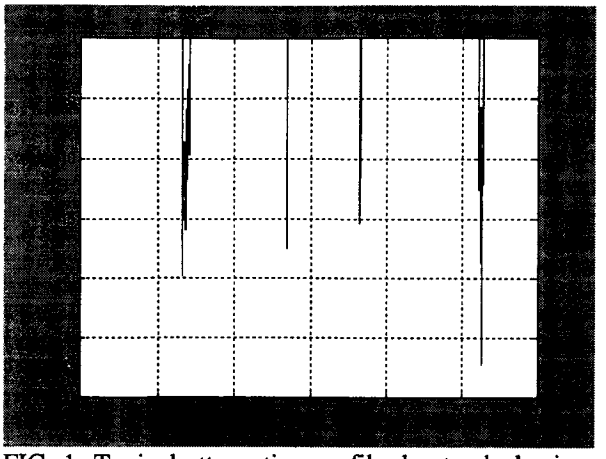


FIG. 1: Typical attenuation profile due to shadowing, on the land mobile suburban channel, at Eb/No = 6 dB (60° elevation).

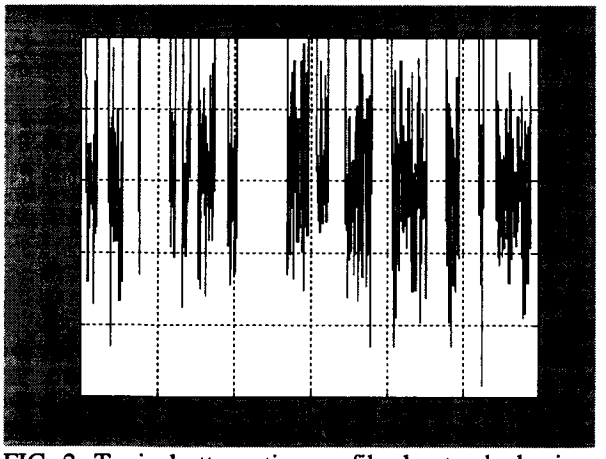


FIG. 2: Typical attenuation profile due to shadowing, on the land mobile urban channel, at Eb/No = 8 dB (60° elevation).

A typical error burst length distribution at the Viterbi decoder output is shown in Fig. 3, for the AWGN channel at Eb/No=3.2 dB. An error burst is defined here as a group of bits in which two successive erroneous bits are separated by less than 12 correct bits [8]. The length of the error bursts is concentrated around multiples of 8 bits, i.e. multiples of the number of bits necessary to return the decoder to the correct decoding path, after an error has occurred. This figure shows that the Viterbi decoder, when operating in a static environment, produces mostly error bursts with a maximum length of 32 bits.

A simulated error burst distribution for the hand-held urban channel is shown in Figure 4. As expected, large burst lengths are common, although the distribution is dominated by the bursts smaller than 32 bits. This indicates that, although it is undesirable (due to the constraints on the transmission delay) to use an interleaver that would spread out the longest bursts, selecting one that would control the bursts typical of the static channel would minimize the degradation in the subjective quality of the telephony services.

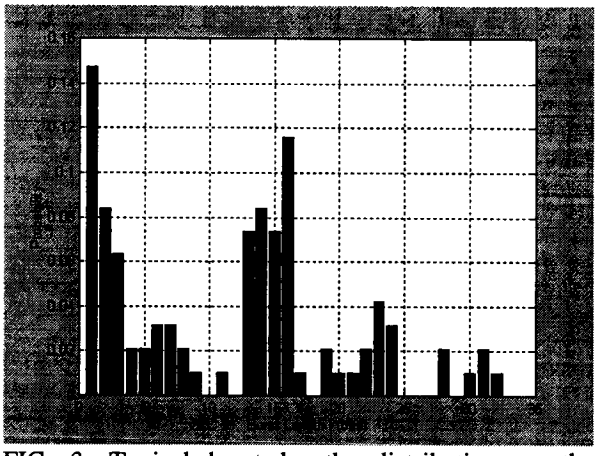


FIG. 3: Typical burst lengths distribution on the AWGN channel, Eb/No = 3.2 dB.

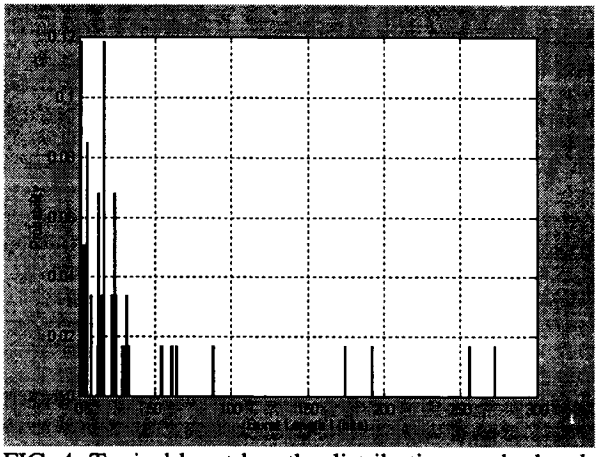


FIG. 4: Typical burst lengths distribution on the hand-held urban channel, Eb/No = 6.5 dB.

5. INTERLEAVER DESIGNS AND SIMULATED PERFORMANCE

In order to combat the effects of error bursts, as described in the previous section, two types of block interleavers are selected, and simulated performance are provided for the two telephony services.

5.1 AUDIO TELEPHONY

In the audio telephony service, the 90-bit frames (31 coded, 59 uncoded) are read in a matrix of the form given in Fig. 5. In this block, one row contains a coded frame, in which every coded bit is separated by two uncoded bit (except for the last two coded bits). These interleaved coded frames are stacked on top of each other, and the bits are transmitted by reading the matrix column by column. Because the design of the communications system is done on a slot basis, with 540 coded bits per slot, an interleaver of the form of Fig. 5 does not introduce any further delay if it encompasses 6 speech frames (6x90 interleaver). A 12x90 interleaver introduces an additional delay of 60 msec.

This type of interleaver spreads the coded bits of a given 31-bit field over a multiple of 6 codewords. It has the effect of interleaving 17 or 35 uncoded bits between every one of the 31 coded bits. The 12x90 interleaver is therefore adequate to conteract all of the error bursts of length 32 or less that dominate the output of the inner Viterbi decoder.

READ IN →

OUT ↓										
	x	-	-	x	-	-	...	x	-	x
	x	-	-	x	-	-	...	x	-	x
	x	-	-	x	-	-	...	x	-	x
	x	-	-	x	-	-	...	x	-	x
	x	-	-	x	-	-	...	x	-	x

FIG. 5: Interleaver table for the audio telephony service. "x" represents a (31,21) BCH coded sensitive bit, and "-" represents an uncoded bit.

The BER, as measured at the output of the (31,21) BCH decoder for the simulated channel configurations of Table 2, is given in Table 4. Except for the AWGN channel, the outer BER is about one order of magnitude higher than the threshold of $5x10^{-5}$ that was determined in [5]. This fact does not necessarily lead to a large subjective voice degradation, as will be demonstrated below. This is the case because the value of $5x10^{-5}$ was determined when the 59 less sensitive bits were subject to an error rate of $5x10^{-3}$. Here, these bits experience a BER of 10^{-3}, which gives some margin for the subjectively tolerable BER on the sensitive bits.

Channel		Operating Eb/No at the 10^{-3} threshold	Outer BER on Coded Bits (12x90)
AWGN		3.2 dB	4.5e-06
Land Mobile	Highway	4.0 dB	2e-04
	Suburban	5.5 dB	6e-04
	Urban	7.5 dB	8e-04
Hand-Held Urban		6.5 dB	8e-04

TABLE 4: The measured BER at the output of the (31,21) BCH decoder (sensitive bits) in the audio telephony service operated at threshold.

The received voice quality corresponding to the conditions of Table 4 has been evaluated by using the segmental SNR measure (SEGSNR). This measure is an average, and tends to mask the localized effects of the bursts of errors. The results, for a one minute audio passage, are given in Table 5. The SEGSNR is the average obtained for a male and a female speaker. It is noted that the SEGSNR is always close to its largest possible value of 1.73. The degradation in voice quality, as evaluated subjectively (in informal tests), is also indicated in this table . This degradation is small for the AWGN and the two fastest fading channels, but it is dominated by the burst of errors still present in the slowest fading cases. Between these error bursts, the subjective quality is high. The speech intelligibility is high at all times.

Channel		SEGSNR (dB)	Subjective Degradation	Intelligi-bility
AWGN		1.67	small	high
Land Mobile	Highway	1.67	small	high
	Suburban	1.71	small	high
	Urban	1.68	medium	high
Hand-Held Urban		1.71	medium	high

TABLE 5: The objective voice quality (SEGSNR) and the subjective degradation for the cases of Table 4. The error-free SEGSNR is 1.73 dB. The degradation scale is: none, small, medium and high.

The BER and voice quality, for channel conditions slightly better than threshold (i.e. at an inner BER of about $5x10^{-4}$) are given in Tables 6 and 7. These results indicate that the service quality quickly improves when the operating point is better than the threshold value. They also show that, although the probability of occurrence of localized degradation, due to error bursts, cannot be reduced to zero, its effect can

be minimized, and a high quality audio telephony service can be maintained.

Channel		Operating Eb/No better than the 10^{-3} threshold	Outer BER on Coded Bits (12x90)
AWGN		3.5 dB	1.1e-07
Land Mobile	Highway	4.3 dB	1e-04
	Suburban	6.0 dB	3e-04
	Urban	8.0 dB	3e-04
Hand-Held Urban		7.0 dB	4e-04

TABLE 6: The measured BER at the output of the (31,21) BCH decoder (sensitive bits) in the audio telephony service operated at a BER *better than* the threshold.

Channel		SEGSNR (dB)	Subjective Degradation	Intelligibility
AWGN		1.71	small	high
Land Mobile	Highway	1.71	small	high
	Suburban	1.71	small	high
	Urban	1.71	small	high
Hand-Held Urban		1.72	small	high

TABLE 7: The objective voice quality (SEGSNR) and the subjective degradation for the cases of Table 6. The error-free SEGSNR is 1.73 dB. The degradation scale is: none, small, medium and high.

5.2 VIDEO TELEPHONY

The powerful 8-bit (255,223) Reed-Solomon code selected for the video telephony service is by nature a burst error correcting code. It has 32 parity symbols for each codeword, and can correct error bursts of up to 121 bits [9]. On the slowest fading channel configurations, this error-correcting capability can be easily exceeded, leading to visually annoying artifacts. This implies that a fairly large symbol interleaver is required in these cases. It was determined through simulations that a 255x6 block symbol interleaver was very good at reducing the decoded bit error rates to acceptable levels (around 10^{-5}) in all but the most extreme channel cases.

The multiplexing of the audio and video bits is performed using a 10 msec time base, at a rate of 66.4 kbits/sec† . Each multiplexed frame contains eighty G.729 bits (audio) and 512 H.263 bits (video). A 255x6 block symbol interleaver therefore spreads the error effects over 6 codewords, and introduces a delay of 16.2 multiplexed frames (162 msec).

The video telephony service was evaluated for the channel conditions of Table 2, for one minute sequences. The BER measured at the output of the (255,223) Reed-Solomon decoder is indicated in Table 8. These results are either better or very close to the BER subjective threshold of 10^{-5} .

Channel		Operating Eb/No at the 10^{-3} threshold	Outer BER
AWGN		3.2 dB	< 1e-10
Land Mobile	Highway	4.0 dB	< 1e-10
	Suburban	5.5 dB	< 1e-10
	Urban	7.5 dB	5e-05
Hand-Held Urban		6.5 dB	5e-05

TABLE 8: The measured BER at the output of the (255,223) Reed-Solomon decoder in the video telephony service operated at threshold.

The subjective degradation corresponding to the cases of Table 8 is indicated in Table 9. Even with the powerful channel coding and interleaving used in the video telephony service, the subjective quality is degraded in the slowest fading cases. This is particularly true for the video portion of the communications, in which even the smallest artifact is annoying. The audio sequence is always reproduced with good fidelity.

These simulations indicate that video telephony is possible in a mobile satellite environment, with high audio quality in all cases, and with acceptable video quality in most scenarios.

Channel		*Audio* Subjective Degradation	*Video* Subjective Degradation
AWGN		none	none
Land Mobile	Highway	none	none
	Suburban	none	none
	Urban	small	high
Hand-Held Urban		small	high

TABLE 9: The subjective degradation for the cases of Table 8. The degradation scale is: none, small, medium and high.

† Note that this is 2.4 kbits/sec in excess of the 64 kbits/sec target, because the system is designed to accommodate a channel bit rate of 76.8 kbits/sec, including signalling bits, and that a combined bit rate of 66.4 kbits/sec gives 75.9 kbits/sec after the (255,223) coder. There is built-in flexibility in the multiplexing standard (H.223) to reduce this rate if necessary.

5.3 SHADOWED CONDITIONS

The two telephony services were simulated in the channel conditions shown in Table 3, resulting in the shadowing profiles of Figs. 1 and 2. Since shadowing corresponds to a drop in channel gain, these profiles indicate that when the shadowing gain is below 0 dB, the channel corresponds to operation much worse than the BER threshold. It is therefore expected that the subjective degradation will be very high within a shadow, but that a good service quality will be attained outside. This behavior was confirmed by the simulations. The average BER on the channel is on the order of 10^{-2} in the suburban case, and 10^{-1} in the urban case. The channel coding, in both services, is then very inefficient, and the resulting subjective degradation is always on average very high, although the audio quality is good outside the shadows. Because of the less frequent occurrence of the shadows in the suburban case, the voice telephony service is still operational in this case. The shadowed urban channel is too noisy to allow adequate voice quality. As for the video service, the error rate within the shadows is so high that it can produce a total loss of the visual information, requiring re-initialization and re-synchronization of the transmission. This is true with high probability for the shadowed suburban channel, while the shadowed urban channel is virtually useless for the current design of the video telephone.

6. CONCLUSIONS

The simulated performance of audio and video telephony over a mobile satellite communications link was presented in this paper. The joint source and channel coding techniques selected for these services were discussed, as well as the resulting objective and subjective quality obtained over a series of channel configurations. The results show that audio and video telephony services are feasible with reasonable quality over most unshadowed channel setups.

REFERENCES

[1] International Telecommunication Union, *Coding of Speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)*, ITU-T Recommendation G.729 (03/96), March 1996.

[2] R. V. Cox, P. Kroon, "Low bit-rate speech coders for multimedia communication," IEEE Communications Magazine, vol. 34, no. 12, December 1996, pp. 34-41.

[3] International Telecommunication Union, *Multiplexing Protocol for Low Bitrate Multimedia Communication, Annex A*, Draft ITU-T Recommendation H.223/Annex A, May 1996.

[4] International Telecommunication Union, *Video Coding for Low Bitrate Communication*, ITU-T Recommendation H.263 (03/96), March 1996.

[5] D. Boudreau, C. Dubuc, J. Lodge, P. Guinand, R. Lyons and L. Erup, "High-quality video and audio telephony for mobile satellite applications," Proceedings of the Japan-Canada International Workshop on Multimedia Wireless Communications and Computing WMWCC'96, Victoria, British Columbia, September 17, 1996, pp. 49-50.

[6] N. Naka, S. Adachi, M. Saigusa and T. Ohya, "Improved error resilience in mobile audio-visual communications," 1995 Int. Conf. On Universal Personnal Communications, Tokyo, November 6-10, 1995, pp. 702-706.

[7] P. S. Guinand, J. H. Lodge, L. Erup, S. Buonomo, R. Lyons and V. S. Urad, "Dynamic modeling of the wideband mobile satellite channel," Proceedings of the Japan-Canada International Workshop on Multimedia Wireless Communications and Computing WMWCC'96, Victoria, British Columbia, September 17, 1996, pp. 47-48.

[8] CCITT *Blue Book*, vol. 1, Fascicle I.3, Terms and Definitions, ITU 1988.

[9] R. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.

A Real-Time Testbed for Satellite and Terrestrial Communications Experimentation and Development

K. Angkasa, J. Hamkins, J. Jao, N. Lay, E. Satorius, A. Zevallos
Jet Propulsion Laboratory, California Institute of Technology,
4800 Oak Grove Drive, Pasadena, California 91109

Abstract

This paper describes a programmable DSP-based testbed that is employed in the development and evaluation of blind demodulation algorithms to be used in wireless satellite or terrestrial communications systems. The testbed employs a graphical user interface (GUI) to provide independent, real-time control of modulator, channel and demodulator parameters and also affords real-time observation of various diagnostic signals such as carrier, timing recovery and decoder metrics. This interactive flexibility enables an operator to tailor the testbed parameters and environment to investigate the performance of any arbitrary communications system and channel model. Furthermore, a variety of digital and analog interfaces allow the testbed to be used either as a stand-alone digital modulator or receiver, thereby extending its experimental utility from the laboratory to the field.

1 Introduction

Blind demodulation of communication signals under a variety of unknown channel conditions is an area of significant and active interest for both terrestrial and satellite based applications. The testbed described in this paper has been designed to facilitate the development, comparison and evaluation of novel detection techniques. The testbed system comprises integrated hardware and software components that may be operated in both real- and non-real-time. Digital signal processors (DSP's) coupled with a general purpose processing unit provide for algorithm implementation and user interface and control. In Figure 1, a functional block diagram of the testbed shows the details of the test signal generation and demodulation functions that are divided between three blocks: the modulator, channel, and demodulator blocks. A graphical user interface (GUI) enables user operation and control of the testbed.

The testbed is targeted towards the development of algorithms that facilitate the demodulation of communications signals that have been distorted by a variety of channel impairments. These distortions include intersymbol interference (static and time-varying with uncorrelated taps), additive Gaussian noise, additive coherent interference (e.g. CW or modulated signals) and time-varying carrier phase.

The paper is organized into two sections consisting of an architectural description and a description of various blind demodulation tests and the corresponding performance results for analog and digital modulations. In the testbed architecture section, both the hardware and software design of the testbed and its intended use within the overall context of an algorithm development methodology are discussed in detail. In addition, baseline capabilities are reviewed including generic digital modulators and demodulators for M-ary PSK, M-ary QAM and similar complex signal formats, blind and decision directed linear equalizers, and blind maximum likelihood based data sequence estimators.

The performance results section provides both algorithmic description and quantitative performance results for different blind equalization techniques. As an example, we describe the O_s^r Rectangular method which has been specifically designed to equalize high order QAM constellations while providing a low residual ISI. Performance results are compared to more conventional techniques such as the constant modulus algorithm. Blind demodulation performance of signals distorted by a spectral null channel is also examined in terms of linear versus maximum likelihood sequence estimation techniques.

2 System Architecture

The real-time communications testbed has been developed for a number of interrelated purposes. It is primarily a platform upon which to test new demodulation algorithms against controlled data and against a "library" of existing techniques. The eventual goal is a fair evaluation to demonstrate

improvement or loss compared to conventional approaches. A secondary purpose is to impose a certain level of practicality upon the algorithm development process by attempting to achieve real-time implementations for narrowband signal demodulators. Furthermore, testbed algorithm implementation is only performed with promising candidate techniques that meet a dual requirement of performance improvement and moderate complexity.

2.1 System Hardware and Software

The communications testbed hardware is composed of different programmable processing elements that implement both control of the testbed and the signal processing algorithms employed for demodulation. By selecting hardware components from commercially available VME-bus board level products, a fairly rapid system integration can be achieved. The elements that implement GUI and system control consist of a Motorola 68040 processor board and a corresponding X-Windows server. The software developed for the GUI and system operation resides in a real-time operating system environment (VX-Works) which enables automated scheduling of various system processes. Some of these processes include remote operation and user displays as well as real-time storage of demodulator parameters and decoded data. An example screen of the GUI is shown in Figure 2.

The graphical user interface of the testbed is designed using Imperial Software Technology's XDesigner application. XDesigner is an interactive tool for building GUIs using widgets of the standard OSF/Motif toolkit as building blocks. The designed GUI is then compiled and runs on a real-time operating system which allows multiple tasks to be running simultaneously with priorities given to the more critical tasks.

Results or output from these real-time tasks may then be displayed using laboratory signal analyzers and/or any terminal that handles X-Windows. The latter was made possible through an integration with SciPlot Widget of the Free Software Foundation, Inc., a widget capable of plotting Cartesian or polar graphs, including logarithmic axes in Cartesian plots. This widget is subclassed directly from the Core widget class and may be freely used with Athena, Motif or the Open Look/Xview widget sets. These plots enable real-time monitoring of the transmitted, distorted or equalized signal quality via constellation plots, for example.

On the signal processing side, the hardware consists of TMS320C40 floating point DSP boards. In addition, low rate (< 200 kHz) dual channel A/D's and D/A's allow analog I/O of complex baseband waveforms as shown in Figures 3 and 4 representing respectively, a multipath distorted 8PSK waveform and a blindly equalized version. This output signal capability coupled with the X-Windows based real-time plots affords the user a good level of flexibility in evaluating system and receiver performance. Furthermore, by developing the DSP software in C, we are able to achieve fairly rapid prototyping of real-time demodulators. In addition to the programmable signal processor hardware, digital demodulation from a more typical IF (e.g. 10.7 or 21.4 MHz) is achieved through the use of a high speed, 10 bit, 65 MHz A/D coupled with a complex downconverter board. The downconverter employed within the testbed is capable of demodulating up to eight medium bandwidth channels in parallel and then interfacing directly to the serial ports of the C40 DSP's.

2.2 System Functionality

In order to develop demodulation techniques under varying channel conditions, a controlled channel simulation environment has been established in conjunction with the associated modulators and demodulators (modems). From the point and click GUI, the user is at liberty to select transmitter parameters such as the complex modulation type and desired spectral shaping. Channel selections correspond to such impairments as bandlimiting, finite-impulse-response (FIR) multipath, additive noise and carrier frequency offsets. Control of both the signal generation and channel blocks allows for well-calibrated evaluation of the performance of the different demodulators. Both modulated and unmodulated co-channel interference may also be synthesized. For digital signal demodulation, fixed or adaptive equalizers can be selected -- either for benchmark or test purposes -- with a "menu" allowing the modification of various real-time parameters including data and sampling rate and internal or external signal interfaces.

3 Blind Demodulator Testing

In this section, we discuss different areas of investigation for which the testbed is employed. These areas are currently subdivided between techniques that address interference suppression and/or cancellation for analog modulations (e.g. FM, AM, SSB) and algorithms that provide solutions to

blind demodulation of digitally bauded signals in the presence of frequency selective fading.

3.1 Interference Suppression Tests

The first suite of testbed implemented interference cancellation applications concentrated on the use of cross-coupled phase locked loop designs [1-3] to mitigate CW and modulated interference upon FM signals. While these implementations proved fruitful for demodulating narrowband FM in an additive CW environment, modulated interferers produced significantly worse performance. The testbed has since been employed in comparative testing of other systems such as the amplitude locked loop [4] and is currently being programmed with a trellis based phase detection technique [5].

3.2 Blind Adaptive Equalizers

Among the more robust methods that also exhibit fairly low processing complexities are the time recursive formulations of CMA, a modulus restoral technique [6], and a cumulant-based signal restoral approach [7]. These blind equalizers represent baseline approaches for the purposes of performance comparison. These and others under development demonstrate an insensitivity to frequency offsets which allows equalization without full carrier recovery. A new class of cost functions was derived in [9], which extended the real-valued, variable norm cost function, originally developed for geophysical deconvolution, to blind complex equalization. One example from this class that is particularly suited for high order square constellations is the $O^r_{s,RECT}$ cost function, (1) and (2). The variables r and s are algorithm parameters that control the adaptation properties (typically, r=2, s>>r).

$$O^r_{s,RECT} = \frac{\left\{\frac{1}{N}\sum_{k=1}^{N}|z_k|^r\right\}^{1/r}}{\left\{\frac{1}{N}\sum_{k=1}^{N}\left(|\Re(z_k)|^s+|\Im(z_k)|^s\right)\right\}^{1/s}} \quad (1)$$

$$O^r_{s,\mathrm{RECT}} \xrightarrow{s\to\infty} \frac{\left\{\frac{1}{N}\sum_{k=1}^{N}|z_k|^r\right\}^{1/r}}{\max\left\{|\Re(z_i)|,|\Im(z_i)|\right\}_{i=1}^{N}} \quad (2)$$

In (1) and (2), z_n denotes the complex equalizer output time samples and N represents a batch sample block size. The maximization of this cost function yields blind signal recovery and a measure of phase coherency. For real-time implementation, we have developed a time-recursive form (3), where $\mathbf{w}_n$ represents the linear equalizer weight vector and $\mathbf{r}_n$ corresponds to the elements of the equalizer tapped delay line.

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu\left\{\Re(z_n)\cdot|\Re(z_n)|^{s-2} + i\cdot\Im(z_n)\cdot|\Im(z_n)|^{s-2} - R_{ORECT}z_n\right\}\mathbf{r}_n^* \quad (3)$$

Real-time demodulator results for both 64- and 256-QAM constellations are shown in Figures 6-7 and 8-9, respectively. Figure 5 details the equalizer-demodulator configuration utilized in conjunction with the $O^r_{s,RECT}$ algorithm. Unlike more standard approaches [8], the PLL is positioned prior to the equalizer, as the greatest performance improvement achieved by the new algorithm occurs in the low frequency offset regime. In Figures 6 and 8, the mean square demodulator error is plotted for CMA, SW and $O^r_{s,RECT}$ equalizers. A significantly lower MSE is achieved by the new algorithm, but at the cost of convergence time. To investigate this behavior, the new equalizer was re-evaluated with an increased step size resulting in the performance plots of Figures 7 and 9. Significant improvement is still demonstrated over the conventional equalizers in both acquisition time and residual demodulator error.

3.3 Blind MLSE Demodulators

In order to treat the problem of data detection in the presence of spectral null channels, a joint channel and data sequence estimator was implemented in the testbed using the PSP algorithm [10]. Generically, an MLSE receiver attempts to minimize the Euclidean distance between the observed signal vector and data symbols convolved with an estimated channel (4). The surviving data sequence selection is mechanized through the use of the Viterbi algorithm. For PSP, the channel estimates are updated recursively, *for each state*, according to (5), where $\mathbf{h}(\mu)$ represents the channel estimate vector associated with each state and β is an adaptation rate step size.

$$\min_{\hat{\mathbf{a}}_N}\left|\mathbf{r}_N - \hat{\mathbf{H}}(\mathbf{r}_N,\hat{\mathbf{a}}_N)\cdot\hat{\mathbf{a}}_N\right|^2$$

where, $\mathbf{r}_N \equiv$ *observations,* $\hat{\mathbf{H}} \equiv$ *channel estimates,* $\hat{\mathbf{a}}_N \equiv$ *data estimates* (4)

$$\hat{\mathbf{h}}(\mu_k) = \hat{\mathbf{h}}(\mu_{k-1}) + \beta \cdot \varepsilon(\mu_{k-1} \rightarrow \mu_k) \cdot \hat{\mathbf{a}}^*(\mu_{k-1} \rightarrow \mu_k)$$
$$\text{where,} \quad \varepsilon(\mu_{k-1} \rightarrow \mu_k) = r_k - \hat{\mathbf{h}}(\mu_{k-1}) \cdot \hat{\mathbf{a}}\ (\mu_{k-1} \rightarrow \mu_k) \quad (5)$$

Quantitative test results are shown in Figures 10 and 11 which show data detection performance for linear equalizers and a blind data sequence estimator, respectively. The modulation type is QPSK with 100% root Nyquist pulse shaping transmitted at a symbol rate of 5 kHz and passed through a two tap, spectral null channel. In Figure 10, the steady-state (after convergence) equalizer and symbol-by-symbol detector error rate is shown as a function of SNR. In Figure 11, different ensemble averaged acquisition curves are shown for four different SNR's. The first point to note is that the blind linear equalizer convergence time was observed to be 5 to 10 times that of the PSP demodulator – even with PSP channel estimates initialized to zero. If the detection performance is also compared, for an 8 dB SNR, the detected symbol error rate is one to two orders of magnitude better.

4 Conclusions

A real-time, digitally based testbed has been developed for evaluating demodulation algorithms in communications systems. In addition to serving as a prototyping tool for algorithm implementation, its integrated design affords the user an unbiased capability in the evaluation of new receiver designs and demodulation techniques.

5 Acknowledgment

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

6 References

[1] F. Cassara, H. Schachter, G. Simowitz, "Acquisition Behavior of the Cross-Coupled Phase-Locked Loop FM Demodulator," *IEEE Trans. Comm.*, v. COM-28, pp. 897-904, 1980.

[2] Y. Bar-ness, F. Cassara, H. Schachter, R. DiFazio, "Cross-Coupled PLL Interference Canceller with Closed Loop Amplitude Control," *IEEE Trans. Comm.*, v. COM-32, pp. 195-199, 1984.

[3] G. Zimmerman, "Applications of Frequency Modulation Interference Cancellers to Multiaccess Communications Systems," Ph. D. dissertation, California Institute of Technology, Pasadena, CA, 1990.

[4] A. Pettigrew et al., "Test Results for the Ampsys FM-201/5 High Performance Demodulator," Product Notes, Ampsys Laboratories, University of Paisley, Paisley, Scotland.

[5] J. Hamkins et al., "A Comparative Study of Co-Channel Interference Suppression Techniques," *Proceedings of the Fifth International Mobile Satellite Conference (IMSC '97)*, JPL 97-11, Jet Propulsion Laboratory, Pasadena, CA, 1997.

[6] D. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Trans. On Comm.*, Vol. COM-28, No. 11, pp. 1867-1875, November 1980.

[7] O. Shalvi and E. Weinstein, "New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)," *IEEE Trans. on Info. Theory*, Vol. 36, No. 2, pp. 312-321, March 1990.

[8] N. Jablon, "Joint Blind Equalization, Carrier Recovery, and Timing Recovery for High-Order QAM Signal Constellations," *IEEE Trans. Sig. Proc.*, Vol. 40, pp. 1383-1398, June 1992.

[9] E. Satorius and J. Mulligan, "An Alternative Methodology for Blind Equalization," *Digital Signal Processing*, Vol. 3, pp. 199-209, July 1993.

[10] R. Raheli, A. Polydoros, C. Tzou, "The Principle of Per-Survivor Processing: A General Approach to Approximate and Adaptive MLSE," *IEEE Trans. on Comm.*, Vol. 43, Nos. 2-4, Part 1, pp. 354-364, February-April 1995.

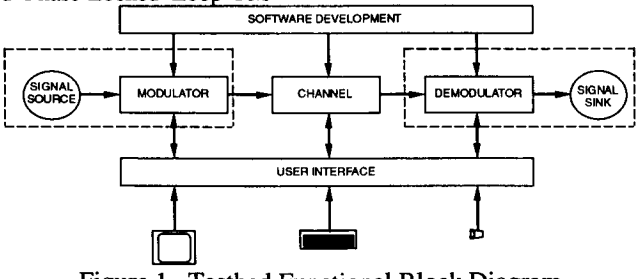


Figure 1. Testbed Functional Block Diagram

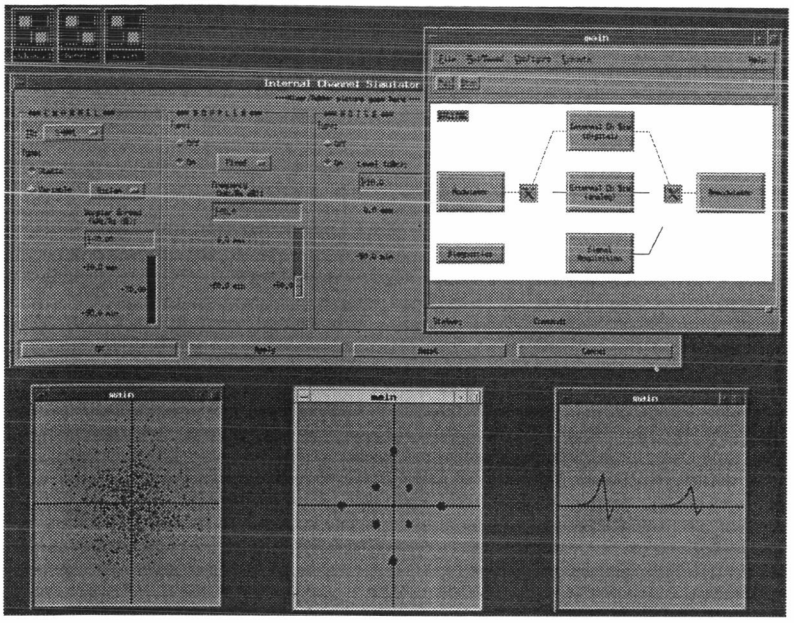


Figure 2. Motif/X-Windows Based Graphical User Interface

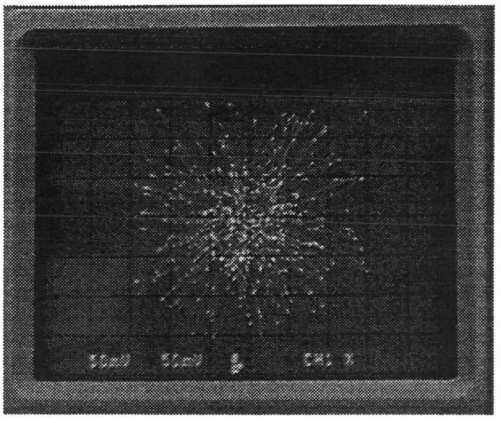


Figure 3. Multipath Distorted 8PSK Signal

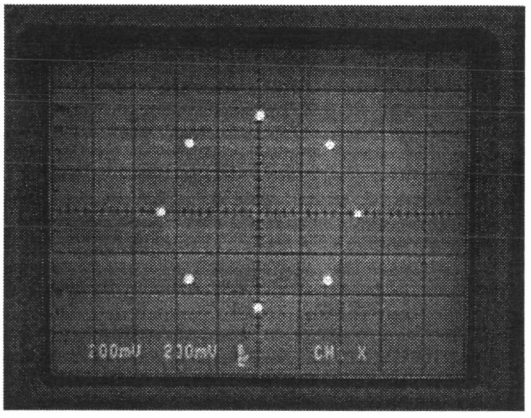


Figure 4. Recovered/Equalized 8PSK Constellation

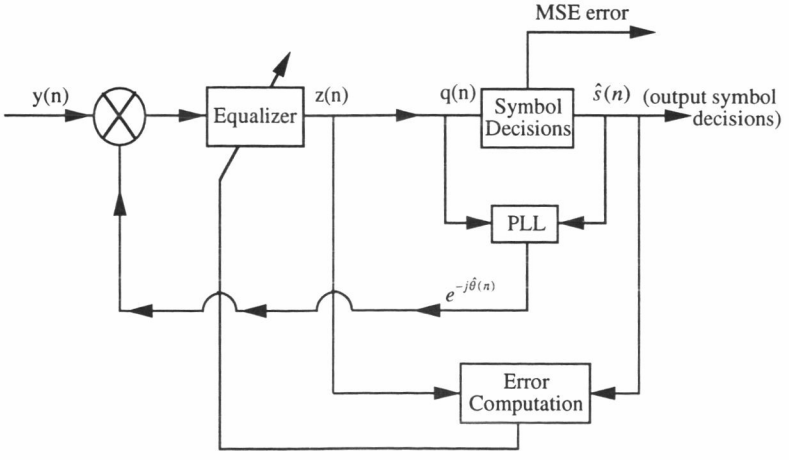


Figure 5. O-rs RECT Blind Equalizer Configuration

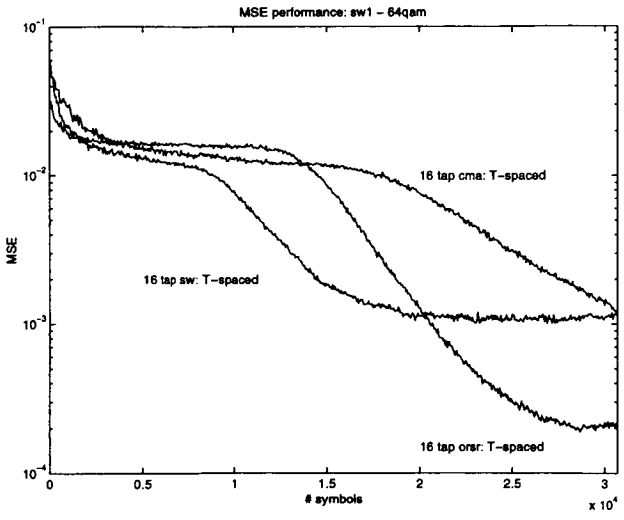


Figure 6. 64-QAM Ors-RECT vs. CMA/SW

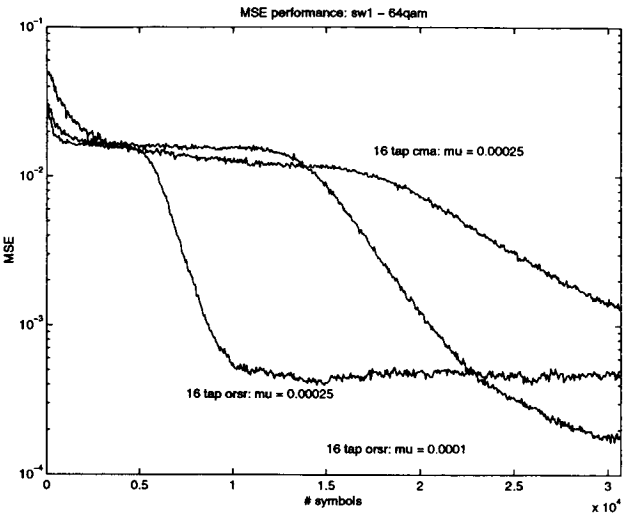


Figure 7. 64-QAM Ors-RECT Varying Step Sizes

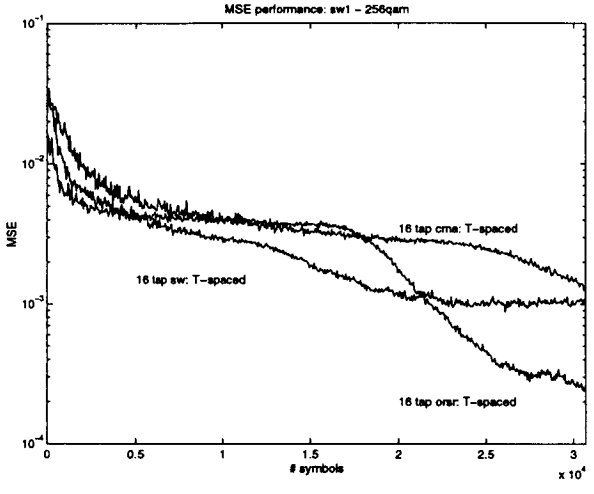


Figure 8. 256-QAM Ors-RECT vs. CMA/SW

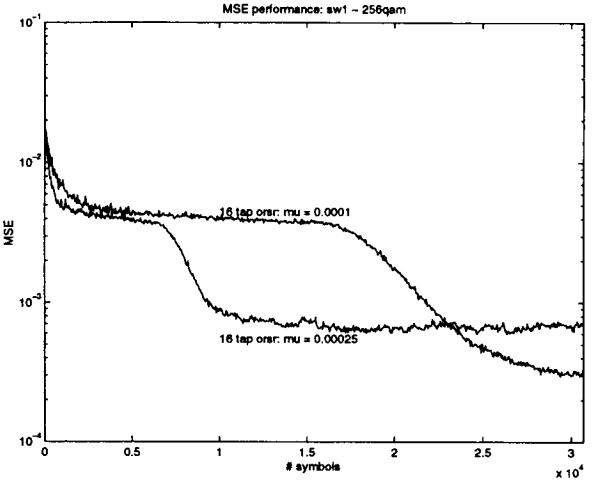


Figure 9. 256-QAM Ors-RECT Varying Step Sizes

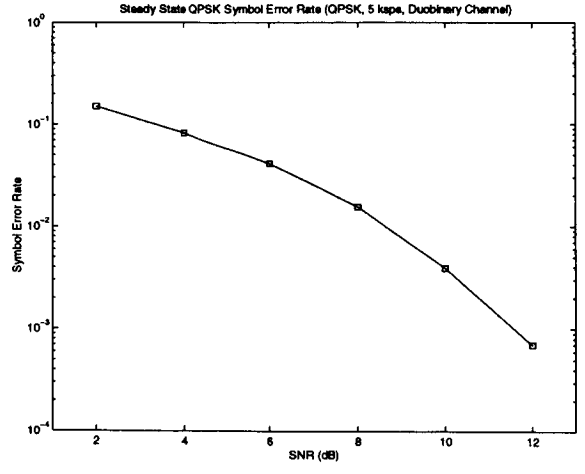


Figure 10. QPSK Linear Equalizer Performance

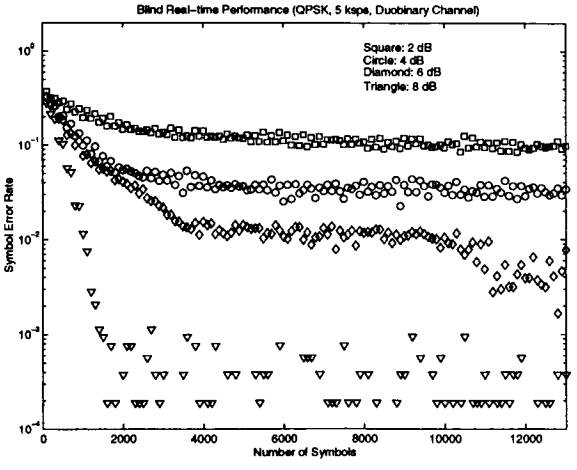


Figure 11. QPSK PSP Acquisition Performance

Session 14
Applications and Experiments

Session Chairperson—*Peter Rossiter*, Calian Communications Systems Ltd., Canada
Session Organizer—*Dave Halayko*, Communications Research Centre, Canada

EUTELSAT
DEVELOPMENT AND PLANS
IN MOBILE SATELLITE COMMUNICATIONS

IMSC '97

By Mr Laurent Vandebrouck

EUTELSAT

Mobile Services Manager

70, Rue Balard

75015 Paris - FRANCE

Phone: +33 1 53 98 46 89 & Fax: +33 1 53 98 46 64

email: lvandebr@eutelsat.fr

EUTELSAT, the leading European satellite operator, launched in 1992, in partnership with ALCATEL, the EUTELTRACS Service in Europe. EUTELTRACS is today's first mobile communication service via satellite dedicated to enhancing the productivity of fleet managers. EUTELTRACS is a service fully tailored for trucking operators and ship owners, offering turnkey management software solutions, through the simultaneous provision of five complementary services available via a single robust and compact on-board communications terminal:

- Location of a vehicule or boat with an accuracy of 100 metres;
- Message exchange between a mobile and its base and/or between mobiles;
- Data collection/transfer from the mobile to its base;
- Transmission of alarm and distress messages from the mobile;
- Access by any mobile to external databases.

The advantages of the system are:

- A tried and tested on-board technology, with more than 175,000 trucks and boats equipped world-wide, of which more than 15,000 are in Europe;
- Tailor-made software mapping, data processing and message handling solutions integrated into the client's management system;
- Fast, confidential and secure transmission;
- Round-the-clock support service;
- Facilities to send messages to the mobiles from fax or telex machines.

The applications of the system are:

- Real-time fleet management across Europe;

- Awareness of load, mobile and driver status from the base;

- Cheap, confidential and secure communication between mobiles and between a mobile and its base or any other authorized third party;

- Access by mobiles to external databases, such as quayside fish market prices;

- Transmission of data to mobiles (e;g; meteorological data);

- Access by mobiles to value added services such as telex and fax services;

- Transmission of distress and alarm messages from the mobile to its base, as well as to other mobiles and to the closest search and rescue centre.

The benefits of the system are:

- Increased productivity (reacting faster than the competition);

- Greater flexibility (orders taken in real-time with identification of the mobiles closest to a new destination and awareness of the current loads);

- Improved customer service (ability to inform customers of the exact time of delivery);

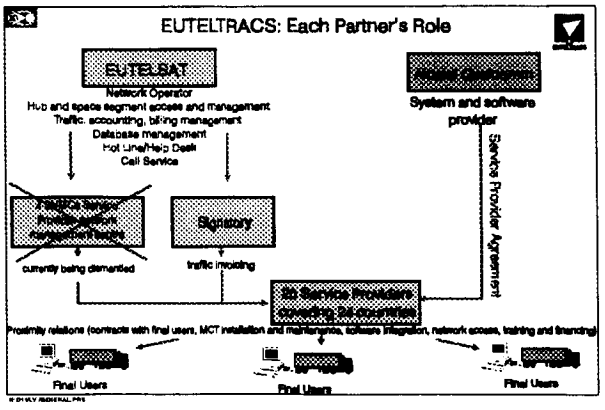
- Faster invoicing;

- Increased security (fight against theft and secure transmission of alarm messages);

- Improved quality (remote checking of crucial parameters such as temperature, vibration and possible leakage);

- Preventive maintenance of the mobiles.

EUTELSAT, which is the satellite network operator, and Alcatel Qualcomm, which is the supplier of the hardware in Europe, developed a network of 20 service providers present in over 24 countries in the course of the last five years to develop, install and implement the Fleet Management solutions. These service providers are all professionals specialised in the world of transport and logistics, they are the unique interface for the customer and can therefore produce turnkey management solutions adapted to the specific needs of a customer in order to optimize his fleet management and increase competitivity.

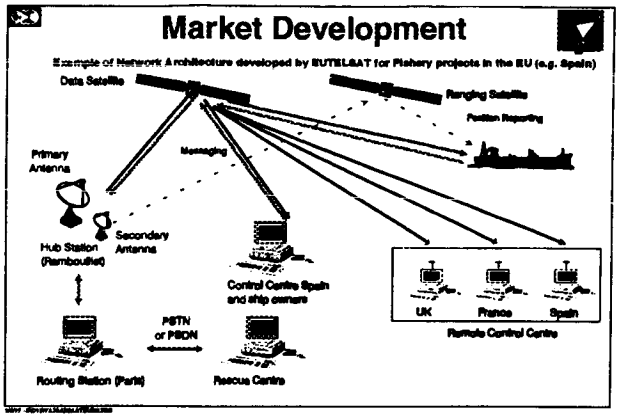


Following the success of the system in the land mobile communication market, EUTELSAT has recently decided to develop new applications and to open the solution to complementary technologies and services. This decision was made in order to maintain the competitivity and growth of the service in the coming years mainly with GSM data but also with other competing systems which will soon be operational in Europe. These recent developments are presented below.

Indeed, EUTELSAT and Alcatel have recently decided to develop and implement new system functionalities and new software solutions. These new solutions and applications were adopted in order for the EUTELTRACS system to remain competitive and maintain the service growth around 40 to 50% per year.

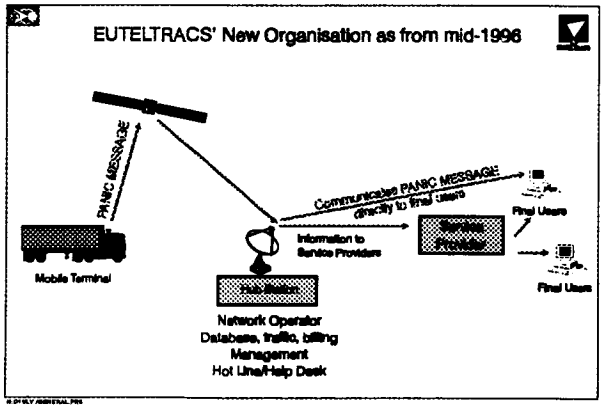
The new system solutions and functionalities which were introduced in the course of 1996 have been the following:

1. The introduction of a new maritime version of Geotrek software, developed and adapted to the specific needs of the fishermen & the ship owners. This software is not only used by leading ship owners and independent fishermen but also by several European Control Centres for Fishing operations and supervision. The French government, using Geotrek for Fishing boat operation surveillance has recently authorised the use of the EUTELTRACS system for the fishing zone changing declarations after a successful trial of 16 intensive months. Also, the system has been installed on board several fast patrol boats belonging to the French Maritime Affairs which will be able to take quick action when locating a vessel which has entered their control zone without declaring itself.



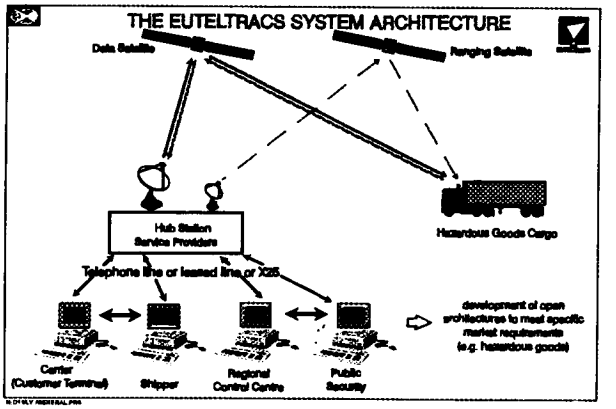
The Geotrek software includes detailed maps of the fishing areas and the delimitation of national fishing zones. It is very user friendly, simple and fast for the fishermen or the ship owner at the base. It can be used for double-hop communications and ship-to-ship communications, as well as for automatic declarations when changing zone and as a fax-telex interface.

2. The EUTELTRACS Call Service provides customized and immediate handling of all distress and panic messages received at the Rambouillet HUB station.



When a distress or panic message is received, a EUTELSAT operator will immediately contact both the fleet manager and the closest search and rescue centre by telephone and fax, transmitting the content of the message and giving the position of the mobile. The service has proved extremely helpful: when tracking down a stolen truck and helping the police to dismantle the organised crime band, or when enabling to call the rescue services when a fisherman has injured himself badly and needs to be rescued.

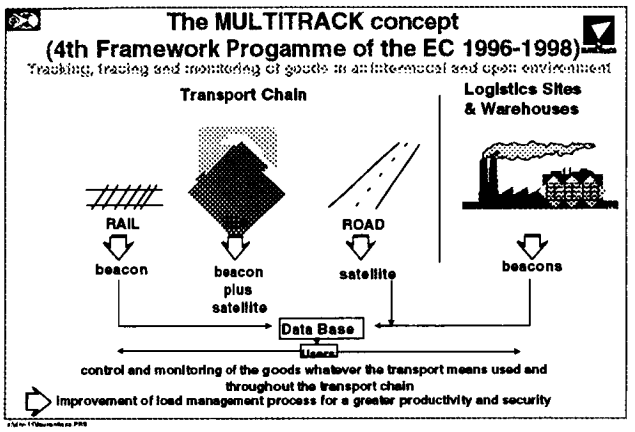
3. QMASS (Qualcomm Multiple Access Software Service) is an application which enables a third party to receive automatically all or part of the data transmitted by a mobile. This service is particularly sought after by shippers working on just-in-time delivery, for example the motor industry which needs to know the condition, status and probable time of arrival of a load containing spare parts. As a result, in 1996, a number of road haulage operators were able, in 1996, to successfully bid for RFPs in the motor industry. QMASS is offered at no extra cost for the shippers.



4. At the end of 1996, a telex and fax interface was installed at the Hub allowing any mobile to send a message to a third party which is not equipped with any of the EUTELTRACS softwares but which does have a telex or a fax machine. Therefore when the customer, mainly the shipowners, opts for this service he shall be able to open his system and communicate with people outside the EUTELTRACS system.

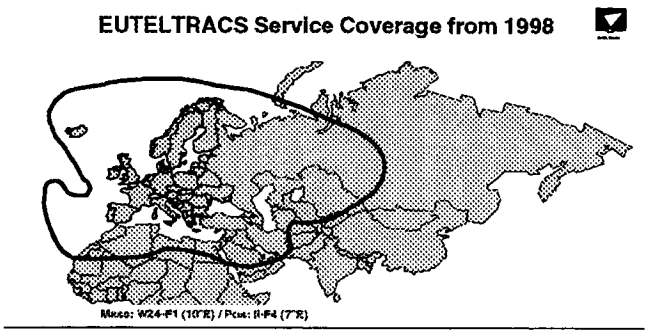
In order to maintain its competitivity in 1997, especially when faced by new and replacement technologies such as GSM Data, EUTELSAT has decided to continue its policy of investing in EUTELTRACS. For example, in conjunction with 14 partners and backed by the European Community financing, EUTELSAT is developing

the MULTITRACK project. This service, by the end of 1997, will allow transporters and shippers to identify at any time the position and condition of their load throughout the whole logistic chain regardless of the means of transport being used (terrestrial, maritime, rail, etc.), in the most transparent way. To achieve this, EUTELSAT is developing the necessary points of interface with other information systems and technologies to complement EUTELTRACS and allow the



client to monitor his load in a multi-modal environment by accessing a unique database in Europe. This is EUTELSAT's load management solution for multi-modal transportation in Europe integrating EUTELTRACS with complementary terrestrial tracking and tracing systems for example passive technologies such as Amtech.

Also, regarding the increasing need for long distance trucking companies to go further East as well as their geographical development in these regions, EUTELSAT has decided, in order to satisfy existing and potential customers, to put the EUTELTRACS service on to the EUTELSAT W1 satellite from early 1998 so as to extend the existing coverage - currently limited to the East by the Urals - as far as Omsk and Alma-Ata.



In preparation of this event and in view of the substantial improvement to the geographical coverage of the service as of 1998, EUTELSAT is currently seeking partners in the Middle East in order to introduce new service providers.

Together, the above developments are expected to allow EUTELSAT to reach its target of 18,000 EUTELTRACS-equipped mobiles by the end of 1997.

Considering the coming launch of global and personal voice communication systems by satellites which will also enter the fleet management market with very cost effective messaging and fleet management solutions, EUTELSAT is analysing and preparing the launch of new technology based systems for mobiles in Europe.

High Data Rate Point-to-Multipoint Transmission Over Mobile Satellite Channels: Aeronautical and Maritime Field Trials

M. Gertsman and R.G. Lyons, Square Peg Communications, and
G. Kinal, Inmarsat.

Abstract

There has been significant interest in higher data rate services to mobile platforms in general, and satellite-serviced mobile platforms in particular. An investigation of high data rate point-to-multipoint transmission over mobile satellite channels has been undertaken, resulting in the deployment of field trial systems for the maritime and aeronautical environments. This paper gives an overview of the transmission system design and describes the configuration and objectives of the trials. Early results are presented.

1.0 Introduction

The previously separate industries of telecommunications, information technology, media and consumer electronics are undergoing significant change as the boundaries between them are reexamined. This convergence has been hastened by the recent explosion of Internet-based applications focused on delivering text and multimedia content via the World Wide Web. This in turn has been made possible by the development of sophisticated digital compression techniques to allow the transmission of audio and video at relatively low data rates. This technology can now be exploited to deliver the same types of services over the power and bandwidth limited mobile satellite channel.

This paper describes the Inmarsat Data Distribution Service (IDDS), an integrated system of multipoint transmission facilities that supports data transmission to mobile platforms at rates above 128 kbit/s. In particular, the configuration of a field trial system aimed at assessing the demand for and performance of the system is introduced, and the content of aeronautical and maritime field trials is described. The objectives and early results of the trials are discussed.

2.0 IDDS System Description

The IDDS is designed to serve diverse land, maritime and aeronautical mobile markets for broadcast-type services. The transmission technique used, Variable-Carrier Coded Orthogonal Frequency Division Multiplexing (VC-COFDM), is bandwidth efficient and extremely robust in the mobile broadcasting environment [1]. Each VC-COFDM channel has a maximum aggregate capacity of approximately 152 kbit/s in a bandwidth of 200 kHz. This capacity is subdivided into datagram and stream-mode services at rates from 500 bit/s to 128 kbit/s with datagram capacity further segmented into 128 individual datagram virtual circuits. Both the total capacity/bandwidth and the service mix can be varied to accommodate the service requirements at any given time. A variety of receiver configurations is possible, ranging from a low-cost audio or text-based personal receiver to a multi-channel receiver providing multimedia services.

2.1 System Characteristics

In designing the system, a number of key requirements were identified and incorporated:

1) Efficient use of satellite resources

To offer the lowest cost services possible, the system must use satellite resources efficiently. To do so, the IDDS allows capacity to be assigned among services sharing the transmission facility such that when more bandwidth is required for one service and less for another, it can be dynamically re-allocated.

Additionally, the COFDM modulation scheme is bandwidth and power efficient for mobile satellite channels and the signal has been designed such that coherent detection without differential encoding is feasible.

2) Service configurations

The system supports two basic types of services. Stream mode services are real-time data applications such as digital audio or video. Datagram services are page mode applications such as news briefs, stock quotes and weather reports. The bit rate for any service can range from 500 bit/s to 128 kbit/s.

3) Universality

The system has been designed to work across the range of environments that Inmarsat currently serves. These include maritime, aeronautical mobile, land mobile and land transportable. It is expected that this universality will lead to lower cost receivers due to the commonality of components over a larger user base.

4) Complexity

The system has been designed for the lowest operational cost that is consistent with low cost receivers. In addition, since the receiver is the system's most cost sensitive component, complexity has been transferred from the receiver to the transmitter where possible. [This also applies to the compression technology that is critical to audio and visual services.]

5) Performance

The system has been designed to provide a maximum bit error rate of 10^{-4} for all services and across all unshadowed mobile environments.

Selectable forward error-correction coding is available to provide a reduced bit error rate of 10^{-6} for error sensitive services, and repetitive transmission is used to ensure the successful reception of critical information.

Signal acquisition time after power-on or channel change is less than 11 seconds.

3.0 IDDS Field Trials

To facilitate testing and evaluation of the system and potential services, Square Peg Communications Inc. developed PC-based transmit channel equipment and radio frequency receivers. These transmitters and receivers support a maximum composite bit rate of 70 kbit/s and a maximum stream data information rate of 64 kbit/s.

Two field trials using this equipment were planned, one addressing the maritime environment and one addressing the aeronautical environment. The objectives are the trials are:

a) to clarify the requirements of the content providers and to assess the impact of the system design on them;

b) to gain insight into the terrestrial infra-structure required to convey programming from the content providers to the satellite transmission system;

c) to confirm the performance of the transmission system under actual operating conditions;

d) to determine potential mechanisms for interfacing to existing entertainment and/or multimedia systems; and,

e) to assess the reaction of users (passengers or crew) to a wide variety of potential services.

The results of the trials can be used to refine the system design and the applications programming interface (API) required to support innovative service offerings.

3.1 Maritime Field Trial

The maritime field trial was initiated by Inmarsat's Swedish signatory, Telia Mobile. It provides daily and weekly Radio Sweden audio broadcasts to an Inmarsat-A equipped merchant ship. The broadcasts are played in real time as well as stored for later playback. The trial equipment is also used to distribute daily electronic copies of the Swedish newspaper Svenska Dagbladet to the ship. Crew members can peruse the newspaper at their leisure over the ship's local area network. This trial commenced in Fall 1996.

A picture illustrating the configuration of the maritime field trial is shown in Figure 1. The downlink is monitored continually at Inmarsat and as required by Square Peg.

3.1.1 Transmitter

The field trial transmitter consists of a Pentium PC with 3 plug-in cards which implement the IDDS physical layer. This unit has the following interfaces:

a) three X.21 serial stream input ports;
b) one datagram input port (PC serial port operating at up to 56 kbit/s).

The transmitter is located at the Eik earth station in Norway and interfaces at a 70±18 MHz IF to the station's block upconverters. During the trial, operation over Inmarsat's AOR-W and IOR satellites is supported.

Monitoring and control of the transmitter is performed from Inmarsat's headquarters in London using a commercial remote control software package over ISDN.

3.1.2 Receiver

The field trial receiver consists of a 19" rack mount Pentium PC located in the radio room of the Wallenius Lines Isolde. Three plug-in cards implement the IDDS physical layer. An audio decoder card and LAN interface card are also present. In the field trial configuration, the unit has the following interfaces:

a) 600Ω balanced audio output;
b) interface to ship's onboard Novell network.

The receiver is connected to the ship's existing NERA Inmarsat-A receiver via a splitter at a 212 MHz IF access point.

The receiver has an application program interface (API) so that an application on the PC can receive datagrams and/or stream data as well as control the operation of the receiver. The maritime application presents a schedule of programs to the crew, and allows them to select which should be played live and/or saved.

The receiver also displays and logs numerous performance parameters, including estimates of:

a) E_b/N_0;
b) uncoded BER;
c) datagram packet error rate;
d) frequency offset; and,
e) Doppler rate-of-change.

This information can be used to correlate crew reports with transmission system performance.

3.1.3 Newspaper Service

Svenska Dagbladet publishes a number of editions each day. An electronic copy is made of the final edition and passed to Nobicon in a desktop publishing format. This format enables pictures, headline text and text body to be associated as an article.

Nobicon compiles this electronic newspaper (currently without pictures) into an EDIFACT

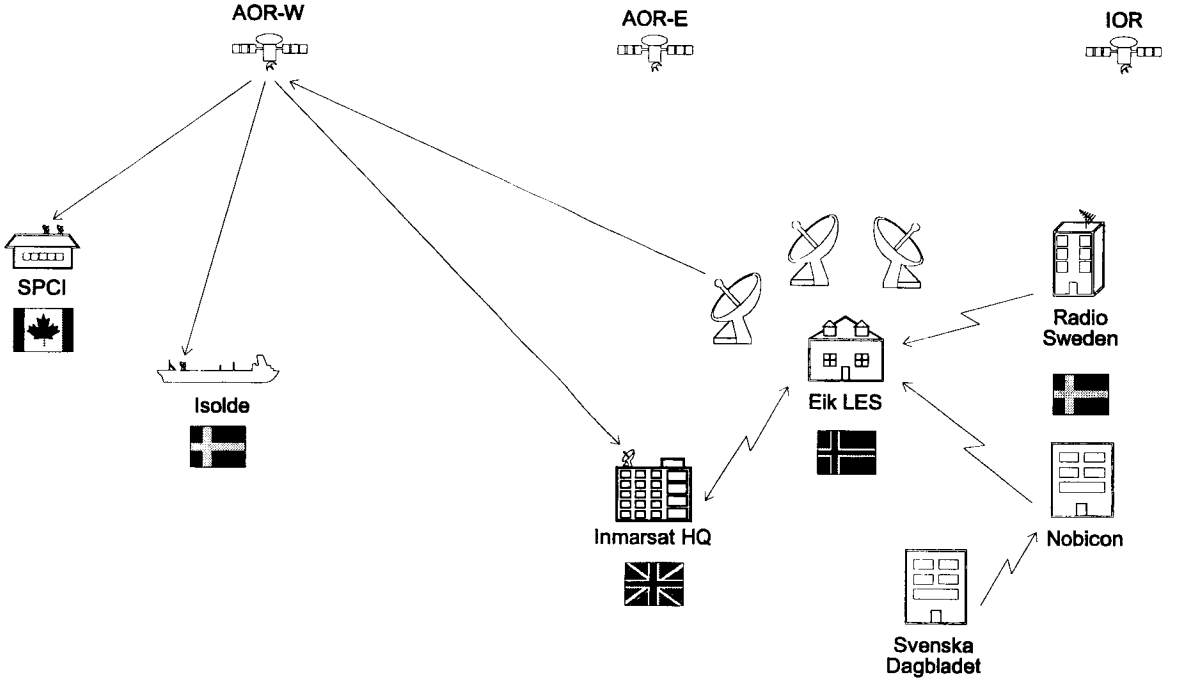


Figure 1 Maritime Field Trial Configuration

format and is responsible for its distribution via network, modem and ISDN. Supported formats include:

a) HTML (for standard Web browsers);
b) .hlp file (for the Windows help utility); and,
c) Lotus Notes database entries.

For transmission to the ship, the EDIFACT file is processed to create a new one compatible with the file transmission protocol of the application and with the datagram protocol of the transmitter. At the scheduled program transmission time, the transmitter reconfigures the ensemble to the required capacity. At the same time, a computer at Nobicon makes an ISDN call to the LES. A terminal adapter with rate adaptation interfaces the ISDN line to the transmitter PC's serial port input.

At the receiver, the application detects that the program is about to start and requests the datagrams for the appropriate virtual channel. The datagrams are assembled into a file which is placed in a directory accessible via the LAN.

When the scheduled timeslot is complete, the transmitter reconfigures the ensemble for the next required service mix.

3.1.4 Radio Service

The Radio Sweden transmissions are existing programs which are originated and encoded by their digital production system. For the trial, monaural MPEG 2 Layer 3 coding with a sample rate of 24 kHz and bit rate of 64 kbit/s was used.

When a program is scheduled to start, the transmitter stream input port asserts its Data Set Ready (DSR) output and prepares to reconfigure the ensemble. At the same time, an ISDN call is placed from Radio Sweden to the transmitter site and transmission of the 64 kbit/s audio data stream begins.

At the receiver, the application detects that the program is about to start and requests the stream data for the appropriate channel. If live playback is enabled for the timeslot, it redirects this stream to the hardware decoder. It also saves the bit stream to a file on the hard disk for later playback upon operator request.

At the end of the scheduled timeslot, the ISDN call is cleared, the ensemble is reconfigured, and the receiver closes its file.

3.2 Aeronautical Field Trial

The aeronautical field trial will be conducted by the British Broadcasting Corporation (BBC) and the Voice of America (VOA). It will initially deliver real-time audio and text news to passengers on a British Airways 747-400 aircraft. As the trial progresses, other multimedia applications, including compressed video, may be evaluated. The Aero-IDDS receiver interfaces at L-band to the existing Inmarsat Aero-H Earth Station (AES) antenna and via audio/video feeds to the in-flight entertainment system. This trial is scheduled to commence in Spring 1997 and last for six to nine months.

A picture illustrating the configuration of the aeronautical field trial is shown in Figure 2.

3.2.1 Transmitter

The field trial transmitter is identical to that used in the maritime trial. It is located at the British Telecom's Goonhilly earth station.

The earth station can see three of Inmarsat's satellites - AOR-W, AOR-E and, experimentally, IOR. Subject to spectrum planning constraints, the 70 MHz feed can be split and sent simultaneously over more than one satellite, allowing the use of a single transmitter for more than one ocean region (the signal structure is such that the receiver can operate in the presence of multiple transmissions of the same signal on a single frequency).

3.2.2 Receiver

The IDDS Aeronautical Receiver is an avionics qualified Line Replaceable Unit (LRU) which packages the receiver PC in an ARINC 8 MCU enclosure. It is designed to operate with an Inmarsat SDM compliant high gain antenna (G/T = −13 dB/K). Internally, it is very similar to the maritime receiver except that its input is at L-band, there is no LAN connection, and its video output is a standard NTSC signal. All interfaces to the aircraft are via a standard ARINC connector in the rear of the unit.

Figure 3 illustrates the installation configuration for the trial aircraft. The only connection between the IDDS receiver and the AES is made via a directional coupler which splits the L-band

feed from the LNA(s) to both the IDDS receiver and the Satellite Data Unit (SDU) of the AES.

The remaining interfaces from the IDDS Receiver to the aircraft are the AC power input and the audio and video outputs to the In-Flight Entertainment (IFE) system.

Upon powering up, the receiver will typically:

a) acquire an IDDS carrier from the satellite to which the antenna is pointed;
b) select the desired service(s) from among those available on that carrier;
c) store service programming to disk for later playback; and/or,
d) output services in real-time to the video and audio outputs of the system.

The operation of the receiver is controlled by script files and schedules which can be uploaded over the satellite. If the AES moves the aircraft's antenna to point at a different satellite, the IDDS receiver will recognize the loss of the signal and then repeat the above steps to acquire the new satellite (the receiver has no explicit indication from the AES of where the antenna is pointing).

The receiver also logs numerous performance parameters which can be used to correlate crew reports with transmission system performance.

3.2.3 Audio Service

The BBC and VOA are sharing an audio channel which will carry the BBC World Service and VOA news and news magazine programs. The purpose is to present updates on major news stories as near real-time as possible. MPEG 2 Layer 3 coding at 16 kbit/s is used.

The audio is encoded at the BBC's London headquarters and then transmitted over their existing European distribution system which utilizes capacity on a Eutelsat satellite. A 62 kbit/s pipe is available for the trials, allowing for the 16 kbit/s audio channel plus an additional 46 kbit/s capacity for other applications.

The signal from the Eutelsat satellite is downlinked to a demodulator and demultiplexer located at the BT Goonhilly earth station. The demultiplexed data streams are fed to the IDDS transmitter's stream ports.

At the receiver, the logic in the script file selects the audio service whenever it is available. The received data stream is passed directly from the demodulator to the audio decoder and then played via the receiver's audio output interface. From the passenger's perspective, the service appears simply as another channel from the aircraft's In-Flight Entertainment (IFE) system.

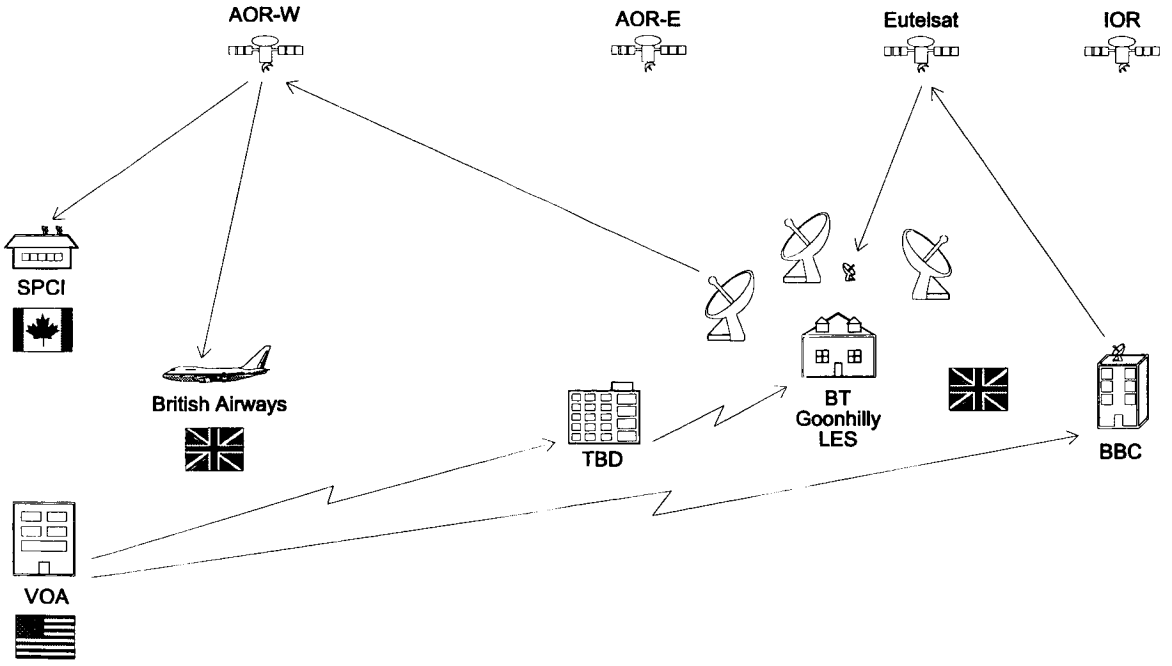


Figure 2 Aeronautical Field Trial Configuration

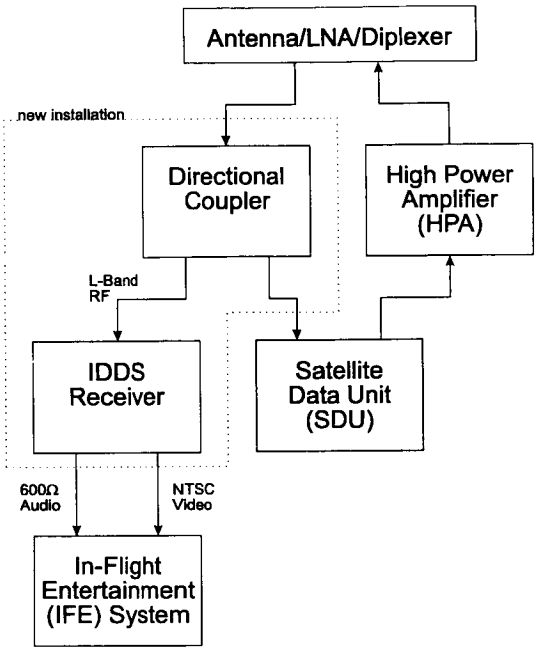


Figure 3 Aeronautical Receiver Installation

While audio is being received, the seatback video display will display stored features, or the VOA text service described below.

3.2.4 Text Service

VOA is providing a text "headline news" type service. News stories appearing on the wire feed from VOA correspondents are sent via the Internet File Transfer Protocol (FTP) to a gateway PC. From there, they are processed by a file transfer application and then sent to Goonhilly. At Goonhilly, the data is presented to the transmitter's datagram input port.

The datagram service will be allocated 1000 bit/s of capacity. News stories will be transmitted in a circular order, and the transmission system will utilize multiple transmissions of packets to ensure error-free reception.

At the receiver, the news files are stored on the hard disk and then displayed in sequence. Out-of-date stories are discarded. Before it is displayed, each news story is searched for keywords which can be used to determine the type of story (for a themed news presentation), and which of a predefined set of backgrounds should be displayed concurrently with the text.

3.3 Preliminary Results

At the time of submission of this paper, only the maritime trial was underway. Lessons learned from the early phase of this trial include:

1. It is important to synchronize the IDDS system time to a precise clock (e.g. GPS). This is crucial given that the most important news stories usually lead off a broadcast.
2. The design of the terrestrial infrastructure for delivering the content to the transmission system is critical. Careful consideration needs to be given to the synchronization and rate adaptation issues as well as to schedule coordination.

As might be expected, the overall performance of the system is dependent upon all of the links in the chain. Problems encountered to-date have mostly been with the terrestrial infrastructure and the interfaces to the content providers. Overall, the reaction of the ship's crew has been positive and the quality of the received audio excellent.

4.0 Next Steps

The priority at this time is to get the aeronautical field trial operational. Concurrently, the feasibility of adding real-time video to the aeronautical trial will be investigated.

In addition to assessing the technical feasibility and suggesting improvements to the technical approach, an important goal of the field trials is to gather the marketing data necessary to determine the viability of potential commercial services. Consequently, another important near-term task is to design and carry out surveys to assess crew and passenger reaction to demonstrated and potential IDDS services. The results will be valuable input for those considering the distribution of multimedia programming to mobile terminals.

5.0 References

[1] W. Zou and Y. Wu, "COFDM: An Overview" in *IEEE Trans. on Broadcasting*, Vol. 41, No. 1, March 1995, pp. 1-8.

K/Ka-Band Aeronautical Experiments

Martin J. Agan, Michael J. Connally, Thomas C. Jedrey

Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, California, USA 91109
Phone: (818) 354-3426, Fax: (818) 354-6825, email: agan@jpl.nasa.gov

ABSTRACT

This paper discusses a series of aeronautical experiments that utilize the Advanced Communication Technology Satellite (ACTS) Broadband Aeronautical Terminal (BAT) [1]. These experiments were designed to explore the uses of K- and Ka-band for aeronautical applications. Planned experiments are also discussed.

INTRODUCTION

A natural extension of the ACTS land-mobile experiments was to investigate the aeronautical K/Ka-band communications channel. The BAT operates at higher data rates than the land-mobile terminal, and demonstrates and characterizes the performance of high data rate aeronautical Ka-band communications [2]. The experiments performed with the BAT are described in this paper.

EXPERIMENTS

The BAT is designed to explore the use of K/Ka-band for high data rate aeronautical satellite communications. The broadband terminal, used in conjunction with the ACTS mechanically steerable antenna, can achieve data rates of 2 megabits per second. The aeronautical terminal is being utilized to test a variety of applications that require a high data rate communications link.

Experimentation with the BAT requires ACTS to be operated in the bent-pipe mode. The ACTS LA/San Diego spot beam is used to establish the communication link between ACTS and the fixed terminal at JPL in Pasadena, CA. The ACTS mechanically steerable dish antenna is used to establish the link between ACTS and the aircraft. Use of the ACTS steerable antenna introduces the additional complication of requiring the antenna to continuously track the aircraft. The ACTS steerable antenna has a 3 dB contour diameter of 280 miles, which coupled with a maximum aircraft ground speed of 700 mph, results in a low dynamic tracking requirement. This tracking is accomplished by multiplexing aircraft positioning information (GPS latitude and longitude) with the data stream transmitted from the aircraft to the fixed terminal located at JPL. At the fixed terminal the positioning information is then demultiplexed and transmitted via the public switched telephone network (PSTN) to the ACTS control station, located in Cleveland, Ohio, where the ACTS is then commanded to point the steerable antenna to the aircraft location.

Three sets of experiments/demonstrations are described in the remainder of this section.

Kuiper Airborne Observatory Live Television Broadcast

During the summer of 1995 the BAT was installed on NASA's Kuiper Airborne Observatory (KAO), a C-141A jet transport aircraft which carries a 0.9-meter reflecting telescope used for infrared astronomy. Four experiments using the ACTS were carried out from the KAO in flight: 1) "Live from the Stratosphere", a multi-media educational program which included live video/audio broadcasts on U.S. PBS and NASA TV, and a wide spectrum of Internet activities, 2) a video teleconference between the KAO and several hundred people at the San Francisco Exploratorium, called "Live Interactive Network to Knowledge" (Project LINK), 3) "telescience", i.e. control of science instrumentation on the KAO from the ground during data flights, and 4) telescope-system health monitoring from the ground during flights. Figure 1 shows the KAO and Figure 2 shows the KAO with the antenna and radome installed.

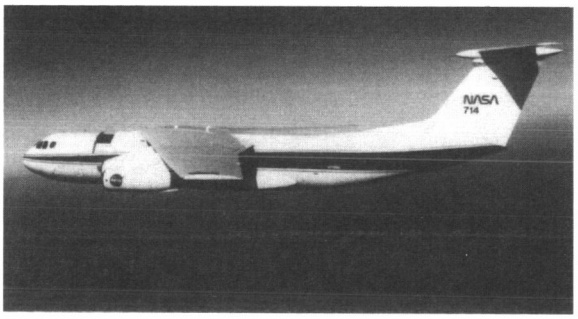


Figure 1 Kuiper Airborne Observatory (KAO) C-141

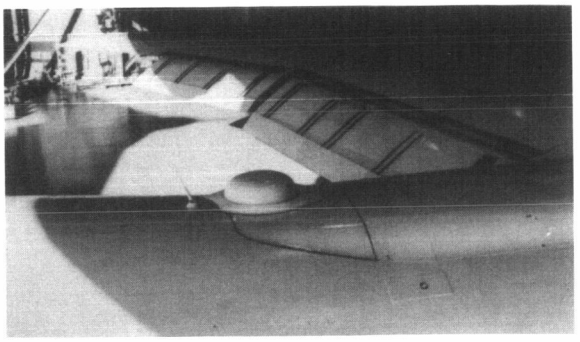


Figure 2 BAT Antenna and Radome Installed on KAO

The configuration for the KAO experiment is shown in Figure 3. As indicated in this figure the aircraft transmitted a combination of video, audio, and Internet data via ACTS to the JPL fixed station. The fixed station relayed this information via terrestrial T1 line to NASA Ames where worldwide Internet connectivity was established, and the audio and video were transmitted via commercial satellite to Public Broadcasting System (PBS) stations in Maryland and New York who incorporated them into live broadcasts. Similar data traveled on the return path from the fixed station to the aircraft. The four components of this experiment are discussed below.

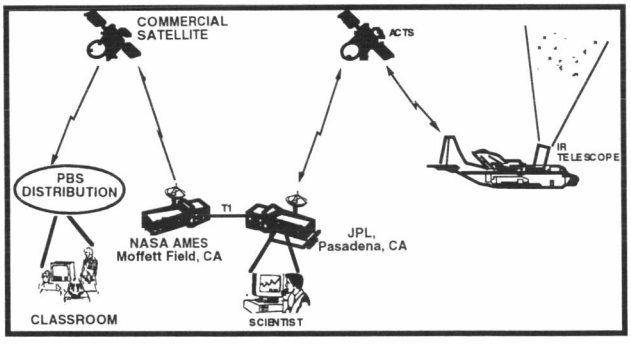


Figure 3 KAO Experiment Configuration

Live From the Stratosphere was the second of three programs in the Passport to Knowledge series, a multi-media educational program designed to take students on "electronic field trips" into the world of scientific discovery. The Passport to Knowledge series seeks to integrate various communications media into a hands-on, interactive educational experience for teachers and students.

Live From the Stratosphere activities centered on two KAO flights on October 12 and 13, 1995. During these flights, two-way video, audio, and data were transmitted, via ACTS, between the KAO and the ground. The video and audio were sent via a combination of land lines and commercial satellites to public TV stations on the East coast, where the program was produced and sent out live for broadcast on PBS channels and NASA TV. Sixty percent of PBS stations around the country broadcast some part of *Live From the Stratosphere.*

During the *Live From the Stratosphere* program, select "uplink" sites at museums and schools were given the opportunity to communicate directly with the crew and scientists aboard the KAO. Students and teachers interviewed the crew and scientists, and participated in on-line activities relating to the KAO flights. For example, students plotted the track of the aircraft on maps, using latitude, longitude and heading data that was transmitted over Internet at regular time intervals during the flights. One group of students at the Adler Planetarium in Chicago was able to control the KAO's telescope via the Internet from their computer. A scientist on-board the KAO controlled and obtained data from a telescope in New Mexico, designed for remote control.

There were 1300 email subscriptions to *Live From the Stratosphere*, and approximately 100 teachers participated in on-line discussions. There were tens of thousands of visitors on the Web site, and fifty KAO experts responded to about 250 email questions from children.

Project LINK: During two KAO research flights in September 1995, teachers and students aboard the KAO conducted "CUSeeMe" teleconferencing across the Internet with an audience of hundreds at the San Francisco Exploratorium. During the flights, educational demonstrations relating to astronomy and to the aircraft environment were conducted. People on the aircraft communicated with the audience at the Exploratorium via Internet teleconferencing, and airphone. The excitement and enthusiasm of the crowd at the Exploratorium was tremendous, as kids had the opportunity to interact in real-time with people aboard the KAO.

Remote Observing: During research flights in September, and during the Live From the Stratosphere flights in October, scientists from the University of Chicago's Yerkes Observatory demonstrated remote control of systems on the aircraft via the ACTS communications link. Activities included control of the telescope system, transmittal of science and telescope data between the aircraft and the ground, and "CUSeeMe" teleconferencing with audio and video capability. In addition, a telescope in New Mexico was operated from the KAO in flight.

Real-time Automated Diagnosis System (RAD): The RAD is a failure monitoring system that receives data from KAO telescope subsystems and uses the data to diagnose and report system problems to an operator on the ground. Communication was over the Internet, via the ACTS link, during KAO research flights in August, 1995. The purpose of this experiment was to demonstrate the feasibility of using an automated diagnosis system for the airborne telescope, and to demonstrate the possibilities that the RAD may provide for monitoring and diagnosing telescope problems from the ground.

The Broadband Aeronautical Terminal was successfully installed on the KAO C-141 aircraft shown in Figure 1. A total of 150 hours of in-flight operation of the terminal during a four month test period were performed and the terminal was found to perform quite well during these tests. The system was able to acquire the satellite signal prior to take-off and remain locked during take-off, cruise, and landing, maintaining a full-duplex compressed video link. Measurements of signal-to-noise ratio for the signal received in the aircraft and at the fixed terminal indicate that the terminal performance is better than predicted by the link budgets, and a full-duplex T1 data rate (1.544 Mbps) could be supported.

The flight plan for a KAO test flight on August 26, 1995 is shown in Figure 4. The flight originated in Moffett Field, CA, position 1, proceeded to position 2 and so on until returning to Moffett Field after passing through position 15. An example of the data recorded during this flight over the Western United States is provided in Figure 5. This plot shows the received pilot power level in the aircraft of a pilot signal that is transmitted from the JPL fixed terminal via ACTS to the aircraft. The received pilot power level is shown during a two minute long aircraft turn in which the heading changes by 330° and the roll angle changes by -30°. The received pilot power indicates that the antenna maintained tracking during this steep roll angle change. The plot also indicates that there is signal variation of up to 0.7 dB during the turn (+/- 0.5 dB RMS for the entire data set), and as a result the antenna tracking algorithm parameters were adjusted to improve this performance for subsequent flights. Figure 6 contains freeze frames of the live compressed video that was transmitted from the in-flight KAO to the JPL fixed terminal from 41,000 feet. The frames show a scientist taking measurements with the telescope.

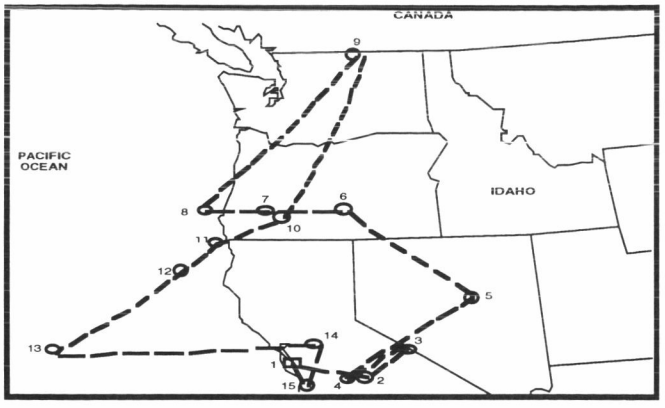


Figure 4 KAO Flight Plan August 26, 1995

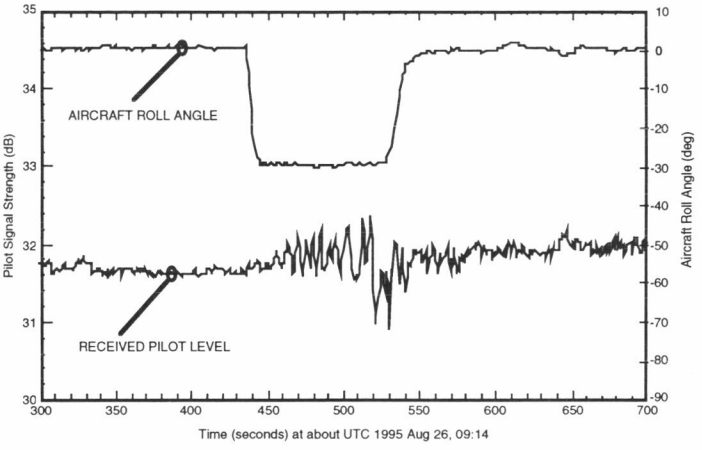


Figure 5 Flight Performance Data for KAO

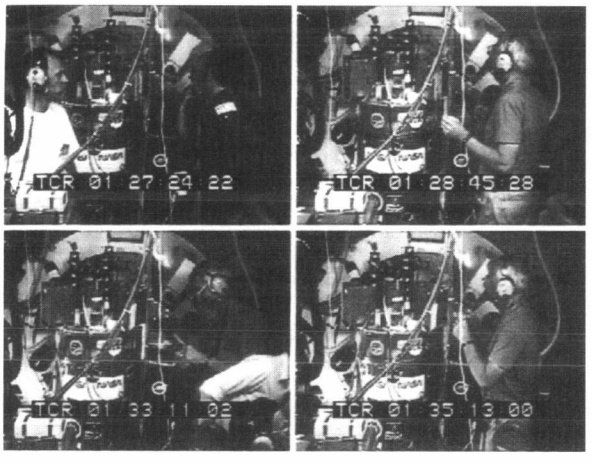


Figure 6 In-flight Video Transmitted from the KAO to JPL Fixed Terminal

The experiments carried out from the KAO using ACTS will serve as proof-of-concept for development of future satellite communication capabilities for airborne astronomy on the KAO or its planned successor, the Stratospheric Observatory For Infrared Astronomy (SOFIA).

Both the KAO and SOFIA projects are committed to hands-on educational programs which get teachers and students directly involved with science activities. The Live From the Stratosphere and Exploratorium experiments demonstrated satellite based communications capabilities for involving larger numbers of participants in research flights, via video/audio broadcasts, teleconferencing, and other Internet activities, than is currently possible due to seating and other limitations.

Results from the RAD experiment may be used in developing similar capabilities for the KAO or SOFIA. Development of a remote system health monitoring and diagnosis capability via satellite communications may allow a reduction in the number of crew members needed on board the aircraft for research flights, which is a goal for both programs.

Remote observing capabilities would open up new possibilities for scientists involved in airborne astronomy. Scientists could participate in KAO or SOFIA research flights without having to be present on the aircraft. In addition to lowering the number of people on board the aircraft, and lowering travel time and costs for the scientists, this feature would be useful to those scientists who cannot fly due to health reasons. Having immediate access to science data from the aircraft may also be helpful for time-critical events, such as the Jupiter-comet collisions that occurred in July 1994, or occultation events where world-wide coordination of science results in real-time may be important.

Rockwell International Aeronautical High Data Rate Experiments

Rockwell International flew an ACTS experiment to characterize the high speed data performance available to/from a turbojet aircraft, and more specifically a small business jet. The motivation in undertaking this experiment was several fold:

- The demand for higher data rates to/from aircraft continues to climb, and presently exceeds the technical capability to provide it. Imagery downlinks constitute most of that demand.

- Government users are primarily concerned with sensor downlinks demanding wide digital bandwidth, plus the ability to upload large databases quickly.

- The commercial airline market wants significant "office-in-the-sky" bandwidth into which multiple two-way digital signals may be multiplexed (e.g., voice, fax, data, etc.) Only limited use is foreseen for video entertainment.

As a supplier to both markets, Rockwell recognizes the potential for Ka-band service to satisfy these needs, and has undertaken an effort to establish a performance baseline to support further commercialization of Ka-band services.

The configuration for the Rockwell experiment is shown in Figure 7. As indicated in this figure the aircraft transmitted a combination of video, audio, and sensor data via ACTS to the JPL fixed station. Similar data traveled on the return path from the fixed station to the aircraft. The configuration details of this experiment are discussed below.

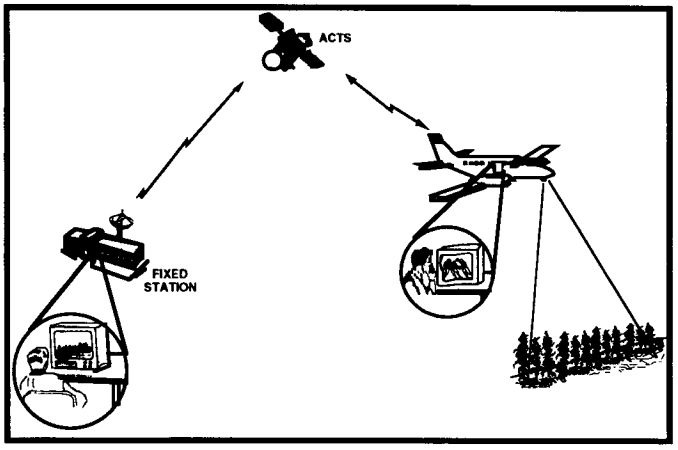


Figure 7 Rockwell Experiment Configuration

Experimental Equipment Configuration: The Rockwell experiment utilized one of the company's Saberliner jet aircraft, and more specifically the aircraft used for engineering and FAA qualification testing of various avionics. The aircraft antenna was mounted on an existing reinforced port of the aircraft approximately over the wing and just slightly off centerline. The antenna was designed to have minimal protrusion, and hence protrudes through the pressure hull adapter into the headroom of the cabin only a few inches. Careful analysis was done by Saberliner Inc. to determine the aerodynamic impact of the radome on aircraft safety and performance, and concluded that the top speed of the aircraft has to be reduced with the antenna mounted. Installation of the antenna was done by Saberliner Inc. in August 1995 and is shown in Figure 8.

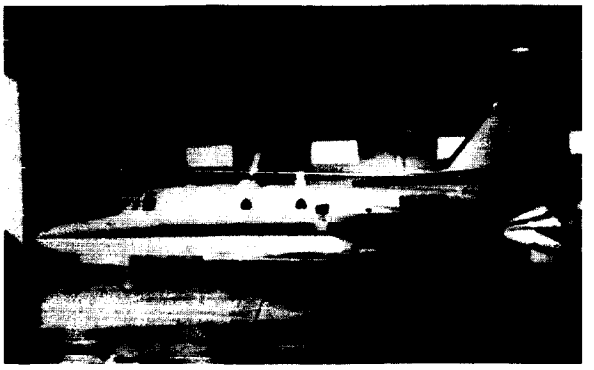


Figure 8 Rockwell Saberliner with BAT Antenna and Radome Installed

The equipment configuration installed in the aircraft by Rockwell's Flight Applications Engineering shop includes the basic JPL BAT: RF converter and waveguide, IF converter, TWT power amplifier, Modem, Video Codec, Antenna control unit, Data logger, and test equipment.

Navigation data is supplied to the antenna controller from the aircraft bus for antenna pointing, and relayed to the laptop for logging and uplinking. The equipment is mounted in a 19 inch rack pallet, which is fastened to the seat tracks in place of a passenger seat.

Flight Tests: The Saberliner flew the test missions in two sessions: August 1995 and March 1996. The tests were deliberately split to allow testing in the humid summer weather and dry winter weather of the midwestern United States. The broad goals of the testing were to characterize link performance as a function of aircraft motion/antenna pointing as well as ambient weather conditions. This was done at a variety of data rates. While precipitation attenuation is a well-known phenomenon on the ground, little detailed data is available on the airborne effects and hence this was a primary objective of this experiment.

To characterize the precipitation conditions, data from the onboard Rockwell weather data together with NEXRAD weather radar data were collected to characterize the precipitation and humidity conditions in the path between the aircraft and satellite. The aircraft flight schedule and course were set to obtain the desired weather conditions.

The final series of tests utilized a "live" television camera in the cockpit window downlinked to the ground via a video codec, replacing the digital test pattern used for the preceding tests. The complete data set for the flight test series is being analyzed to establish the link performance patterns to guide system and equipment design for the next generation of airborne Ka-band capability [3].

Flight tests found the terminal to perform quite well. The system was able to acquire the satellite signal prior to take-off and remain locked during take-off, cruise, and landing, maintaining a full-duplex data link the entire time. During these flights that used a 10 W TWTA and the ACTS steerable antenna in a fixed point mode, full duplex links of 64 kbps were established between the Saberliner flying in the Iowa/Nebraska/South Dakota area and the fixed terminal at JPL. Good signal-to-noise ratios were maintained for a variety of aircraft maneuvers and varying weather conditions in both the summer and the winter.

Figures 9 and 10 show a sampling of the flight test data that was recorded during the Rockwell experiments. Figure 9 covers a period when the aircraft was taxiing on the ground prior to take-off and the subsequent climb to an elevation of 30,000 feet. The parameters displayed include the pitch, roll, yaw, and altitude of the aircraft, the azimuth and elevation angles of the antenna relative to the aircraft, and the aircraft's received pilot signal-to-noise ratio. A point of interest is the pilot SNR which is displayed as a voltage ratio. When this scale is converted to dB-power the maximum peak-to-peak variation is found to be less than 1.5 dB. This indicates that the aeronautical antenna was able to consistently track in both azimuth and elevation and receive the transmitted signal uninterrupted for a variety of aircraft maneuvers on the ground and while climbing to altitude. Figure 10 shows the same parameters for the level flight phase following immediately after the climb to altitude. This plot again shows that the maximum peak-to-peak variation is 1.5 dB as the antenna tracks through various pitch, roll, and yaw perturbations.

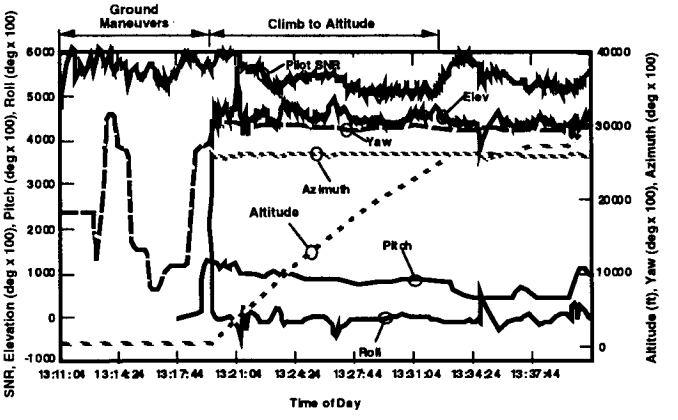


Figure 9 Rockwell Saberliner Flight Test Data for Take-off and Climb to Altitude

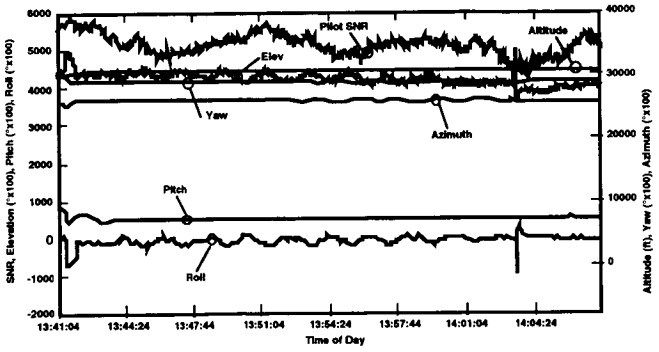


Figure 10 Rockwell Saberliner Flight Test Data for Level Flight

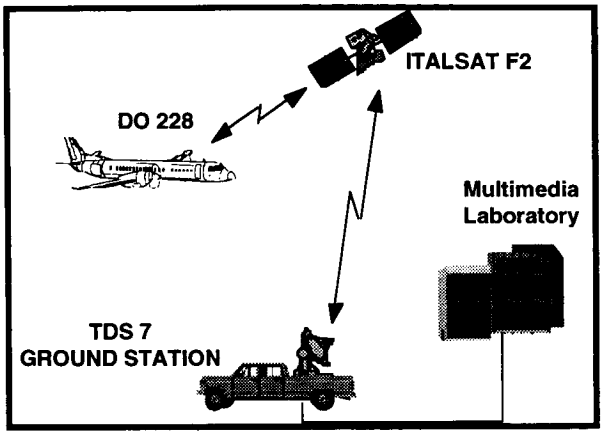


Figure 11 ABATE Experiment Configuration.

ABATE

Early in 1998, components of the BAT will be used to support flight trials as part of the European Space Agency's (ESA) Advanced Communications Technology and Services (ACTS) project, specifically the ACTS Broadband Aeronautical Terminal Experiment (ABATE). The ABATE project will assess technologies, leading to definition of a future, advanced aeronautical mobile satellite system operating in the K/Ka bands and capable of providing broadband telecommunications services as well as classical services associated with current low bit rate technologies.

Two sets of flight trials are planned but, the BAT equipment will only be used during the second set. The first set of flight trials will take place in the fall of 1997 and will not utilize the JPL equipment. During those trials, the aircraft will be equipped with a receive only system to make channel measurements and test different antenna pointing techniques. Deutsche Forschungsanstalt fur Luft und Raumfahrt eV (DLR) will modify their land mobile antenna and RF equipment for this phase of flight trials.

The second set of flight trials will be conducted in March 1998. The experimental configuration is shown in Figure 11. These flights will serve to demonstrate a full duplex audio/video/internet capability between the aircraft and the multimedia laboratory in Rome, Italy via the ITALSAT satellite. The JPL equipment (i.e., BAT antenna, RF and IF converters, support equipment) will be used during this phase of the flight trials to provide a transmit and receive capability onboard the aircraft. During these trials, channel measurements (e.g., BER, Eb/No, received power, pilot power) will be made.

Four to six flights, each of about four hours duration, will be performed. The same aircraft, DLR's DO-228 turboprop, will be used for all flight trials. It is expected that the plane will fly out of Ciampino, a military airbase south of Rome.

REFERENCES

[1] **B. Abbe, M. Agan, and T. Jedrey**, *ACTS Mobile Terminals*, International Journal of Satellite Communications, John Wiley and Sons, May-June 1996, Vol. 14, No. 3, pp. 175-189.

[2] **M. Agan, D. Nakamura, A. Campbell, R. Sternowski, W. Whitting, and L. Shameson**, *ACTS Aeronautical Experiments*, International Journal of Satellite Communications, John Wiley and Sons, May-June 1996, Vol. 14, No. 3, pp. 233-247.

[3] **E. Perrins, and M. Rice,** *Propagation Analysis of the ACTS Maritime Satellite Channel*, Proceedings of the International Mobile Satellite Conference 1997 (IMSC'97), JPL 97-11, Jet Propulsion Laboratory, Pasadena, CA.

ACKNOWLEDGMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration in conjunction with Rockwell International Corporation, NASA Ames Research Center, Geoff Haines-Stiles Productions, Inc., the National Science Foundation, Maryland Public Television, WNET/New York, Alenia Spazio, DLR, Dassalt Electronique, and INMARSAT.

Use of Mobile Communication Satellites in the Canadian Ice Service: Past, Present and Future

Ken W. Asmus[1] , Amir Haghgooie[1] and Russ Renaud[2]

[1]Canadian Ice Service
373 Sussex Drive, 3rd Floor, Block "E"
Ottawa, Ontario Canada K1A 0H3
Tel: 613-947-1866 Fax: 613-996-4218
E-Mail: Ken.Asmus@ec.gc.ca

[2]Canadian Coast Guard
Canada Building
344 Slater St
Ottawa, Ontario, Canada K1A 0N7
Tel: 613- 993-2479 Fax: 613- 995-4700

ABSTRACT

The Federal Mandate of the Canadian Ice Service (CIS), an arm of the Department of Environment weather service, is to provide accurate and timely ice information for the safety of Canadians, the protection of life and property, the prevention of environmental disasters and for the economic benefit of the country. Our main client is the Canadian Coast Guard (CCG) and their fleet of Icebreakers deployed year round, at times in extremely remote areas.

An important and integral part of safe and efficient navigation in ice covered waters is the timely reception onboard the vessel of high quality ice information products[1]. This information can be in the form of satellite images (visual and microwave) or data products (ice and weather charts, forecasts, messages) produced at the Canadian Ice Service Operations Centre located in Ottawa, Canada. One of the major difficulties facing the marine operator is the reception of these data products (generally large data files) on the vessels in a timely and cost effective manner while operating in remote areas and, at times, outside the range of traditional communications links.

INTRODUCTION

In the past, all ice and weather information products were received onboard the vessels via high frequency (HF) radio facsimile transmissions (or even via morse code!). These products were limited to chart and text data only as it was not possible to send image files over this communication medium. The reception of data was subject to ionospheric conditions and, especially in the Arctic areas, periods of "radio blackouts". In the last few years the Canadian Ice Service has been successfully using mobile satellites to transmit these data products to the marine users. Inmarsat A analogue voice channels and "off the shelf" 9600 bps data modems have been used for a number of years. The first experiment in the use of Inmarsat A high speed (56 kbps) data transfer took place during an expedition to the North Pole in 1994. The first trials using a low earth orbit store and forward satellite (Health Sat-1) occurred during this historical voyage. Since the fall of 1995, four Icebreakers have been upgraded to allow the use of the Inmarsat A High Speed 56 kbps service and in the spring of 1996 one vessel was equipped with Inmarsat B high speed 56/64 kbps service. During the summer operations of 1996 in the Arctic, a number of vessels were equipped with MSAT terminals.

HISTORY

The Canadian Ice Service provides ice information support to the fleet of Coast Guard Icebreakers on a year round basis. It was recognized in the early 1990's that the traditional way of delivering ice information (radiofax) was not going to be satisfactory especially with the planned launch of the new Radar satellite which provides tremendous amounts of new high resolution (high volume) data not previously available.

The Canadian Coast Guard installed it's first COMSAT Marisat terminal onboard the CCGS JOHN A. MACDONALD in July 1976 to coincide with the availability of COMSAT voice and telex services. In the early 80's, Magnavox MX211 terminals were phased in for larger icebreakers, followed by the Magnavox MX2400 in 1989.

In 1993 experiments were undertaken testing the use of cellular telephones to transfer data files to ship's operating in the Gulf of St. Lawrence. These initial proof of concept tests used off-the-shelf PC's and modems. Although limited by the coverage of the cellular system, many data files were successfully transferred and the experiments were deemed a success.

The next phase of evaluation was to use Inmarsat to transfer image data files. During the summer of 1993 Inmarsat A service using 9600 bps off-the-shelf modems was utilized. Inmarsat was essentially the only mode of communications available in the High Arctic where the icebreaker's were deployed and we were successful in delivering image data files to the vessels.

A Standard M terminal was fitted onboard the CCGS GORDON REID in 1994 soon after Standard M services were made available via the INMARSAT Pacific Ocean satellite

In preparation for a scientific voyage to the North Pole in the summer of 1994, a 56 kbps modem was installed on the CCGS Louis S. St. Laurent for use with Inmarsat A service. Numerous data files were delivered to the ship as far north as 82 degrees (under optimum conditions). Although this was an experimental installation the success of the data delivery lead to the installation of 56 kbps modems on 4 other icebreakers in the following year.

It was realized that during the scientific voyage to the North Pole that the ship would be outside the mask of Inmarsat (or any geostationary communications satellites) for a large extent of the trip. In order to try to enable a link to deliver text and/or data files, an experimental low earth orbit (LEO) satellite system was evaluated [2]. A store-and-forward LEO satellite terminal, utilizing the near-polar-orbiting HealthSat-1 system, was installed on the ship. Nearly all attempts to transfer data to/from the ship were successful (file sizes ranging from .1 to 60 kbytes). Although this was an experimental non-operational system, it did demonstrate the potential of non-real time store-and-forward communications in a remote marine environment.

During 1995 and 1996, two Inmarsat B terminals were installed c/w 56/64 kbps modems. These systems (along with the reduced cost/minute) have proven to be very successful and will lead to further Inmarsat B upgrades.

MSAT terminals were first fitted in 1996 and with 32 terminals in service in the Fleet, the number of MSAT terminals deployed is already comparable to Inmarsat voice terminal installations. During the summer of 1996 the first trials in transferring ice information services using the MSAT system were undertaken. Although limited to only 2400 bps transfer rate, the MSAT service proved to be quite reliable, in particular for low volumes of data (i.e. text and small graphic files).

There are currently over 31 Inmarsat equipped vessels (A, B and M) in the CCG Fleet.

PRESENT OPERATIONAL SCENARIO

The current delivery system for Ice Information Products to the Coast Guard Icebreaker fleet uses an in house developed system called Ice-Vu. The Ice-Vu System is based on the MS Win NT 3.51 operating system running on an Intel-based PC. Ice-Vu integrates commercial software packages and in-house developed applications to provide the user with the necessary tools to

communicate with the Canadian Ice Services data warehouse, download imagery and alphanumeric products. As well, Ice-Vu allows the user to view and analyze the image products. It also incorporates a navigational tool called NavAid, which allows the user to plan ship routes on a regional image.

Ice-Vu uses the Inmarsat High Speed data link (56kbps) as it's primary mode to communicate with the host computer, which is the repository for all products of the Canadian Ice Service. A diagram of the Ice-Vu communications data flow is shown in Figure 1. The link is established via an onboard hardware interface (Digi Syncport) simulating a Frame Relay network interface card. Another NT Server equipped with the similar hardware handles the routing of packets to the Local Area Network at the Canadian Ice Service in Ottawa.

RADARSAT and SAR/SLAR imagery is downloaded to the IceVu system. The images then are automatically decompressed, processed and catalogued by a background application. This data can then be viewed and analyzed.

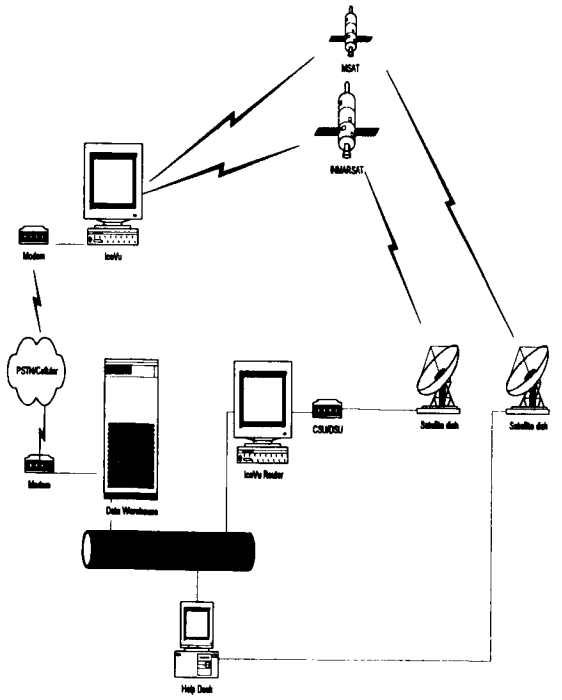


Figure 1. Ice-Vu Communications Data Flow

The primary satellite used for the transfer of Ice Information Products today is Inmarsat (A and B service). Table 1 shows the usage of Inmarsat over approximately a one year period (1996).

	AVERAGE MONTHLY USAGE/ SHIP	AVERAGE YEARLY USAGE ALL SHIPS	AVERAGE COST PER YEAR
Inmarsat A 56 kbps $25/min	**300 minutes**	**14400 minutes**	**$360,000**
Inmarsat B 56 kbps $15/min	**300 minutes**	**7200 minutes**	**$108,000**
Total	**600 minutes**	**21600 minutes**	**$468,000**

Table 1. Inmarsat Usage

Starting in the summer of 1996, the CIS began to experiment using the MSAT system to transfer data to vessels not equipped with Inmarsat. Initial results were good although the slow transfer rate (2400 bps) is not really suitable for the transfer of large imagery products. In an effort to lower costs, investigations into increasing the use of MSAT are ongoing. For example, it may be feasible to offload small files from Inmarsat onto MSAT.

The CIS is monitoring new developments in satellite communications technologies in particular those that provide service to the marine community. It is anticipated that there will be a need for faster data links (i.e. higher bandwidth) with the increased use of satellite imagery data along with possible multi media files being transferred to our clients. As well, with the continuing reduction in operating budgets, ways of reducing these costs must be investigated.

Due to the fact that our operational area encompasses extremely remote regions of the earth (i.e. the High Arctic) outside of the range of traditional geostationary satellites, there is interest in the new suite of polar orbiting LEO satellites.

WHAT'S AHEAD ?

Project approval is currently being sought to equip the greater part of the CCG Fleet with MSAT terminals. This is not surprising, in that during it's requirements definition phase, MSAT designers consulted closely with Canadian government fleet operators. MSAT is capable of satisfy many of the same requirements of the CCG Fleet presently provided by Inmarsat A and B terminals, with lower capital costs, lower operating costs, superior domestic coverage and substantially smaller above and below-decks equipment suitable for smaller vessels.

As well, project approval for fitting a number of Inmarsat Standard C terminals for CCG vessels engaging in International voyages is being sought under the Global Maritime Distress and Safety System (GMDSS) project. Inmarsat C installations are planned primarily to satisfy the regulatory requirements of the GMDSS.

Other developments are monitored by the CCG to evaluate their applicability for use in the operation of the CCG Fleet. As an example, VSAT services via various GEO systems have been considered, as they are suitable for high-bandwidth applications such as Ice-Vu. However, the potential lower operating cost currently cannot offset the capital cost of adapting these systems to the shipboard environment, precluding them from the communications equipment mix in our Fleet. Various LEO systems, with their smaller, omni-directional antennas and proposed low capital and operating costs have potential for adaptation to the shipboard environment.

CONCLUSION

Mobile satellite communications are an integral component of CIS operations. Although the currently available mobile satellite systems meet most of our operational requirements, we are monitoring developments with the following in mind:

1) the requirement for higher bandwidth
2) reduction in operating costs
3) accessibility in remote areas outside of traditional satellite coverage

In this regards, the Canadian Ice Service and the Canadian Coast Guard are working together to investigate potential new mobile satellite systems that will meet our joint operational requirements.

REFERENCES

[1] K.W. Asmus, A. Koonar and S. MacDonald. *Ice-Vu: An Integrated Data Communications and Image Analysis System,* proceedings of the IEE Fifth International Conference on Satellite Systems for Mobile Communications and Navigation, London, UK , May 1996.

[2] T. Keith Elms, Ken A. Butt and K.W. Asmus. *A Low Earth Orbit Satellite Marine Communication System Demonstration* , proceedings of the Fourth International Mobile satellite Conference 1995, Ottawa, Canada, June 1995 pp. 170-175.

Improved Coast Guard Communications using Commercial Satellites and WWW Technology

LT Gregory W. Johnson and ET1 Mark D. Wiggins
USCG Research and Development Center
1082 Shennecossett Rd., Groton, CT 06340
Phone: 860-441-2671 Fax: 860-441-2792
gjohnson@rdc.uscg.mil http://comms.rdc.uscg.mil/

Abstract

Information collection and distribution are essential components of most Coast Guard missions. However, information needs have typically outpaced the ability of the installed communications systems to meet those needs. This mismatch leads to reduced effectiveness of Coast Guard operations. One current need is for Coast Guard aircraft to communicate information on vessels sighted to the shipboard commander quickly and efficiently. The shipboard commander needs to be able to access this information in real-time as well as retrieve related information from historical databases. This paper describes an R&D initiative to demonstrate a new concept of operations to meet this need using COTS technology. A database server installed at the R&D Center which can be accessed via the Internet collects the information and provides it to authorized users using web-based forms. A sighting report (from aircraft and cutters) is entered using the minimum amount of information; this information is then combined by the server with information from historical databases to make a complete record. The users (Coast Guard aircraft, stations, and cutters) access the information using standard WWW browser software. All users connect to the server using either fixed network connections to the Internet, Coast Guard Intranet, or dial-up PPP connections into a remote access server. Commercial satellite systems (AMSC and Inmarsat) provide the communications links for the mobile users (aircraft and cutters). This system was demonstrated in the New England area during March of 1997. This paper describes the concept, implementation, demonstration, and a preliminary analysis of the performance of the communication links to the mobile users.

Background/Problem Statement

The U.S. Coast Guard's law enforcement operations in the First District (New England area—Figure 1) utilize a medium-endurance cutter (WMEC) to provide the operational control for patrol boats (WPB) and other surface and air units within the District Area of Responsibility (AOR). During these operations, aircraft conduct flyovers of fishing vessels and report the identity, activity, and position of all vessels sighted to the WMEC, which then makes decisions as to which vessels to board and inspect. The accuracy and timeliness of the these sighting reports varies widely depending upon the type of aircraft, range to the WMEC, and weather conditions. Frequently there is no active communications path between the aircraft and the WMEC. Sometimes the information is not reported until the aircraft lands and uses shore-based communications systems. When the sighting information is received by the WMEC, the operational commander makes use of information in operational databases as factors in his boarding decision. [1]

To make effective use of the available resources, the aircraft must have a mechanism for capturing information about the sighted vessels. The aircraft must also have a "real-time" communications path to transmit the sighting information to the WMEC and shore-based operational databases. In addition to the sighting information, the operational commander needs access to the information in the operational databases ashore (historical information as well as current lists of wanted vessels).

Figure 1— Map of USCG Districts [2]

Concept

The Coast Guard has complex communications requirements in that most of the operational units are mobile (ships and aircraft) separated from the information they need to have access to by many, sometime hundreds of miles of open water. Although the amount of data that needs to be moved is relatively small by today's standards, it is bridging the gap to the mobile users that provides the challenge to Coast Guard communications. Under current scenarios, it is a tedious and time consuming task to retrieve law enforcement information from various databases located on shore. It is even harder to add information to these databases. Many times, the process is reduced to handwritten notes being manually entered into the database

The opinions or assertions contained herein are the private ones of the authors and are not to be construed as official or reflecting the views of the Commandant or the Coast Guard at large.

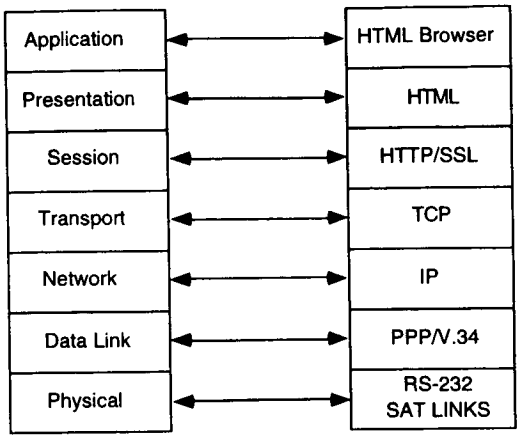


Figure 2—OWL Project mapped to the OSI model

after the information is gathered, adding hours to the latency of the information. The retrieval of the information by the mobile client, when possible, is most often by means of an RF link that is at best slow and unreliable. Also the presentation and usability of the information is a challenge since there are many different platforms being used to store, transmit and retrieve the data.

In today's austere budgetary climate, commercial off-the-shelf (COTS) solutions are very much in favor since they are usually the lowest cost to the government. This applies as well to communications, especially satellite communications. In fact, commercial satellite solutions are often the cheapest communications solution, and sometimes the **only** solution to meet Coast Guard communications needs [3].

For this proof-of-concept demonstration, a system was needed to provide over-the-horizon communications to Coast Guard aircraft and ships. The system needed to provide moderate data speeds and be economical to use. The system also needed to be installed and operational on these units in a short period of time. These requirements lead to the selection of the SkyCell™ system from American Mobile Satellite Corp. (AMSC) for the aircraft and patrol boats, and the use of the already installed Inmarsat Standard-A on the medium endurance cutters.

Once the satellite system was determined, the methodology of the data retrieval and transmission and its presentation had to be decided on. A user-friendly environment that required as little end-user training as possible was necessary since Coast Guard personnel are already tasked with performing multiple jobs and have many responsibilities. Most units cannot afford to dedicate one or more person to operate and maintain the communications equipment.

A dial-up Point-to-Point Protocol (PPP) connection using Hyper-Text Transfer Protocol (HTTP) to a world wide web (WWW) server that presented the data in a Hypertext Mark-up Language (HTML) protocol was selected for this project. In other words, providing Internet access to the mobile users and using a web server to collect, store, and distribute the data. The database resides on a server that is accessed via a Transmission Control Protocol/Internet Protocol (TCP/IP) connection and distributed in a Common Gateway Interface (CGI) environment. The TCP/IP connection is provided by a PPP dial-up link via the commercial satellite systems (Inmarsat-A and AMSC). A laptop with an HTML browser such as Netscape™ Navigator or Microsoft Internet Explorer™ is the user interface for the mobile units. The OSI Seven Layer Model [4] mapped fairly well to the communications network established for this project (Figure 2).

The proof-of-concept project is named Operational Web Link (OWL) and the system concept is illustrated in Figure 3. Some of the goals of the project were to:

- minimize user actions (minimal typing, quick, easy)
- provide onetime data entry; eliminate voice relays and multiple entries
- minimize the amount of data to be transferred (utilize links to static databases)
- provide a simple user interface for the mobile users (user shouldn't need to memorize or lookup codes)

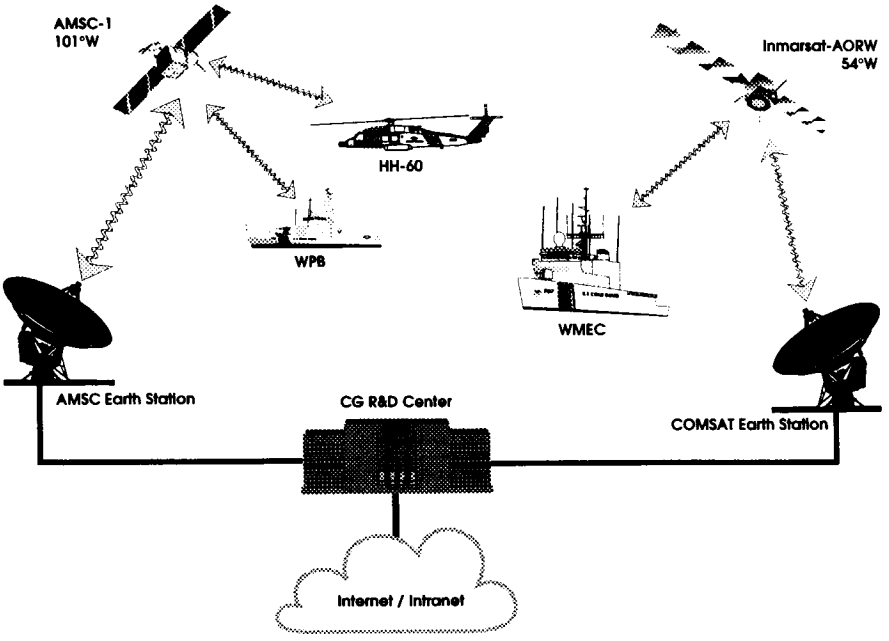


Figure 3—Operational Web Link Proof-of-Concept Demonstration

The United States Government does not endorse products or manufacturers. Trade or manufacturers names appear herein solely because they are considered essential to the object of this report..

- develop a system that is easy to modify as needed, keeping complexity at a central server
- provide real-time response
- give users only the information they need, when they need it
- provide automatic data transmittal into the main Coast Guard law enforcement database (LEIS)

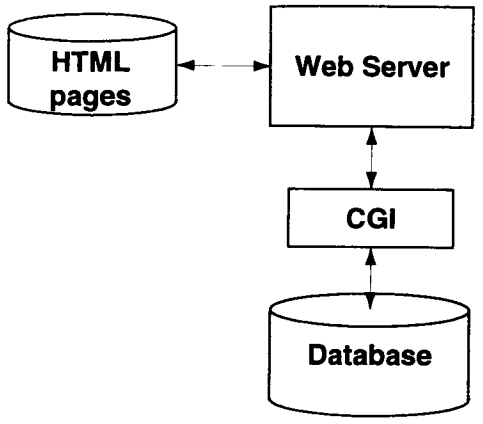


Figure 4—Generic web server configuration. Each box is a separate application or file(s). They could all reside on the same computer or be running on separate machines.

Implementation

With the concept described above, all of the "intelligence" resides in the server; each client (ship, aircraft, or station) merely uses a simple web browser. The server is actually several linked parts as illustrated in Figure 4. The web server is an application that serves up (transfers upon request) various pages (files), typically text files in hypertext markup language (HTML) format. The functionality of the web server is extended using small programs known as CGI (common gateway interface) scripts. The final component, and the most important part for this experiment, is the database. Each of these components is described in more detail in the following sections.

Web Server

The web server is the heart of the system. The information in the database is worthless unless it can be distributed and updated with ease. A Macintosh platform was chosen for this rapid-prototype project due to the ease of software configuration and development. The hardware consisted of a 7500/100 PowerPC running System 7.6 with 64 MB of RAM. The web server application used for this experiment was StarNine's WebSTAR™ SSL. The server actually ran two versions of WebSTAR concurrently: one version non-secure and the other secure. The non-secure side serves as our general Advanced Communications Project site [5] and is also used for the training data base that was used to test and train end-users. The secure side was the server used for the experiment.

The database program used was Claris's FileMaker Pro™ 3.0 (discussed in detail below). A CGI acts as the interface between the web server and the database application. The recent trend has been to implement CGIs as plug-ins to the web server software which allows for better performance. There are many COTS solutions to interfacing a FileMaker Pro 3.0 database with the WebSTAR server. The WebFM™ plug-in was used because it provided the ability to use the calculation fields in FileMaker Pro™ to manipulate the HTML which allowed extremely flexible access to the database. Since the HTML is generated "on the fly" as the "GET" and "POST" commands are carried out, the presentation of the data can be tailored according to it's content.

The HTML pages control most of the "look and feel" of the application. The structure of the web site for this application, as defined by the web pages is illustrated in Figure 5. The top level of the site is a designated main page which serves as an entry point for the site. This page offers the user four choices; each choice leads to a different linked page. Each linked page is actually a web form that specifies some interaction with the database—either adding new records (vessel sightings) or searching the database. The pages in the level below this are actually generated by the database in response to the web form. At the bottom of most of the pages is a navigation bar which can be used by the user to go directly to any of the pages.

Sighting Entry Forms: The first two choices both relate to the entry of vessel sightings. There are two forms; a short form (Figure 6) designed primarily for the aircraft to enter the minimum amount of information quickly and easily, and the full form which allows for a full information report to be entered. Each of these two forms contains hidden fields which specify various attributes and the database name. The rest of the form consists of the information fields that the user can fill in. Pull-down menus have been used as much as possible to minimize the amount of typing by the user. In all cases where the entry needs to be coded for transmission into the main Coast Guard law enforcement database (LEIS), a pull-down menu has been used so that the user is shielded from the codes. The user is presented with a plain language list, which the form translates into the appropriate code.

The form is posted by the user clicking on the SUBMIT button (labeled "Send Sighting" on the form). This directs the

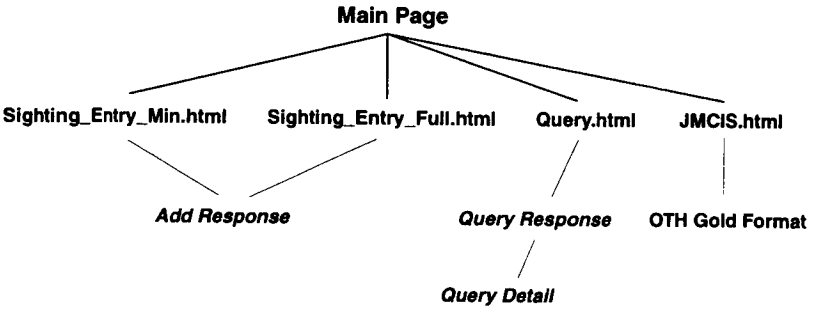


Figure 5—Web site structure.

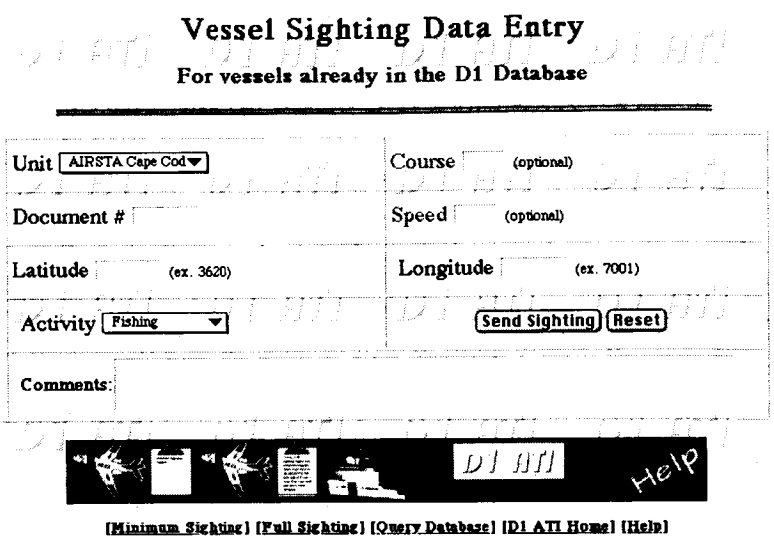


Figure 6– A screen shot of the Minimum Sighting Entry page. The appearance will vary slightly depending upon the computer platform and browser software.

browser to transfer the data from the fields of the form (not the entire form) to the web server which hands the information off to the database through the CGI. The database then generates the appropriate reply (actually this is specified by one of the hidden fields on the form), hands off the HTML formatted text to the web server, which then transfers the data to the browser. Generally, the response is simply the statement that the sighting has been successfully added to the database and the record number. However, if an error occurs, the response is the error code and description. Also, when the sighting is added to the database, checks are made of the linked databases (described below). If the vessel is on the Lookout List, that fact is returned to the user. If the vessel has been previously boarded by the Coast Guard, the date the vessel was last boarded is also indicated. This gives the user instant feedback on the success of the entry as well as pertinent law enforcement information.

Query Form: The other main form is the Query form (Figure 7) which is used to request information from the database. Like the sighting entry form, the query form also contains hidden fields which specify various attributes such as database name and sort order for returned records. The visible fields on the form are used to specify the search criteria, such as all vessels of a certain type. If multiple fields are used the search is performed using a logical AND of all the criteria. Again, pull-down menus are used as much as possible to minimize typing—especially for fields in the database that are coded. In this case the pull-down menu contains plain language which is converted to the appropriate code behind the scenes.

The query is sent to the database by the user clicking on the SUBMIT button (labeled "Find Sightings"). This directs the browser to transfer the data from the selected fields to the web server which hands it off to the database through the CGI. A search of the database is done using the specified criteria and the records found are returned to the browser in the format specified by the query form. In order to minimize the amount of data transmitted, only a small amount of information is returned for each found record (Vessel Name, Type, Document Number, Position, Date and Time of Sighting, and whether the vessel is on the Lookout List). Part of the data returned for each record is a web link (through the vessel number) which the user can click on to retrieve the full record from the database. This allows the user to quickly see all of the vessels in the database meeting the search criteria and then download a full report only on the vessels desired.

JMCIS Download form: The final form was included to enable the large Coast Guard cutters to transfer the information from the database into their Command and Control (C^2) systems. The C^2 system used on these cutters is the same as used in the Navy; the Joint Maritime Command Information System (JMCIS). This system only accepts vessel reports in a certain format (OTH Gold). This form has a single button on it; all of the functionality is specified in hidden fields. When the user clicks on the button, a special search of the database is done, and all of the vessel sightings that have not been downloaded yet, are found and transferred to the browser in OTH Gold format. This information is then saved as a text file by the user and transferred to the JMCIS computer.

Database

The operation of this experiment is controlled by the database. FileMaker Pro works with several database files that are linked together (Figure 8). The OWL main database contains all of the information for this experiment. The sighting information is entered into this database by the WebFM CGI, and this is the database that is queried by the user. The linked databases are used to reduce the amount of data that the user needs to

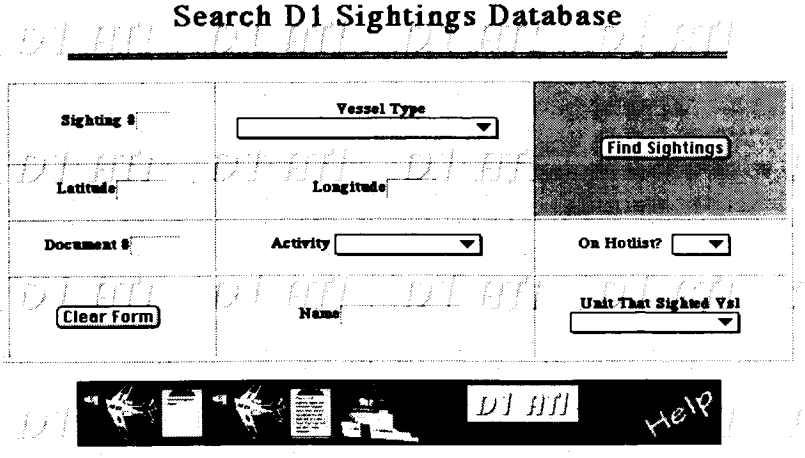


Figure 7—Screen shot of the Query Database page.

enter, since much of the information about a given vessel is static, and is already known. The user simply enters the dynamic data (position, course, speed, and activity) along with the vessel's document number. The document number is the unique identifier which is used to retrieve the static information from the linked databases. Currently, each of these linked databases resides along with the main database. Ideally, in the long-term, these database would be linked over the Internet or an Intranet, with each database being maintained by the appropriate agency.

NMFS Database: This is a database of fishing vessels obtained from the National Marine Fisheries Service (NMFS) which contains information on all documented fishing vessels (about 8000 entries). The database contains fields for document number, name, homeport, state, length, and the fisheries permits and gear codes issued to each vessel. When a vessel is sighted, all of this information for that vessel is transferred into the OWL database.

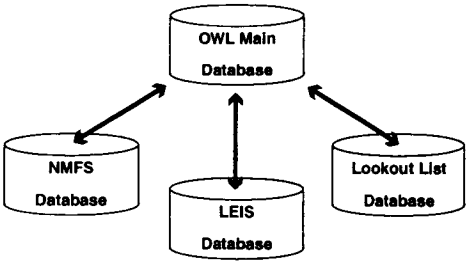


Figure 8—Linked databases used in this experiment.

LEIS Database: This is a database of information obtained from the USCG law enforcement database (LEIS) which contains information on all vessels sighted and boarded by the Coast Guard. The information duplicated locally is a subset of the corporate database. The database is used to supply the following fields into the main database: hull and superstructure color, vessel type, flag, and the date last boarded by the Coast Guard.

Lookout List Database: This is a database of all vessels of interest to the First Coast Guard District. Each time a vessel sighting is entered into the main database, it is cross-checked with this database to see if it is of interest (a "hot" vessel).

Database Fields: There are several types of fields in the database. In addition to those that are manually entered or looked-up in the linked databases, some fields are automatically entered and some are calculated by the database. The auto-entered fields are date and time of sighting entry and the sighting number. The calculation fields are used as program flags, internal calculations for special formatting, and most importantly, to create the HTML code returned to the user. Many of the web pages viewed by the user are actually created "on the fly" by the OWL database. The main ones (described above) are the Add Response, Query Result, Query Detail, and OTHG Format pages.

Communications Links

Providing the mobile units with a "real-time" communications link was one of the major goals of this experiment. For the medium endurance cutters (WMEC) the existing Inmarsat Standard-A terminals were used. Inmarsat-A is a commercial L-Band mobile satellite system that provides analog voice and data capability for about $6/min. The maritime antenna is 1.244m in size, and tracks in both azimuth and elevation [6]. External V.34 modems (Supra 33.6 external models) were used with the analog channel that Inmarsat-A provides, for the data connections. The patrol boats (WPB) do not have the topside real estate to mount an Inmarsat-A antenna, so we installed an AMSC terminal on board. This is also an L-Band mobile satellite system, but with a smaller antenna (.6m) and usage cost of approximately $1.50/min. It is a digital system that initially provided a data rate of 2400 bps, and is now transitioning to 4800 bps [7]. For the aircraft (HH-60 Jayhawk helicopters) an aeronautical version of the AMSC terminal, manufactured by CAL Corp. was used.

Each of these terminals was connected to a Toshiba P75 laptop running Windows 95™ via the RS-232 serial port. The end user would dial into one of the Remote Access Servers (RAS) located in our lab in Groton, CT. Two different RAS's: a SonicSystem Quickstream™ which has three ports, and a Shiva LanRover™ with eight ports were used. The LanRover was chosen because it supports all of the platforms and protocols we were using and it had some proprietary features that improve performance such as PowerBurst™ technology and STAC™ compression [8].

There were several challenges to be overcome due to the delays inherent with all satellite communications. The Inmarsat-A connection had a very long delay before the call went through and the Windows 95 dial-up networking application could not be configured to take this delay into account and would drop the connection before the modem connection could be established. With the help of Shiva technical support, we were allowed to use a beta version of Shiva's new PPP dial-up application which is more configurable and provides additional features such as detailed session statistics. This software allowed for a successful dial-up PPP connection through Inmarsat-A. The units using the AMSC terminals are required to dial into the QuickStream RAS because they were unable to successfully connect through the Shiva RAS using the AMSC terminals. This issue is still unresolved.

Security

Since this experiment involved actual law enforcement operations, security was a very important issue. The web server used is in general, very resistant to outside manipulation. The root level of the system is not accessible; whatever directory the application resides in appears to be the root level, and only files in the same directory or subdirectories are accessible. The system is also configured to reduce vulnerability by denying all FTP, TELNET, and SMTP connections which are respon-

sible for some of the most common server security breaches. Also, the system does not allow file uploads or directory indexing.

In order to restrict access to the sensitive law enforcement information, additional security layers were added to the system in three levels. The first is access protection. The secure web server is configured to only allow certain users to access the OWL web site. Any user can access the main page, but to go deeper into the site requires authorization. This is done using IP address filtering; the server is configured to only recognize authorized IP addresses. Any other user, with an IP address not on the authorized list will not be granted access to the web pages unless they have a valid user name and password. All others attempting to connect to the site will receive an Access Denied message. Dial-up access to the RAS is also protected by user name and password.

The second level of security is to protect the information during transmission. This is accomplished using standard software data encryption. Since the encryption is end-to-end, protection is provided both across the terrestrial network and the two satellite systems (AMSC & Inmarsat-A). The secure server is implemented using Netscape's Secure Socket Layer (SSL) protocol [9] to encrypt each session using RSA [10] encryption algorithms. The secure server supports DES (56 bit), 40 bit (international) and 128 bit (domestic) algorithms; the 128 bit version which provides the highest level of security is used for this experiment. Each session is encrypted separately so breaking one session does not allow other sessions to be decrypted. Netscape 3.0.1 domestic version was used as the web browser because it supports the high-grade encryption key (128 bit).

The third level of security is server authentication. Server authentication uses RSA public key cryptography in conjunction with ISO X.509 digital certificates [11] and lets the user (client) verify that the server is valid and not an imposter. Each connection to the secure server checks the secure server's site certificate with a trusted third party; for this server the security certificate was provided by Verisign™.

Demonstration/Results

The system described above is currently undergoing testing on several ships and aircraft in the First Coast Guard District. The operational units participating in the experiment are the *USCGC Monomoy*, a 110 ft. patrol boat (WPB) operating out of Woods Hole MA, the *USCGC Escanaba*, a 270 ft. medium endurance cutter (WMEC) based in Boston MA, the *USCGC Vigorous*, a 210 ft. WMEC out of Cape May NJ, and several HH-60 Jayhawk helicopters flying from Air Station Cape Cod, MA. Each unit received a laptop running Windows 95 and Netscape Navigator for the client software. Since the testing is ongoing at the time this paper is being written, final results are not available. Some information on the communications links is provided in the table below; additional information will be available at the conference.

In general, the system has worked as desired, and been well received by the operators. All of the project goals are being met. The system is easily used by Coast Guard personnel from junior Seamen up to senior Commanders and Captains. The work required by the user to enter sightings into the LEIS database has been reduced considerably—a savings of 10–15 minutes for **each** sighting! Data can be shared between the aircraft, cutters, and Operational Commanders ashore in a real-time fashion without requiring extensive user actions or voice relays. In addition, the web-based architecture allows the client-side software to be very simple (a COTS web browser) while all of the complexity is in the central server. This allowed the HTML code to be quickly and easily modified in response to input from the end-users and the changes automatically distributed to all users.

	Inmarsat-A	AMSC
Typical time required to:		
Connect communications channel	67 sec	25 sec
Accomplish RAS log-on	10 sec	20 sec
Enter sighting and receive response	7	14 sec
Perform query and receive response	15	20 sec
Typical connection bit rate	12,000 bps	2,400 bps
Link data compression?	Hardware (V.34)	Software (STAC)

References

[1] USCG R&D Center proposal for Pulse OPS Communication Project to First Coast Guard District, 12 December 1996.

[2] USCG web site, http://www.dot.gov/dotinfo/uscg/units.html.

[3] *Technology Assessment of U.S.C.G. Long-Range Communications Alternatives*, USCG R&D Center Advanced Communications Technology Report, March 1995.

[4] W. Stallings, *Data Computer Communications 5th ed.*., New Jersey: Prentice Hall, 1997, pp. 19–21.

[5] USCG R&D, Advanced Communications Technology web site, http://comms.rdc.uscg.mil/.

[6] Inmarsat web site, http://www.inmarsat.org/inmarsat/.

[7] AMSC web site, http://www.skycell.com/

[8] Shiva web site, http://www.shiva.com/

[9] Netscape's web site, http://home.netscape.com/newsref/std/SSL.html.

[10] RSA Data Security, Inc. web site, http://www.rsa.com/.

[11] Verisign's Web Site, http://www.verisign.com/pk_intro.html.

K/Ka-Band Maritime Mobile Satellite Communications

R.A. Axford[1], G. Evanoff, R.C. North, K. Owens, J. Toy and G. Bostrom (NRaD, San Diego, CA)
T.N. Englund, W. Schmalgemeier, B.C. Hopkins and J. Griffin (USS PRINCETON (CG 59))
M. Kelly (NAWC/AD, Tracor, San Diego, CA)
P.H. Moose (Naval Postgraduate School, Monterey, CA)

ABSTRACT

The Naval Command, Control and Ocean Surveillance Center's RDT&E Division (NRaD) and the Jet Propulsion Lab (JPL) installed JPL's Advanced Communications Technology Satellite (ACTS) Mobile Terminal (AMT) on USS PRINCETON (CG 59) in July 1996. The AMT remained installed for one year during which time PRINCETON deployed in the Eastern Pacific, Sea of Cortez and Caribbean on a six-month drug interdiction patrol in support of the U.S. Coast Guard. Via ACTS, NASA's experimental geostationary K/Ka-band communications satellite at 100 West Longitude, and the AMT, PRINCETON enjoyed the benefits of a full-duplex T1 data rate (1.536 Mbps) pipe back to CONUS during her deployment. This paper describes the installation and testing of the AMT on PRINCETON, the communications services that PRINCETON used via ACTS while on deployment and the lessons learned by PRINCETON which have implications for the future of full-duplex, high data rate K/Ka-band SATCOM for surface combatants and all ship classes.

INTRODUCTION

The U.S. Navy desires to increase the capacity of the communications links available to its deployed forces. The naval communications developer faces numerous challenges when addressing this desire for surface ships. This is especially true in the case of the so-called "small deck" ships - anything that isn't an aircraft carrier. The mission effectiveness of U.S. Navy surface combatants is currently "antenna limited." We are not limited by how many more computers we might put below decks, we are limited by how many more antennas we can put topside. Look at any U.S. Navy surface combatant and you see a forest of antennas. While each one performs useful functions, each one also contributes to one or more undesirable ship's characteristic:

- Topside weight and moment
- Radar cross-section (RCS)
- Electromagnetic interference
- Physical obstruction of other antennas
- Electromagnetic distortion of other antenna patterns

The U.S. Navy is investing in carefully selected, *"desired ship's characteristics preserving"* technologies for increasing surface combatant communications capacity. A critical part of the technology selection process is at-sea testing of early prototypes. Using this approach, the U.S. Navy has recently increased its confidence in the efficacy of K/Ka-band satellite communications for the AEGIS Class surface combatants: the USS TICONDEROGA (CG 47) Class cruisers and the USS ARLEIGH BURKE DDG 51 Class destroyers.

Under the joint sponsorship of the Office of Naval Research (ONR), the National Astronautics and Space Administration (NASA) and the Program Executive Office of the Surface Combatant AEGIS (PEO SC AEGIS) Program, the Naval Command, Control and Ocean Surveillance Center's RDT&E Division (NRaD) and the Jet Propulsion Lab (JPL) installed JPL's Advanced Communications Technology Satellite (ACTS) Mobile Terminal (AMT) on USS PRINCETON (CG 59, a member of the CG 47 Class) in July 1996. Taking advantage of NASA's investment in ACTS and JPL's investments in mobile satellite communications, the "AMT/CG" project clearly demonstrated the contributions that a reliable, full-duplex T1 data rate K-/Ka-band SATCOM link can make to the mission effectiveness of, and the crew's quality of life aboard a CG 47 Class cruiser. Furthermore, the combination of a high powered K/Ka-band satellite and the AMT's small antenna allowed this capability to be demonstrated in a small shipboard package capable of preserving many desired ship's characteristics.

This paper describes the installation and testing of the AMT on USS PRINCETON. The testing was carried out over a full year which included a six-

[1] axfordra@nosc.mil, 619-553-3729

month drug interdiction patrol in the Eastern Pacific, Sea of Cortez, Gulf of Tehuantepec, Caribbean and two transits through the Panama Canal..

AMT ON USS PRINCETON

Prior to its installation on PRINCETON, the AMT and its direct predecessors were employed in aircraft and terrestrial vehicle experiments with ACTS [1]. In these installations all components of the AMT were located in immediately adjacent racks, within a few feet of the antenna and easily accessible to the operator. For the installation on PRINCETON, because of the large distance between the operator's station in Radio Central (02 level, frame 220) and desirable locations for the antenna on the weather deck, the AMT was divided into a topside unit and a below-decks equipment suite. Figure 1 is a high level block diagram of the AMT that shows how its components were sub-divided for the installation on PRINCETON. Exterior signal and power cable runs were enclosed in separate flexible shielded conduit from the skin of the ship to the topside unit weatherized enclosure. These exterior conduit runs were approximately 35 feet long. Once inside the skin of the ship (at AEGIS Radar Room #5, 04 level, frame 330) the cable runs to Radio Central were approximately 150 feet.

Five topside locations for the weatherized enclosure were considered and these are shown in Figure 2. Because of its relatively unobstructed view of the sky and the availability of the space, the location identified as "#5" in Figure 1 was selected. This location is on the 05 level weather deck, on the starboard side at frame 333. The blockage diagram for location #5 is shown in Figure 3. This position gives the AMT antenna a 360 degree unobstructed view of the sky above 45 degrees elevation except for slight intrusion by HF whip 2-4 at ~20 degrees relative bearing. On 21 October 1996 this slight blockage was observed to cause a ~3 dB fade in the received signal level from ACTS. PRINCETON's location was 32.6° N, 118.1° W from which ACTS bears 150° AZ, 48° EL. No other blockage problems were observed during the deployment.

The AMT's slotted waveguide array antenna is shown in Figure 4. It is mechanically steered in both azimuth and elevation, and is designed to enable mounting on a variety of mobile platforms. The radome is shaped with a peak height of 6.65 inches and a 27.42 inch diameter. The actual dimensions of the combined transmit and receive array apertures inside the radome are ~16 inches wide and ~4.5 inches high. The antenna is capable of tracking 360° in azimuth and 0° to zenith in elevation. The antenna transmit gain is 29 dB and the G/T is 0 dB/°K. Although ACTS is linearly polarized, the AMT antenna uses circular polarizers to avoid the deep fades that would occur during linear polarization misalignment in the mobile environment. The ~3 dB linear/circular polarization mismatch penalty is constant with orientation and accounted for in the link budget calculations. Although the antenna is capable of handling up to 120 Watts of transmit power, a 100 Watt TWTA was used on PRINCETON. The measured EIRP from the transmit aperture was 48.5 dBW. The transmit array 3 dB beamwidths are 5° and 2.5° in elevation and azimuth, respectively. The receive array 3 dB beamwidths are 7° and 4° in elevation and azimuth, respectively. The antenna tracking algorithm utilizes three sources of information, an inertial 3-axis rate sensor, an Inertial Navigation System (INS) and pilot signal strength feedback from circular dithering of the beam. The rate sensor provides the primary

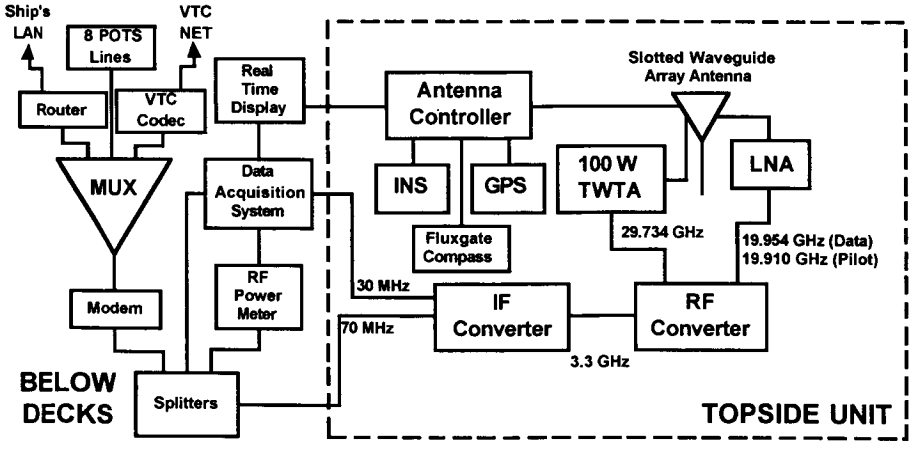


Figure 1 High level block diagram of AMT as installed on PRINCETON.

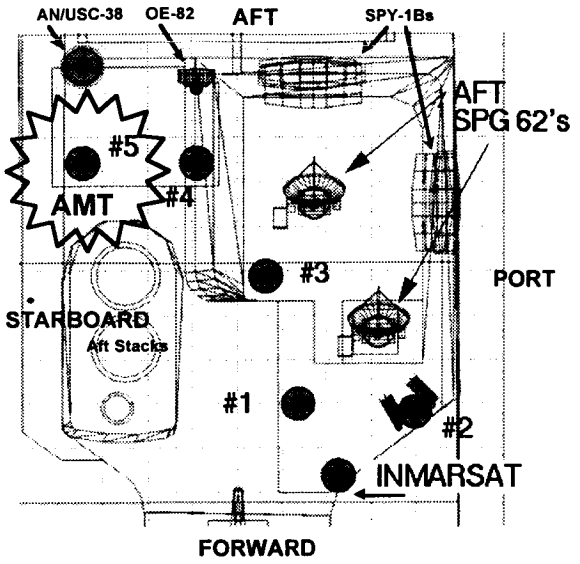


Figure 2 The five topside locations considered for the AMT antenna and weatherized case.

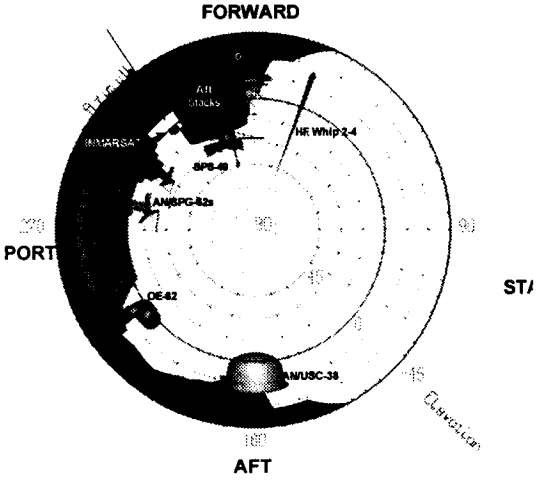


Figure 3 Blockage plot for location #5 of Figure 2. Except for the slight intrusion of HF whip 2-4, there is no blockage above 45 degrees elevation.

information for accurately pointing the antenna, while the INS and the dithering mechanisms provide adjustments for long term drift of the rate sensor.

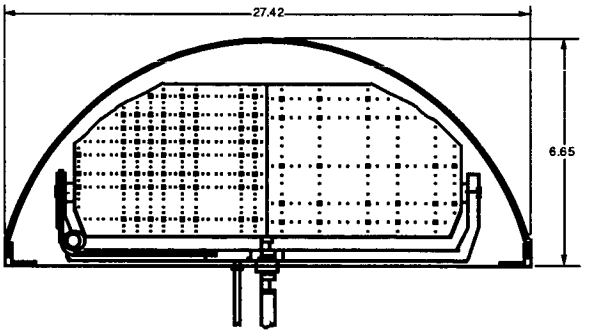


Figure 4 AMT slotted waveguide array antenna and radome. The 30 GHz transmit aperture is on the left and the 20 GHz receive aperture is on the right. Dimensions are in inches.

EMC TESTING

During the week of 14 October 1996 extensive electromagnetic compatibility (EMC) tests were conducted to determine if the AMT was adversely effected by any of PRINCETON's radar, electronic warfare (EW) or communications equipment (and vice versa). Table 1 lists the ship's systems that were employed during the testing. A total of 12 different scenarios were run which consisted of assorted combinations of these systems transmitting at various frequencies and power levels. The 12th scenario, the "big bang," consisted of all systems in Table 1 simultaneously transmitting at their maximum power levels. During the entire testing sequence, the AMT was sending and receiving data via ACTS, including video-teleconferencing (VTC) traffic, at an aggregate data of 512 kbps. No degradation in service quality, exhibited as disruption in the VTC video or audio signals, or spurious signals detected by the AMT's data acquisition system (including a spectrum analyzer on the received signal) were observed on PRINCETON.. In subsequent months of operations, no EMI from any of PRINCETON's systems was ever observed on the AMT. Furthermore, the AMT never induced any EMI in any of PRINCETON's electronic systems.

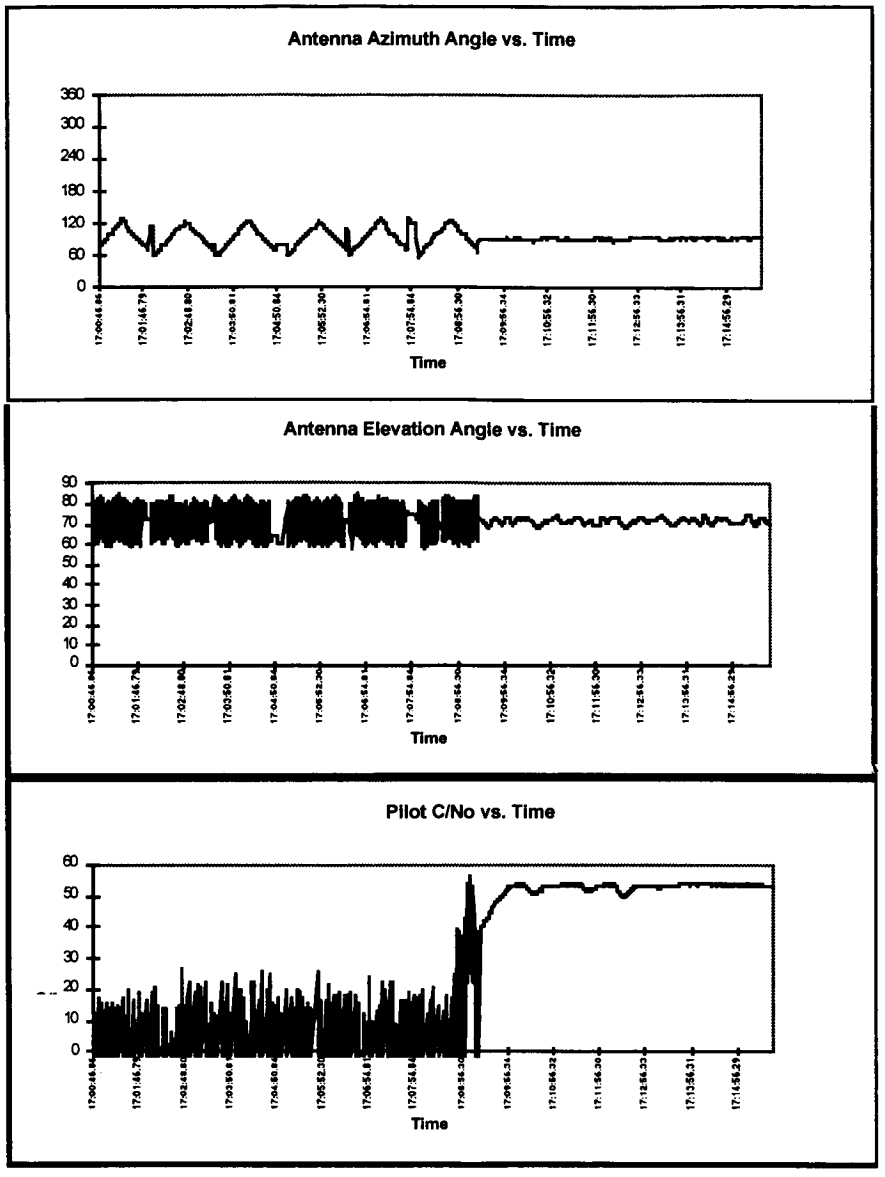


Figure 5 Signal acquisition event on 15 April 1997. Antenna search pattern in azimuth and elevation. Corresponding time record of pilot signal C/No.

SIGNAL ACQUISITION

Figure 5 shows a typical signal acquisition event. Using its internal GPS and compass systems, the antenna controller scans the correct portion of the sky. In 'Acquisition Mode,' the antenna scans a 60 degree section in azimuth, centered on the current bearing to ACTS, while the elevation angle is rapidly scanned +/-10 degrees about the current elevation to ACTS. Typically the signal will "pop up" once or twice (here once) before final acquisition and tracking are accomplished.

COMMUNICATIONS SERVICES

The communications services that PRINCETON used via ACTS and the AMT fall into three classes:

- Plain Old Telephone Service (POTS)
- TCP/IP Applications
- Video Teleconferencing (VTC)

PRINCETON used this connectivity to access medical, logistical, maintenance, operations, training, personnel and weather forecasting services. PRINCETON's officers and crew also used ACTS to stay in touch with their families via POTS and E-Mail. The crew diligently logged their use of ACTS

during the deployment. Monthly reports were issued via GENSER message traffic that documented the total number of ACTS access hours scheduled for, and actually used by PRINCETON. The usage was further documented in the number of e-mails sent and received, number of phone calls made, number of VTCs conducted, number of files downloaded from the WWW or via FTP and highlights such as Navy communications "firsts" and high visibility events. Here is an excerpt from PRINCETON's monthly ACTS usage message for March 1997 [2]:

INTERNET USAGE.

- REMOTE LOG-INS WITH TELNET: 5
- FILES AND COMPUTER PROGRAMS DOWNLOADED WITH FTP: 102
- E-MAIL LOGGED AS RECEIVED: 1527
- E-MAIL LOGGED AS SENT: 1321
- INFORMATION DOWNLOADS LOGGED FROM WWW: 500

Table 1 List of systems used in the EMI testing of the AMT on USS PRINCETON.

Tx TYPE	ANTENNA	FREQ. (MHz)
AN/URT-23D	OE-367(V)	2-6
AN/URT-23D	OE-214B/U	4-12
AN/URT-23D	OE-323B/U	10-30
AN/URT-23D	AS-4014A/U	2-30
AN/WSC-3(V)7/11	AS-4163/URC	225-400
AN/VRC-46A	AS-3226/URC	30-32
AN/VRC-46A	AS-3226/URC	32-76
AN/GRC-211	AS-2809/SRC	116-152
AN/WSC-3(V)3	AS-3018A/WSC	225-400
AN/WSC-3(V)7	AS-1735/SRC	225-400
AN/URC-93(V)	AS-1735/SRC	225-400
AN/URC-80(V)5	AS-3015/URC	156.025-157.425
AN/URC-107(V)5	AS-4127/URC	969-1206
JTIDS	AS-177B/UPX	969-1206
INMARSAT MX-2400	81641-802	1635-1645
AN/URN-25	AS-3240A/URN	962-1215
AN/SPS-64(V)9	AS-3194/SPS	9345-9405
AN/SPY-1B(V)	AS-3572/SPY	3100-3500
AN/SPS-49(V)8	AS-3263/SPS	851-942
AN/SPS-55	AS-2953/SPS	9050-10000
AN/UPX-29	AS-3134/UPX	1030
AN/UPX-29	AS-2188/UPX	1030
AN/UPX-28	AS-177B/UPX	1090
AN/UPX-29	AS-177B/UPX	1030
AN/SLQ-32(V)3	CW-1186/SLQ	Classified
AN/SPG-62	AS-3444/SPG	Classified

PRINCETON's capability to VTC over ACTS was used to render emergency medical assistance to the Master of a Greek freighter off the coast of southern Baja California. On Feb. 9, PRINCETON received word from the U.S. Coast Guard that the 66-year-old Master of a Greek cargo vessel was in medical distress with severe abdominal pains and cramping. PRINCETON immediately proceeded to rendezvous with the Greek ship. At approximately 6:00 p.m. on Feb. 10, after getting the patient onboard, PRINCETON's medical officer rendered immediate aid and then used ACTS to VTC with Naval Medical Center San Diego. Two Navy urologists further examined the patient and determined that the Master had a blocked urinary tract. By projecting their expertise to the ship, the urologists were able to guide PRINCETON's medical officer through the best possible course of treatment until the patient could be flown to La Paz, MX on Feb. 11 for hospitalization. The VTC link over ACTS enabled the patient to receive specialized medical attention a day earlier than would have been possible otherwise.

CONCLUSIONS

The fact that the AMT was unaffected by the AEGIS EMI environment is extremely good news for the DoD's Global Broadcast Service (GBS) Program. GBS will launch K/Ka-Band transponders, hosted on UHF satellites beginning in 1998. However, from the perspective of the sailor, GBS will not provide the quality of service that AMT/CG did. While it is useful to have a "big pipe" from shore to ship, as GBS will provide, it is at times even more useful to have *full-duplex* high data rate connectivity from the ship to shore. PRINCETON's TeleMedicine episode with the Greek freighter Master is a classic example. As the Master will attest, it is much more important to have a high resolution picture of the shipboard patient available to the shore-based doctor than it is for the patient to see a high resolution picture of the doctor! Perhaps the Navy will also be able to make use of some of the emerging *full-duplex* T1 commercial K/Ka-Band SATCOM systems that will appear over the next several years [3].

REFERENCES

[1] T.C. Jedrey (Ed.), *The International Journal of Satellite Communications*, vol. 14, no. 3, May-June 1996 (ACTS experiments special issue).

[2] R 080120Z APR 97, FM PRINCETON, ACTS Monthly Usage for March 1997.

[3] J. Careless, "Ka-Band Satellites, A Guide to the Proposed U.S. Systems," *Via Satellite*, pp. 62-74, Feb. 1996.

Session 15
Advanced System Concepts II

Session Chairperson—*Vijaya K. Gallagher*, Globalstar L. P., USA
Session Organizer—*Polly Estabrook*, Jet Propulsion Laboratory, USA

Satellite Constellations for Millimeter Wave Communication A Comparison

Paul Christopher
Stanford Telecommunications, Inc.
1761 Business Center Drive, Reston, VA 20190, USA
Phone: 703-438-8017 FAX: 703-438-8112
email: pchristopher@sed.stel.com

ABSTRACT

Three satellite constellations which offer high ground elevation angles are examined. The first is a hybrid combination of Molniya and geostationary satellites, followed by the Motorola M-Star contellation, and finally a minimal constellation of elliptic LEOs. Elevation pdfs are found for each. They are found to be promising for millimeter wave communication, and are compared for expected performance in the 30-50 GHz region.

1. BACKGROUND

Satellite constellations with continuous global coverage have been intensively studied since the early '60s. J. Walker (1) showed that only five phased geosynchronous satellites at 47 degrees inclination could provide continuous worldwide coverage. He also appreciated the unique advantages of low earth orbiters (LEOs) and he conceived the 48 satellite circular LEO system which appeared in a later incarnation as the Loral Globalstar system. J. Draim (2) later showed that only four hypersynchronous satellites could provide continuous worldwide coverage. All of these concepts assumed that ground antennas could achieve good communication link performance at low elevation angles. This typically meant choosing frequencies which suffered little atmospheric attenuation. This assumption became less satisfactory in the early '90s as satellite frequencies increased rapidly above 14 GHz. High elevation angles and short atmospheric paths began to enjoy more emphasis.

Three constellations are outlined here which would supply high ground elevation angles over large areas. Elevation probability density functions (pdf) are found for each after exhaustive area and time searches. The first is a hybrid system of three phased Molniya and two antipodal geostationary satellites. It has been known to offer high elevation angles for all time (3) in the Northern Hemisphere. The 72 satellite Motorola LEO system (4) is shown next. An elliptic 48 satellite LEO system is then described. The final section uses the elevation angle pdfs to show that each of the constellations is well suited to millimeter wave communication, even for frequencies greater than 40 GHz. A quantitative comparison based on atmospheric loss is also given.

2. GEOSTATIONARY AND MOLNIYA CONSTELLATION

Geostationary satellites have been known to offer high elevation angles in the tropics, but their elevation is compromised in the Temperate zones. On the other hand, the productive Soviet Molniya series offers high elevation at high latitude. Figure 2-1 shows the positions of a Molniya satellite at one hour intervals of its 12 hour orbit. It offers both high elevation and convenient ground antenna pointing, as it appears to be nearly stationary for many hours near its apogee. The combination of two antipodal geostationary satellites and three phased Molniya satellites has been examined (3), and the constellation offers excellent elevation angle statistics throughout the Northern Hemisphere. Figure 2-2 was generated with the aid of the Stanford Graphics Analysis Tool (5) and indicates a snapshot of the 30 degree elevation coverage with the 5 satellite constellation. Note the excellent coverage offered by the two Molniya satellites near apogee, with one over Norway and the other over the Bering Straits. Figure 2-1 also shows that key parts of the U.S. also have the possibility of angle diversity to either of two Molniya or to a geosynchronous satellite. The third Molniya is waiting in the wings over Antartica to relieve the Norway satellite as all of the Molniya advance eastward. The apogee positions are always

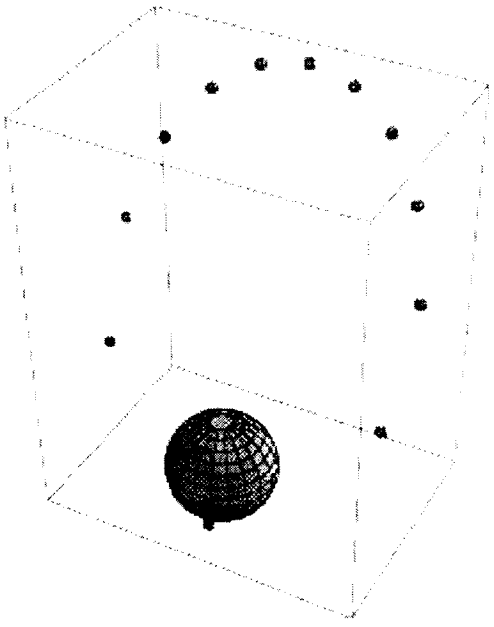


Figure 2-1. One Hour Snapshots of Molniya

filled, or nearly filled. This concern about the long term adequacy of the satellite system can be relieved by checking all locations in the Northern Hemisphere for elevation angle at all times. Elevation angle statistics may then be represented as in Figure 2-3, where the density function is plotted vs. latitude and elevation angle. A cut at a given latitude will give a bell shaped curve which is suitable for that latitude for all time. The overall elevation statistics for the Northern Hemisphere have mean and standard deviation (52.45, 15.24) degrees, respectively. The constellation can be seen to give outstanding elevation angle to a large population of ground stations. The nodal regression of the Molniya satellites, at a rate of 4.03 deg/month, must also be accounted for by the ground stations or with scheduled orbit corrections.

3. MOTOROLA M-STAR

As other telecommunication corporations moved vigorously into the 28 GHz region, Motorola recognized that frequencies greater than 40 GHz would be even more promising (4). They filed a plan for a 72 satellite system with the FCC in September, 1996. The 50 GHz uplink/40 GHz downlink system would have high ground antenna elevation angles and low atmospheric losses throughout the Temperate regions. The circular orbits are proposed as 12 planes of 6 satellites/plane at 1350 km altitude. Inclination is set at 47 degrees (this might be compared with the emphasis on the polar regions with the earlier Iridium polar orbits). Figure 3-2 allows the excellent coverage to be visualized at epoch. A pdf for elevation angle as a function of latitude may be generated for all time, as Figure 3-3. Elevation angles can be seen to be high for all latitudes up to 60 degrees, and to deteriorate at 70 degrees N.

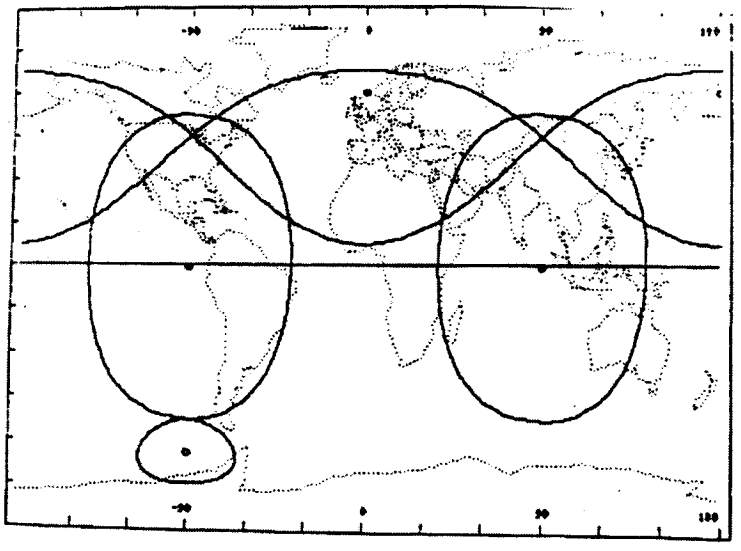


Figure 2-2. 30° Elevation Contours of Hybrid Satellite System at Epoch (3 Molniya + 2 Geosynchronous at 90E, 270E)

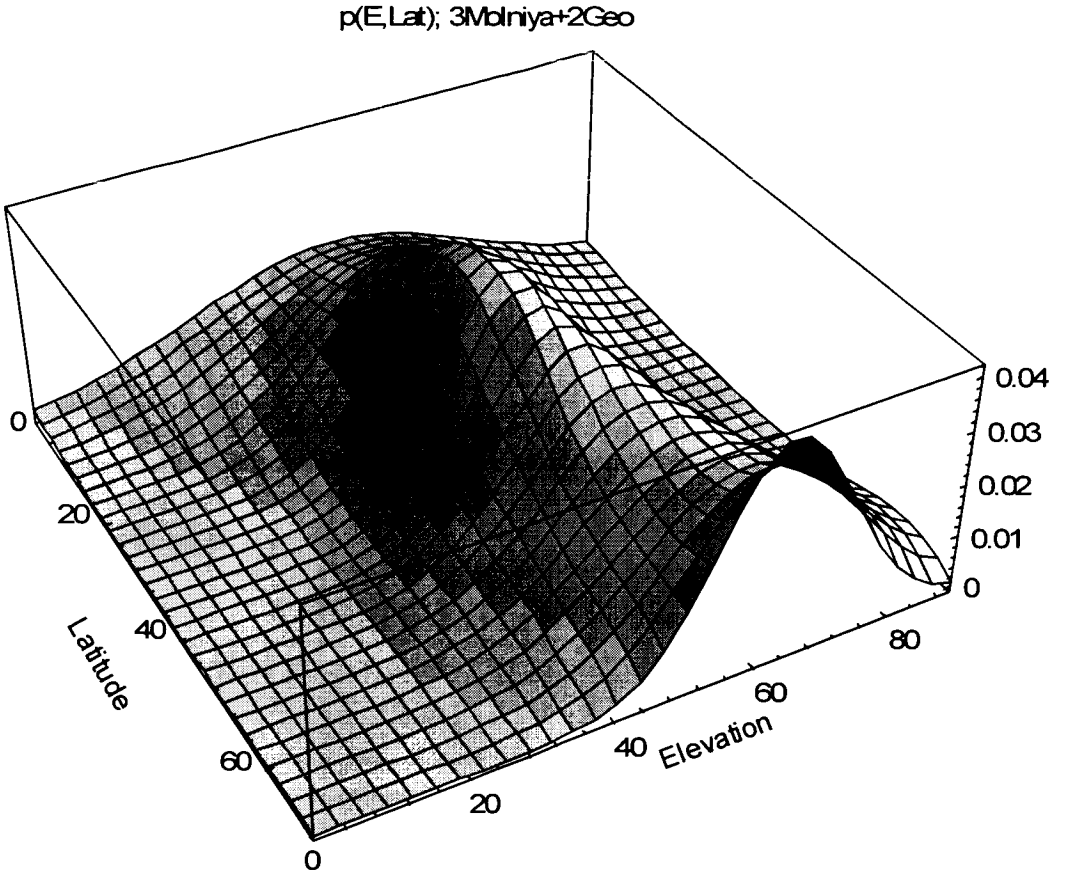


Figure 2-3. Probability Density Function for Elevation as a Function of Latitude 3 Molniya + 2 Geostationary

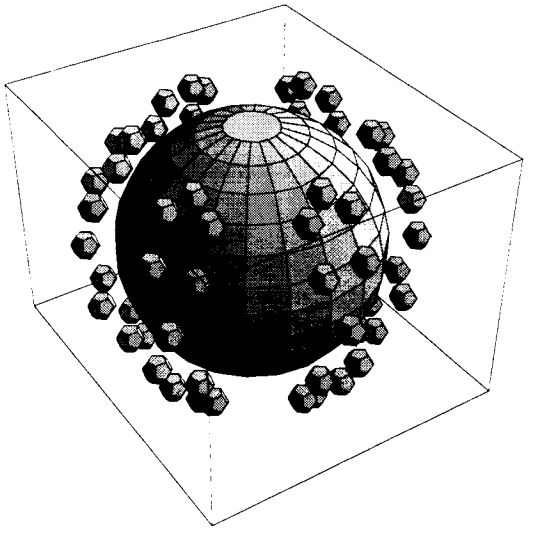


Figure 3-1. Motorola M-Star at Epoch

The elevation angle density function may be concisely approximated with the aid of Mathematica's Fortran form (6).

```
Out[20]//FortranForm=
      1/ (E** ((-37.67078605295647 -
     -            2.2035171098530856-12*LAT -
     -            0.013147127535769416-1*LAT**2 +
     -            5.575875699427762e-16*LAT**3 +
     -            3.955286161206665e-6*LAT**4 + x)**2/
     -          (2*(17.72772486955788 -
     -            0.0002899907822926401*LAT -
     -            0.0013716935489725776-1*LAT**2 +
     -            7.185352764841304e-8*LAT**3 -
     -            4.170493335351661e-7*LAT**4)**2))*
     -        (17.72772486955788 -
     -          0.0002899907822926401*LAT -
     -          0.0013716935489725776-1*LAT**2 +
     -          7.185352764841304e-8*LAT**3 -
     -          4.170493335351661e-7*LAT**4)*Sqrt(2*Pi))
```

where LAT = North Latitude (deg.)
x = elevation angle (deg.)

Figure 3-2 can also help to visualize the worldwide coverage of the M-Star system. The emphasis is clearly on the large community of communication users in the Temperate regions.

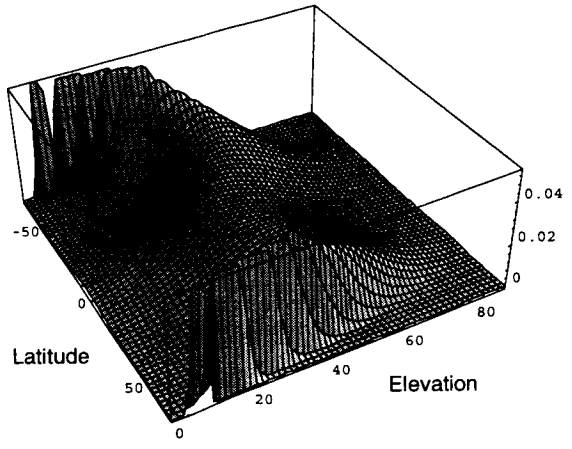


Figure 3-2. M-Star Elevation Probability Density Function

4. MINIMAL CONSTELLATION OF ELLIPTIC LEOs

A set of 48 phased elliptical satellites was examined (7) as an attempt to minimize the number of LEOs required for millimeter wave communication. The 2 hour orbits are critically inclined and intended to miss the most intense portion of the inner Van Allen Belt, as shown in the 10 minute snapshots of Figure 4-1. The 3010 km apogee offers a convenient communication altitude with relatively long viewing times.

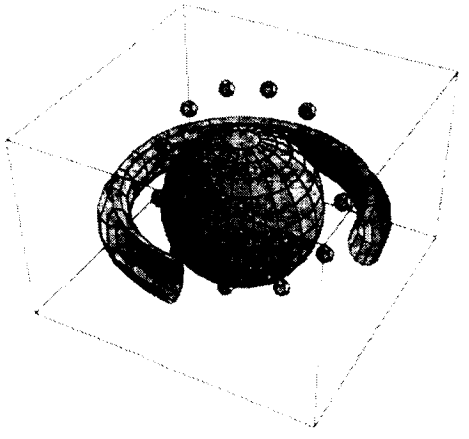


Figure 4-1. 10 Minute Snapshots of Elliptic LEO Inside Inner Van Allen Belt

The satellites would be arranged as 3/plane and 8 planes for high elevation angles in the Northern Temperate zone. Figure 4-2 shows the first 24 satellites at epoch, as they give an indication of the high elevation angles they afford for the Northern Hemisphere. The small arrowheads indicate the velocity vector, and similar velocity vectors are seen for separations in right ascension of 90 deg. This implies that easy crosslinks should be possible between sister satellites at 90 deg. East and West.

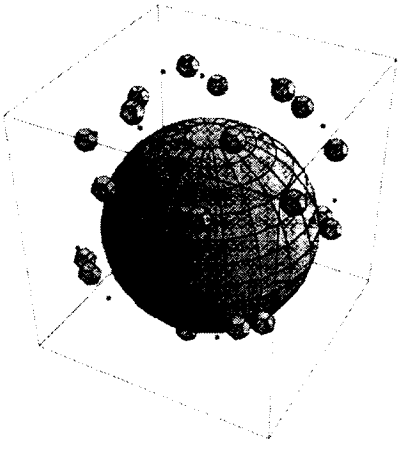


Figure 4-2. First 24 Elliptic LEOs for Northern Hemisphere Coverage

Figure 4-3 indicates the 30 deg. ground elevation contours at epoch. Figure 4-4 shows the results of exhaustive searches for elevation angle statistics for the Northern Hemisphere. For ground locations north of 40 deg., elevation is seen to be very promising. South Florida, however, at 20 deg. North, has sharply deteriorating elevation angle. A complementary set of 24 satellites is introduced in Figure 4-5 for a total of 48. The 30 deg. elevation contours are shown on Figure 4-6. The asymmetry between the northern and southern satellites can be seen at epoch. A full search for elevation angle statistics for all time, however, shows the long term symmetry for elevation statistics (Figure 4-7).

The density function may be found as:

The density function may be found as:

```
Out[41]//FortranForm=
   1/(E**((-32.520821624196 36 -
        -          5.039715796377119e-13*LAT -
        -          0.015229320473122 05*LAT**2 -
        -          1.274872465286e-16*LAT**3 +
        -          2.237417639119951e-6*LAT**4 + x)**2/
        -       (2*(15.42016475784284 +
        -            1.583391651436639e-12*LAT -
        -            0.001422826592914807*LAT**2 -
        -            4.005909975141205e-16*LAT**3 +
        -            2.445116802724547e-8*LAT**4)**2))*
        -    (15.42016475784284 +
        -      1.583391651436639e-12*LAT -
        -      0.001422826592914807*LAT**2 -
        -      4.005909975141205e-16*LAT**3 +
        -      2.445116802724547e-8*LAT**4)*Sqrt(2*Pi))
```

The 2 hour orbit is intended to offer convenient repeat cycles at 3 hours for ground antenna pointing.

5. MILLIMETER WAVE PERFORMANCE FOR THREE SYSTEMS

Figure 5-1 shows a sum of rain (8) and gaseous attenuation (9) for a family of elevation angles vs. frequency for a typical temperate region as Washington, D.C. Only 3 mm/hr rain rate is included for a 99% link availability. Attenuation is seen to increase almost monotonically with frequency, with the only exceptions at the 22 GHz water vapor absorption line and the overwhelming 60 GHz oxygen attenuation. This high attenuation would imply that high frequencies should be unattractive, but it does not include the (f^2) gain advantage of a fixed aperture antenna. Fixed aperture sizes are important for mobile communication, especially aircraft communication, where available real estate is very limited. Antenna system cost is also strongly related to size, even across a large range of frequencies (10). Net loss (loss-antenna gain for a fixed aperture) is shown on Figure 5-2. Characteristic minima appear at attractive frequencies. The entire 30-48 GHz region offers low net loss for high elevation. The system performance is not this obvious, however, because elevation is probabilistic for each of the three systems. The net loss must be integrated (9) across on the system pdf to achieve a system net loss.

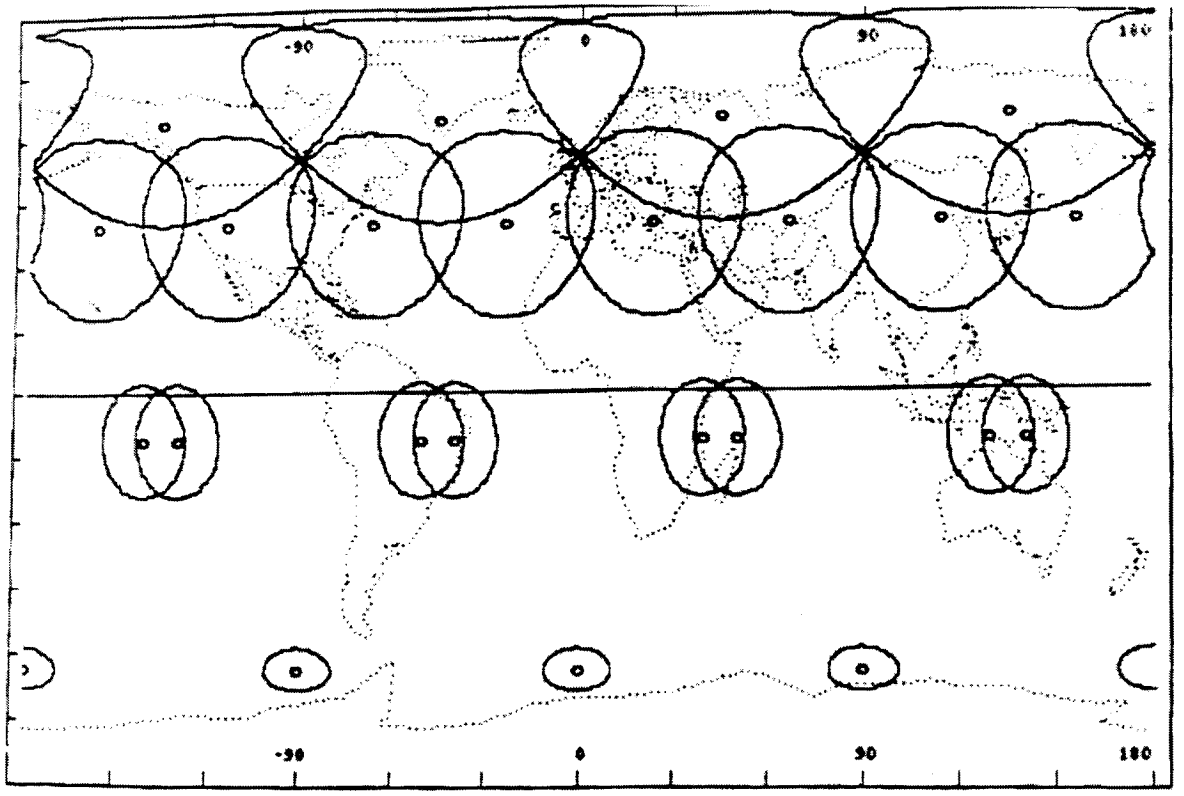


Figure 4-3. High Elevation Footprints for 24 LEOs at Epoch. Contours at 30° Elevation

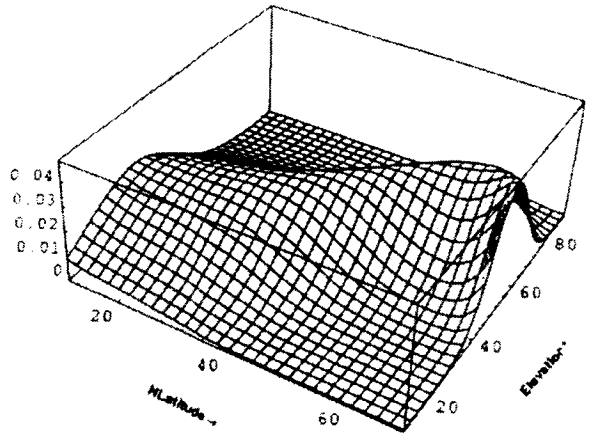


Figure 4-4. Elevation PDF vs Nlatitude (24 sats.)

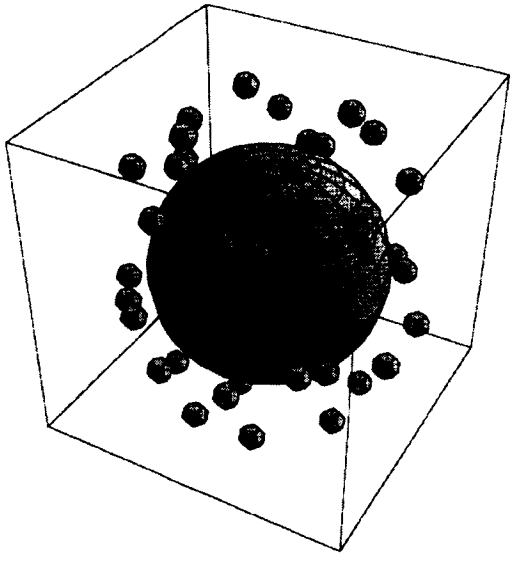


Figure 4-5. 48 Elliptic LEOs at Epoch

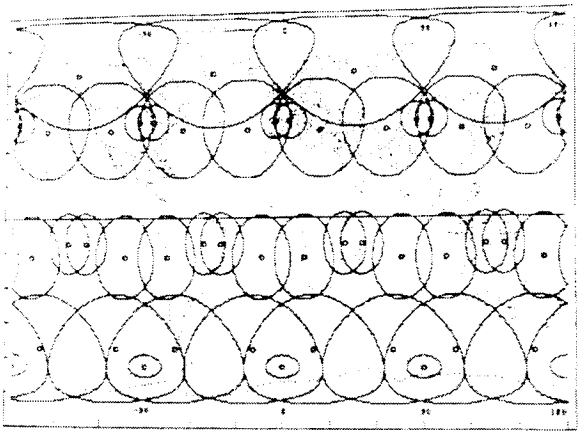


Figure 4-6. High Elevation Footprints for 48 LEOs at Epoch Contours at 30° Elevation

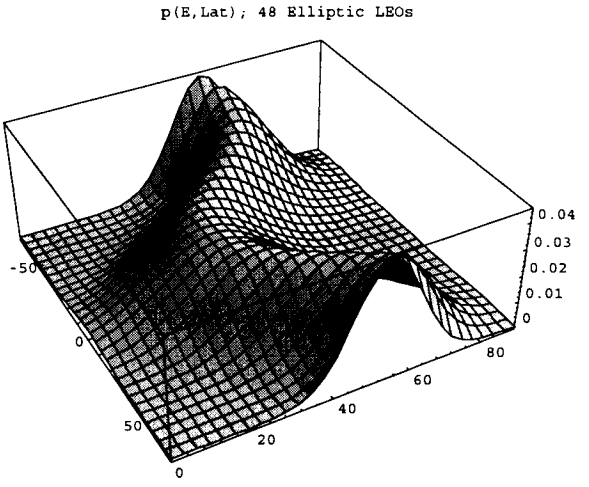


Figure 4-7. PDF (Elevation, Latitude) for 48 Elliptic Satellites

The integration for a hybrid Molniya system is shown as Figure 5-3 (the ordinate is only a relative number, which depends on actual antenna area). Elevation is so high that a system optimum is suggested at 44 GHz. Figure 5-4 has the additional net loss functions for the elliptic LEO system and the M-Star system listed from bottom to top, respectively. The area of coverage, from 30N to 70N, is outside the area of interest for M-Star. All three systems can be found to offer promising millimeter wave performance when the region is confined to 30N to 60N. This becomes immediately important for the M-Star system, because it will function in the 50/40 GHz bands.

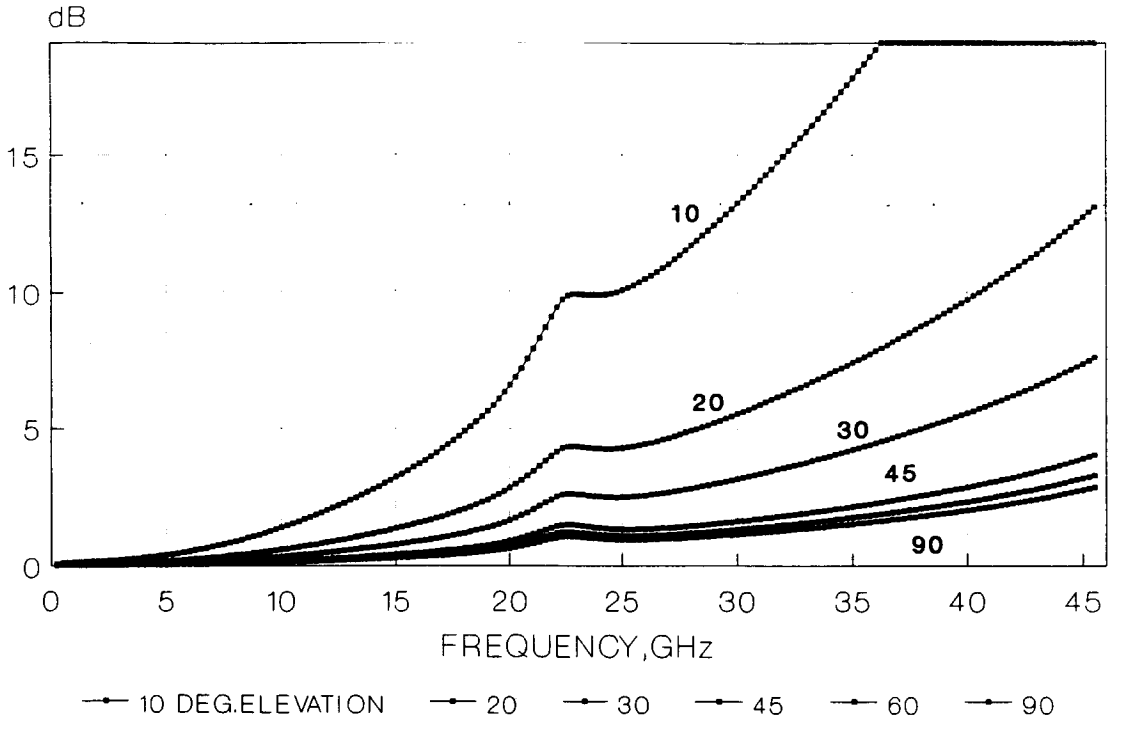


Figure 5-1. Attenuation at Washington, D.C. 3.0 mm/hr Rain Rate

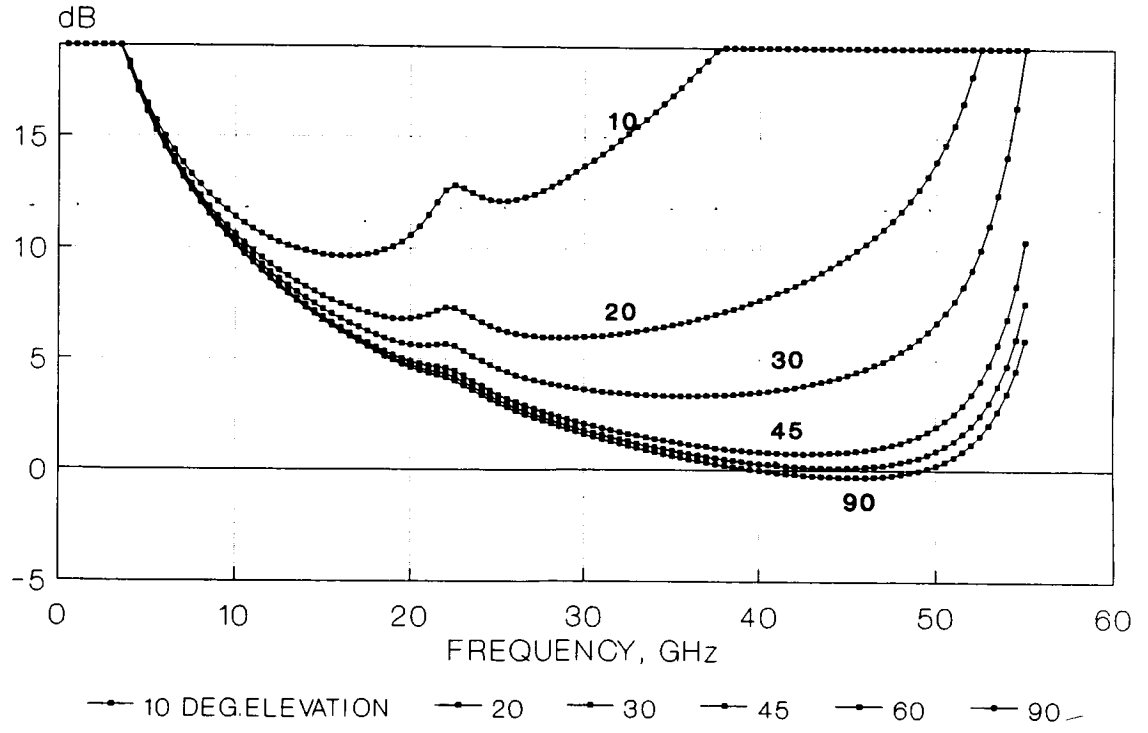


Figure 5-2. Net Loss at Washington, D.C. with Constant Aperture Antenna

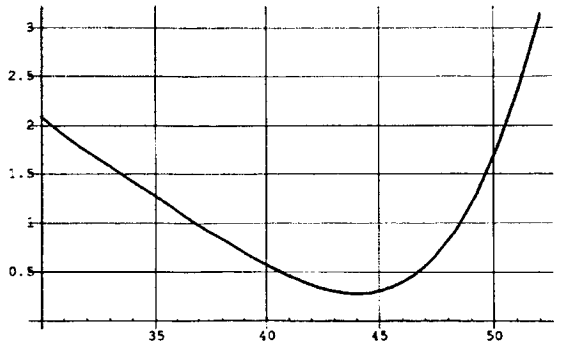


Figure 5-3. Integrated Molniya Net Loss, dB

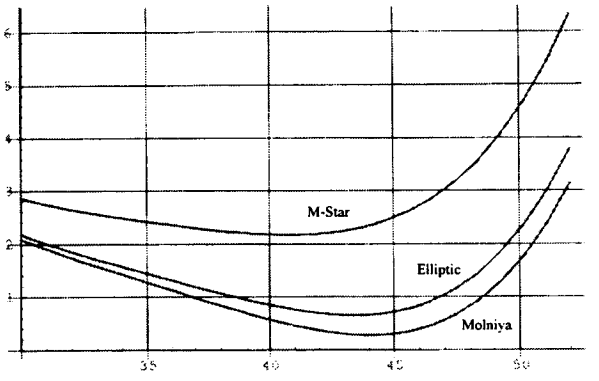


Figure 5-4. Integrated LEO Net Loss, dB Temperate Zone P=0.99

ACKNOWLEDGMENTS

A. H. Jackson provided valuable discussions of millimeter wave antenna developments and power sources. Angie Bowman formatted the paper while fitting unreasonably oversized figures into narrow columns.

REFERENCES

1. J. G. Walker, "Circular Orbit Patterns Providing Continuous Whole Earth Coverage", Royal Aircraft Establishment, Technical Report 70211 (UDC 629.195:521.6), November, 1970.
2. J. E. Draim, "A Continuous Period Four-Satellite Continuous Global Coverage Constellation," Journal of Guidance, Control, and Dynamics, Vol. 10, No. 5, 1987.
3. Dorothy C. Rodgers, P. Christopher, "Satellite Orbits to Relieve Ionospheric Scintillation", Proc. of Ionospheric Effects Symposium, Alexandria, VA, April, 1981.
4. Joseph C. Anselmo, "Motorola Unveils New Satcom Plan", Aviation Week and Space Technology, September 16, 1996.
5. Francesco Linsilata, Bruce Gribble, Barry Gribble, "SGAT User's Manual", Stanford Telecom Report, Reston, VA, 1988.
6. S. Wolfram, "Mathematica, A System for Doing Mathematics by Computer", Second Edition, Addison-Wesley, 1991.
7. A. H. Jackson, P. Christopher, "Angle Diversity for Millimeter Wave Satellite Communications", Proc. IEEE Symposium on Mobile Satellite Communication, London, May, 1996.
8. R. K. Crane, "Prediction of Attenuation by Rain", IEEE Transactions on Communication, Vol. Com-28, No. 9, September, 1980.
9. A. H. Jackson, P. Christopher, "A LEO Concept for Millimeter Wave Satellite Communication", Proc. of International Mobile Satellite Conference, Ottawa, June, 1995.
10. D. L. Pope, "Parametric Representations of Ground Antennas for Communication System Studies", Bell System Technical Journal, Vol. 47, No. 4, December, 1968.

Constellation Coverage Analysis

Martin W. Lo
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
mwl@trantor.jpl.nasa.gov

ABSTRACT

The design of satellite constellations requires an understanding of the dynamic global coverage provided by the constellations. Even for a small constellation with a simple circular orbit propagator, the combinatorial nature of the analysis frequently renders the problem intractable. Particularly for the initial design phase where the orbital parameters are still fluid and undetermined, the coverage information is crucial to evaluate the performance of the constellation design. We have developed a fast and simple algorithm for determining the global constellation coverage dynamically using image processing techniques. This approach provides a fast, powerful and simple method for the analysis of global constellation coverage.

1. INTRODUCTION

Ever since the first satellites entered orbit, people have been concerned with their coverage. When can we observe the satellite from a point on the ground? What is the best location for a ground station? With the advent of satellite constellations, the complexity of the coverage problem increased combinatorially. Even for simple, highly symmetric constellations, coverage questions are not easily answered. Generally, the algorithms are not hard, it's just that the problems are time-varying and there are so many things to keep track of and to compare. As a result, planners rarely are able to obtain global system performance estimates that are rigorous and meaningful. What-if studies are difficult to perform if at all possible.

The difficulty of static discrete counting problems, also known as combinatorics, is well known; for example, the travelling salesman problem or the four color problem. Coverage analysis is inherently difficult because it is a dynamic infinitesimal "counting problem". One might think of this as "infinitesimal combinatorics". The idea of infinitesimal counting suggests that measure theory should be applicable to coverage analysis. This in turn suggests that averaging theory may be used to handle the dynamic combinatorics. These mathematical methods are by no means difficult, but may not be familiar to this community. Nevertheless, they are conceptually simple and provide rigorous metrics for global system performance which are easy to apply.

We present two measure-theoretic methods for global coverage analysis in this paper which are geometric in nature and simple in concept. By simple we mean the resulting algorithms are easier to implement than the standard brute-force approach which is computationally and data intensive. For our purposes simply consider "measure" as a measure of area or volume. Mathematically, we mean the Lebesgue measure on the plane and 3- dimensional space.

The first method we call "Visual Calculus" because it is based on the computations associated with the visualization of the coverage problem. We demonstrate this method on the global performance of a bistatic SAR (synthetic aperture radar) system with a double constellation. In this analysis, in addition to providing the global coverage metric for the system, our method also provides the performance analysis of this new SAR concept. It clearly demonstrates the power of computer visualization for analytic purposes.

The second method is an application of ergodic theory to the satellite view period problem: Given a fixed point on the ground (such as a station or a user), how frequently is the satellite in view of this point on average? The standard approach to this question is to compute the actual view periods from propagating the satellite orbit with at least J2 perturbation assumed. Suppose now we wish to compare the coverage of a mission set of over 100 different orbits for 20 years. For planning purposes, most of the satellite orbits are ill defined. The combination of possibilities can quickly make this problem intractable, especially if the ground element is mobile. The amount of raw view period data generated alone is considerable. Add to this the calculations of the statistics such as view period conflicts; the combinatorial problem quickly becomes intractable. Our approach, using ergodic theory, reduces this problem to solving definite integrals which is easily accomplished in packages

such as Mathematica in just a few seconds. No large data set, no orbit propagation, but rigorous statistics and metrics.

2. THE VISUAL CALCULUS

The coverage of the 2-sphere by satellites is the key question for satellite coverage analysis. Figure 1 defines the satellite to ground station visibility geometry where we have assumed a minimum elevation angle alpha. Note, we use the term "ground station" loosely to mean any element on the ground which wishes to view the satellite. The figure shows that when the satellite groundtrack enters the circle of radius theta about the ground station, it is in view of the station. We call this circle the station mask. One way to visualize the coverage of a complex constellation is to plot the station masks on the globe with some map projection or in 3D. This enables the analyst to quickly get a sense of the performance of the constellation from the geometry.

It is well known that most projections greatly distort the area or measure of the sphere. For example, in the cylindrical projection, the polar regions are artificially enlarged. In order to provide the analyst with a better sense of the coverage, an equal-area projection such as the sinusoidal or the Mollweide projection can be used. In this paper, we use the sinusoidal projection for simplicity to demonstrate this technique.

An equal area projection is also known as a measure-preserving transformation. In other words, it preserves the area of the sphere transformed onto the plane. Given any region in the sinusoidal map, if we transform it back to the sphere, we should get a region with exactly the same area. In particular, this is true at the pixel level, ignoring the discretization error for the moment. Therefore, we can compute area simply by counting pixels.

What are the advantages of this approach? It provides a very simple algorithm for the discretization of the sphere and for computing coverage.

To compute the coverage, for each instant in time, one simply draws the instantaneous limb of the planet with a minimum elevation angle of alpha. One then performs a polygon fill centered on the satellite nadir for the limb circle (radius theta), which is also the station mask at the nadir. This represents all points of the ground which can see the satellite. Notice, no visibility verification is necessary. Suppose we add 1 for each pixel in the limb circle of the satellite and add 0 everywhere else. Suppose we do this for each satellite. The resulting map provides a tabulation of the instantaneous coverage of the constellation. The value of each pixel is precisely the number of satellites in view at that point on the map. In this way, not only have we quickly generated the coverage information, but we have also generated an image which can be displayed for visualization purposes by appropriately defining the color map to reflect the number of satellites in view at each point.

Now begins the fun and games we can play with these data. Let I(t) denote the coverage image produced in the above fashion at the t-th time step. Define

total(S(I)) = total number of pixels in image I where the statement S(I) is true.

Then total(I(t) > 0) provides the total number of pixels > 0 in the image I(t). This counts the number of points on the map which can see the satellite. Let N denote the total number of pixels of the map. Then, since we are using an equal area projection,

$$C(n) = 100 \times \text{total}(I(t)>0)/N$$

gives the percentage of total coverage at time t.

Now suppose we wish to compute the total coverage over land. Let LAND denote the image of the map obtained by setting the pixels over land to 1 and over sea to 0. Similarly, we can define SEA to be the {0,1} matrix for the oceans. Let A*B denote pixel-wise multiplication, i.e. if we think of the images A and B as matrices, this means:

$$A*B(i,j) = A(i,j) * B(i,j).$$

Then, I(t)*LAND gives the coverage over land, and I(t)*SEA gives the coverage over sea. For the coverage over any particular region of the planet, simply create the {0,1}-image, R, for the region (where R is 1 over the region and 0 otherwise). I(t)*R defines the coverage over the region defined by R.

To compute the total coverage from time 1 to n, compute

$$C = 100*\text{total}(\{I(1)+I(2)+ \ldots I(t)\} > 0)/N$$

which defines the total percent of the planet covered by the constellation over time t. To compute the total coverage of the planet by 3 satellites simultaneously, compute

$$C3 = 100*\text{total}(I(t) > 2)/N.$$

For a few satellites, these calculations are easily computed within a package such as IDL. But for larger constellations, custom codes are required, but they are easily implemented. Care must be taken with the image resolution and transformations. To perform the calculus above, all images must be of the same resolution. Resolution can be lost during projection transformations since they are "not reversible" from the pixel point of view.

3. SARCON: APPLICATION OF MEASURE PRESERVING MAPS TO CONSTELLATION DESIGN

This method can be applied to instruments with more complex footprints and pointing. In this section, we present an application to the bistatic SAR (synthetic aperture radar) defined by a double constellation. For this system, the first constellation emits the signals while the second constellation receives the signals. In general, the emitters are cheaper to build and the receivers are much more expensive. By dividing the two, one can provide 50 emitters and 4 receivers, for example, thereby creating an efficient and cost-effective SAR with global coverage. However, typical SARs flown to date have both emitter and receiver on the same satellite. In fact, the scatter geometry provided by this configuration is essential. When the path of the receiver satellite is orthogonal to the emitter satellite, the geometry for the SAR is very poor and the signals are greatly degraded and are useless.

We developed a rapid prototype, SARCON (SAR Constellation Analysis Tool), to analyze this system and to demonstrate the utility of the Visual Calculus. SARCON provided the first global visual verification of the performance of the bistatic SAR. Fig. 2 is an image of an instantaneous coverage from SARCON using a double constellation consisting of 4 satellites each. We refer to the panels numbered 1 through 4, starting from the bottom left going clockwise. Panel 1 provides the raw coverage map without SAR performance. Panel 2 provides the intersection of the footprints where SAR calculations are to be made. In this figure, the "time data" are also plotted. Panel 4 provides the four SAR performance indices: Resolution, SNR, Area, and Time. Panel 3, in the right top corner, provides the image of the pass/fail index obtained from the sum of the indices of Panel 3.

As it is a rapid prototype, little care was used in the numerical and image processing algorithms in the development of SARCON. Aside from total coverage between user specified latitudes, no other analysis was provided in the current version which was mainly used as a first order proof of concept for the bistatic SAR. Once the coverage information is obtained, the statistical analysis outlined above can be used to obtain a great deal of information and metrics on the performance of the system. A second generation tool using a distributed architecture with cluster computing is currently under development to provide just these kinds of analytical capabilities. A multiresolution approach using wavelets to model the sphere is also being considered.

The performance of the bistatic SAR is extremely complicated to analyze. The SARCON visualization provided a quick, intuitive, and quantitative evaluation not only of the bistatic SAR, but of the entire system. The importance of visualization for the analysis of complex geometric problems cannot be over emphasized. The use of measure preserving projections can further leverage the visualization process to provide quantitative information as in the bistatic SAR problem.

4. CONCLUSIONS

This paper presented a new approach to coverage analysis which yields quantitative measures of the constellation performance. The use of a measure preserving map projection enabled the use of simple image processing techniques to compute the coverage of complex missions while providing the visualization at the same time. The SARCON prototype demonstrated the utility and power of this concept.

ACKNOWLEDGMENTS

The work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration. The SARCON prototype was a collaboration with David Imel, Marc Pomerantz of JPL.

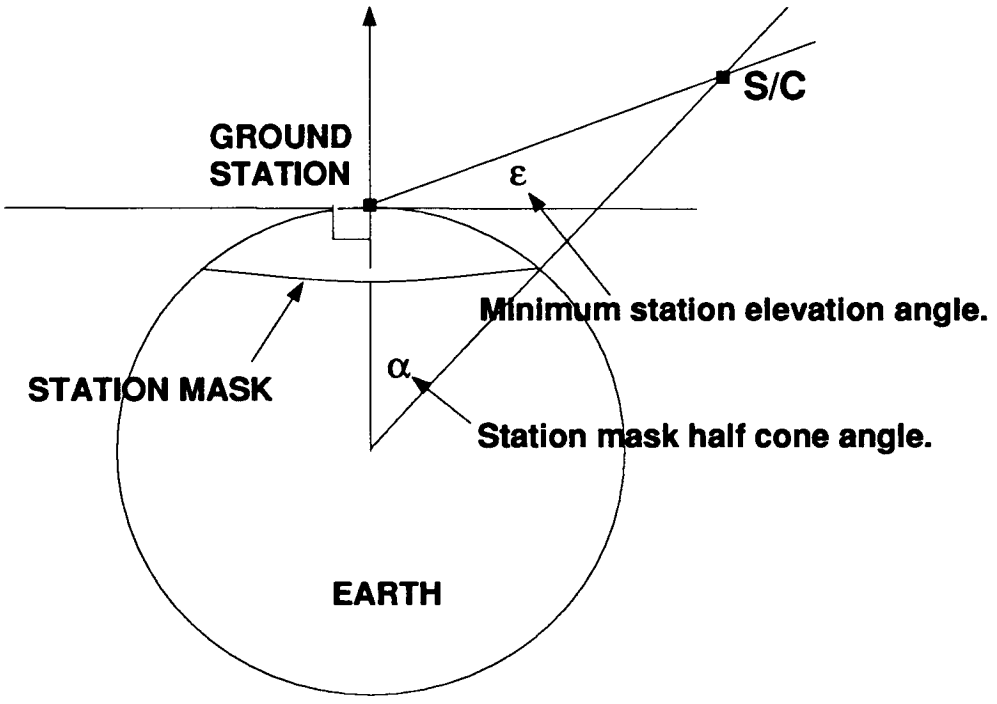
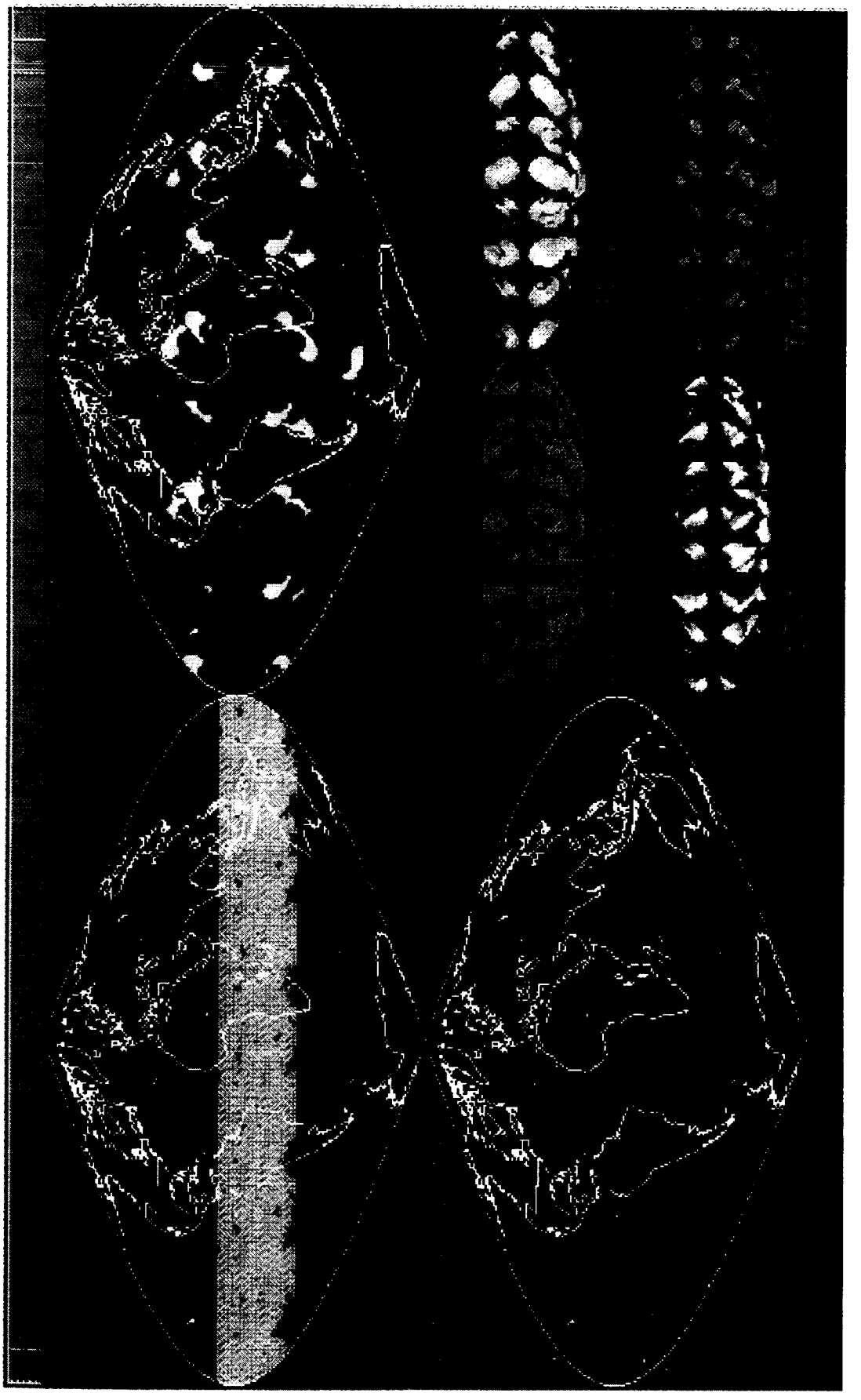


Figure 1. Ground Station to Satellite Visibility Geometry



Power and Bandwidth Tradeoffs for a Third Generation Satellite System

Michael Moher
Communications Research Centre, P.O.Box 11490, Station H, Ottawa, Canada
Tel. 613-998-8669, Fax. 613-998-1686, Email. michael.moher@crc.doc.ca

R.G.Lyons
Square Peg Communications, P.O.Box 11490, Station H, Ottawa, Canada
Tel. 613-820-7817, Fax. 613-993-3122, Email. http://www.crc.doc.ca/innovation/squarepeg.html

L.Erup
European Space Agency (ESTEC/TST), P.O.Box 299, Noordwijk, NL2200 AG, The Netherlands
Tel. +31 71 5653161, Fax. +31 71 5654598, Email. lars@t.estec.esa.nl

Abstract

One of the distinguishing characteristics of the next generation of communications satellites will be the multitude of spotbeams. Currently planned systems have from tens to over a hundred spotbeams per satellite. A key to the economic viability of these systems will be their ability to reuse the limited available frequency spectrum. In this paper we present a tool for the analysis of the interference in a multibeam satellite and some examples showing its application to the system design.

1.0 Introduction

This paper describes a tool, developed in the course of an ESA-sponsored transmission study for a third generation mobile satellite system [1], to evaluate different frequency reuse strategies for a multibeam satellite. The object of [1] was the investigation of methods of providing multi-media services to mobile terminals ranging from paging services (1 kbps) to video-telephony (64 kbps) over a multibeam satellite. Other aspects of this study are covered in [2],[3].

The analysis of intrasystem interference in a multibeam satellite system depends upon a number of parameters. These parameters include the geographical location of the users sharing the channel, the number of such users, the power levels of the users, the characteristics and design of the satellite spotbeams. The effect of the interference, described by the above parameters, in turn depends upon the modulation and coding strategy used. Given the variability of the interference environment, we take a statistical approach to the analysis of multibeam interference [4].

We then show how this tool can be used to examine tradeoffs between system capacity and i) satellite EIRP in the forward direction, and ii) satellite G/T per spot beam and terminal EIRP in the return direction. Insight is derived into a fundamental power-bandwidth tradeoff associated with the choice of modulation and coding strategy, namely single user bandwidth efficiency versus multibeam frequency reuse factor.

2.0 Derivation of C/I pdf

The C/I probability distribution function (pdf) resulting from all contributing random variables determines the probability that the interference will exceed the tolerable level for any given user. Conversely, the allowable service unavailability due to interference determines the tolerable interference pdf, i.e., the tolerable frequency reuse.

Consider interference in the forward direction at a location z caused by a co-channel transmission to another terminal at location y. This is denoted by $I_y(z)$. Define a normalized link-loss function, $\Phi_y(z)$, which represents the link losses at location z when using the spotbeam associated with position y. The link-loss function Φ is a composite of a number of effects,

$$\Phi_y(z) = \frac{G^y_{sat}(z) G_{mob}(z)}{\Phi_o L_p(z)} \qquad (1)$$

where G^y_{sat} is the antenna gain of the spotbeam corresponding to location y in the direction of z,

G_{mob} is the antenna gain of the mobile terminal, and L_p is the pathloss. The constant Φ_o is chosen to normalize the link-loss function to one at the subsatellite point. The link-loss function is the relative loss experienced by any mobile which is not at the subsatellite point. The satellite antenna gain is the product of spotbeam gain and the scanning loss.

With this definition of the link-loss function, ideal power compensation corresponds to

$$P_c(z) = \frac{1}{\Phi_z(z)}. \qquad (2)$$

In practice it should be possible to almost perfectly compensate the static link losses but not the dynamic components due to multipath and shadowing. Thus, the power compensation could also include a position-dependent margin to compensate for these dynamic components.

Let P_o be the nominal power transmitted by the satellite for a user at the subsatellite point. Then the co-channel interference $I_y(z)$ is

$$I_y(z) = (P_o\Phi_o)P_c(y)\Phi_y(z) \qquad (3)$$

where the first term represents the nominal received power at the subsatellite point, the second term represents the power compensation, and the third term is the isolation provided by the spotbeam. One can generalize the above result by adding a factor to represent the dynamic power variations due to voice activation.

The first object is to calculate the probability distribution of the interference at location z, $p[I_z]$. The interference at a position z due to the co-channel users is given by

$$I(z) = \sum_k I_{y_k}(z) \qquad (4)$$

The locations of the co-channel interferers, y_k, may or may not be in the same spotbeam as z; this will depend on the frequency reuse strategy. The probability density function of the interference at location z is given by

$$p[I(z)] = \otimes_k p[I_k(z)] \qquad (5)$$

where $\otimes$ represents convolution, and $p[I_k(z)]$ represents the pdf for the kth interferer obtained by considering all possible positions for that interferer.

The C_z/I_z distribution is a function of both the interference distribution and the received carrier power at the desired mobile. With ideal power compensation, the received power level is constant, $C_z=P_o\Phi_o$, independent of position. Computing the distribution $p[C/I_z]$ is not tractable analytically, but it is straightforward numerically, given $p[I_z]$. The overall C/I distribution is then given by, using Bayes' rule,

$$p[C/I] = \int_V p[C/I|z]p[z]dz \approx \sum_Z p[C_z/I_z]p[z] \qquad (6)$$

where V is the coverage area of the satellite and $Z=\{z_i\}$ represents a sampling of the coverage area. The distribution of terminals, $p[z]$, includes the geographical distribution of terminals but may also include information regarding traffic density.

These results are generic in that they apply to both wideband and narrowband modulations. The numerical computation of the final distribution (6) using steps (3) and (5), can be simplified computationally by taking into account the arrangement and distribution of users among the beams for particular frequency reuse patterns.

The interference mechanism on the return link differs from that of the forward link. In the forward link, the interfering co-channels are those which radiate into the antenna of the desired mobile terminal. In the return link, the interfering co-channels are those mobiles which radiate into the spotbeam of the desired mobile. The same set of mobiles is generally involved but the interference calculation is different. In particular, on the return link, the normalized interference power from a co-channel user at location y into a user signal from location z is

$$I_y(z) = P_o\Phi_o P_c(y)\Phi_z(y) \qquad (7)$$

This is analogous to the forward direction except that $P_o\Phi_o$ is the nominal received power from a mobile at the subsatellite point; and the last term, Φ_z, is now the link-loss function for the return direction. Note that Φ has its dependencies exchanged compared to (3). Replacing equation (3) with equation (7) is the only difference between the forward and the return link analysis.

3.0 C/I cdf results: narrowband and wideband cases

The tools described in the previous section have been applied to a number of scenarios but, as an example, we consider the following two. In the first scenario, frequencies are reused in every beam

but users within a beam are assumed orthogonal and interbeam interference is Gaussian. This model is often applied to wideband CDMA. In the second scenario, frequencies are used in every third beam. This model is often applied to narrowband FDMA.

In Figure 1, the cumulative distribution function (cdf) of the forward link C/I is shown for two frequency reuse strategies with 37 and 61 spotbeams. The total traffic is the same for both narrowband and wideband cases. The figures use the energy per channel symbol to interference power density, E_s/I_o, as a measure of C/I since it reduces confusion when there is FEC encoding and signal spreading applied. The vertical axis represents the probability that the C/I on the horizontal axis is not exceeded, as averaged over the service area of the satellite. These results correspond to a satellite at the apogee of an 8-hour highly elliptical orbit.

The first observation is that there is only a slight dependence of the interference distribution on the number of spotbeams. This indicates that even with just 37 spotbeams, results are close to the asymptotic results one would expect with an infinite number of beams.

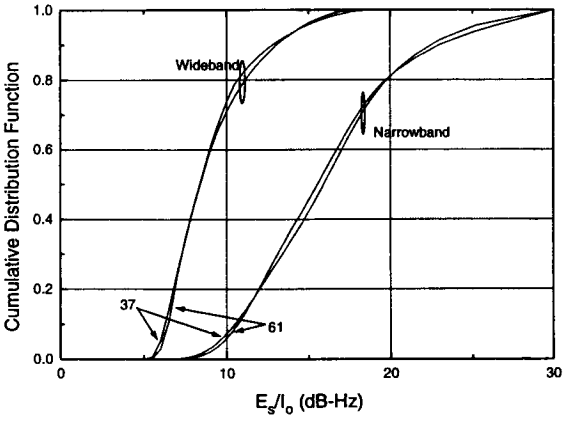


Fig. 1 Wideband and narrowband forward link C/I distributions for 37 and 61 spotbeams.

The second observation is that there is a considerable difference in the C/I for the two strategies. The C/I is much greater in the wideband case. The system-level effect of this behaviour will depend on the modulation and coding strategy. Also in the wideband case, there is a much smaller variance in the C/I as indicated by the steepness of the distribution. In the narrowband case, 5% of the locations/frequencies will have a C/I ratio of 10 dB or less but more than 50% will have a C/I ratio exceeding 17 dB.

An important factor affecting the co-channel interference is the traffic distribution. In addition to the uniform traffic assumption, two "hotspot" traffic models are considered. The first corresponds to what might be considered a typical case with 50% of the traffic in five out of 37 beams and the rest of the traffic uniformly distributed. The second hotspot model corresponds to what might be considered worst case, with 80% of the traffic in ten beams. The results assume that the hotspots are approximately uniformly distributed over the service area and are not in adjacent beams. The wideband forward link C/I distribution for the three different traffic scenarios are shown in Figure 2.

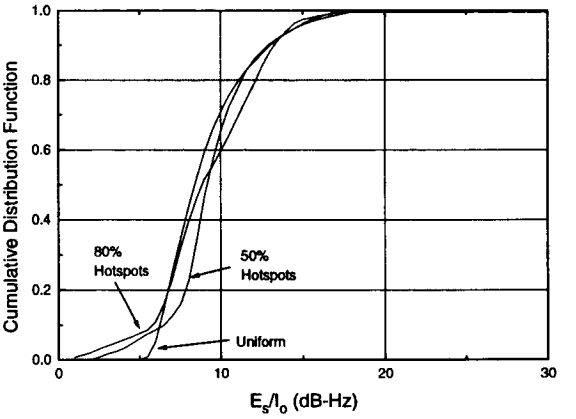


Fig. 2 Forward link wideband C/I distributions for different traffic scenarios.

For the forward link, the major difference is an extension of the lower tail of the distribution. This can have a significant impact on the frequency reuse capabilities.

4.0 Capacity results and tradeoffs: wideband case

In [1], services were to be provided to a variety of terminal types ranging from handheld, vehicular, to transportable. However, not all services are expected to be provided to the lower end terminals, and higher end terminals are presumed to dominate. Three issues that had to be addressed were:

i) the power-bandwidth tradeoff for the various combinations of modulation and coding techniques and multiple access strategies;
ii) the dependence of this tradeoff on terminal types; and
iii) the dependence of this tradeoff on the BER requirements of the different services.

The tradeoff includes the comparison of the various modulation and coding techniques and is affected significantly by the presence of interbeam and intrabeam interference.

4.1 Forward link example

The forward link satellite power requirements will depend upon the satellite antenna gain, the path loss, and the propagation conditions. All of these parameters will vary as a function of the spotbeam, i.e., the elevation angle of the terminal. Of prime interest, however, is the total EIRP requirement of the satellite, or the average EIRP requirement per spotbeam. The capacity results were parameterized in terms of satellite EIRP and mobile-link bandwidth. An example of the forward link tradeoff between power and bandwidth for a multibeam satellite is shown in Figure 3. The results illustrated in Figure 3 correspond to a BER of 10^{-3}.

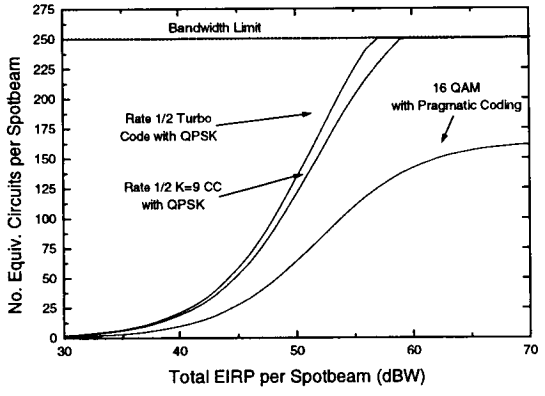


Figure 3: Capacity with 75% highway land mobile terminals, 25% urban handheld terminals, and 2 MHz mobile-link bandwidth.

This figure shows the power and bandwidth tradeoff for three modulation and coding strategies. For each strategy, it assumes a mixture of 75% highway land mobile terminals and 25% urban hand held terminals with the E_b/N_0 requirements shown in Table 1. All codes use a 512 bit block and corresponding interleaving.

Coding Strategy	*AWGN*	*Urban*	*Highway*
QPSK r= ½ , K=9	2.3	5.0	2.6
QPSK r= ½ , Turbo	1.5	4.75	2.0
16-QAM, K=9	4.75	7.4	6.2

Table 1. Comparison of E_b/N_0 (dB) requirements for different modulation and coding strategies.

The propagation models were as follows: i) an additive white Gaussian noise (AWGN) channel was assumed for line of sight operation; ii) a Rician channel with a K factor of 8 dB and a Doppler spread, f_d=10 Hz was assumed for a handheld terminal in an urban environment; and iii) a Rician channel with a K factor of 15 dB and a Doppler spread, f_d=160 Hz for a vehicular terminal in a highway environment. Consistent with the results of [5], the time dispersion is assumed to be negligible. Shadowing was not considered as it was not anticipated to affect the choice of modulation and coding.

The results in Figure 3 assume a frequency reuse strategy that may be called orthogonal CDMA. The frequency band is reused in each beam, users within a beam are assumed to be orthogonal, and interference between beams is assumed to be Gaussian with 50% voice activation. To compare the different modulation and coding strategies without going into the details of the different service offerings, an *equivalent circuit* was defined as an 8 kbps service, and the system capacity was calculated in terms of the number of equivalent circuits. The equivalent circuit bandwidth is a function of the modulation and coding strategy. For the QPSK strategies it is assumed to be approximately $10\log_{10}$(8 kHz). For the 16-QAM strategy it is assumed to approximately $10\log_{10}$(4 kHz).

These results assume an operating BER of 10^{-3} and a total mobile link bandwidth of 2 MHz. For all three modulation and coding strategies the results show a progression from a power-limited situation to a bandwidth-limited situation as the total available EIRP per spotbeam increases. The power-bandwidth tradeoff for the two rate ½ strategies, QPSK with convolutional coding and QPSK with Turbo-4 coding, is quite similar. With the 16-QAM pragmatic coding scheme and low satellite EIRP, capacity is power-limited to a greater extent than the other schemes, as one might expect. However, one might also expect that, once there was sufficient power, the capacity of the 16-QAM pragmatic scheme would exceed the other two schemes. However, in reality, the 16-QAM pragmatic scheme becomes interference-limited before it becomes bandwidth-limited, and it never achieves the bandwidth limit. Consequently, the 16-QAM pragmatic scheme is clearly the inferior of the three techniques in terms of system capacity in this scenario.

4.2 Return link example

In the return link, the total EIRP per spotbeam is not a relevant parameter. There are two new parameters to take its place. The first is the EIRP of the mobile, which, of course, will be limited. The second is the G/T of the satellite. The latter is often primarily determined by the field of view of the satellite and the number of spotbeams.

For the return link, we consider an example of using a frequency reuse strategy corresponding to quasi-synchronous CDMA. The return link modulation is $\pi/4$-QPSK for the vehicular terminals and the compatible constant envelope scheme PFM4[6] for the handheld terminals. Both types of terminals use a rate ½, constraint length 9, convolutional code. The results shown in Figure 4 assume a mixture of 75% vehicular terminals and 25% handheld terminals, with E_b/N_0 requirements of 2.8 and 5.2 dB respectively, for a BER of 10^{-3}.

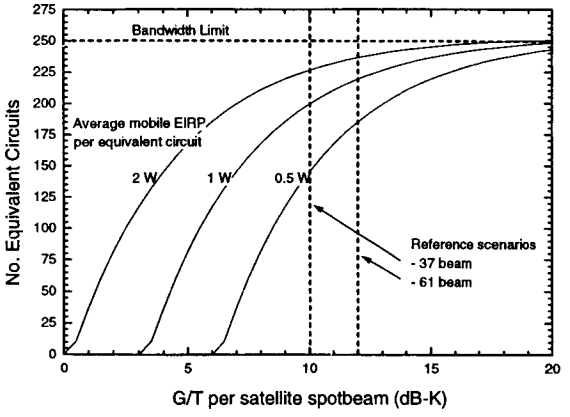


Figure 4: Quasi-synchronous return-link capacity in equivalent circuits as a function of satellite G/T assuming 75% vehicular terminals, 25% handheld terminals, and 2 MHz of mobile bandwidth.

In Figure 4, we show the return link capacity in equivalent circuits as a function of the satellite G/T for the case of asynchronous users. Three curves are presented corresponding to average EIRPs of handheld terminals of 0.5, 1 and 2 watts, respectively. The vehicular terminal will need less EIRP to achieve the same performance, and the calculations assume that the vehicular terminals only use the minimum EIRP required, in order to minimize the interference. Also indicated on the curves are the approximate satellite G/Ts for the two reference scenarios of 37 and 61 spotbeams. These reference scenarios assume a satellite at an altitude of 26000 km with coverage down to 40° elevation. The capacity is the number of equivalent circuits per beam, and assumes that a total mobile bandwidth of 2 MHz is available in each beam. Clearly, the capacity increases significantly with mobile EIRP and satellite G/T.

Relative to the case where users are perfectly orthogonal, the use of a quasi-synchronous strategy will have two effects on performance. The first is an increase in the intrabeam interference due to the fact that the spreading codes are not perfectly orthogonal. The second effect will be a greater sensitivity to timing error. The mobile EIRP corresponds to the worst case terminal, in this case, the handheld.

5.0 Conclusions

In this paper, we have presented a generic tool for the interference analysis in a multibeam satellite. Examples have been presented to show how this tool can be used to determine this interference, and how the results can be used to evaluate the tradeoffs between various system level parameters. From the examples, it is clear that the best approach depends upon how much satellite power and bandwidth are available, and the power and bandwidth required by the application.

References

[1] R.G.Lyons et al, "A transmission system for future satellite mobile and personal communications," Square Peg Communications Final Report, ESTEC Contract No.: 11515/95/NL/US, 1997.

[2] S. Crozier, B.Mazur, L.Erup, "Design of a synchronous CDM forward link and a quasi-synchronous CDMA return link for future satellite based mobile and personal communication systems", Proc. IMSC'97, Pasadena, California.

[3] D.Boudreau et al, "The performance of video and high-quality audiotelephony services for mobile satellite applications," Proc. IMSC'97, Pasadena, California.

[4] M.Moher, L.Erup, and R.G.Lyons, "Interference statistics for multibeam satellites", Proc. 2nd European Workshop on Mobile/Personal Satcoms, Rome, Italy, 1996.

[5] A.Jahn et al, "A wideband channel model for land mobile satellite systems," Proc. IMSC'95, Ottawa, Canada.

[6] S.Crozier, P.Guinand and B.Mazur, "Design and performance of precompensated frequency modulation (PFM) for use with a quasi-synchronous CDMA return link," Proc. IMSC'97, Pasadena, California.

Intelligent Paging in Dynamic Satellite Personal Communication Networks

C.Meenan, R.Tafazolli, B.Evans
Mobile Communications Research Group
Centre for Communications Systems Research
University Of Surrey
Guildford, England, GU2 5XH
Phone: +44(0)1483 259808 Fax: +44(0)1483 259504
Email: c.meenan@ee.surrey.ac.uk

Abstract

In this paper we present a novel spotbeam selection criteria for the first paging step in a dynamic satellite personal communication network. This paging method is based on the creation of Virtual Paging Cells. We demonstrate how the first paging step success rate and load depends on the number of virtual paging cells used. We then apply the algorithm to two satellite paging methods in two types of satellite fading channel. The results are shown for two different types of satellite constellation. These are a Globalstar-like LEO system with 19 spotbeams and a ICO-like MEO system with 163 spotbeams.

Introduction

When a mobile communications network receives a mobile user terminated call, the system pages the mobile terminal. In a dynamic satellite personal communication network paging is performed through a collection of spotbeams that provide coverage over the location area that relates to the mobile terminal being paged. The number of spotbeams varies depending on the size and latitude of the location area. Reducing the number of spotbeams used in paging is highly desirable as this can save satellite power, increase system capacity and reduce interference.

To reduce the number of spotbeams used in paging, the paging can be performed in multiple steps. For example in two step paging, the most likely spotbeams are paged through, if no response is received from the terminal the remaining spotbeams are paged. In multiple step paging it is important to maximise the success rate of the first paging step, as a low success rate will cause an increase in call set-up time as further paging steps will be required. In this paper we a present a novel paging strategy and apply it to the dynamic location area approach for mobility management. Spotbeam selection for paging in this new intelligent paging strategy is based on the creation of virtual paging cells within the users location area. The probability of successful paging through a given set of spotbeams is evaluated by calculating the users position probability within each of the virtual paging cells. This algorithm allows factors such as user mobility profiles, point and time of last contact etc., that have be shown to reduce the paging load in terrestrial PCN in [1] &[2], to be applied to a dynamic satellite PCN.

Mobility Management and Paging Redundancy

A mobility management scheme suitable for dynamic S-PCN's was introduced in [3] & [4]. This mobility management protocol avoids terminal location updates due to the satellite spotbeams continuous motion over the surface of the earth. In this mobility management scheme mobile terminals make a location update after moving a pre-determined distance (*Lar*) from the point of last contact with the network. Therefore the Location Area (LA) associated with each terminal becomes a circle of radius *Lar* centred at the point of last contact with the network.

When the network requires contact with the mobile terminal it uses knowledge of the point of last contact with the terminal and its location area radius (*Lar*), to page through the paging channel on all the spotbeams that the terminal maybe monitoring. This corresponds to all spotbeams that are totally or partially covering the terminals location area.

This paging technique guarantees a high paging success probability. However a significant amount of redundant paging signalling occurs when this technique is adopted. The source of the paging signalling redundancy can be identified as follows;

1. Significant areas outside of the terminals location area will be paged over, this will always occur when spotbeams are larger than the location area.

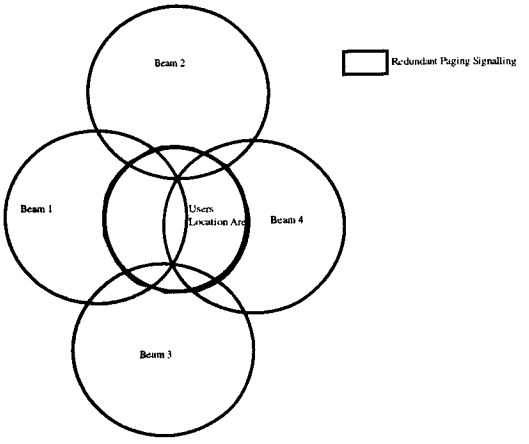


Figure 1 Redundant Paging Outside of location area.

Consider the scenario shown above in Figure 1. Spotbeams 1,2,3 and 4 are all on the same satellite and are providing coverage over the users location area. Therefore at paging time each spotbeam is paged through. As demonstrated in the diagram above large areas outside of the users location area are also paged.

2. The two satellite constellations considered in this paper provide multiple satellite visibility to combat shadowing in an attempt to provide a higher quality of service to the user. Therefore a high percentage of the time a location area will be covered by more than one satellite. If we consider a single terminal roaming in the location area, using the paging method described above, all satellites above the minimum elevation angle of the terminal will be paged though.

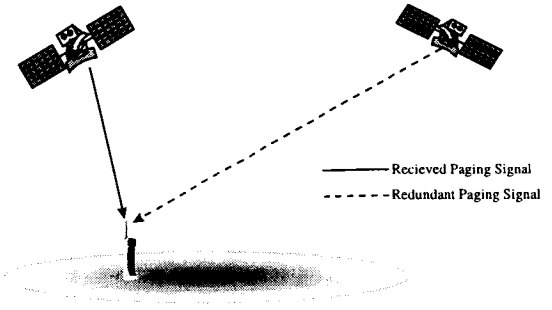


Figure 2 Redundant satellite paging.

3. Paging will occur through spotbeams that the terminal is not monitoring due to its position inside the location area. This type of redundancy is similar to that which occurs in GSM[5] paging. Considering the example in Figure 1, assuming that the terminal is monitoring only spotbeam 4, then although spotbeams 1,2 and 3 are providing coverage over regions of the location area, paging the terminal through these beams is redundant.

The work presented in this paper proposes a novel technique of eliminating type 1 and 3 redundancy by the creation of *virtual paging cells* to compute the probability of paging through a given spotbeam being successful. Two methods of satellite selection for paging are presented to reduce type 2 redundancy, these are ;

Optimum Satellite Paging

The system pages the mobile terminal through spotbeams on the optimum satellite (i.e. the satellite with the highest elevation angle w.r.t. the mobile terminal). As shadowing becomes more prevalent at lower elevation angles, the highest satellite will have the highest probability of being unshadowed[6].

Satellite Diversity Paging

This technique exploits satellite diversity when it is available, i.e. the mobile is paged through the highest satellite and through the second highest satellite also. It is envisaged this paging scheme would be used when the probability of each channel being shadowed is high.

System and User Mobility Parameters

In this paper we consider the two types of dynamic satellite constellation described below in table 1.

Name	GlobalStar-like[7]	ICO-like[8]
Number of Sats	48	10
Altitude (km)	1414	10349
No. Of Spotbeams	19	163
Min. Elevation	10°	10°

Table 1 Constellation Parameters.

Approximate spotbeam models were developed for each of the constellations, these are shown below in Figure 3.

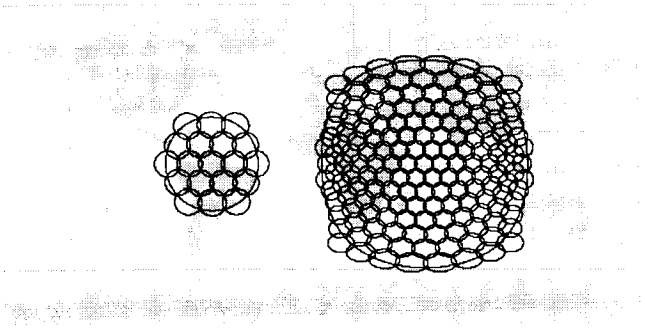


Figure 3 Spotbeams models for Globalstar-like and ICO-like systems.

The mobility and call arrival rate assumptions for users are shown below in table 2.

Number of Users	1000
Service Area	20°- 60° latitude
Direction of Travel	Uniform between 0 and 2π
Speed	LAR/2 km/hr ± 10%
Location Update Rate	0.5 updates/hr
Call Arrival Rate / user	2 calls/hr
Simulation Time	24 hours

Table 2 User Parameters

Shown below in Figure 4 is the full paging load for the Globalstar and ICO-like systems. It is clear as the location area size increases, the number of spotbeams required to page the location area also increases

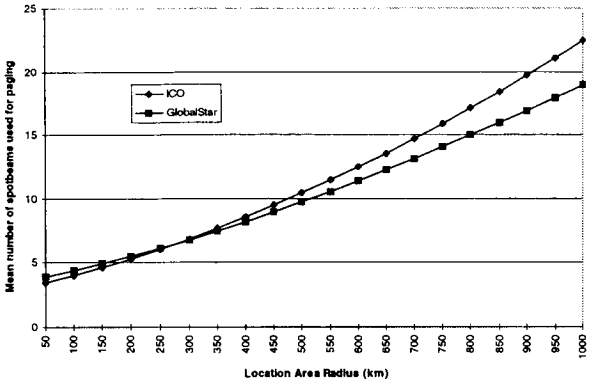
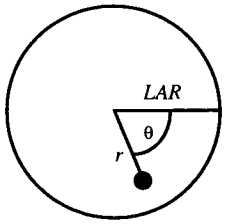


Figure 4 Globalstar-like and ICO-like Spotbeam paging requirement.

Spotbeam Selection for First Paging Step

Consider a mobile terminal at angular co-ordinates (r,θ) from the centre of its location area at time of paging (t_p). We define the function $Oc(r,\theta, t_p)$ that identifies the spotbeam on a given satellite in which a mobile at the point (r,θ) at instant t_p is located.



$$\text{Spotbeam} = Oc(r,\theta, t_p) \quad (1)$$

We also define a user position probability density function ,P, at the point (r,θ) from the centre of the location area at time t_p.

$$P(r,\theta, t_p) \quad (2)$$

As the terminal must make a location update after moving a distance *LAR* from the centre of its location area, the p.d.f. boundaries are;

$$\int_{r=0}^{LAR}\int_{\theta=0}^{2\pi} P(r,\theta,t_p)\, r\, d\theta\, dr = 1 \quad (3)$$

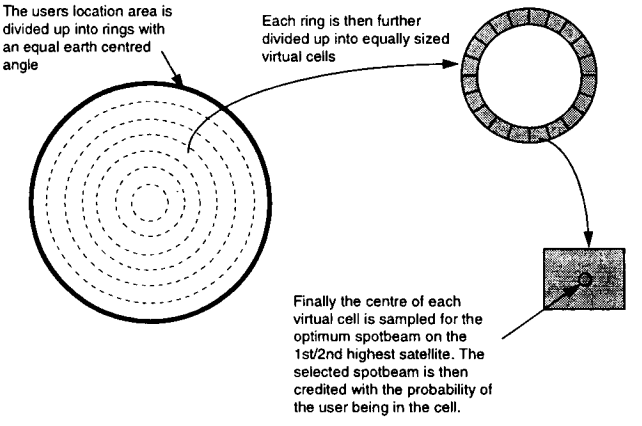
Therefore the probability of the user being located in a given spotbeam S is;

$$PS = \int_{r=0}^{LAR}\int_{\theta=0}^{2\pi} \delta\big(Oc(r,\theta,t_p), S\big)\, P(r,\theta,t_p)\, r\, d\theta\, dr \quad (4)$$

Where the delta function, $\delta_{(m,n)}$, is;

$$\delta_{(m,n)} = \begin{cases} 1 & m = n \\ 0 & otherwise \end{cases}$$

To calculate the probability of spotbeams required for diversity paging, equation (1) needs to be modified to include the spotbeam on the second highest satellite for a mobile terminal the point (r,θ) from the centre of the location area at the instant t_p.
Equation (4) is evaluated by the creation of Virtual Paging Cells (VPC) shown below;



The probability of a user being in a given cell at paging time is highly dependant on the mobility of the user. In this paper we assume that the all users travel at a constant speed uniformly distributed in azimuth. The cell creation algorithm and cell occupancy probability for users of this mobility is presented in [9].

Shown below in Figure 5 is the first step paging load vs. the number of VPC's used. Spotbeams on the optimum satellite (i.e. highest elevation) in the Globalstar-like system, with a probability greater than zero were selected for the first paging step.

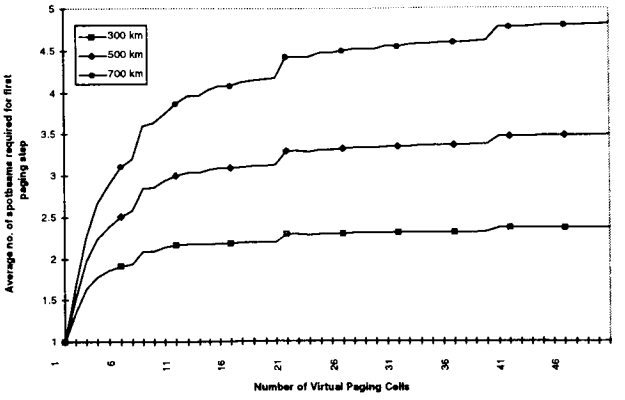


Figure 5 First step page spotbeam load vs number of VPC.

As expected as the number of VPC is increased more spotbeams are selected for the first paging step. In the case of one VPC, the spotbeam covering the centre of the location area is the only one paged through. The sudden increases in paging load result from the cell creation algorithm creating an additional ring of VPC.

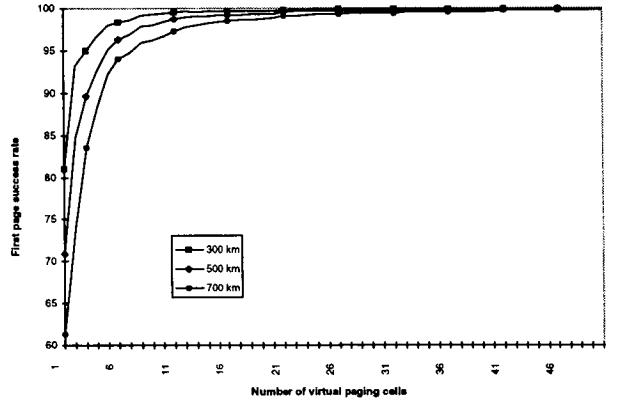


Figure 6 First step page success rate vs number of VPC.

As more paging cells are included and more spotbeams are selected for paging the first page success rate also rises. The results shown above could be used as an intelligent paging algorithm as they exhibit a trade-off between paging load and success rate. However it is more appropriate to base the selection of a spotbeam for the first paging step on its probability of paging success.

First Paging Step Signalling Load

In this section we evaluate the first step paging load in paging scenarios where spotbeams are included in the first paging step if their probability of paging success is above a certain threshold. As shown above in Figure 6 as the location area size increases more VPC are required to maintain a given success rate. Therefore the number of VPC used in this paper is equal to the location area radii in km.

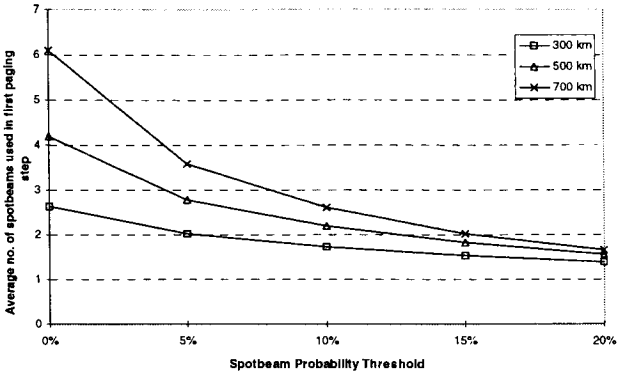


Figure 7 ICO-like optimum satellite first step paging load vs. spotbeam probability threshold.

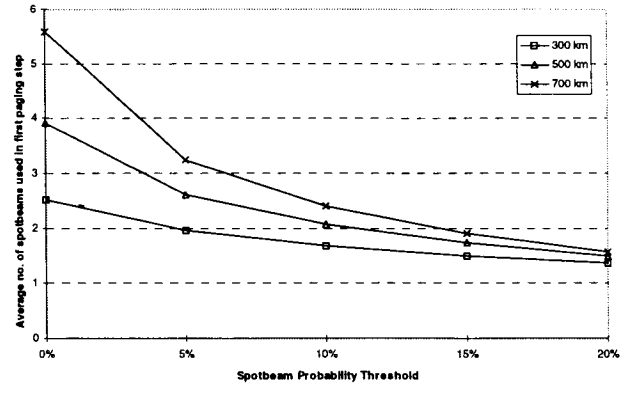


Figure 8 Globalstar-like optimum satellite first step paging load vs. spotbeam probability threshold

The results demonstrate that the application of a probability threshold results in an increasingly smaller number of spotbeams being selected for the first paging step.

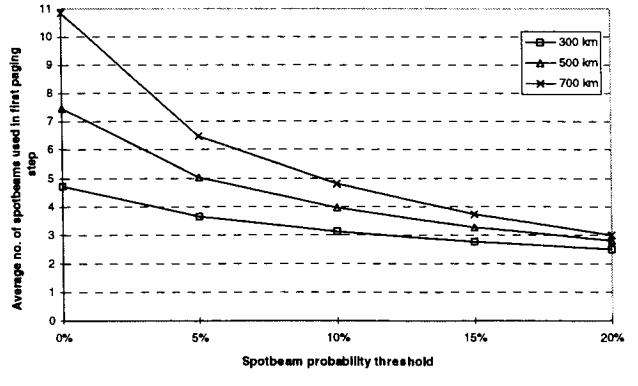


Figure 9 ICO-like satellite diversity first step paging load vs spotbeam probability threshold.

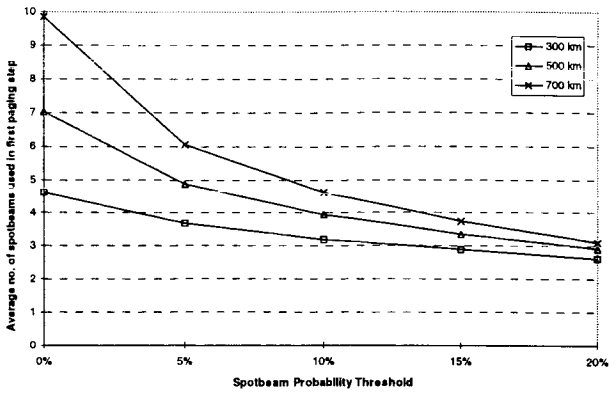


Figure 10 Globalstar-like satellite diversity paging load vs. spotbeam probability threshold.

The results in Figures 9 & 10 show the number of spotbeams selected for the first paging step in the diversity satellite paging scheme.

Channel Characteristics

To evaluate the paging success rate we must first consider the satellite fading channel. A simplified approach to the satellite channel model[10] is a two-state Markov process in which the channel is either 'good' (Ricean type fading) or bad (shadowed). It is assumed that a paging packet is lost if the channel is bad. The mean duration of the good and bad states depends on the satellite elevation w.r.t. the mobile terminal, and the environment in which the terminal is located. Shown below in table 3 are the channel parameters used in this paper.

Elevation	Environment	Good State Duration (m)	Bad State Duration (m)
13°	City	9	70
18°	City	8	32
24°	City	27	52
34°	City	24	33
43°	City	42	49
13°	Highway	90	29
24°	Highway	188	62
34°	Highway	1500	12
43°	Highway	8300	17

Table 3 Channel Characteristics

First Paging Step Success Rate

In this section we present the first step paging success rates for the two S-PCN systems being considered. The results demonstrate how the first step paging success rate varies as a result of the applied spotbeam probability threshold and satellite paging technique.

Optimum Satellite Paging

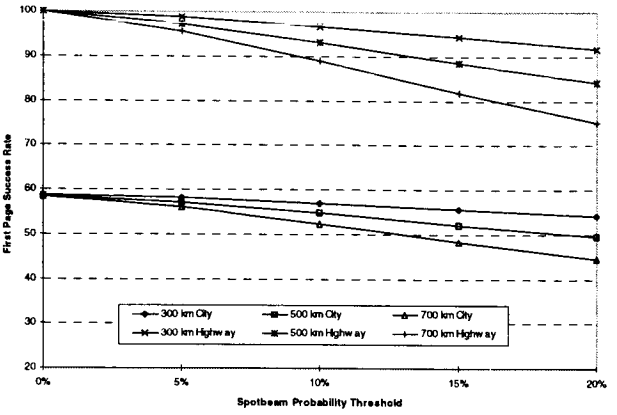


Figure 11 ICO-like optimum satellite paging success rate vs. spotbeam probability threshold.

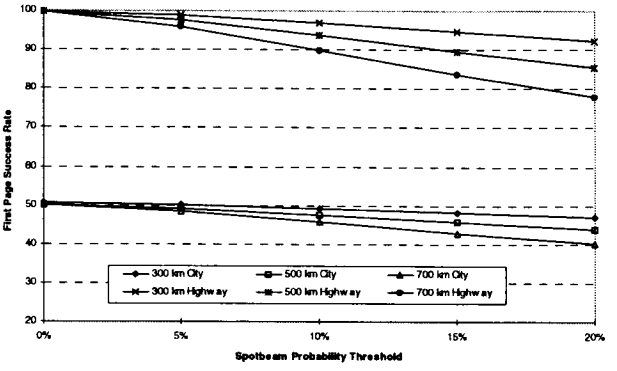


Figure 12 Globalstar-like optimum satellite paging success rate vs. spotbeam probability threshold.

From the results shown above is clear that the application of a spotbeam probability threshold does not significantly reduce the paging success rate. As expected the first step paging success rate in the city environment is significantly lower than that in the highway. However in the city environment the shadowing has the effect of reducing the paging failure rate due to the application of a larger spotbeam probability threshold.

Diversity Paging

Shown below are the results for the first page success rate using diversity satellite paging vs. spotbeam probability threshold.

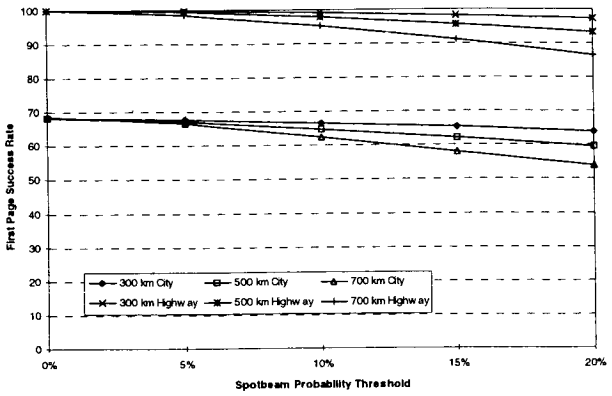


Figure 13 ICO-like satellite diversity paging success rate vs. spotbeam probability threshold.

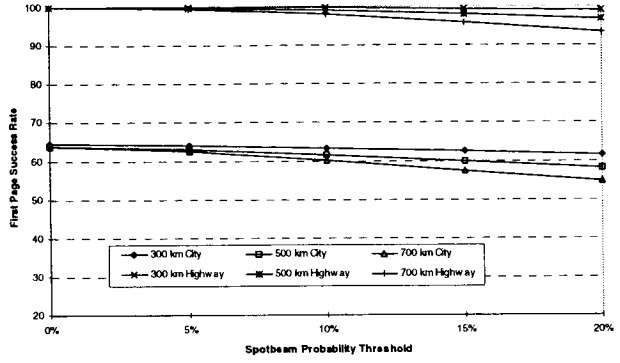


Figure 14 Globalstar-like satellite diversity first page success rate vs. spotbeam probability threshold.

The results demonstrate that satellite diversity paging results in a higher first step paging success rate. The increase in success of the first paging step is larger in the Globalstar-like system than in the ICO-like system. This is due to the fact the satellite diversity is more available in the Globalstar-like system, however the ICO-like system success rate remains higher as the mean satellite elevation angles are higher in this system. This fact also explains why the application of a spotbeam probability threshold with diversity paging in the highway environment results in a noticeably larger decrease in success rate for the ICO-like system than it does for the Globalstar-like system.

Conclusions

In this paper we have discussed the different forms of paging redundancy that result from paging in dynamic satellite systems. We have presented a novel spotbeam selection algorithm and applied it to two paging schemes. The paging load of the first paging step was presented, for a range of spotbeam probability thresholds. The success rate of each of the paging schemes was evaluated in two types of satellite fading channel. The results demonstrate the first step paging load can be much smaller than the full paging load while not making a significant impact on the paging success rate.

The *virtual paging cells* spotbeam selection method can be extended to include knowledge of the user's direction of travel and speed to reduce the first paging step signalling load further.

References

[1] G. Lyberopoulos, et. al. , 1995, "Intelligent Paging Strategies for Third Generation Mobile Communication Systems", IEEE Transactions on Vol.44, No.3,543-553.

[2] Sami Tabbane, 1995, "An Alternative Strategy for Location Tracking", IEEE Journal on Selected Areas in Communications, Vol. 13, No.5, June 1995.

[3] A. Sammut, et. al., 1994, "Mobility Management Related Signalling for a MAGSS-14 Based Satellite Personal Communications Network (S-PCN)" ,COST227D, National Technology University Of Athens, Greece.

[4] Saint, Race Project 2117, Deliverable No. 15.

[5] M.Mouly and M.B.Pautet, "The GSM System for mobile communications", ISBN:2-9507190-0-7.

[6] M. Parks et. al., 1993,"High Elevation Angle Propagation Results, Applied to a Statistical Model and an Enhanced Empirical Model", Electronic letters, Vol.29,No.19.

[7] Alcatel Espace,'Globalstar', France. Submission to RACE 2 / SAINT project, June 1994, Oslo, Norway.

[8] J. Singh.'Inmarsat and Personal Mobile Satellite Communications', 10th European Satellite Communications Conference, December 1993, London, England.

[9] C.Meenan et. al., 1997, "Intelligent Paging Schemes For Non-GEO Satellite Personal Communication Networks", VTC 97, Phoenix Arizona.

[10] E.Lutz et. al., 1991 "The Land Mobile Satellite Communication Channel-Recording, Statistics, and Channel Model", IEEE Trans. On Veh. Tech., 40, No.2

An *Ad Hoc* Wireless LAN to Enable Satellite Paging and Ringing in Urban Areas

Wolfhard J Vogel
EERL/The University of Texas at Austin
10100 Burnet Road, Austin, TX 78758
Phone: +1 (512) 471-8608, FAX: +1 (512) 471-8609
email: Wolf_Vogel@mail.utexas.edu

ABSTRACT

Because satellites soon to be used for personal hand-held communications are power limited, the fade margin to be allocated to forward services such as paging and ringing may not be sufficient to achieve message delivery with a satisfactory level of probability when users are inside large, attenuating buildings. Especially in urban areas, where the user density may be high, one may take advantage of the spatial diversity of the signal level over the distance scale of a few hundred meters. For this purpose, it is proposed to incorporate packet radio features into satellite pagers and handsets. Customer terminals so equipped will be able to assist each other in receiving satellite messages. Basic design features and possible applications are given.

INTRODUCTION

Communications satellites employed in multi-user, bi-directional services are power limited. In orbit, the finite power resource of these commercial spacecraft has to be shared among many users. On the ground, the power transmitted by the mobile customer terminal is constrained by battery life and safety considerations. For satellite-to-customer (forward) services, such as paging and ringing, the data rate can be slowed down below that required for voice transmissions. Hence, the bandwidth can be reduced with a concomitant reduction of the noise power and increase in signal-to-noise ratio. Even with this strategy, the attenuation of the satellite signal can reduce its strength to below the detection threshold. This is expected to be the case especially in urban areas, where customers spend a significant percentage of their time inside large office buildings.

Measurements of slant path attenuation at L- and S-Band into buildings have shown that typical mean fades into illuminated, peripheral rooms of office buildings are about 20 dB with a standard deviation of about 8 dB and that the fades (in dB) are distributed normally [1]. Consequently, fades are expected to exceed 30 dB, 33 dB, and 39 dB at the 90, 95, and 99 percent probability levels, respectively and thus may exceed the "punch-through" capability of satellite transmissions. Figure 1 shows a comparison of into-building fading experimental data to the normal distribution. The data were collected at 1626 MHz, using an airplane to simulate a low-earth-orbit satellite and a narrow-band receiver with a patch antenna to simulate the customer equipment. During the "satellite" pass, the receiver was walked back and forth along an interior 11th

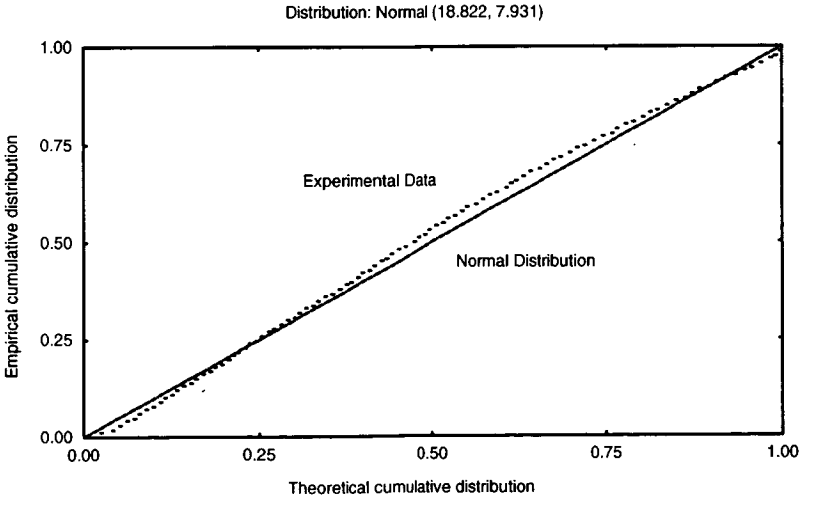


Fig. 1. Patch antenna fades in dB inside an old office building compared to the normal distribution.

floor hallway of an old office building. Other data, obtained for instance in a modern 45-story building, were similar.

This paper describes a method (patent pending) of exploiting local medium-scale spatial diversity by the *ad hoc* formation of a packet radio network between a user located in a restricted environment and a user who receives a satellite signal. It is expected that this method can significantly improve the chance of success for two new satellite personal communications network (S-PCN) services, i.e., paging and call-notification (ringing) in dense urban areas. These features are important elements in the spectrum of satellite personal communications services which are to be offered very soon. For most planned systems, satellite power limitations will prevent the transmission of signals of sufficient strength to ensure high-probability penetration into steel-concrete-glass buildings. This, however, is where business persons who are a very important customer group for these services and who have to be served with very high dependability tend to spend much of their time.

COOPERATIVE PAGING AND RINGING

The concept of and techniques for routing of messages to addressable nodes in a wireless mobile network have already been studied in the development of packet switched radio networks for military applications; for example, [2] describes a low-cost packet radio testbed. The aim of these types of systems is to provide a dynamic wireless network consisting of multiple stations and one or more servers over which one can coordinate the actions of the members of the network. The terminals may be deployed in restricted environments, such as forests, mountains or cities, and their relative positions may be continually changing. The system depends on a topology of overlapping regions of radio connectivity among user terminals which allows any terminal to send messages to any other terminal via a chain of interconnected pairs of terminals. Much work has been done on the methodology to develop hardware and protocols for optimal message traffic flow through such a network [3, 4].

The new paging/ringing amplification by cooperative emission relay (Pacer/Racer) system introduced in this paper differs from the packet radios mentioned above in several aspects: (1) the service is intended for the forward link only, i.e., it is unidirectional, (2) the basic topology involves only two customer terminals forming an *ad hoc* network, and (3) the service can be grafted onto a system not necessarily designed as a packet switched network.

Each Pacer/Racer customer terminal (CT), either a pager and/or telephone, incorporates the capability to both transmit and respond to requests for cooperative forwarding of pages or rings. When a CT fails to detect any satellite paging or ringing activity for approximately 4 seconds, as counted by the CT's internal clock, it will issue, with very low power, a request for assistance. A CT may fail to detect activity, for example, because it is in a signal-blocking environment. This request will include the blocked CT's unique address and other information, such as whether the request is for paging, ringing, or both. Any other CT within the general vicinity of the blocked CT, generally on the order of 300 meters, which receives the request will store the address of the requesting CT in a memory stack.

From then on this receiving CT, in addition to its normal functions, such as responding to its own messages and rings, will also respond to pages or rings directed to any CT address in its cooperative address stack. The receiving CT responds by re-transmitting such messages, also with low power, to the blocked CT. Thus, the blocked CT will receive a relayed message which may be decoded by the CT as if it had been issued directly from the satellite.

A functional block diagram of a Pacer/Racer CT is shown in Figure 2. The CT includes an RF unit with associated antenna, a digital signal processor, a beeper/vibrator/ringer, a display unit, user interaction keys, an internal clock, and a battery. The RF unit contains components used to receive satellite transmissions and to transmit, with low power, satellite-like transmissions. The digital signal processor contains components used to determine whether a message is intended for the present CT, or should be relayed to another CT.

The RF unit in Figure 2 includes a low noise receiver, a transmit/receive switch, and a low power transmitter. In operation, the CT receives signals from the satellite or another CT through the antenna. The antenna couples the electromagnetic waves to the transmit/receive switch. During the receive mode, the switch output connects to the low-noise receiver and disconnects the transmitter output. The low noise receiver filters and amplifies weak received signals in the desired frequency band and converts them to baseband without adding significant thermal noise power.

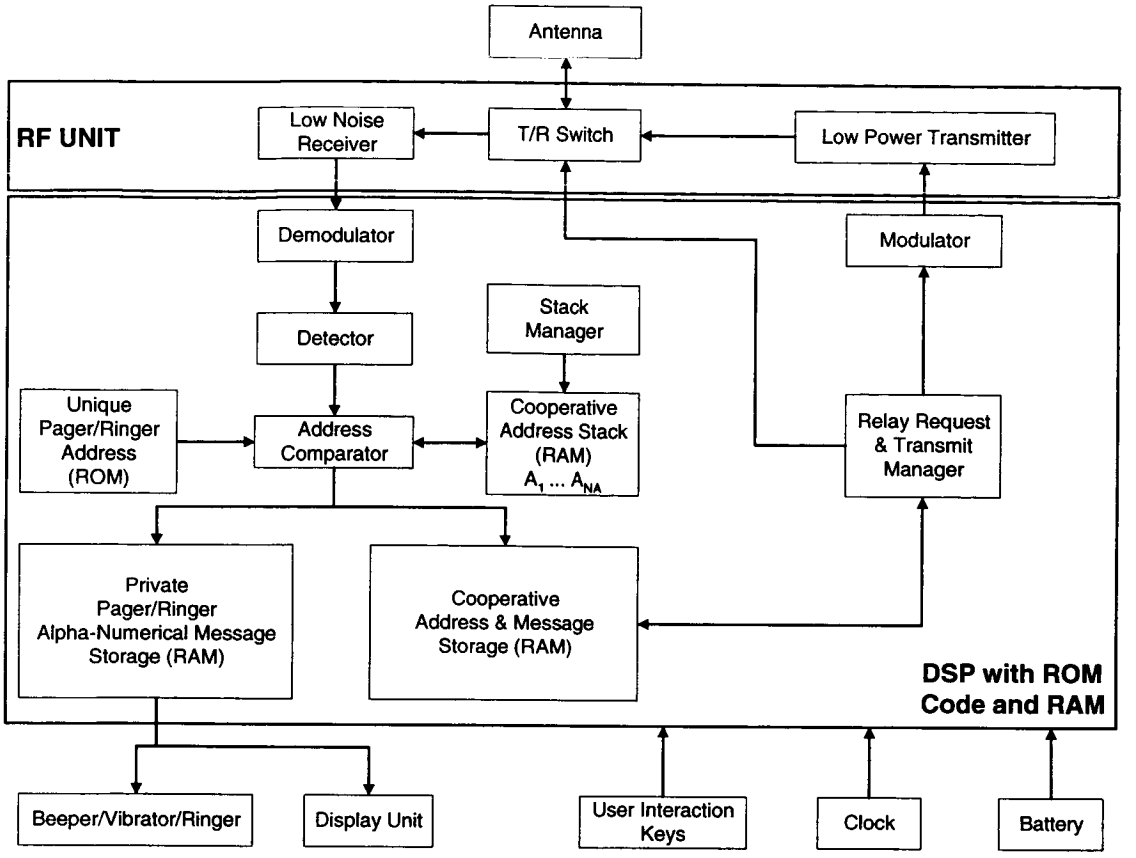


Fig. 2. Functional block diagram of the Pacer/Racer customer terminal.

After passing through the receiver, received signals enter the digital signal processor. The received signals are first converted from modulated baseband signals to logic levels in the demodulator. The logic signals then pass to the signal detector which decodes and forwards the decoded address and any message content to the address comparator. If the signal detector notes the absence of any useable satellite signal for a continuous period of approximately 4 seconds, it will notify the relay request & transmit manager to issue a request for paging or ringer relay (RPR). As long as the blocked condition continues, even intermittently, the request will be repeated approximately every 4 minutes.

The address comparator compares the received message's address to the unique pager/ringer address, which is stored in read-only memory. If there is a match, the comparator sends the related received message to the private pager/ringer message memory. In addition to the message, the time of arrival of the message is also stored in memory. If a new message comes in, it can cause the beeper/vibrator/ringer to be activated to alert the CT user to the presence of a new message. If the received message is not for the receiving CT, the address comparator will check to determine whether the message is an RPR from another CT. If the message is a request, the address comparator will send the requesting CT's address and type of request (ringer, pager amplification, or both) to the cooperative address memory (CAM). This memory stores the address of the requesting CT and the time at which the request was received most recently. The CAM may be just one read/write structured memory location, but a larger memory of 16 or 256 locations may be utilized. Over-sizing any memory will have only minimal impact on system efficiency. The stack manager is a device or algorithm that deletes old addresses from the cooperative address memory. After approximately 4 to 5 minutes with no additional request from a blocked CT, the stack manager will delete the blocked CT's address from the cooperative address memory.

If the received message is neither for the receiving CT nor an RPR, the address comparator will compare the received address to the addresses stored in cooperative address RAM. If there is a match, the address, message and time of arrival are stored in the cooperative address and message memory (CAMM). The size of this may be just one location, if the message is re-transmitted immediately during the next message slot and no repeated re-transmissions are made. If, however, re-transmission is delayed or repeated, then the CAMM may be larger, on the order of 16 or 256 locations. Over-sizing the CAMM

will have minimal impact on system efficiency. If the received message is not an RPR or a match, then the address and message are discarded.

Messages to be relayed to other CT's are sent from CAMM to the relay request & transmit manager (RRTM). This is a device or algorithm that repackages the relayed message in the same format used by the S-PCN. The RRTM transmits the relayed message to the modulator at the proper time. Alternately, when ordered by the signal detector, the RRTM will send a request for cooperative emission, also at the proper time. The proper time is derived from the satellite transmissions if available, or the internal clock, which may be, for instance, a stable crystal oscillator. The timing of all actions is derived from the clock when the CT is not receiving synchronizing satellite transmissions.

The output of the RRTM enters the modulator, which converts the logic signals to a modulated baseband signal. This signal is output to the low power transmitter which converts the baseband signal to the operating frequency of the S-PCN. The transmitter also filters out any undesirable frequencies and provides sufficient amplification so that approximately 90% of any UT's within a range of approximately 300 meters can receive and decode the signal. The transmitter type depends on the frequency and modulation employed by the S-PCN, as a satellite-like transmission is desired.

The amplified signal from the transmitter passes through the T/R switch, which is controlled by the RRTM to connect the transmitter output to the antenna and disconnect the input to the low noise receiver while in the transmit mode. The antenna then converts the signal into propagating electromagnetic waves to be received by another CT.

Other features of the Pacer/Racer CT shown in Figure 2 include the display unit, which is an alpha-numerical display for user interaction with the CT. Also, user interaction keys provide for one or more switches or keys for user interaction with the CT. The battery stores and supplies the electrical power needed to operate the CT.

One method of implementing the cooperative emission relay is for an S-PCN user entering a restricted environment, such as a steel-concrete-glass building, to bring a second CT to be utilized as a relay station. This second CT is placed at a location more likely to receive satellite signals, for example, outside of the building or at a different location within the building. However, as the S-PCN system attracts more subscribers, such a method should become unnecessary, especially in high-density urban areas. Other methods of implementing the relay system depend on the environment. For example, an office building, hotel, or airport operator may deploy roof-top Pacer/Racer CT's as local repeaters. Alternatively, satellite operators may "seed" strategic locations with Pacer/Racer CT's.

The drain on the CT battery does not present a major problem, as the transmitting CT need only re-transmit selected messages over a short range with minimal impact on battery life. The power required for the close-range bursty communication is approximately several milli-watts, permitting low-cost implementation. Further, this drain does not present a problem for satellite telephones, which already have a transmitter present, as well as a large capacity battery.

Since the Pacer/Racer system is transparent to the satellite, it has minimal impact on satellite design. However, for optimal application of the relay scheme some impact on the paging and ringing signaling protocols can be expected. The problems of building penetration of the relay signals can be analyzed with the tools already developed for terrestrial paging. These tools include measurement campaigns or simulations based on measurements published in the literature. The timing protocols of the relay signals can also be analyzed with already existing terrestrial design tools. The pager/ringer emissions may use, for example, a slotted ALOHA protocol, or any other common random messaging scheme. Preferably, the timing protocol should match the protocol used by the S-PCN. Depending on the protocol implemented, the short range nature of the relay allows coherent addition of multiple relay emissions, thus further increasing the chance of reception.

Strategies may include that the RPR includes an indication of the generation level of the request, to enable more sophisticated relay managing. For example, a blocked CT's RPR can be repeated by another blocked CT's RPR, and so on, until it reaches a clear CT. Earlier blocked CT's addresses are sent as part of the later blocked CT's RPR. Including a relay depth in the memory stacks and requests would allow limiting the number of relay hops by a relay depth manager, not shown in Figure 2. Further, a clear CT may send an acknowledgment message to a blocked CT, indicating its receipt of the RPR. This acknowledgment message may include the address and an acknowledgment code and provide a timing signal. This timing signal is used by a blocked CT for time synchronization so that it can accurately predict when to re-transmit a message.

To enable optimal system performance, CT addresses in the memory stack of a clear CT should have a limited lifetime (latency), maybe of the order of 4 to 5 minutes. This latency will prevent a clear CT's memory stack from becoming filled. Because of this latency, blocked CT's must re-issue RPR's within a time interval less than that lifetime to allow for continued cooperative emissions.

CONCLUSION

As an alternative, high-power paging and ringing repeaters could be placed into selected high-density cities such as New York, Hong Kong or Tokyo, but these would produce a negative image for a global satellite system, besides having their own technical and logistical drawbacks. The CT's outfitted with pager/ringer amplification by cooperative emission (PACER/RACER) technology are only activated when needed and have a self-limiting range. An S-PCN user visiting a building can also bring a second CT and deploy it as a relay station at a location in or near the building more likely to receive satellite signals, for instance in a car or a window. As the satellite service attracts more subscribers, however, such a strategy should become unnecessary, especially in high-density urban environments. Traditionally, pagers are one-way devices. Two-way pagers for messaging have been considered, but as they have to reach back to the base station and because of battery-power limitations, they are impractical. The PACER/RACER CT's receive signals from satellites but need to retransmit only a few of these over just a short range with minimal impact on battery life. Depending on the protocol implementation, the short range also allows coherent addition of multiple relay emissions, thus further increasing the chance of reception.

BIBLIOGRAPHY

[1] **W. J. Vogel and N. Kleiner**, *"Propagation Measurements for Satellite Services into Buildings,"* Proceedings of the 1997 IEEE Aerospace Conference, Snowmass at Aspen, Colorado, Vol. 1, pp. 91-107, Feb. 1- 8, 1997

[2] **W. C. Fifer and F. J. Bruno**, *"The Low-Cost Packet Radio,"* in Proc. Of the IEEE, Vol. 75, No. 1, pp. 33 - 42, Jan. 1987

[3] **B. M. Leiner, D. L. Nielson and F. A. Tobagi**, *"Issues in Packet Radio Network Design,"* ," in Proc. Of the IEEE, Vol. 75, No. 1, pp. 6 - 20, Jan. 1987

[4] **R. Dube, C. D. Rais, K. Y. Wang and S. K. Tripathi**, *"Signal Stability-Based Adaptive Routing (SSA) for Ad Hoc Mobile Networks,"* IEEE Personal Communications, Vol. 4, No. 1, pp.36 - 45, Feb. 1997

Enhancing the Coverage of Mobile Satellite Systems: The Local Signal Repeater

Tony Spizzichino, Enrico Salvatori
Elsacom
Viale Maresciallo Pilsudski 92, 00197 Rome, Italy
Tel: +39 6 80902062 Fax: +39 6 8088104
E-mail address: spizzichino@elsacom.finmeccanica.it

ABSTRACT

Mobile Satellite Systems (MSS) will provide, in the near future, global mobile communication services virtually everywhere in the world. However, due to the ITU Power Flux Density limitations, such services may suffer from harmful blockage and radio-wave penetration phenomena that is characteristic of man-made structures. When users are inside such structures, these propagation phenomena can severely affect the quality of service of those potential MSS users like executives and businessmen who, more than others, need continuity of service. In particular, office buildings, airports, but also trains and ships are places where these mobile satellite systems might experience link outages depending upon material characteristics (e.g., cement, coated glasses) and receive antenna position.

The solution to this problem could be the implementation of a *local signal repeater.*

INTRODUCTION

Amongst the future MSS systems, Globalstar services will be the ones probably suffering less than others from propagation phenomena. The combined use of CDMA technology, Rake receiver multipath re-construction techniques and satellite diversity should provide a very good signal quality and service coverage. However, blockage phenomena inside office buildings and airports can still occur. Service continuity can not be augmented in all circumstances by the fact that MSS user terminals are inherently multi-standard. For example, a U.S. executive just landed in some European airport, traveling by train or participating in a business meeting inside an office tower will still have difficulties making calls back to his family or staff; his MSS terminal with the multi-standard built-in capability will be of little use either for lack of international roaming agreements, different analogue and digital standards, or tail charges that are too expensive.

A *local signal repeater* collects the signal received from an outside receive antenna and transmits it inside the man-made structure to be covered (vice versa for the out-bound traffic). Depending upon the multiple access scheme adopted by the MSS system, the implementation of a *local signal repeater* will need different technologies and, as a general rule, different costs. In satellite systems like Globalstar such implementation represents an easy technical task with almost no impact on the overall network costs. Due to the intrinsic properties of path diversity and Rake receiver multipath reconstruction technique found in Globalstar, a very simple repeater can be designed. In fact, with the adopted CDMA, it is possible to avoid taking into account the possible interference among the repeated signal and the direct or reflected signal. The repeated signal will be treated by the Rake Receiver as a normal signal and will be simply added to the other possible signals (with the same Walsh code, but in general delayed) if any.

BACKGROUND

The Globalstar Space Segment

The Globalstar space segment consists of 48 satellites in a 1410 Km circular low earth orbit distributed in eight orbital planes with six satellites spaced equally at 7.5 deg

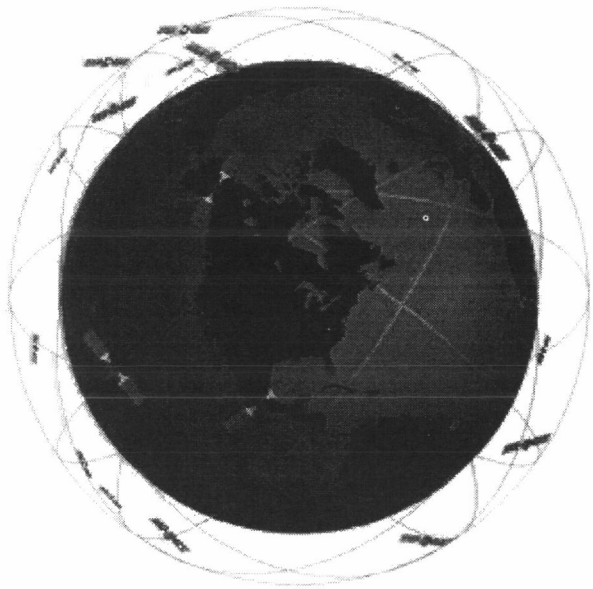


Figure n. 1: Globalstar constellation [6]

per orbital plane (satellite nominal phase angle, 48/8/1 Walker Delta Pattern constellation, shown in figure n. 1). Each orbit has a duration of approximately 114 minutes. The orbital planes are inclined 52 degrees.

Therefore, the satellite orbits are optimized to provide highest link availability (redundant coverage) in the area between 70 deg north latitude and 70 deg south latitude, as shown in the following figure n. 2.

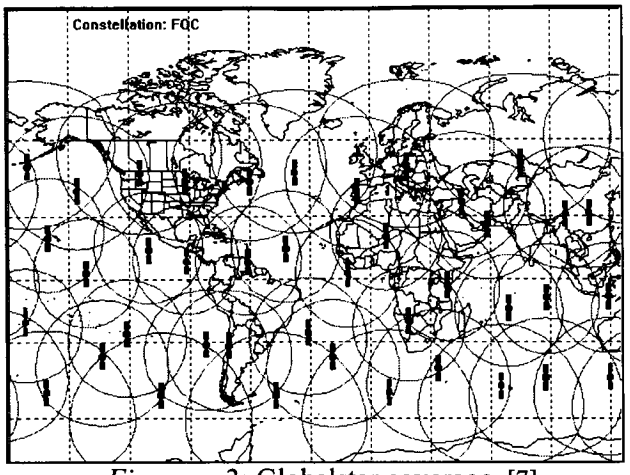


Figure n. 2: Globalstar coverage [7]

Service is feasible in higher latitudes with decreased link availability. User terminals in a particular location on the earth surface are illuminated by a 16 beam satellite antenna (see figure n. 3).

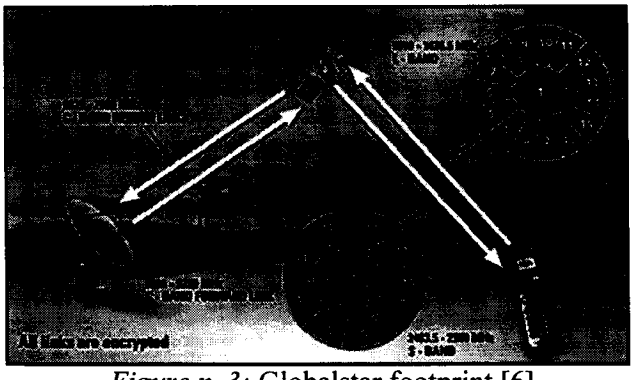


Figure n. 3: Globalstar footprint [6]

The Globalstar system uses the C-Band from the Gateway to satellite and the S-Band from the satellite to the UT (Forward Link). The L-Band is used from the UT to the satellite and the C-Band is used from the satellite to the Gateway (Return Link).

User terminals can be served by a satellite 10 to 15 minutes out of each orbit. Futhermore, subscriber terminals will be able to operate with a single satellite in view, although typically two to four satellites will be overhead (figure n. 2). Therefore, the loss of an individual satellite will usually not result in any gap in global coverage. Each mobile terminal will communicate with as many as three satellites simultaneously, combining the signals received to ensure maximum service quality. As satellites are constantly moving in and out of view, they will be seamlessly added to and removed from the calls in progress, thereby reducing the risk of call interruption. In figure n. 4 the Globalstar diversity concept is shown for a user in Germany connected through three satellites to the Elsacom Italian gateway.

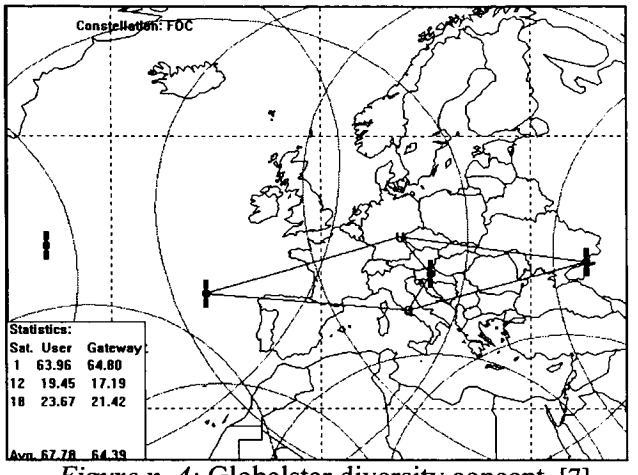


Figure n. 4: Globalstar diversity concept [7]

The Globalstar CDMA Rake Receiver

The Rake receiver (figure n. 5), using its fingers (three in the Globalstar realization), is able to add the different signals from multiple satellites, beams or reflected due to multipath propagation [3].

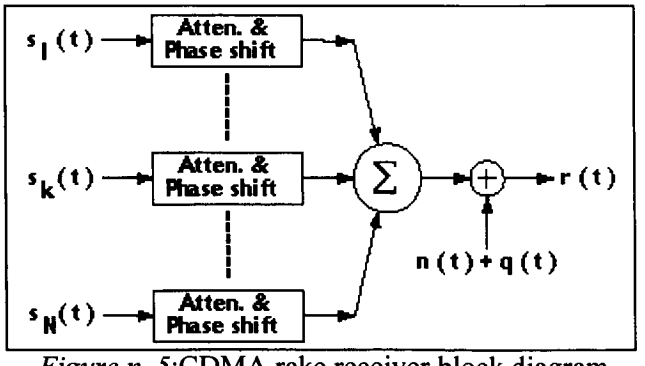


Figure n. 5:CDMA rake receiver block diagram

The multipath components are "resolved" by the CDMA waveform. When their delays are separated by at least the decorrelation time of the spreading, then they can be separated by the despreading correlator in the receiver. They do not interfere because each component correlates at a different delay. When the multipath components are separated by less than the decorrelation time, they cannot be separated in the receiver, and they do interfere with one another, leading to what is sometimes called flat fading. The duration of one spreading chip is 1/1.2288MHz = 814 ns, or at the speed of light, 244 meters [3]. This is the

minimum path difference between two signals that can be resolved.

The Globalstar CDMA Variable Rate Trasmission

The Globastar CDMA channel will support variable data rates (frame by frames, at 2400÷9600 bps, in conjunction with a variable rate vocoder) in order to minimize the user terminal power and battery consumption. It will also allow service, with a slight decrease in voice quality, in marginal areas where the environment conditions do not permit the link to close, even with the User Terminal, the Satellites and the Gateway at their maximum allowed power [4].

GLOBALSTAR SERVICE INDOOR BEHAVIOR

Due to the ITU Power Flux Density limitations, mobile satellite services may suffer from harmful blockage and radio-wave penetration phenomena characteristic of man-made structures. It is expected that MSS will have limited capacity to assure communication inside a building, a train, a ship or a tunnel. Communication is considered feasible only within the line of sight of the satellite, or at the most, in presence of a reflected path with poor attenuation.

Globalstar has limited ability to penetrate buildings. Operation is possible in wood frame buildings or near windows [4]. A link communication has been simulated to underline the behavior in such a condition.

The considered geometry

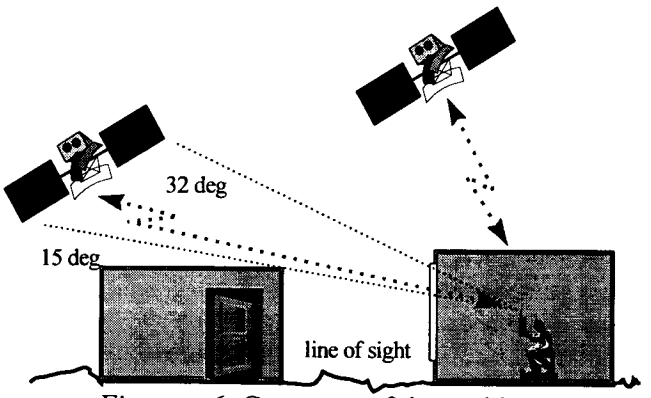


Figure n.6: Geometry of the problem

The geometry shown in fig. 6 has been taken as an example to evaluate Globalstar performances inside a building. The line of sight for a user in the middle of the room is inside 15 and 32 degrees. The structure is concrete; an average attenuation of 20 dB in the Forward direction (Satellite to user, S-Band 2.5 GHz) and 15 dB in the Return direction (User to Satellite, L-Band, 1.6 GHz) can be assumed [5].

The Link Simulation

A link simulation program has been implemented to evaluate the performances of the Globalstar system.

The present software release takes into account all the Globalstar system parameters; however the simple ray tracer implemented doesn't consider the reflected rays. This assumption is quite pessimistic for our purpose. In fact, multipath could be useful for operation when no satellite is in line of sight. The rays that cannot be added because they are delayed less than the decorrelation time are in general several dB attenuated w.r.t. the main signal, and can be managed effectively with the hand-held antenna shaping.

The results of the simulation are shown in figure n. 7 and n. 8, respectively, for the Forward and Return directions.

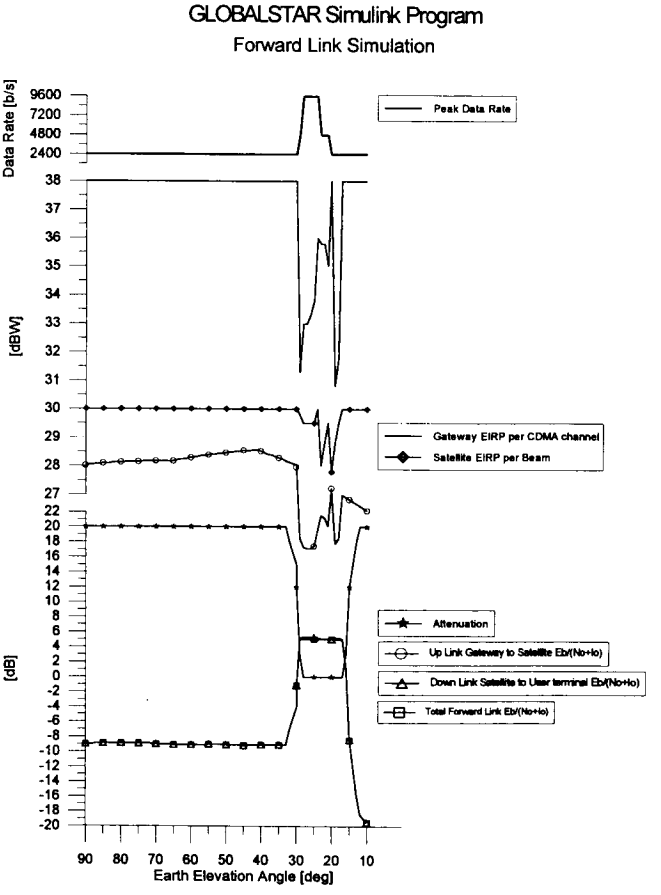


Figure n.7: Forward link simulation

The peak data rate and the $E_b/(N_0+I_0)$ at the user terminal are then computed. The satellite, the gateway and the user terminal EIRPs are also presented. A maximum system capacity (2400 contemporary users per satellite, 1000 contemporary user per gateway) and a quasi uniform

distribution on Earth, (i.e., 150 contemporary user per beam) have been considered.

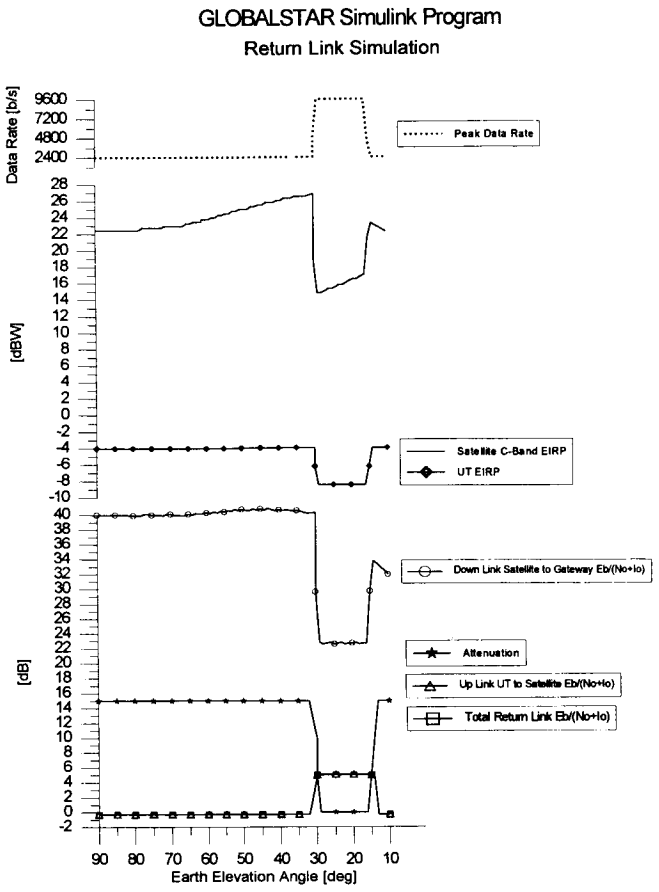


Figure n.8: Return link simulation

Furthermore, a 9600 bps data rate has been assumed; this is the worst condition because it is equivalent to consider a data transmission with an external modem at the maximum data rate. In general, in voice operation the data rate is 4800 bps and only for 10% of the time is it 9600 bps (peak data rate is used at beginning of word or in the emphasized syllables, not in adjacent frames , ensuring high voice quality).

The Globalstar CDMA full performance operations are ensured with a value of

$$E_b/(N_0+I_0)=5 \text{ dB [4]}$$

where,

E_b : Energy per bit
N_0: Thermal Noise Density
I_0: Total Interference Density

This is equivalent to a BER of 10^{-6} and, in conjunction with the retransmission of the corrupted frame, a quasi error free transmission is assured.

With the maximum load capacity, even at the maximum allowed power, the operation inside the room should be feasible only through the window, that is, within the line of sight (without considering the multipath due to reflected rays).

THE LOCAL SIGNAL REPEATER

The Globalstar Local Signal Repeater could be an active or passive device. The passive version has an outdoor antenna with a hemispherical pattern (a more directive antenna can be useful to minimize the performances degradation at high latitudes, where the Globalstar constellation forecasts less satellite overlapping), able to receive at S-Band and transmit at L-Band with satellites up to 10 degrees of view angle, an RF coaxial cable, a delay line and an indoor antenna. Since the polarization is circular, the outdoor antenna can be a simple quadrafilar helix (both for S and L band) with a total height of 18 cm. The overall gain results 0.5 dBi (see table n. 1).

	L-Band	**S-Band**
Antenna height	18 cm	13.6 cm
Antenna Directivity (average)	3.2 dBi	3.2 dBi
Polarization losses	-0.4 dB	-0.4 dB
Mismatch losses	-0.4 dB	-0.4 dB
Ohmic losses	-1.9 dB	-1.9 dB
Antenna Gain (average)	**0.5 dBi**	**0.5 dBi**

Table n.1: Antenna Characteristics

The delay line is needed to allow the Rake receiver the addition of the repeated signal with the eventual direct signal coming from the satellite. The delay value must be at least one chip (814 ns) [3].

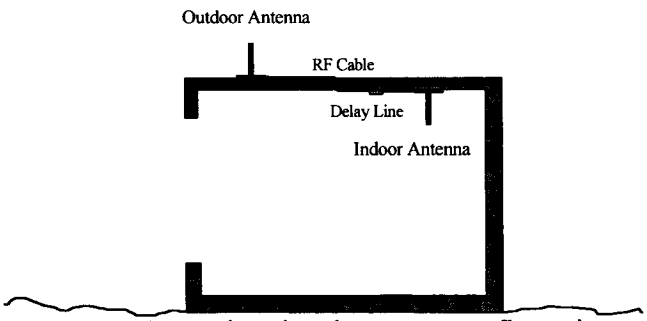


Figure n.9: Passive signal repeater configuration

The table n. 2 shows the budget for the proposed passive local signal repeater. An overall attenuation of 3.1 and 3.7 dB can be respectively considered at L and S-Band.

Other simulations have shown the possibility to split the signal received from the outdoor antenna unit to feed two separate indoor antennas positioned into two different rooms or floors, with an assured $E_b/(N_0+I_0)$ of 5 dB, accepting a decreased peak data rate (2400-4800 Kbps).

	L-Band	S-Band
Outdoor Antenna Gain	0.5 dBi	0.5 dBi
Cable losses	-2.0 dB	-2.5 dB
Delay line losses	-0.6 dB	-0.6 dB
Insertion/connector losses	-1.5 dB	-1.6 dB
Indoor Antenna Gain	0.5 dBi	0.5 dBi
Passive Signal Repeater Att.	**-3.1 dB**	**-3.7 dB**

Table n.2: Passive Signal Repeater Budget

A certain accuracy should be reserved to indoor antenna positioning, in order to minimize the effects of stationary waves that could be present in some close structure with poor absorbance.

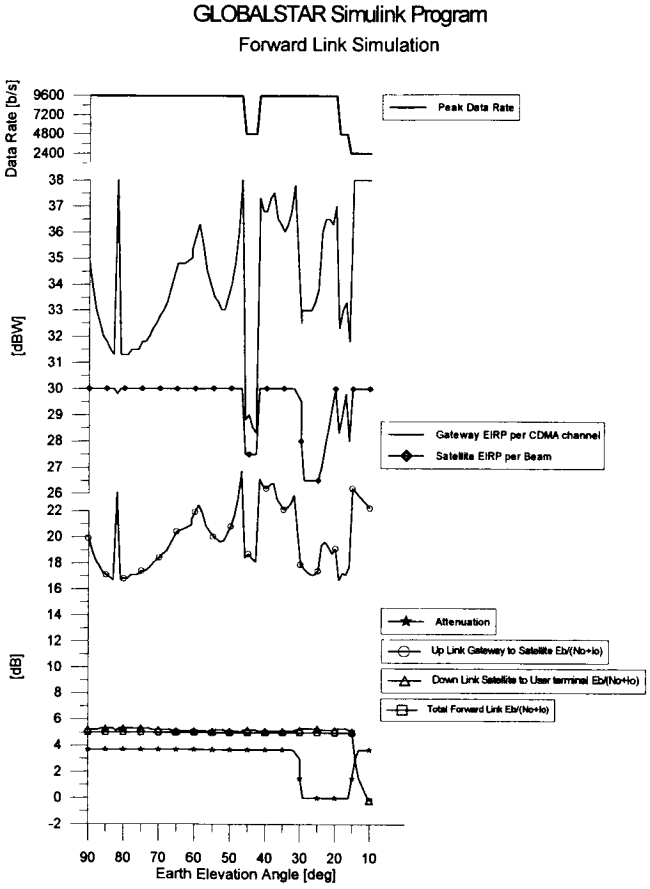


Figure n.10: Simulation with passive signal repeater (Forward Link)

Further quality service improvement can be achieved optimizing the outdoor and the indoor antennas and fine tuning the overall local signal repeater system in order to decrease the overall local signal repeater attenuation.

Finally, it is evident that an active local signal repeater should be able to ensure better quality of service, performing as a full pico-cellular mobile service. Of course, the price to be paid is the slightly increased complexity and cost, and the need of an electrical power supply.

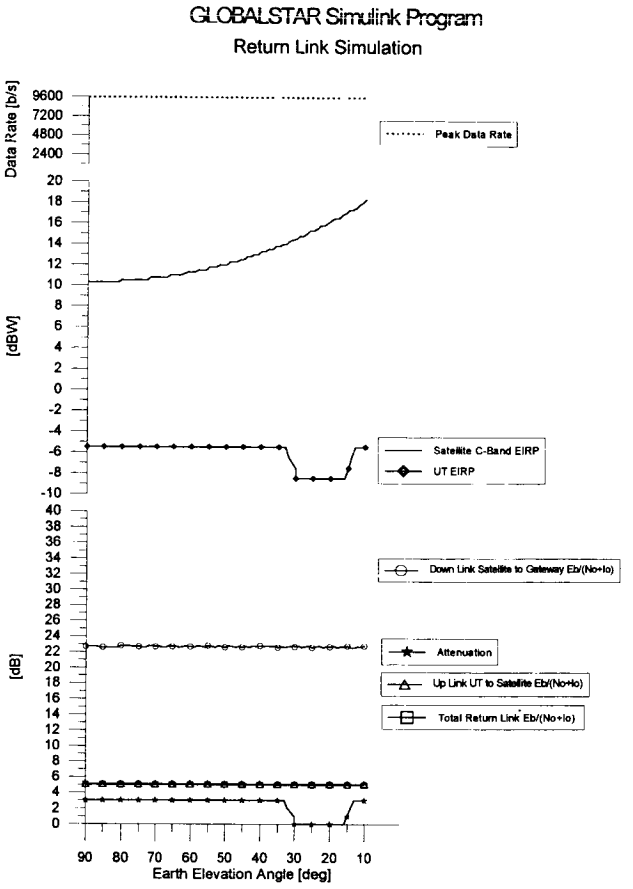


Figure n.11: Simulation with passive signal repeater (Return Link)

SUMMARY

In this paper, we have considered the possibility of employing a local signal repeater in order to avoid the possible harmful blockage and radio-wave penetration phenomena characteristic of man-made structures that can severely limit mobile indoor operation with the satellite system at L and S-Band. In particular, the Globalstar system has been considered and a completely passive local signal repeater has been taken into account. Due to the intrinsic diversity of the Globalstar CDMA system, the feasibility of such a device can result in a simple and low cost design.

The performed simulations have shown that, even with a less-than-optimal local signal repeater device, good performance can still be obtained, particularly inside a simple building structure. In presence of a complex structure with several rooms and floors and/or high RF cable attenuation, an amplifier can be added to implement an active local signal repeater device.

REFERENCES

[1] **William C. Y. Lee**, *Mobile Cellular Telecommunications, Analog and Digital Systems*, 2nd Edition, McGraw-Hill, Inc., New York, 1995, ch 15.3, pp. 503-550.

[2] **A. H. M. Ross and K. S. Gilhousen**, *"CDMA Technology and the IS-95 North American Standard" in The Mobile Communications Handbook*, CRC Press in cooperation with IEEE Press, 1996, pp. 430-448.

[3] **A. H. M. Ross**, *The CDMA Technology Site*, http://www.cdg.org:80/a_ross/index.html

[4] *Description of the Globalstar System*, August 10, 1995, GS-TR-94-0001, Revision C.

[5] *ETSI Technical Report n. 103*, October 1993.

[6] *Official Globalstar internet site*, http:/www.globalstar.com

[7] Simulation performed with the GINA program, Globalstar INteractive Analysis, ver. 5.61, Globalstar LP

Author Index

Author Index