On Wide Area Networking via ATM
over Ka-band SATCOM
with CDMA
On Wide Area Networking via ATM
over Ka-band SATCOM
with CDMA

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.F. Wakker,
in het openbaar te verdedigen ten overstaan van een commissie,
door het College voor Promoties aangewezen,
op dinsdag 1 december 1998 te 10.30 uur

door

John Roland FARSEROTU,
Electrical Engineer,
geboren te Washington D.C., USA.
Dit proefschrift is goedgekeurd door de promotor:
Prof.dr. J.C. Arnbak

Samenstelling promotocommissie

Rector Magnificus, voorzitter
Prof.dr. J.C. Arnbak, Technische Universiteit Delft, promotor
Prof.dr. R. Prasad, Technische Universiteit Delft, toegevoegd promotor
Prof.dr. G. Stette, University of Trondheim, Trondheim, Norway
Prof.dr.ir. G. Broussard, Technische Universiteit Eindhoven
Prof.dr.ir. L.P. Ligthart, Technische Universiteit Delft
Prof.dr.ir. D. van Willigen, Technische Universiteit Delft
Dr. Y. Depeursinge, CSEM, Neuchatel, Switzerland, reservelid

Published and distributed by:

Delft University Press
Mekelweg 4
2628 CD Delft
The Netherlands
Telephone: +31 15 2783254
Fax: +31 15 2781661
E-mail: DUP@DUP.TUDELFT.NL

ISBN 90-407-1785-0/CIP

Copyright © 1998 J.R. Farserotu

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any for or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval systems, without permission from the publisher: Delft University Press.

Printed in The Netherlands
# Table of Contents

## Preface

## Acknowledgments

1. **Introduction**
   
   1.1. *Preview of thesis* ........................................ 1
   1.2. *ATM over SATCOM* ........................................ 4
   1.3. *Third generation mobile communication and UMTS* ....... 13
   1.4. *CDMA for SATCOM and wireless communication* .......... 19
   1.5. *GPS and potential applications* .......................... 25
   1.6. *Commercial CDMA perspectives* ........................... 27

2. **A concept for ATM over SATCOM based on CDMA** ............ 31
   
   2.1. *Introduction* ............................................. 31
   2.2. *The ATM-SATCOM protocol stack* .......................... 33
   2.3. *CDMA for ATM-SATCOM links* ............................. 35
   2.4. *SATCOM system link analysis* ............................. 38
   2.5. *Capacity as a function of the user service distribution* ........................................ 41
   2.6. *CDMA based ATM-SATCOM connection admission control* ........................................ 50
   2.7. *Summary and concluding remarks* ........................ 54

3. **MPSK TCM** .................................................. 57
   
   3.1. *Introduction* ............................................. 57
   3.2. *Modeling of Pragmatic MPSK TCM* .......................... 58
   3.3. *Pragmatic MPSK TCM performance* ........................ 67
   3.4. *Pragmatic MPSK TCM with symbol diversity* ............... 72
   3.5. *Concatenated Reed-Solomon MPSK TCM* ..................... 73
   3.6. *Summary and concluding remarks* ........................ 75
4. ATM-CDMA transceiver

4.1. Introduction ................................................................. 79
4.2. ATM-CDMA transmitter .................................................. 81
4.3. ATM-CDMA receiver ...................................................... 82
4.4. Time and frequency uncertainty over satellite links .............. 86
4.5. The impact on CDMA acquisition and synchronization .......... 96
4.6. ATM-SATCOM delay variance analysis for multimedia services ... 102
4.7. Summary and concluding remarks ..................................... 109

5. Error correction and control for IP/ATM-SATCOM .......... 111

5.1. Introduction .................................................................. 111
5.2. TCP/IP over ATM-SATCOM and the importance of error control ... 112
5.3. Concept for error control tailored to the ATM service class ...... 115
5.4. A CPCS based truncated SRQ protocol for ATM AAL 5 ........... 116
5.5. FEC options for ATM over SATCOM ................................. 121
5.6. Multicast with ARQ diversity for local error recovery ............ 127
5.7. ARQ protocol delay analysis ............................................ 130
5.8. Summary and concluding remarks ...................................... 135

6. SATCOM efficiency ...................................................... 139

6.1 Introduction .................................................................. 139
6.2 Analysis of the results .................................................... 139
6.3 Summary and concluding remarks ...................................... 141

7. Discussion, conclusions and recommendations ................ 145

7.1 Discussion and conclusions ............................................. 145
7.2 Recommendations for future work ................................. 146

Resume and publications
Preface

Information services are becoming more bandwidth intensive and traffic is rapidly increasing on the Internet. Most users of the world wide web (WWW), if not all, have experienced delays due to traffic congestion at one time or another. Long delays may be simply an inconvenience to individual home users, but they may cost time and money to a business. The demand for more bandwidth and more efficient use of existing capacity is steadily growing.

Satellite communication (SATCOM) can provide wide area coverage, which makes it well suited to facilitate the implementation of wide area networks (WANs). SATCOM can be employed to provide connectivity where the terrestrial infrastructure is thin or lacking, as well as, to enhance user mobility over large geographic regions.

ATM is rapidly emerging as the backbone of future information networks. It is designed to support high speed, high bandwidth communication, on-demand. The combination of ATM over SATCOM is a logical extension of technologies for supporting broadband services over WANs. However, questions remain concerning how to effectively combine ATM and SATCOM, such as, ATM-SATCOM integration, resource allocation and management, the ability to achieve and maintain the desired quality of service and the overall efficiency.
Acknowledgements

I wish to gratefully acknowledge my indebtedness to Professor Dr. Ramjee Prasad of the Delft University of Technology in The Netherlands whose support and encouragement were invaluable in the course of the work reported here.

I wish to thank my wife, Veronique Farserotu, for her support, patience and understanding over all of the years that I have studied, my children, Alix and Nicholas, for their understanding and my parents, Michael and Sylvia Farserotu, for instilling in me a desire to learn.
Chapter 1

Introduction

1.1. Preview of thesis

A preview of the thesis is provided in this introductory chapter. Additionally, four key technologies with respect to the concept described in this thesis are reviewed: ATM-SATCOM, UMTS, CDMA and GPS. Key issues were identified and interrelationships were explored.

In Chapter 2, a concept for broadband wide area networking via ATM over Ka-band SATCOM is presented. The concept is designed to combine the soft capacity limit of CDMA [1] with the flexibility of ATM, as well as, the use of bandwidth and power efficient modulation and coding, in order to enhance the overall efficiency of the SATCOM WAN. Broadband bearer services are considered to begin in the range of a hundreds of Kbps to a few Mbps [2 and 3]. The focus is on the ability to support broadcast or multicast transmission of data and video in the range of about 64 Kbps to 8.448 Mbps. However, access rates from 10's of Kbps to 34 Mbps, as consistent with the requirements for services ranging from low rate data and voice, to the broadcast of high quality video signals, are also considered. The concept exploits the multicasting capability of ATM and the wide area coverage of SATCOM to reduce unnecessary transmission of duplicate information and it facilitates multimedia communication to small SATCOM terminals by separating multimedia services into their component parts and transmitting each on a distinct time-shifted PN code.

The concept calls for IP/ATM multicast over Ka-band SATCOM. A high data rate forward link with a low data rate return is envisioned. The goal is to support applications such as broadcast of data and video, fast Internet access and teleconsultation with interactive white board. Operation from mid-size terminals located at a corporate headquarters to small highly transportable terminals at remote facilities is envisioned for broadband communication. Operation to very small mobile terminals, consistent with the Universal Mobile Transmission System (UMTS), is considered for a subset of services.
The wide area coverage of SATCOM is coupled with the broadcast and multicast capabilities of asynchronous transfer mode (ATM) to facilitate the implementation of a WAN with high rate forward link and a low rate return.

The satellite coverage in this concept is intended to support communication anytime and nearly anywhere. Less than global coverage is envisioned. The primary users are expected to be businesses and the coverage would be constrained to the latitudes in which the demand for business communication is greatest. The intent is to keep the constellation to a minimum and the overall system cost down. Although the primary modes of operation are intended to be broadcast and multicast, the use of a multiple beam antenna can facilitate both high rate and low rate operation. A constellation of 5 satellites in geosynchronous orbit is postulated as sufficient to satisfy the coverage requirements and meet a link availability requirement of 99%, with an acceptable margin for rain and other atmospheric effects.

A "bent-pipe" satellite transponder relay is proposed for simplicity of design and ease of interoperability. In the future, the concept may be enhanced through the use of an ATM "switch in the sky". The use of code division multiple access (CDMA) is envisioned to enhance flexibility and facilitate network access and frequency management. The use of separate transmit and receive frequencies is assumed to reduce interference at the receiver.

1.1.1. Focus

The focus of this thesis is on flexible and efficient wide area networking via IP over ATM using Ka-band SATCOM. The emphasis is on support of multimedia communication to a wide range of users. Apart from necessary allowances to exploit the features of ATM, the work is guided by the desire to keep the design as simple, practical and low cost as possible.

1.1.2. Key issues to be addressed

Key issues addressed in this thesis include:
• flexible wireless WAN in support of multimedia services,
• effective integration of ATM, SATCOM and CDMA and
• modulation and coding and efficient use of SATCOM resources.

These are important issues within the wireless communication industry today and are the basis for the concept and analysis that follow.

1.1.1. Statement of thesis
The benefits of ATM over EHF SATCOM via CDMA exceed the sum of the parts.

1.1.2. Organization of thesis
In order to exploit the synergy between emerging technologies and combine them effectively it is necessary to understand the key technologies and their interrelationships. In the remainder of Chapter 1, the potential benefits and issues associated with the operation of ATM over SATCOM, the UMTS, CDMA and GPS are examined. Key issues with respect to the operation of ATM over Extremely High Frequency (EHF) SATCOM are identified and the relationship between the traffic profile and design of an ATM based SATCOM system is discussed.

A concept for broadband networking via IP/ATM over EHF SATCOM based on CDMA is presented in Chapter 2. The concept is designed to combine the flexibility of ATM, with the wide area coverage of SATCOM and the efficiency of CDMA.

Bandwidth and power efficient MPSK TCM is considered in Chapter 3. Aspects of the ATM-CDMA receiver necessary to implement the concept are described in Chapter 4. Error correction and control techniques for IP/ATM-SATCOM links are evaluated in Chapter 5. Efficient use of SATCOM resources is addressed in Chapter 6. Conclusions are summarized in Chapter 7 and recommendations for future work are presented in Chapter 8.

1.1.3. Novel elements
The novel elements the work described herein include the ATM over SATCOM concept, the Pragmatic MPSK TCM analysis and the error control and correction
techniques proposed for ATM over SATCOM. The focus is on CDMA for ATM-SATCOM links and error control and correction for ATM-SATCOM links. Performance issues, limitations and requirements for reliable operation of ATM over SATCOM are examined. Emphasis is placed on efficient use of SATCOM resources. Specifically, the novel aspects of the concept and analysis in this thesis include:

- concept for integration of ATM and CDMA over SATCOM,
- ATM multicasting via CDMA,
- analysis of capacity as a function of user service distribution,
- improved analytical results for "Pragmatic" TCM in AWGN,
- 32-PSK "Pragmatic" TCM in AWGN, Rician and Rayleigh fading,
- analysis of concatenated RS/TCM and symbol repetition,
- concept for error control and correction by ATM service class,
- CPCS based SRQ protocol for ATM AAL 5,
- concept, options and analysis of FEC for ATM over SATCOM,
- multicast with ARQ diversity for local error recovery,
- the concept for code-aided carrier recovery and
- the analysis of ATM-SATCOM efficiency.

The novel elements of this thesis will be described in detail in the chapters that follow. An examination of the key technologies with respect to this work is provided in the remainder of this chapter.

1.2. ATM over SATCOM
Flexible wide area networking and wireless wide area coverage.

1.2.1. The potential benefits of ATM over SATCOM

*International standard* - The ATM protocol is based on international standards and it is designed to support multimedia information services. A combination of features including flexible bandwidth allocation, statistical multiplexing, priority queuing and multicasting make ATM potentially more efficient in terms of bandwidth utilisation than conventional switching technologies, despite the additional overhead of 9.4%
due the 5 byte header per 53 byte ATM cell. Flexible bandwidth allocation is a key reason.

**Flexible bandwidth allocation and bandwidth on demand** - The ability to allocate whatever bandwidth is required, whenever it is required, can outweigh the reduction in peak throughput associated with the ATM header. In essence, the peak throughput is sacrificed in favor of flexibility with the goal of greater overall bandwidth efficiency and the ability to accommodate multimedia services on-demand.

**High speed** - ATM is designed for high speed low delay operation. Commercial-off-the-shelf (COTS) equipment exists that is capable of supporting operation at standard rates including T1 1544 Kbps and E1 2048 Kbps, DS3 45 Mbps, OC3 155 Mbps, well as, higher rates.

**Multimedia support** - The ATM protocol is designed to support multimedia communication. This is accomplished through a combination of high speed, low delay hardware implementation, tailoring of the protocol at the ATM adaptation layer (AAL) to the unique requirements of each service class and the ability to support priority queuing by service class. At present, four AALs are defined, which roughly correspond to the service classes supported by the ATM protocol: AAL 1 for voice, AAL 3/4 for video and AAL 5 for data. AAL 2 is not yet mature. A detailed description of the service classes and the AALs is provided in [4].

**Multicasting** - Multicasting is an important feature of the ATM protocol. The ability to multicast information and avoid unnecessary transmission of duplicate information, such as copies of e-mail messages or video transmissions, may be especially relevant to wide area coverage, bandwidth constrained satellite links.

**Priority queuing** - ATM supports the preferential treatment of information based on service class. For example, within the ATM switch, delay sensitive voice (AAL 1) has priority over data (AAL 5). As a result of priority queuing by service class, ATM possesses has an inherent ability to accommodate an effective increase in the traffic load, such as may occur due to a surge in traffic or, conversely, a reduction in the available capacity due to channel degradation. The ability to accommodate traffic
surge, or conversely tolerate periods of reduced capacity, while maintaining the quality of high priority services, facilitates graceful degradation of services by type and enhances network robustness in terms of delivering priority services.

The ability to realize the potential benefits associated with priority queuing is dependent on the maturity of ATM related hardware and software. Priority queuing by service class must be fully supported by the switch and applications must be written to take advantage of priority queuing by service class. At present, the necessary hardware and software are still being developed in the commercial sector. In the future, when ATM hardware and software are more mature, both accommodation of traffic surge and graceful degradation of service may be of interest from an ATM-SATCOM perspective.

A summary of the key features of ATM and their potential benefits from a SATCOM perspective is provided as Table 1.1. For these and other reasons to be discussed, ATM over SATCOM, where bandwidth is often at a premium, is considered an effective combination.

**Table 1.1: Key features of ATM and potential SATCOM benefits**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Potential benefits</th>
</tr>
</thead>
</table>
| international standard| • non-proprietary  
                        | - seamless integration, interoperability and lower cost            |
| flexible bandwidth    | • bandwidth efficiency                                          |
| allocation            | - only use the required capacity                                  |
| bandwidth on-demand   | • bandwidth efficiency                                          |
|                       | - only use the capacity when required                             |
| high speed            | • low delay and delay variation                                   |
| support for multimedia | • efficiency, convenience and potentially lower cost            |
|                       | - fast hardware implementation                                   |
|                       | - tailoring of the protocol to service classes                   |
|                       | • priority queuing by service class                               |
| Multicasting          | • bandwidth and power efficiency                                 |
|                       | - no unnecessary repetitions                                     |
| priority queuing      | • accommodation of traffic surge                                 |
|                       | • potential for graceful degradation                             |
Despite the numerous potential benefits, ATM is a new technology. Although it is available now, the real benefits are only beginning to emerge and may not be obvious from a conventional SATCOM perspective. ATM can have a major impact on SATCOM access, efficiency, interoperability and cost when the technology is mature and the full range of features are available. An examination of the possibilities associated with ATM over SATCOM is an integral part of this thesis.

1.2.2. Key issues with respect to the operation of ATM over SATCOM

A summary of the key issues with respect to the operation of ATM over SATCOM is provided as Table 1.2.

**ATM protocol maturity and information service maturity** - At present, ATM technology is not fully mature. Future developments are likely to be driven by the proliferation of new and increasingly bandwidth intensive information services and the requirements imposed by these services. The requirements for wideband services and applications are evolving and ATM standards also continue to evolve. Traffic and congestion control remains an area of development. Integration with existing networks is another.

**Table 1.2: Key issues with respect to ATM over SATCOM**

<table>
<thead>
<tr>
<th>Category</th>
<th>Key issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATM protocol</td>
<td>• maturity of the technology and standards</td>
</tr>
<tr>
<td></td>
<td>• maturity of information services and applications</td>
</tr>
<tr>
<td>LAN-WAN integration</td>
<td>• traffic control and congestion</td>
</tr>
<tr>
<td></td>
<td>- QoS negotiation, flow control, priority queuing</td>
</tr>
<tr>
<td></td>
<td>• IP over ATM, UDP over ATM (legacy networks)</td>
</tr>
<tr>
<td></td>
<td>• physical interface and rate conversion (SATCOM)</td>
</tr>
<tr>
<td>ATM over SATCOM</td>
<td>• degradation due to increased delay, delay jitter and error rate</td>
</tr>
<tr>
<td></td>
<td>• mitigation via error control and correction</td>
</tr>
<tr>
<td></td>
<td>• the availability of bandwidth</td>
</tr>
<tr>
<td></td>
<td>• rain margin and recovery from outage</td>
</tr>
<tr>
<td></td>
<td>• SATCOM access technique</td>
</tr>
<tr>
<td></td>
<td>• dynamic resource allocation</td>
</tr>
</tbody>
</table>
The ability to realize the potential benefits of ATM, is dependent on the maturity of the ATM protocol and the availability of information services designed to exploit the full capabilities of ATM protocol, or "native mode" information services. The current trend seems to be away from native mode ATM in the purest sense and towards the conversion of commercial-off-the-shelf (COTS) information services at the users facility and, effectively, native mode equivalent operation. A simple modular interface and protocol converter is anticipated. A key assumption is that either ATM, or a modified version of TCP/IP, which includes ATM-like features for multicasting and priority queuing, will emerge commercially.

LAN-WAN integration and congestion control - The object of ATM traffic and congestion control is to achieve performance objectives in terms of cell loss ratio, cell delay and cell delay variation [4]. A connection will be refused if the appropriate capacity and QoS are not available. This raises the question as to the basis for evaluating whether capacity is available. The ability to determine if capacity is available at the appropriate QoS, over an ATM-SATCOM link, is an important part of ATM-SATCOM LAN-WAN integration. The mechanism is an open issue and seems likely to remain unresolved until ATM-SATCOM LAN-WAN integration is more mature. Of course, satisfaction of the QoS requirements over a SATCOM link is only part of the problem. The broader issue concerns satisfaction of the QoS end-to-end and network wide and may require further developments in terms of network management, as well.

LAN-WAN integration and priority queuing - The problems associated with traffic and congestion control and QoS negotiation are complicated by the need to integrate ATM with legacy networks that do not support priority queuing by service type or QoS based traffic and congestion control.

The lack of information concerning priority and QoS may be especially significant over satellite links due to the combination of generally lower transmission rates and longer delay than over F-O links. Even a simple "all or nothing" algorithm, where the priority is simply high or low, and flow control is little more than turning the transmission on or off, may be useful. Similarly, it is especially important to understand the traffic profile. The solution must be tailored to the information flows,
which may be point-to-point, point-to-multipoint or a combination of both. A solution which works well in one case is not necessarily the best for all. The flexibility of ATM not only facilitates tailoring of the solution to the user requirements, it demands so, if the full measure of benefit is to be obtained.

**Integration of IP/ATM over SATCOM** - In the foreseeable future, ATM is expected to coexist with conventional networks based on IP. Logically, IP over ATM (IP/ATM) is one way to integrate ATM with legacy networks.

The integration of ATM into a WAN via SATCOM provides obvious advantages in terms of the ability to extend communications to geographically distant or remote locations, but it also poses challenges due to the longer delay, delay jitter and potentially higher error rates over the SATCOM link, than over the terrestrial fiber-optic (F-O) links for which the ATM protocol was originally conceived. As will be shown in Chapter 5, this can be especially true for IP over ATM. This raises issues concerning error control and correction. Efficient error control and recovery, tailored to the ATM-SATCOM protocol stack, are required. Additionally, the integration of ATM with SATCOM requires interface and rate conversion, as SATCOM links typically do not support rates as high as those of terrestrial F-O links.

**ATM over SATCOM** - Even if SATCOM links did support data rates commensurate with terrestrial F-O links, broadband networking via SATCOM would place significant demands on already limited spectral resources at 14/12 GHz K_u-band and lower frequencies. Fortunately, more bandwidth is available at extremely high frequency (EHF) in the commercial 30/20 GHz Ka-band. However, losses due to rain and atmospheric attenuation can be considerable at EHF. This raises issues concerning recovery from outage and availability, which may place limitations on the type of services that can be supported.

**Resource allocation** - Flexible resource or bandwidth allocation is one of the key features of ATM, but with flexible resource allocation comes a need for more sophisticated resource allocation techniques that take into consideration user QoS and bandwidth requirements [5]. From an ATM perspective, both static assignment of the required bandwidth and dynamic bandwidth allocation have merit. Issues associated
with static and dynamic resource allocation are described in [5]. The question of resource allocation becomes more complex over satellite links, where access is often limited to a few pre-determined rates.

1.2.3. The traffic profile and it's relationship to ATM-SATCOM links

The performance of the ATM protocol is heavily dependent on the traffic profile. The traffic profile defines the traffic mix and intensity by service type. It also defines the traffic flow, which determines the type of satellite link or links that are required.

Sensitivity to service type and traffic intensity - The ATM cells may be thought of as packets and the performance can be evaluated by applying queuing theory. An M/D/1/N queue [4] provides a simple performance model, where M stands for Poisson arrivals, D for packets of fixed size, 1 for a single server and N for the available buffer or queue length. Delay, delay jitter, error rate and queue length are key factors in determining the cell loss rate. Assuming that an acceptably low error rate can be achieved and maintained, then delay and delay jitter become the primary sources of degradation. In general, the delay increases with traffic intensity, as illustrated by Fig. 1.1 [6] for either an M/M/1/∞ or an M/D/1/∞ queue, with delay normalized to the mean packet size.

SATCOM implications - Satellite links do not ordinarily operate at transmission rates as high as those of terrestrial fiber-optic networks. Although they can facilitate LAN-LAN integration into a WAN, they tend to constrain the end-to-end transmission rate between interconnected LAN, as shown by Fig. 1.2, which illustrates the rate conversion problem between, for example, interconnection of a 155 Mbps OC3 carrier to a 2.048 Mbps E1 SATCOM carrier. Idle cells can be deleted, but user traffic cannot.

Bandwidth constriction may be viewed as analogous to an increase in traffic intensity given a fixed arrival rate. This corresponds to a shift in the operating point on the curves of Fig. 1.1. Delay can rapidly increase. However, a reduction of the transmission rate does not necessarily imply a significant degradation in performance. For example, if the reduction in data rate is accompanied by a proportional reduction in the arrival rate, then the traffic intensity remains unchanged.
Figure 1.1: Traffic intensity and the need for realistic traffic profiles

Figure 1.2: Rate conversion and bandwidth constriction

The performance, as measured in terms of delay, is dependent on $\rho_{\text{SATCOM}}$, which represents the fraction of the total traffic be carried over the satellite link. An illustration is provided as Fig. 1.3, where "native ATM" traffic is combined with traffic from legacy systems for transmission over a satellite link. The term native ATM is used herein to denote ATM cells, which have neither been converted to, nor encapsulated by, any other protocol for transmission over the satellite link.
Traffic of all types, including packet switched $\rho_{ps}$, circuit switched $\rho_{cs}$ and native ATM traffic $\rho_{ATM}$, are combined into $\rho_{SATCOM}$ for transmission over the satellite link. This applies to virtual channel connections (VCC) and groups of VCCs or virtual path connections (VPC). It also applies to static resource allocation of semi-permanent private virtual circuits (PVC), as may be assigned using the virtual path identifier (VPI) field in the ATM cell header, as well as, to dynamic resource allocation of switched virtual circuits (SVC), using the virtual channel identifier (VCI) field of the ATM cell header. An understanding $\rho_{SATCOM}$ is important in the process of identifying the traffic profile.

**Traffic flow** - Another important aspect of the traffic profile is the traffic flow. This is more than merely the connectivity. The traffic flow on some links may be symmetric in nature. Full duplex 64 Kbps voice provides an example. The transmission rate in each direction is 64 Kbps. Logically, a SGT of the same size is expected on either end of a symmetric link. Time division multiplexing (TDM) of frequency division multiple access carriers (FDMA) is commonly employed on such links.
However, a system that is optimized to support primarily symmetric traffic flow is not necessarily the best when it comes to supporting asymmetric traffic flow, or mixed traffic flows. Examples of potentially asymmetric traffic include broadcast or multicast of video or data to one or more groups of users. For the case of asymmetric traffic, a large SGT on one end of the link may communicate with one or more small SGTs on the other end. This scenario is likely be advantageous in a system where large numbers of small, low cost, mobile terminals are envisioned. This is the case of interest in this thesis, where the focus is on high rate wideband communication in the forward direction, with a low rate return. Typically, the maximum transmission rate in the forward direction will be “downlink” limited by receiver noise in the small terminal. As will be shown in Chapter 2, this may be employed to advantage.

A thorough understanding of the user traffic profile is always important if the communication system is to be designed effectively. It is especially important in ATM-SATCOM based systems given the sensitivity to the traffic mix, intensity and flow. Although ATM provides an efficient mechanism to combine voice, data and video traffic, knowledge of the traffic mix and intensity is necessary in order to allocate the available capacity equitably and efficiently and to size the SATCOM system. The user traffic profile is a key factor in determining whether ATM over SATCOM is an appropriate solution. Others include the mix and type of traffic.

1.3. Third generation mobile communications and UMTS
The Universal Mobile Transmission System (UMTS) is the concept proposed for European third generation mobile systems. ISDN compatibility is a requirement for the UMTS, which is intended to support voice, data and video services. Future broadband ISDN networks will be based on ATM, which was designed to support multimedia communication. One way to integrate third generation mobile systems into high speed multimedia ATM networks may be through low rate ATM cell relay. Native mode ATM operation is envisioned from source-to-sink. Seamless integration into broadband WAN is an advantage. Furthermore, native mode ATM operation could help minimize the end-to-end delay if fewer protocol conversions are actually performed in the WAN. More fundamentally, native mode operation is required in order to exploit the full capabilities of the ATM protocol. For example, the ability to support a subset of multimedia component services [4] to less advantaged users with
small terminals, such as UMTS users. The second point is an important element of the concept presented in Chapter 2. An overview of third generation mobile communications, with emphasis on the UMTS, follows.

1.3.1. Overview of UMTS

**FPLMPTS** - The International Consultative Committee on Radio (CCIR) has developed a reference model for third generation mobile telecommunication systems based on the Future Public Land Mobile Telecommunication Systems (FPLMPTS). The FPLMPTS model is based on the Groupe Special Mobile (GSM) European cellular standard, and an illustration is provided as Fig. 1.1.

**Architecture** - Two types of wireless terminals are defined: mobile stations (MS) and portable stations (PS). All terminals communicate with a base station (BS) via an air interface. Each BS is connected to a mobile switching centre (MSC). The MSC provides connectivity to the Public Switched Telecommunications Network (PSTN), or fixed network, as well as, other MSCs and special registers containing data required to support mobility and handover functions, such as, customer location.

![Diagram of FLMPTS model](image)

**Figure 1.1: The FLMPTS model**

- **UE** = user equipment
- **BS** = base station
- **MS** = mobile station
- **MSC** = mobile switching centre
- **HLR** = home location registry
- **VLR** = visitor location registry
- **PSTN** = public switched telecom network
The UMTS concept relies on a combination of picocells, microcells and macrocells to support voice and data services [7]. The architecture is sometimes referred to as a mixed cell concept. The purpose of the mixed cell concept is to improve capacity. This is accomplished by tailoring the cell size to the traffic demand; the greater the expected demand, the smaller the cell size.

Picocells support very localized indoor wireless communications, such as portable terminals of which a cordless telephone may be an example. Picocell size is nominally a few tens of meters. Microcells support outdoor mobile terminals, such as cellular radio. Microcell size is nominally a few tenths of a kilometre (km). Communication with the base station is via the air or radio interface. Macrocells provide wide area coverage and support outdoor mobile users. Macrocell size is nominally a kilometre or more. Both picocells and microcells exist within the larger macrocells, which may extend across operating boundaries and national borders.

Satellite links are envisioned to link remote cells, and provide wide area coverage [7]. The UMTS terminal may be capable of dual SATCOM and terrestrial operation, however, the degree of equipment commonality remains to be determined. The key points with respect to the UMTS in this thesis is that SATCOM is envisioned within the UMTS and the use of SATCOM may be relevant to any of the elements shown in the FLMPTS model of Fig. 1.1. With respect to the concept presented in Chapter 2, the use of SATCOM is considered to be most applicable to the UE and the MSC.

Services - Voice is expected to remain the primary service; however, data and video services are also envisioned [7]. The combination of voice, data and video services, will have a major impact on the requisite QoS. Additionally, the UMTS is intended to function within the context of broadband integrated services digital networks (BISDN) [7]. ISDN functionality implies that the UMTS terminal must be capable of supporting a transmission rate of at least 64 Kbps, and may operate at 144 Kbps (i.e., 2B + D channels), or higher rates. The maximum user transmission rate, or access rate, is an important factor in terms of the number of simultaneous users that can be served by a CDMA system and, consequently, the size of cells in the architecture.
Increased bandwidth efficiency may be particularly important in order to support newer more bandwidth demanding services given that SATCOM capacity is a finite resource; especially, relative to terrestrial F-O links. Multimedia communication with voice, data and video may not be viable without improved bandwidth efficiency, at least where relatively low rate mobile communication systems are concerned. Graphics intensive applications, such as the WWW, are just the beginning.

UMTS terminals are intended to be highly mobile and may be very small, which places significant limitations on user access rates and influences the load $p_{SATCOM}$, as seen by the ATM-SATCOM system and described in section 1.1. Power and bandwidth efficiency are at a premium, especially, if large numbers of users and bandwidth demanding information services are to be supported.

**Signalling and handover** - As consistent with ISDN, it is assumed that the UMTS will use CCITT Common Channel Signaling System No. 7 (C7), or a variant thereof, to support call establishment and control functions such as handover. Presumably, signaling will be accomplished via the 16 Kbps D channel on the basic ISDN 2B + D access. However, this remains to be determined. Regardless of the signaling system, out-of-band signaling and control are envisioned, possibly on a separate channel. In the case of a CDMA system, a separate signaling channel could be provided by a special reserved code sequence, which is analogous to a control channel.

**Control** - The UMTS is expected to be controlled by an Intelligent Network (IN) based [7]. The assumption of IN based control is very significant as it implies a much higher degree of sophistication and complexity than for either first or second generation systems. It also implies increased processing capabilities throughout, in order to gather and process the requisite information, and a highly distributed control of routing and channel access is envisioned [8].

**Delay** - Excessive delay can seriously degrade performance. In addition to the propagation delay there is also processing delay. The processing delay includes delays due to interleaving and queuing. While propagation delay must be tolerated, the processing delay is design dependent. A processing delay budget of not more than 30
milliseconds (ms) has been proposed for UMTS terminals based on the recommended limits proposed by the CCITT for a single hop synchronous orbit satellite link [9]. This value is apparently derived from requirements for ISDN and assumes up to 10 ms of processing delay due to interleaving [9] and, presumably, in the case of an ATM based system, queuing delay.

**Access technique** - Key elements of the UMTS air interface include the channel access technique, modulation and coding and frequency allocation. Code division multiple access (CDMA) is a leading candidate for implementation in third generation mobile systems, including, the UMTS [10]. Many experts consider that spectrum spreading CDMA techniques are more efficient than either Time Division Multiple Access (TDMA), or Frequency Division Multiple Access (FDMA). CDMA offers immunity to some forms of interference, low power density and the potential for simplified frequency planning. Despite the potential advantages, there are issues to be resolved with respect to third generation mobile systems. Power control and synchronization are examples [2], which can affect the integration of SATCOM CDMA based systems with terrestrial networks. The integration of ATM, CDMA and SATCOM is a focus of this thesis.

**Modulation** - Gaussian Minimum Shift Keying (GMSK) is envisioned by various sources as the modulation for the UMTS air interface. GMSK is a variant of MSK, which may be viewed as a form of Offset Quadrature Phase Shift Keying (OQPSK) with sinusoidal pulse shaping [11]. GMSK is MSK with Gaussian pulse shaping. GMSK, like MSK, is a bandwidth efficient constant envelope modulation. Constant envelope modulation offers improved performance in a hard-limited system where significant out-of-band interference may otherwise result [11]. Bandwidth efficiency has obvious advantages in terms of the practical throughput and useable capacity.

**Bandwidth and power efficiency** - Radio frequency (RF) bandwidth is a limited resource. Given the frequency allocations for SATCOM and the increasing demands on available resources, greater bandwidth efficiency is expected to become increasingly important in the future. This is not just due to the growing number of users, but also due to the newer more bandwidth demanding applications, such as graphics intensive communication and video. Although efficiency is always an
important aspect of any communication system, it is considered to have a special significance for future third generation mobile systems, as well as wireless communication, due to RF bandwidth limitations. The use of "Pragmatic" M-ary PSK TCM [12 and 13] is proposed for operation over SATCOM and the performance is examined in Chapter 3. In the future, Turbo Codes may be an alternative. Turbo Codes offer potentially superior performance for high rate transmission, but they are not yet as mature as some forms of MPSK TCM and their effectiveness at low rates may be reduced due to the requirements for interleaving [14].

Regulatory issues - The World Administrative Radio Congress (WARC) allocated 230 MHz of bandwidth in the 1.885 to 2.2 GHz band for the European UMTS in March of 1992. The 230 MHz of bandwidth is not contiguous. This is likely to have an impact on a CDMA system design in the sense that it drives the solution towards narrowband CDMA, vice broadband, unless it can be shown that interference to other bands will be within acceptable limits. The difference between broadband and narrowband CDMA is not clearly delineated, however, a system with a spread bandwidth of 5 MHz or more is generally considered to be broadband, while one with a spread bandwidth of less than 5 MHz becomes increasingly narrowband. In addition to the WARC allocations, the use of Extremely High Frequency (EHF) satellite links has also been proposed [15]. EHF is attractive due to the potential for very small aperture terminals and high bandwidth. However, availability due to rain remains a problem at EHF. EHF broadband CDMA operation is postulated and a link analysis is provided in Chapter 2.

1.3.2. Integration of UMTS into BISDN via ATM
BISDN networks of the future will be implemented using ATM techniques. Although narrowband, the FLMPTS calls for the UMTS to be based on ISDN. Integration of third generation mobile systems into BISDN via ATM requires an understanding of the basic principles of ATM.

ATM is a high speed packet-switching technique. The underlying data transfer rates may be either fixed or variable; however, the ATM carrier operates at a constant rate, much like Ethernet. In a third generation mobile system, a user could, hypothetically, transmit a variable combination of data at 9.6 Kbps or more, voice at 16 Kbps and low
rate video at 64 to 128 Kbps, on a SATCOM carrier operating at N x 64 Kbps.

The essential details of the ATM protocol, with respect to this thesis and the potential for ATM-SATCOM, are described in Chapter 2 and Chapter 5. For additional information, the reader is referred to the ITU-T recommendations below, as well as, the ATM Forum.

- 1.150 B-ISDN ATM Functional Characteristics,
- 1.361 B-ISDN ATM Layer Specification,
- 1.362 B-ISDN ATM Adaptation Layer (AAL) Description,

1.3.3. Summary of third generation mobile systems and the UMTS
Based on the discussion in the preceding paragraphs, key features of third generation mobile systems in general and UMTS in particular, relative to this thesis, include:

- terrestrial and SATCOM links,
- broadband ISDN functionality (ATM based),
- CDMA is a leading candidate and
- supports voice, data, video and possibly multimedia services.

1.4. CDMA for SATCOM and wireless communication systems
Direct sequence (DS) CDMA is a key element of the proposal in Chapter 2 and a leading candidate for third generation mobile systems [1]. Key issues include:

- the DS-CDMA technique (e.g., broadband vs. narrowband),
- UMTS considerations (size, power, frequency allocations, etc...)
- synchronization and rapid acquisition,
- power control,
- frequency allocation and regulatory issues,
- interference reduction or tolerance,
- receiver complexity and
- code generation and reuse.
An understanding of these issues is useful as they influence the proposal described in Chapter 2. The need for ever greater efficiency is a key driver.

**Techniques** - Implementation of CDMA relies on spread spectrum signals with low time correlations. Pseudo random sequences, or codes, are employed to spread the user signals. One or more codes are assigned to each user or group of users [16]. Signals may be spread in frequency or time. Various forms of CDMA exist:

- broadband DS-CDMA,
- narrowband DS-CDMA,
- frequency hopping and
- hybrid DS-CDMA frequency hopped waveforms.

Broadband DS-CDMA has larger processing gain (PG) than narrowband DS-CDMA. Larger PG increases the number of simultaneous users that the system can support and provides improved performance against interference. Another potential advantage of broadband DS-CDMA, relative to narrowband DS-CDMA, is the ability to tolerate frequency selective fading [17]. Note that although fading is less common over SATCOM links than terrestrial, it does occur.

When the extant of the fade in the frequency domain is less than that of the DS-CDMA spread signal, which is more likely to be the case in a broadband DS-CDMA system than in a narrowband system, then only part of the signal power is lost, and the effect of the fade is diluted. By comparison, the fade may completely overlap the DS-CDMA signal in a narrowband system, in which case, the full DS-CDMA signal is subject to the fade. The authors of [17] indicate that a broadband mobile CDMA system will suffer fading losses of approximately 2 dB for a 48 MHz signal and a 15 MHz coherence bandwidth, while fading losses in excess of 10 to 30 dB are expected for a narrowband DS-CDMA system with a 1 MHz spread. The results do not necessarily indicate that broadband CDMA has a clear advantage over narrowband CDMA because the effects of coding and interleaving are not considered and, furthermore, the results are based on greater spread bandwidths than may be practically achievable given the nature of the WARC allocations in Europe.
UMTS considerations - The FPLMTS/UMTS frequency allocations in Europe are not contiguous over the full 230 MHz bandwidth. There are limitations on broadband operation in Europe that remain to be determined. Although DS-CDMA may be employed for third generation mobile systems in Europe, it is not yet clear whether it will be narrowband DS-CDMA or broadband DS-CDMA. Narrowband DS-CDMA is used to describe a DS-CDMA system operating at less than the full available spread bandwidth. Typical spreads might be on the order of 1 MHz to 5 MHz. Given that both broadband and narrowband communication are envisioned, the focus will be on broadband DS-CDMA. However, narrowband DS-CDMA could conceivably co-exist with broadband CDMA.

Hybrid waveforms - A DS-CDMA signal may be further spread, increasing the processing gain and capacity, by frequency Hopping [18]. This creates what is known as a hybrid waveform. Hybrid waveforms may be especially interesting given discontinuities in the available spectrum.

The case of a hybrid waveform is interesting because of the potential to increase the spread bandwidth without increasing the instantaneous bandwidth and to enhance the ability to reuse limited pseudo random codes. The narrowband nature of instantaneous waveform lends itself more readily to the use of non-continuous frequency allocations, as are found in Europe. Frequency-hopped DS-CDMA (FH/DS-CDMA) could also alleviate the requirements for interleaving, or even eliminate the necessity for an interleaver, for example, provided that the hop rate is fast relative to the fade duration. Although the concept proposed in Chapter 2 is focused on broadband DS-CDMA and frequency-hopping is not considered, the concept is compatible with FH/DS-CDMA and an analysis is planned for the future.

Self interference - For both broadband and narrowband CDMA, self interference is a potential problem. In the simplest case, all of the user signals have the same average power and the resultant interference may be viewed as an additional contribution to the noise. In fact, it is most unlikely that all of the signals will be the same average power at the receiver in a multi-user, multi-service, multi-operator, multipath fading environment, where the impact of interference may be compounded by the near-far
effect. The near-far effect occurs when a signal near the receiver overwhelms the desired signal from a transmitter located at a more distant location. It is most relevant to a DS-CDMA system because all of the users share the same frequency at the same time. Interference cancellation may be necessary.

Synchronized CDMA offers a means of reducing self interference between users, however, it may be difficult to implement, at least without benefit of GPS timing. Without interference cancellation or synchronized CDMA, the system must be robust enough to tolerate interference. Even with interference cancellation, the system should be capable of accommodating a certain degree of interference. Interference cancellation is addressed in the literature and will not be considered further. The potential of synchronous CDMA and related issues will be considered in Chapter 2, Chapter 4 and Chapter 5.

**Interference cancellation, reduction and tolerance** - Interference can occur from many sources and can limit or degrade performance. Mutual interference in a direct sequence CDMA system takes the form of noise. In a frequency hopped or hybrid system, it may be viewed as analogous to pulsed partial band noise jamming, especially, in the presence of the near-far effect. In either case, it can reduce capacity in a CDMA system, and it can have a subtle affect on other aspects of the system, such as acquisition and synchronization. Certain modulation and coding techniques are more tolerant of interference. The same is true for acquisition and tracking. For example, non-coherent detection is expected to be more robust than coherent in a high interference environment [16].

There are basically two ways to deal with interference; cancellation, or some combination of reduction and tolerance. Interference cancellation is probably most relevant where the other signal is known, as it may be in the case of mutual interference between users in a CDMA system. Interference cancellation can improve capacity beyond the noise equivalent limitation, because the contribution of interference to the total noise is actually eliminated, even though the noise is present. Thus, interference cancellation can increase capacity and performance.

The ability to tolerate a certain amount of interference is also important. It can
enhance the robustness of the communication system. Furthermore, it may be necessary to tolerate a degree of interference in order to maintain an acceptable minimum quality of service. Although not considered in this thesis, interference tolerance through modulation and coding for CDMA systems will be considered in the future.

**Receiver design and complexity** - There are many possible CDMA receiver implementations. On one end of the spectrum, there is the familiar, relatively "simple" and inexpensive, correlation receiver. On the other end of the spectrum, there is the "optimum" RAKE receiver, as described in [16] and implemented by QUALCOMM [18]. The RAKE receiver employs a tapped delay line to obtain multiple samples of the signal and lends itself to a variety of sophisticated techniques for improving performance. Maximal ratio combining of diversity signals is possible with the RAKE receiver. It may be thought of as a form of soft decision diversity combining.

The diversity chips (i.e., not code chips) can be normalized against channel power variations, such as may be incurred in a high interference or multipath fading channel. Normalization is accomplished by de-weighting contributions to the code decision metric dependent on the availability of side information, such as, power measurements during the period of interest. There is little question that the RAKE receiver can outperform a simple PSK receiver. However, cost and complexity are legitimate issues. The simple PSK receiver remains of practical interest and serves as the initial point of departure.

**Synchronization and rapid acquisition** - The main considerations with respect to synchronization and acquisition of CDMA for third generation mobile systems are network entry (coarse acquisition), maintaining synchronization (fine acquisition and tracking), as well as, special considerations for SATCOM and high speed mobiles and higher level (protocol) synchronization issues.

Network entry is an important factor in any CDMA system. In a commercial mobile communication systems, the network is continuously being reconfigured. New terminals are constantly entering the network while others are leaving. Network entry and re-entry must be fast and reliable.
As will be shown in Chapter 4, analysis suggests that a meaningful reduction of the acquisition time is possible using GPS, as opposed to quartz. Further, the impact of GPS may be more significant when interference and fading are considered, as greater system time accuracy can be traded against power.

**Code generation and reuse** - In a CDMA system, the users are separated in code space, as opposed to frequency (FDMA) or time (TDMA). Although code space is not physically limited in the sense that frequency bands or time slots are, limitations exist due to self-interference, as well as, the availability of codes.

Separate frequency bands for transmit and receive communications can help and are assumed, but code reuse may be an issue where a mobile is subjected to interference by other operators using the same code, or where correlation is higher than desirable, such as due to multipath conditions.

PN code sequences may be generated in a number of ways. One method that may be especially interesting in a high performance system is based on a sort of synchronous sliding window [19]. Each mobile terminal is assigned a unique key, which is analogous to a telephone number. Any mobile wishing to transmit to another tunes to the unique key of that terminal by introducing the appropriate time shift, which is a multiple of the basic chip time, as designated by the access control terminal (e.g., base station). Provided that the receive terminal is aware of the time shifts, it can correlate the delayed versions of the common code sequence, and receive each incoming signal.

Potential benefits of the sliding window concept technique include: simplified receiver design, as only a single receiver is required, the possibility of simultaneous communication with multiple mobile terminals, and the possibility of distributed control using a mesh network. This technique may be especially interesting to link mobile terminals with base stations.

The sliding window technique requires a time accuracy to within a fraction of a chip. For a 50 MHz chip rate, with a 20 nanosecond chip duration, a 2 nanosecond chip accuracy requirement can be postulated based on a tolerance of 1/10 of a unit interval.
This should be achievable using differential GPS based on a position accuracy of less than 1 meter, although there may be other problems to overcome, such as, multipath fading, Doppler shift and satellite motion and delay variation due to processing. Time accuracy and delay variations will be examined in Chapter 4.

**Power control** - Multi-user, multi-rate, multi-operator, mixed cell size operation requires power control for efficient operation of the system. Power control also saves battery life. These are important issues for both terrestrial wireless and SATCOM systems. Multiple users increase the interference, as described earlier in this section. Operation at different rates means that some users may contribute more to the interference. It also means that certain users may be more susceptible to interference. Multi-operator environments imply that there is a greater chance of PN code correlation, which effectively increases the noise and may necessitate interference cancellation.

**Regulatory and frequency management issues** - The WARC frequency allocations affect the available spread bandwidth in a CDMA system. In some cases, narrowband DS-CDMA may be more practical than broadband DS-CDMA, for example, in Europe at UHF frequencies. EHF Ka-band is the option considered.

### 1.5. Potential applications of GPS

The Global Positioning System (GPS) is a source of precise time, frequency and position. GPS may be useful in many ways. Precise time is required for network entry in a CDMA system. As a common precise time source, GPS could facilitate cross-border handover. It may also be possible to use GPS to exploit the code sharing potential of CDMA signals offset by more than one time chip or hop interval (i.e., synchronized CDMA).

**Pre-correction of satellite motion** - The delay over a typical satellite link is generally much longer than over a terrestrial F-O link. The path is typically much longer. Furthermore, satellite motion can contribute to the Doppler shift and the delay variations over the link. As will be shown in Chapter 4, this increases the frequency and time uncertainty and leads to longer acquisition times. GPS can provide information to pre-correct for these effects.
Position and handover - As the mobile terminals move from one cell to another, the network is continuously transformed. The location registers must be updated frequently. GPS position data, transmitted by the mobile terminals to the base station, could provide highly accurate position data to support location dependent functions such as handover. Additionally, a good initial estimate of position is required for rapid network entry in a CDMA system.

Accuracy - It is well known that GPS can provide range to within 100 meters, using the short code. If necessary, accuracy's of less than 1 meter are possible using differential processing techniques. GPS provides time accuracy to within 300 nanoseconds of UTC. Data suggests that performance is actually much better most of the time, perhaps as little as 10 nanoseconds [20], however, there may be occasional problems associated with reception from a single satellite, therefore, use of differential GPS is preferred.

A common frequency reference effectively eliminates frequency drift within the user community, which is important in terms of maintaining timing and synchronization throughout the network. A common frequency reference is especially useful in a CDMA network as a means of facilitating entry and re-entry into the network, also referred to as coarse acquisition and reacquisition.

The GPS satellites have on-board cesium oscillators, as such, measurements indicate that short term frequency variations are less than 1 part in $10^{12}$ given as a root normalized Allan Variance over any 1 second interval [20]. The frequency variation between users deriving timing from GPS is virtually negligible. By comparison, high stability quartz oscillators could accumulate several hundred hertz of frequency offset due to aging per month, however, the short term stability is about the same. GPS provides precise system time to the user community to within 300 nanoseconds (ns) of Universal Time Coordinated (UTC) [20].

As will be shown in Chapter 4, inexpensive quartz oscillators of good quality, with an aging rate of 1 part in $10^6$ per month, could be expected to accumulate a time uncertainty of about 3 ms over a 24 hour period. A mobile terminal crossing between
two base stations timed by separate high stability quartz oscillators could experience much longer delays in terms of initial acquisition and handover than terminals employing GPS. Even with GPS, a time difference of several microseconds could exist between operators, or across national borders, even where the source of timing is referenced to UTC [20] and this may affect DS-CDMA acquisition and reacquisition.

GPS and CDMA - In general, the lower the time and frequency uncertainty are, throughout the DS-CDMA network, the faster acquisition and reacquisition will be. This results in shorter call setup time, improved response time and, potentially, smoother handover. Performance improvements may be particularly significant in high interference and fading environments, where reductions in time and frequency uncertainty may compensate for a lower signal-to-noise (S/N) ratio. Precise time and frequency are also useful for pre-correction of Doppler and delay variations, which may have an impact on PN code acquisition, voice and the performance of higher level protocols. As a minimum, the use of GPS is envisioned at all SATCOM injection sites.

1.6. Commercial perspectives
In order to improve a system, it necessary to understand the strengths and the weaknesses. Having identified the key issues, with respect to third generation mobile systems, with emphasis on CDMA, ATM and SATCOM, the intent is to combine the pieces into a concept that, in some small way, improves the performance. An examination of an existing CDMA system is instructive. Based on research, a brief description of two commercial CDMA systems, and the associated technical issues, are provided.

1.6.1. OmniTRACS
The OmniTRACS system is a mobile satellite communications and vehicle tracking system. In particular, it provides for two-way mobile messaging and is compatible with an external GPS input, although one is not required. Research indicates that the system employs two satellites. A 54 MHz transponder is leased on one satellite for communications. A second satellite is used for ranging. The antenna is shared between communications and ranging.
The forward link (hub-to-mobile) is power limited, and does not use CDMA techniques. The term hub is used interchangeably with base station in this discussion. The return link (mobile-to-hub) is bandwidth limited. CDMA techniques are employed, in fact, the waveform is a hybrid with some form of MFSK on top of a DS-CDMA. The use of FSK is apparently intended to promote fast acquisition. Clearly, faster acquisition is of practical interest, and this may be especially true where coherent modulation is concerned.

Acquisition is obtained in about 20 seconds at a data rate of 9.6 KBPS, based on a search window of about 200 milliseconds and a S/N of around 8 dB. The figure of 200 ms for acquisition also assumes that the system is already up and running. Initial entry takes much longer, perhaps 10 to 20 minutes. The problem is apparently related to antenna pointing.

1.6.2. CDMA cellular

The CDMA cellular system is designed to operate spread spectrum on both the forward and return link. In both cases, the spread is 1.23 MHz, which is considered to be narrowband CDMA. Every cell has a GPS receiver, presumably at the base station. However, the timing and frequency of the individual terminals is expected to be derived from an internal quartz oscillator.

A RAKE receiver is employed, and side information is obtained from a pilot channel for the purposes of power control, and perhaps Doppler correction. Each base station has a fixed number of receivers, which are implemented two per card. The maximum number of receivers is around 64, one for each simultaneous user that can be supported.

Apparently, power control is a technical concern, at least with respect to the speed of reaction. Presumably, the acquisition times are similar to those of the OmniTRACS system. Information in the literature indicates that a pilot signal was transmitted by the base station to ease acquisition by the mobile terminals [21]. The pilot signal is short PN sequence. PN acquisition was accomplished using a technique similar to the one described in the next section, except that coherent chip combining was implemented to improve the performance (i.e., presumably greater S/N).
An acquisition time of roughly 2 to 10 seconds was achieved, at an 8 MHz chip rate. This suggests that there may still be room for improvement. The system appears to be for voice only, and for typical calls of roughly 3 to 4 minutes in duration, a few seconds lost is probably acceptable, as long as the response time is tolerable. However, for bursty data traffic with small message sizes, reduced acquisition time may have a more significant impact on performance.

Use of the pilot signal for the mobile-to-base station was not considered to be feasible. Instead, short preambles were used to acquire frequency, phase, timing and signal levels, presumably for power control. The frequency reference during the test is believed to be high stability quartz given that the emphasis on simplicity, and that modems were only wire-wrap versions.
References

Chapter 2

A concept for ATM over SATCOM based on CDMA

2.1. Introduction

The key advantages of operation at Ka-band (i.e., 30/20 GHz), as opposed to lower frequencies, are the increased availability of bandwidth and the potential for smaller terminals and greater mobility. The major disadvantages are increased rain and atmospheric attenuation and the potential for link outages. The objective is to capitalize on the increased availability of bandwidth, without compromising the requisite QoS requirements of the system. Certain types of SATCOM systems may be more suitable than others for implementation of ATM over SATCOM in general and Ka-band in particular.

Multicast via ATM over SATCOM is a natural combination that matches the wide are coverage of SATCOM with transmission of common information to groups of geographically separate users. An illustration is provided as Fig. 2.1.

Figure 2.1: Overview of broadband WAN via ATM over EHF SATCOM
user groups receive only the information they need
In principle, only one copy of an e-mail message would need be transmitted to a group of users in a multicast via ATM over SATCOM. This eliminates wasteful transmission of duplicate data over the already bandwidth constrained satellite link.

The potential benefits increase with the bandwidth of the service common to the user group and may be very significant when the amount of common user traffic, for example e-mail, is a substantial percentage of the total over the satellite link. Especially significant savings are possible in the case of the broadcast or multicast of broadband video transmissions to a user group, or groups.

Flexibility is a key difference between broadcasting and multicasting. Ideally, multicasting would be accomplished via ATM. IP multicasting over ATM networks may be more practical at this time. Issues associated with IP multicasting over ATM networks are described in [1]. In the future, multicasting may be supported by both IP version 6, as well as, ATM and key aspects of the concept may apply in either case. However, the focus is on ATM in this chapter.

A broadcast transmission is one that is sent to all users. A multicast transmission may be transmitted to all users, or any subset thereof. A broadcast or multicast transmission on the forward link can be coupled with a low rate return link from individual users, which may be employed for error recovery of data transmissions.

A more interesting application of ATM multicast over SATCOM, with a high rate forward link and a low rate return link, might be to support the full broadband multimedia service on the forward link and only a subset of the components of a the broadband multimedia service on the return link. A multimedia service could conceivably be decomposed into voice, data, video and whiteboard. This requires separate Virtual Channel Identifiers (VCI) per component service as described in [2]. Thus, multicasting can provide an extra dimension of flexibility. Potentially, it would be possible to transmit only a subset of multimedia services to different user groups. If so, this would allow privileged users with larger SGTs to receive the “big picture” (i.e., voice, data, video and whiteboard), while disadvantaged users with small SGTs could still participate in a multiparty conference, for example, via whiteboard or voice service.
The idea is to facilitate interactive communication between users of all shapes and sizes. For the sake of discussion, a corporate headquarters with a large terminal could transmit broadband multimedia with high quality video to other users with large terminals and, at the same time, transmit a subset of the component services to users at remote sites with very small or mobile terminals. Similarly, a low rate return could be facilitate remote database access. Thus, the SATCOM link need not be symmetric and, hopefully, the transmission rate over the satellite link can be kept to the minimum necessary to support the essential information transfer. This can help conserve capacity, as well as, facilitate the use of smaller, less expensive SGTs.

Priority queuing is another feature of ATM that makes it a potentially good match with SATCOM. Priority queuing can help accommodate surges in the traffic offered to the satellite link or, conversely, it can facilitate graceful degradation by service of the capacity over the satellite, in the sense that priority users and services are less affected by short term channel disturbances. The last proposal may be especially interesting where EHF Ka-band operation is concerned, given the potential for fades due to rain.

The concept introduced in this section is idealistic. It assumes that ATM technology is mature and that the full measure of capability exists from end-to-end. This is not currently the case. An examination of the ATM-SATCOM protocol stack is useful to for the purposes of refining the concept in the sections that follow, as well as, to help clarify current practical considerations.

2.2. The ATM-SATCOM protocol stack
An illustration of the basic ATM-SATCOM protocol stack is provided as 2.2 [2 and 3]. TCP/IP over AAL 5 is employed as an example case because it is relatively common today. The ATM layer is analogous to the link layer in the Open Systems Interconnect (OSI) model [2]. It is divided into the ATM layer itself and the AAL. The AAL is further subdivided into the convergence sub-layer (CS) and the segmentation and re-assembly (SAR) sub-layer. Four AAL protocols have been defined: AAL 1 for constant bit rate services (e.g., voice), AAL 2 for variable bit rate service, AAL 3/4 for data sensitive to loss, but not to delay and AAL 5 for high speed connection oriented data service [2].
Figure 2.2: Example ATM-SATCOM protocol stack

For satellite links operating over either a standard E1 2.048 Mbps carrier or a T1 1.544 Mbps carrier, G.704 framing applies. For the case of G.704 framing, the ATM cells are mapped into the SATCOM carrier via a physical layer convergence procedure (PLCP). In this case, the ATM cells may be mapped byte-by-byte into the G.704 frames via the G.804 PLCP. G.704 framing, may also apply to higher order carriers, such as, the E2 8.448 Mbps carrier or E3 34 Mbps carrier. Other PLCPs are possible and rates are not limited to T1 or E1 rates; however, standard COTS options are limited at this time.

Functionally, the PLCP is considered to be part of the upper portion of the physical layer [2]. However, additional capabilities for error correction may be incorporated into the PLCP in the future, which push at least part of the functionality into the lower levels of layer 2 of the OSI model. At present, there is little in the way of error correction and control within the ATM layer. Options for error correction and control will be examined in Chapter 5.

Currently, there are very few applications or information services available that operate in native mode ATM. The vast majority of information services that are employed with ATM today actually operate via TCP/IP over ATM, UDP/IP over
ATM, or some combination thereof. This is a practical consideration based on the availability of equipment and is expected to continue well into the future. The result is that AAL 5 has become a de facto standard. Other AALs are not widely available. Until they are, the benefits of ATM over SATCOM may be artificially limited. The discussion of ATM in the remainder of this paper will focus on the use of AAL 5.

2.3. CDMA for ATM-SATCOM links

The basic concept for broadband networking via ATM over EHF SATCOM can be enhanced by combining ATM with CDMA. The reasons include the efficiency of CDMA, the potential to support flexible resource allocation, the potential to support multicasting and the potential to facilitate ATM traffic and congestion control. An illustration of the concept is provided as Fig. 2.3, from which it can be seen that an association is made between ATM connections and CDMA codes.

Ideally, connections to a common user group would be made by service and, to the extent possible, multimedia services would be separated into their component parts [2]. For example, consider multicasting of a multimedia service from one to many users. The multimedia service would be separated into voice, data and video services and a separate virtual channel (VC), or virtual path (VP), could be established over the satellite for each component service.

Figure 2.3: ATM over a CDMA based SATCOM link

One code per ATM VC or service class
A single VC or VP might carry multiple connections in order to accommodate the different multimedia services and their components from multiple users, where each connection corresponds to a separate CDMA code. Some of the connections may be point-to-point, while others are point-to-multipoint, within the same multiparty conference.

Users receiving common information, whether it is a complete multimedia service, or a component of a multimedia service, share one or more common CDMA codes. Note that an association is also possible between beams of a multiple beam antenna (MBA) and VC's or VP's, but this requires a processed satellite.

The concept allows for separate and very flexible handling of voice, data and video traffic over the bandwidth limited satellite link. Communications requirements would be driven more closely by information flow between users and less by the requirements of the most demanding of the services to be supported. To the extent that the information flow is asymmetric, the use of smaller terminals would be possible on one end of the satellite link. Aside from the obvious advantages of smaller terminals, the concept could facilitate interactive multiparty communication between a wider variety of users.

The concept is also bandwidth efficient in that the composite transmission rate over the satellite link is reduced, since only the essential information is transmitted to each user. However, the concept favors "native mode" ATM operation and it may not be possible to fully integrate ATM with CDMA over SATCOM until ATM applications are more mature, or means of conversion at the user premises are more fully developed. A possible interim solution is to assign a CDMA code per service class. This also allows for averaging of the QoS requirements over the satellite, which can enhance the power efficiency.

Additionally, the fact that all users receive the information in a CDMA broadcast or multicast over a satellite, within the coverage of the satellite, could be exploited to facilitate QoS based traffic and congestion control. As long as the system is noise limited by the interference from other CDMA users and the traffic can be identified by service class, then it is possible to evaluate the capacity by service class at any
given time. Specifically, information about the energy to noise density ratio could be employed for the purposes of connection admission control (CAC). It should also be possible to estimate the QoS, in terms of the error rate.

Decisions concerning congestion control may be most relevant from the perspective of the satellite, for example, on a future processed satellite with an on-board ATM switch; however, information can also be obtained for the purposes of congestion control at ground terminals that are similarly noise limited. An assumption in this case is that the noise is user limited not thermal noise limited. This is typical of CDMA systems. By comparison, it is much more complicated to discern similar information about a multi-carrier FDMA based system.

The obvious disadvantage of the concept is that multiple CDMA codes would be required to communicate to a single user or group of users (e.g., one for each service class in a multimedia transmission). However, this is considered to be a digital signal processing trade-off and not a technical limitation. Furthermore, the burden is on the larger terminals, or hubs. The smallest terminals need only support operation on one frequency using one code for a single service.

CDMA comes the closest to matching the flexibility of ATM with regard to using only the capacity that is required and only for as long as it is required. In fact, any data rate can be supported, up to the capacity of the system, at any time, by any user, in support of any service. Where higher rate transmission is required, than can be supported over a single satellite link, multiple carriers may be possible. Only single carrier operation is considered herein.

Additionally, the ability to support voice, data and video on separate virtual circuits or codes facilitates the graceful degradation in the event of an intermittent outage, as may occur due to rain at EHF. In a multimedia teleconference, the white board could remain active albeit at low rates even though video and even voice are unavailable, as long as a marginal capacity remains. It also facilitates averaging of the QoS requirements. Whereas a conventional satellite link supports all services at the same level (e.g., bit error rate and delay), ATM over SATCOM with separate virtual circuits for each service class, each with their own QoS requirements, allows the QoS
to be averaged over all connections and services, according to the traffic mix and intensity. Thus, only the necessary QoS is supported over the satellite.

In summary, combining the multicast capability of ATM and priority queuing by service class with CDMA, allows the voice, data and video components of a multimedia transmission to be transmitted separately over the satellite link. As a result, not all terminals need to be sized to handle multimedia services in order to participate in at least some aspect of the same conference. Without combining ATM multicast with CDMA, multimedia service and component services might have to be carried on separate channels or timeslots. A conventional multiplexing hierarchy based on TDM on FDMA for point-to-point links provides an example.

Point-to-multipoint communication might require that voice, data and video be placed on either separate static carriers or physical links with FDMA, or burst at a fixed rate into timeslots in a TDMA based system. Without the ability to separate information flows by service class and preferably per connection, many of the aforementioned potential benefits of ATM would be lost from a SATCOM perspective.

2.4. SATCOM system link

Capacity is a key measure of system performance. It is dependent on many factors. A link budget is useful to help understand the problem and one is provided as Table 1.3. A pair of steerable multiple beam antennas (MBA) would be employed to provide flexibility and gain: one for reception and the other one for transmission. The receive MBA in this conceptual design is 0.75 m in diameter and has 169 elements (Ne). The field-of-view (FOV) is 16°, which provides near earth coverage (EC).

The transmit MBA is 1.0 m in diameter. Again, the FOV is 16° and the number of elements is 169. Coverage is limited in the far northern and southern latitudes. Additional gain may be possible through beam shaping.

A rain margin of 13.2 dB is considered, based on 99% availability in rain region D2 at Ka-band and an elevation angle of 30° [5 and 6]. The vast majority of users are expected to be within a band from about 55° north to 55° south. Communication
outside this band is possible, but coverage is limited and the satellite antenna gain decreases. However, elevation angles less than 30° generally correspond to regions that are cooler and drier than rain region D2 (e.g., regions B and C), in this concept. As a result, 13.2 dB of rain margin is considered to be more than adequate for the intended purpose.

Table 1.3: 30/20 GHz Ka-band forward link budget from a large SGT to a smaller terminal (4.8 m dish to a 2.4 m dish)

<table>
<thead>
<tr>
<th>Forward link</th>
<th>Value</th>
<th>Units</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Satellite type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Up-link beam</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Down-link beam</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Frequency (up)</td>
<td>30.0</td>
<td>GHz</td>
<td></td>
</tr>
<tr>
<td>5. Frequency (down)</td>
<td>20.0</td>
<td>GHz</td>
<td></td>
</tr>
<tr>
<td>6. Transponder bandwidth</td>
<td>400.0</td>
<td>MHz</td>
<td></td>
</tr>
<tr>
<td><strong>Up-link</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Earth station EIRP (peak)</td>
<td>82.6</td>
<td>dBW</td>
<td></td>
</tr>
<tr>
<td>8. Path loss</td>
<td>-214.2</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>9. Gain of 1 m²</td>
<td>51.0</td>
<td>dB/m²</td>
<td>4π / λ²</td>
</tr>
<tr>
<td>10. Operating flux density</td>
<td>-80.6</td>
<td>dB/W/m²</td>
<td></td>
</tr>
<tr>
<td>11. Saturation flux density (edge)</td>
<td>-76.0</td>
<td>dB/W/m²</td>
<td></td>
</tr>
<tr>
<td>12. Input back-off</td>
<td>-4.6</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>13. Spacecraft G/T (beam edge)</td>
<td>16.1</td>
<td>dB/K</td>
<td></td>
</tr>
<tr>
<td>14. C/T thermal (up)</td>
<td>-115.5</td>
<td>dB/K</td>
<td></td>
</tr>
<tr>
<td>15. Boltzmann’s constant</td>
<td>-228.6</td>
<td>dB/K Hz</td>
<td></td>
</tr>
<tr>
<td>16. C/N₀ (up)</td>
<td>113.1</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td><strong>Down-link</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Saturated EIRP (beam edge)</td>
<td>60.9</td>
<td>dBW</td>
<td></td>
</tr>
<tr>
<td>18. Output back-off</td>
<td>-1</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>19. Path Loss</td>
<td>-210.6</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>20. Earth station G/T</td>
<td>34.4</td>
<td>dB/K</td>
<td></td>
</tr>
<tr>
<td>21. C/T thermal (down)</td>
<td>-116.3</td>
<td>dB/K Hz</td>
<td></td>
</tr>
<tr>
<td>22. Boltzmann’s constant</td>
<td>-228.6</td>
<td>dB/K Hz</td>
<td></td>
</tr>
<tr>
<td>23. C/N₀ (down)</td>
<td>112.3</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td><strong>Margin</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. Modern implementation</td>
<td>1.0</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>25. Gaseous attenuation</td>
<td>2.4</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>26. Rain margin</td>
<td>13.2</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>27. Scintillation</td>
<td>1.0</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>28. Depolarization</td>
<td>0.0</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>29. Sand and dust or clouds</td>
<td>1.0</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>30. System margin</td>
<td>2.0</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. Eₘ/N₀ required</td>
<td>6.0</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>32. Eₘ/N₀</td>
<td>6.0</td>
<td>dB</td>
<td></td>
</tr>
<tr>
<td>33. Channel symbol rate Rₙ</td>
<td>85.7</td>
<td>dB/Hz</td>
<td></td>
</tr>
<tr>
<td>34. Maximum data rate R</td>
<td>85.7</td>
<td>dB/Hz</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- Ne = 169, D = 0.675 m, 30° K
- Ne = 169, D = 1.01 m, 100 W
- P₉ < 10⁻⁶, r = ½, QPSK
- Eₘ = rEₙ₀, r = ½, k = 2
- R = Rₙ / r = 371.5 Mbps (maximum)
A nominal constellation of five satellites in geosynchronous orbit is envisioned. An illustration of the satellite constellation is provided as Fig. 2.4a. The link geometry is shown in Fig. 2.4b. Crosslinks would be required to extend the coverage around the world in the absence of terrestrial interconnections. The design of the crosslinks has not been examined. An analysis is planned; however, crosslinks are not considered to be a key element of the proposed concept.

Figure 2.4 a: Satellite constellation and coverage

Figure 2.4 b: Satellite link geometry
The link budget provides a reference point for the maximum capacity in bits per second (bps) that could be supported between a 4.8 m dish and a 2.4 m dish. The actual size of the ground terminal may be either larger, or smaller, dependent on the user.

A capacity of 371.5 Mbps is estimated for a single point-to-point link assuming the use of QPSK modulation with rate 1/2 forward error correction (FEC) and a nominal operating point of $10^{-4}$, based on TCP/IP over an ATM SATCOM link [7]. The bandwidth needed to support a transmission rate of 371.5 Mbps, with the aforementioned modulation a code combination, is 371.5 MHz, which is well within the coherence bandwidth for Ka-band of 1 GHz, or more [6]. When narrow beam coverage is employed on the satellite uplink (1.1°) and a 1.6° regional spot beam is employed on the satellite downlink, 163.8 Mbps could be supported. Similarly, a 3.5° timezone coverage beam could support 34.4 Mbps and an 8° CONUS coverage beam could support 6.6 Mbps, which is sufficient for broadcast video.

2.5. Capacity as a function of the user service distribution

The number of bits per second is not the only measure of capacity and it is arguably not the best suited to evaluate the performance of an ATM based system, where future users may be allocated any portion of the available capacity, on demand, based on priority. A means of evaluating capacity in terms of users and traffic is needed. Erlangs can provide a measure of how efficiently capacity is used and is proposed here. The maximum number of users and capacity in Erlangs can be evaluated based on the methodology of [8].

Let the maximum number of users in a CDMA system be $M$. Each user is transmitting at a bit rate $R_j$ where $j = (1, 2, ..., M)$. The total signal-to-noise density ratio at any receiver is

$$S_r/N_0 = \sum_{j=1}^{M} S_j/N_0.$$
$S_t/N_0$ is the signal-to-noise density ratio from any given user. All signals are spread over the full spread bandwidth $W$ (broadband DS-CDMA). The total (composite) rate over the satellite link is [9]

$$R_T = \sum_{j=1}^{M} R_j.$$  

From [9], using Shannon’s formula, with all users operating at equal data rates, such that $R_j = R$, the maximum number of users in a power controlled CDMA system may be upper bounded

$$M \leq \frac{W}{R} \log_2 \left(1 + \frac{S_T}{N_0 W}\right).$$

This expression may be converted to a more useful form [9]

$$M \approx \frac{(W/R)F}{E_b/I_0}.$$  

The term $W/R$ is known as the processing gain (PG), $F$ is a reduction factor to account for backoff (e.g., $F = 0.8$), which is already considered in the link budget, and $I_0$ is the total noise from all sources (i.e., thermal noise plus interference from other users)

$$I_0 = N_0 + \sum_{j \neq k} E_{cj}.$$  

Based on the assumption that an $E_b/N_0$ of 6 dB is needed to meet the QoS requirements, which will be examined further in Chapter 3, Chapter 5 and Chapter 6, assuming a spread bandwidth of 204.8 MHz and a data rate of 2.048 Mbps per user, the maximum number of users can be approximated from
\[ M \approx \frac{(204,800,000/2,048,000)}{4} \approx 25. \]

Thus, in this example, about 25 users could simultaneously access at 2.048 Mbps rates. However, this result is heavily dependent on the actual size of the satellite ground terminals. Furthermore, it does not consider the possibility of interference cancellation. Another important point is that an approximation of SATCOM system capacity based on the maximum number of "users" is not necessarily a good estimate of the "traffic" that can be supported over the satellite during a "busy hour", nor does it reflect the nature and mix of the services supported. Conversion to Erlangs or some other measure of the traffic offered to the SATCOM system is necessary. Knowledge of the services and traffic distribution is also required.

The proposed approach is to consider the satellite ground stations as if they were individual "users" and to model the arrival of service requests by the ground stations as a Poisson process, where the service requests are exponentially distributed in duration AND where the capacity requested by the ground stations per service request is either fixed, which is consistent with traditional CDMA, or varies according to a distribution, which is more germane to the combination of ATM and CDMA.

The problem of evaluating the capacity of an ATM based CDMA system in terms of Erlangs is now revisited. Using the Erlang-B formula, the blocking probability of an M/M/S/S queue can be computed [8 and 10],

\[ P_{\text{blocking}} = \frac{(\lambda/\mu)^S / S!}{\sum_{k=0}^{S} (\lambda/\mu)^k / k!} \]

In this expression, \( S \) is the number of servers. Ideally this would correspond to the number of users, but as a minimum, there should be at least 1 server per ATM service class. The quantity \( \lambda/\mu \) represents the traffic in Erlangs, where \( \lambda \) is the information service request arrival rate and \( \mu \) is the mean service duration.
The blocking probability of a CDMA system may be expressed as the probability of exceeding a threshold, which is a function of the number of users, the processing gain and the total equivalent noise density due to AWGN and interference from all other CDMA users.

\[ P_{\text{blocking}} = \Pr[Z > \text{Threshold}] \].

The blocking probability of a CDMA based cellular radio system is given in [8]. The number of potential users is assumed to be large. In the case of satellite communications, the impact of “other cell interference” [8] may be disregarded, assuming transmission and reception in different frequency bands, in which case,

\[ P_{\text{blocking}} = Q \left( \frac{A - E(Z')}{\sqrt{\sigma_Z^2}} \right). \]

\[ E(Z') = (\lambda/\mu) \exp[(\beta \sigma)^2/2] \]

and

\[ \sigma_Z^2 = (\lambda/\mu) \exp[(\beta \sigma)^2]. \]

where \( \beta = (\ln 10)/10 \) and \( \sigma \) is the standard deviation in decibels of the power control loop. For \( \sigma = 0.0 \) dB, power control is ideal. Combining the above equations, the blocking probability may be solved and the capacity in Erlangs may be evaluated from equation (2.1).

\[ \frac{\lambda}{\mu} = \frac{(1 - N_o/I_0)(W/R)\exp[-(\beta \sigma)^2/2]}{(1 + \frac{B/2}{2} \exp[3(\beta \sigma)^2/2] \left[ 1 - \sqrt{1 + 4 \exp[-3(\beta \sigma)^2/2]/B} \right]}) \]  

(2.1)

where,

\[ B = \left[ Q^{-1}(P_{\text{blocking}}) \right] = \left[ \frac{(E_h/I_0)_{\text{median}}}{A} \left[ Q^{-1}(P_{\text{blocking}}) \right]^2 \right] \frac{(W/R)(1 - N_o/I_0)}{(1 - N_o/I_0)} \].
Note that the blocking probability is dependent on the requisite QoS, through the $E_b/I_0$.

In the analysis that follows, the capacity in Erlangs will be a) evaluated for common fixed SATCOM access rates and b) compared to the average capacity in Erlangs given variable SATCOM access rates. The variable SATCOM access rates will be modelled using a geometric distribution, with a mean access rate equal to the selected fixed rate. Data concerning individual service rates in [2] suggests that the geometric distribution is reasonable starting point. The actual composite SATCOM access rate is a function of the mix of information services supported. Operation at fixed pre-determined rates is typical of many, if not most, communication systems. Variable rate access is a better match with the ability of an ATM based system to flexibly allocate the available capacity. It is also a good match with CDMA from the perspective that the processing gain can be allowed to vary while the spread bandwidth remains fixed.

For the purposes of the example that follows, $E_b/I_0$ is assumed to be the same for all users. In practice, it will vary for reasons including the service to be supported and the accuracy of the power control. In terms of power control, note that SATCOM is relatively immune to the near-far problem commonly associated with terrestrial wireless communication, where a near terminal wide angle transmission can readily drown out a neighboring terminal wide angle reception from a third, more distant, terminal. In the case of SATCOM, even when satellite ground terminals are near to each other, a narrow beam is generally focused on the satellite.

The average capacity in Erlangs was calculated using the expression

$$\text{Erlangs} = \sum_{n=0}^{\infty} \text{Erlangs}(n) \cdot \text{distribution}(n).$$

For the case of the geometric distribution, $\text{distribution}(n) = p(1-p)^{n-1}$, where $p$ is the probability of incidence of a given access rate and $n$ specifies the access rate as an integer multiple of 64 Kbps (i.e., $n \times 64$ Kbps).
The results for the analysis of constant vs. flexible rate access are provided as Fig. 2.5, from which it can be seen that the capacity in Erlangs is very sensitive to the access rate. Ideal power control was considered, along with a blocking a bit error rate of $10^{-8}$ and a blocking probability of 1%. These factors will be discussed further, shortly.

At a constant access rate of 512 Kbps, the capacity is 268 Erlangs. By comparison, the "average" capacity when the access rate is geometrically distributed, with a mean of 512 Kbps, is 762 Erlangs. The analysis shows that the average capacity in Erlangs is superior given flexible SATCOM access rates (i.e., flexible bandwidth allocation) than it is assuming constant, or static, SATCOM access rates. This is true even though the average access rate is the same (i.e., 512 Kbps in the above example). There is an improvement, or gain, when flexible access is considered. From Fig. 2.5, it can also be seen that the relative advantage increases with the mean access rate.

Mathematically, the improvement is due to the fact that the relationship between the capacity of the CDMA system in Erlangs and the user access rate is not linear, as can be seen by examining equation (2.1). Intuitively, it is simply wasteful to use more bandwidth than is necessary to support the requested

![Figure 2.5: Capacity in Erlangs constant rate vs. flexible rate SATCOM access](image)
service, or services. If the users do not all require the same SATCOM access rate, then it is better to provide them with exactly what they need; no more, no less. With ATM, any leftover capacity may be allocated to other "SATCOM users", on a priority basis. Conversely, the cost associated with users operating at data rates in excess of the minimum required is greater than merely the overhead associated with the unused portion of the SATCOM carrier.

The above results may vary with the user service distribution, QoS requirements and access rates, but they have general application and they illustrate the point that the soft capacity limit of CDMA [11] and the ability to flexibly allocate capacity via ATM are complimentary. The dependency on QoS will now be examined in terms of $P_b$ and the blocking probability.

In telephony, a busy signal indicates when a call is blocked. The concept of blocking is clear and a blocking probability of 1% to 2% is generally considered acceptable [8]. The concept of blocking is not as clear when we consider data communication. Delay becomes more important. However, blocking can still occur if, for example, service is denied due to lack of available capacity anywhere in the system. Blocking due to a lack of capacity over the satellite link is considered herein.

The data service is assumed to be large file transfer via TCP/IP over SATCOM (e.g., e-mail with large enclosures). For ease of discussion, "blocking" is considered to occur upon refusal of a TCP connection by the "SATCOM" server due to lack of resources on the first try. Blocking can also occur if the end-to-end delay becomes excessive or, in the case of ATM, if a users QoS requirements can not be met.

From Fig. 2.6, it can be seen that the capacity in Erlangs is not very sensitive to the probability of blocking for values of $P_{\text{blocking}}$ greater than about 1%. However, the capacity degrades quickly when a blocking probability of less than 1% is required. The probability of blocking is related to the QoS requirements. The probability of bit error is a common measure of the QoS. The available capacity of the system at any given time will vary dependent on the distribution of access rates, as well as, the QoS that the users require.
Figure 2.6: Sensitivity to the blocking probability

Fig. 2.7 illustrates the sensitivity of the system to QoS in terms of bit error rates in the range of $10^{-3}$ to $10^{-9}$. Over this range, the requisite $E_b/N_0$ needed to close the link may vary several dB dependent on the modulation and coding employed. However, the capacity varies only about 5% for over this range, given a blocking probability of 1%, as can be seen from Fig. 2.7.

Thus, the capacity of the system is not very sensitive to changes of a few dB in terms of $E_b/N_0$. By comparison, the impact of a few dB variation in the mean access rate, or processing gain, has a much more dramatic affect on capacity, as can be seen by comparison with Fig. 2.5, which suggests that the capacity is more dependent on bandwidth than on power.

Given the sensitivity to bandwidth, a comparison is made between QPSK with rate $\frac{1}{2}$ FEC and the more bandwidth efficient 8-PSK TCM. With 8-PSK TCM, an $E_b/N_0$ of about 8.2 dB would be required in order to support operation at a bit error rate of $10^{-8}$, or less. The bandwidth efficiency is 2 information bits/channel symbol. With QPSK and rate $\frac{1}{2}$ FEC, an $E_b/N_0$ of about 6.4 dB is needed to support operation at a bit error rate of $10^{-8}$, but the bandwidth efficiency is only 1 information bit/channel symbol.
Figure 2.7: Sensitivity to the bit error rate requirement

Thus, for the same bandwidth occupancy, the SATCOM access rate can be doubled. The result is a 3 dB improvement in bandwidth at a cost of 1.8 dB in power, for a net 1.2 dB gain. Further, the results indicate that the net improvement may be greater from the perspective of capacity in Erlangs, than it is in terms of bandwidth and power. The potential for increased gain can be seen from Fig. 2.5. For example, at 512 Kbps, the capacity of QPSK with ½ rate FEC is 267 Erlangs, as measured using the curve for constant rate access.

In order to use Fig. 2.5 to estimate the capacity in Erlangs for 8-PSK TCM, the abscissa must be translated into channel symbols per second (Ksps). For QPSK with ½ rate FEC, the data rate equals the channel symbol rate \((rk = 1)\). However, in the case of 8-PSK TCM, the data rate is twice the channel symbol rate \((rk = 2)\), which means that it takes only 256 Ksps to transmit 512 Kbps. Thus, in terms of processing gain, transmitting 512 Kbps using 8-PSK TCM is equivalent to transmitting 256 Kbps using QPSK with ½ rate FEC. As such, the capacity given 8-PSK TCM is estimated to be 722 Erlangs, which is 4.3 dB better than for QPSK, at a cost of about 1.8 dB in power. The net difference is 2.5 dB (4.3 dB - 1.8 dB), which represents a 1.3 dB
improvement. The improvement applies primarily at higher average access rates and only if Erlangs of capacity can be effectively allocated from the "pool" of resources.

The sensitivity of ATM-SATCOM links to the access rate and distribution was shown in this section. The sensitivity to key measures of QoS was also illustrated. Error correction and control can help achieve and maintain the requisite QoS and improve end-to-end efficiency. Techniques for error correction and control over ATM-SATCOM links will be considered in Chapter 5. In the remainder of this chapter, issues concerning the allocation of available capacity in an ATM-SATCOM employing CDMA will be examined.

2.6. CDMA based ATM-SATCOM connection admission control

The analysis of capacity in Erlangs described in the previous section suggests a possible method of implementing virtual circuit connection admission control for CDMA based ATM-SATCOM systems. This method is also based on Erlangs.

Connection admission control (CAC) - CAC is the set of actions taken at call set-up to determine whether to accept or reject an ATM connection [2]. In the case of an ATM-SATCOM system, an analogy can be made between connection admission control to the ATM network and connection access control to the SATCOM system. In either case, a connection will not be allowed if the resources are not sufficient to support the required QoS.

The QoS is dependent on a combination of factors including the error rate, delay, delay jitter and the acceptable chance that a service request will be refused, or blocking probability. Only QoS in terms of the error rate over the SATCOM channel and the blocking probability are examined in this section. This is considered appropriate given that information services suitable for transmission over a satellite link must be relatively tolerant of the magnitude of the delay, within reasonable constraints, and that the focus of the proposed ATM-SATCOM concept is on limited networks based on SATCOM broadcast or multicast transmissions, such that the primary source of delay is transmission over the satellite link, which is the same for all services. The essential assumption is that given transmission over a satellite link, it makes no sense to refuse service based on the expected delay, as long as the expected
delay is the same for all services to be supported.

Delay jitter remains an important QoS parameter, at least for voice and multimedia communications and special aspects will be addressed in Chapter 4 and Chapter 5. The focus in this section concerns the ability of the ATM-SATCOM system to support a service dependent access request, where the access request is a function not only of SATCOM bandwidth and power, but also of the QoS requirements, the traffic load and the access technique. A general comparison of SATCOM access techniques is useful to the discussion that follows and one is provided as Table 2.1. The key point is that the flexibility of CDMA, the fact there is no hard limit on the capacity [11], makes it a good match with the ability of ATM to support scaleable, flexible resource allocation. The greater the flexibility, the more readily capacity can be allocated in any desired amount.

This is a basic assumption in the analysis of optimal capacity assignment [12] and, therefore, the efficient use of the ATM-SATCOM system. Given the dependency of ATM on traffic load, it is only natural to consider traffic intensity and capacity as the basis for resource allocation. A discussion of how this might be accomplished in an ATM-SATCOM-CDMA system follows.

Table 2.1: A comparison of SATCOM access techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Resource allocation</th>
<th>General comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDMA</td>
<td>frequency</td>
<td>• block of bandwidth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• full time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• adjacent channel interference limited</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• same performance for all users</td>
</tr>
<tr>
<td>TDMA</td>
<td>time</td>
<td>• full bandwidth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• part time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• full channel performance for all users</td>
</tr>
<tr>
<td>CDMA</td>
<td>code</td>
<td>• full bandwidth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• full time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• self interference limited</td>
</tr>
<tr>
<td>ATM-CDMA</td>
<td>available capacity</td>
<td>• uses Erlangs or other measure of capacity</td>
</tr>
<tr>
<td>proposal</td>
<td>capacity</td>
<td>• function of traffic, power, bandwidth and QoS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• only use what is required, when it's required</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• consistent with optimal capacity assignment</td>
</tr>
</tbody>
</table>
Capacity based ATM-SATCOM CAC for the "switch-in-the-sky" - All connections are established by the ATM switch on-board the satellite. This means that the satellite is "all knowing" with regards to the available capacity, at all four service classes. Thus, the satellite is free to flexibly allocate any portion of the capacity, or all of it, upon user request.

Capacity based ATM-SATCOM CAC at the SGT - The key assumption here is that access has either been pre-coordinated with the ATM "switch-in-the-sky", or it has simply been pre-scheduled over an ordinary satellite. In either case, a block of capacity has been granted to the ground terminal. The ground terminal is "all knowing" with respect to it's available capacity and may allocate any portion, upon request.

The proposed ATM-SATCOM CAC algorithm - Requests for service are considered to arrive within a period of time (T) in seconds. Although requests are assumed to be handled on a first come first serve basis, they could be accommodated preferentially. The capacity would be allocated by service class and user. An illustration is provided as Fig. 2.8.

![Diagram of CAC algorithm](image)

**Figure 2.8: A CAC proposal for ATM-SATCOM**
The available capacity in Erlangs (Available_Capacity) could be estimated at time of service request by subtracting the capacity in use from the total capacity (Total_Capacity) in Erlangs, where the known or assigned capacity is the Total_Capacity.

\[ \text{Available}_\text{Capacity} = \text{Total}_\text{Capacity} - \text{Capacity}_\text{in-use} \] .

The capacity requested during time period T (Capacity-requested) could be computed by summing over all service classes and users

\[ \text{Capacity-requested} = \sum_{\text{class}=1}^{d} \sum_{\text{user}=1}^{\infty} \text{Capacity}_{\text{class},\text{user}} \] .

If the Available_capacity is greater than the Capacity-requested, then all of the new requests are satisfied. Otherwise, requests are granted on a first come first serve basis.

With ATM, there are four service classes. Since there is no hard limit on the capacity in a CDMA system, the number of users is limited by the Capacity required per user (Capacity_{class, user}). If each user requires an infinitesimally small capacity, then the number of users that can be supported approaches infinity.

From equation (2.1), it can be seen that the Capacity_{class, user} is a function of power (E_b/I_0), bandwidth (R and W) and QoS (E_b/I_0 and P_{blocking}):

\[ \text{Capacity}_{\text{class, user}} = \frac{\lambda}{\mu_{\text{class, user}}} = f\left(\frac{E_b}{I_0}, W / R, P_{\text{blocking}}, \sigma\right) . \]

CAC would be dependent on the above function. The parameter (\sigma) represents the standard deviation of the accuracy of the power control loop. The accuracy of the power control loop is a limiting factor on the performance of the CAC proposal, but it is not expected to be pose any more or less of a limitation on the proposed ATM-SATCOM-CDMA system than the transponder back-off required in a much less flexible conventional SATCOM system employing multiple FDMA carriers. An allowance for power loop accuracy represents a back-off in capacity from the
maximum available. It may be viewed as an overhead, instead of a back-off in power.

It may be possible to use side information obtained on the control channel to obtain an estimate of the accuracy of the power control loop over the time period of interest. This may be useful or necessary; however, the requirements for power control should be taken in the context of ATM, where operation at average traffic loads greater than about 75% can quickly lead a significant increase in the average delay, which is not desirable and tends to impose about a 1 dB de facto back-off from the maximum capacity. Thus, the power control need not be ideal.

2.7. Summary and concluding remarks
In this chapter, a concept for ATM over SATCOM via CDMA was presented. The concept is focused on broadcast and multicast from a large terminal to smaller terminals. The unique means of integrating ATM and CDMA with SATCOM is a key and novel element of the proposal. The flexibility of ATM is combined with that of CDMA through the use of a separate PN code for ideally each ATM connection and, as a minimum, per ATM service class. The decomposition of multimedia communications into their component services for broadcast or multicast via a family of CDMA codes, generated by from a parent on a time delayed basis, is a second novel element of the proposed concept, which will be examined further in Chapter 4. Decomposition of multimedia services is considered to help small disadvantaged users participate in one or more aspects of a multimedia conference, where they can not, and may not need to, support the full bandwidth intensive service requirements.

The topic of CAC was examined and a technique proposed based on capacity allocation and traffic load, as opposed to merely power and bandwidth. The special relationship between ATM and traffic load was examined and it was shown that the benefit of bandwidth efficiency is dependent on the user access distribution.

In the next chapter, bandwidth and power efficient modulation is examined. ATM specific error control and correction techniques will be examined in Chapter 5.
References

7. J. Farserotu and A. Tu, "TCP/IP over Low Rate ATM-SATCOM Links", IEEE MILCOM '96, McLean Va., USA.
Chapter 3

MPSK TCM

3.1. Introduction

The need for bandwidth and power efficiency - As greater demands are placed on available resources, the need for bandwidth and power efficient modulation becomes increasingly important. Trellis Coded Modulation (TCM) offers bandwidth and power advantages over other modulations. TCM is of particular interest to users of small aperture terminals, where bandwidth and power efficiency are at a premium. In addition to increasing the number of users that can be supported, the bandwidth efficiency of TCM may facilitate the use of newer more bandwidth intensive multimedia applications, over mobile satellite and wireless radio links, where the available bandwidth and power are limited.

Under ideal conditions, the advantage of higher-order MPSK, in terms of bandwidth efficiency relative to BPSK or QPSK modulation, is clear. However, mobile communication channels are not ideal, fading is common and delay and delay variation may also be significant. The overall performance benefit is a function of the improved bandwidth efficiency and power, as well as, the channel-error rate and delay and their influence on higher-level error-recovery protocols, such as TCP/IP.

Outline of the chapter - In this chapter, the performance of 8-PSK, 16-PSK and 32-PSK TCM are examined in Additive White Gaussian Noise (AWGN), Rician fading and Rayleigh fading channels. The focus is on the use of TCM for mobile SATCOM, however, the results can also be applied to other forms of mobile wireless communication. The "Pragmatic Approach" to TCM is assumed, as described in [1] and [2], where the code is constructed using a common rate $\frac{1}{2}$, constraint length 7, 64-state convolutional code.
The bit error rate (BER) is evaluated analytically, based on the approach of [2], modified as described in section 2. The analytic performance model has been enhanced to account for the unencoded bit(s) resulting from implementation of TCM using the “Pragmatic Approach”. Averaging of the Euclidean error weight profiles is also employed. The results provide a tight upper bound on published data from simulations and measurements.

The results for 8-PSK TCM and 16-PSK TCM are a match with the published measurements and simulated results of [1] and [4] for TCM in AWGN. The results for 32-PSK TCM are novel. The performance of 8-PSK, 16-PSK and 32-PSK TCM are compared with those of uncoded QPSK, 8-PSK and 16-PSK, respectively. The results are then employed to facilitate the evaluation of a concatenated Reed-Solomon (RS) MPSK TCM code, as well as, TCM with diversity in the form of symbol repetition. Symbol diversity is considered for its potential to improve performance in fading. The concatenated Reed-Solomon code is considered for its potential to improve performance under normal operating conditions, regardless of the channel, and to maintain performance in the presence of burst errors due to interference from various sources.

**Why Pragmatic MPSK TCM?** - The use of “Pragmatic” MPSK TCM has been proposed for operation over SATCOM [5]. TCM is a known codulation technique, which can be constructed using and existing commercial-off-the-shelf codec [1]. In the future, Turbo Codes may be an alternative to TCM. Turbo Codes offer potentially superior performance, but the technology is not yet as mature as that of Pragmatic MPSK TCM, Turbo codes may require a substantial interleaver, they are geared towards high rate transmission and may not be as effective at lower data rates [6], such as may be encountered over a SATCOM. For these reasons and given the emphasis herein on the use of existing technologies, Pragmatic MPSK TCM is proposed and its performance is analyzed.

### 3.2. Modeling of Pragmatic TCM

A block diagram of an MPSK TCM transmitter is provided as Fig. 3.1. Unencoded bits are input to the TCM encoder. The TCM code is implemented using the “Pragmatic Approach” to TCM described in [1] and [2].
Figure 3.1: Block diagram of an MPSK TCM transmitter

In this case, the TCM is constructed from the common rate ½, constraint length 7, 64-state convolutional code. Only one of the $k$ information bits is actually input into the rate ½ encoder, the remaining $k - 1$ bit(s) are unencoded.

Interleaving is necessary to break up channel memory and maintain the effectiveness of the code, in the presence of fading, or interference. Symbol interleaving is assumed, as consistent with higher order M-ary modulation and symbol coding. Bit interleaving is a possible alternative.

3.1.1. General MPSK performance model

Define a coded symbol sequence

$$X = (x_1, x_2, \ldots, x_N),$$

where $x_n$ represents the $n^{th}$ transmitted PSK symbol. The MPSK modulator produces one of the symbols

$$x_n = \sqrt{E_c} \exp(j \theta_n),$$

$$\theta_n = \frac{2\pi n}{M}, \text{where } n = 0, 1, 2, \ldots (M - 1).$$

$E_c$ is the energy per channel symbol, which equals $r k E_b$, where $r$ is the code rate, $k$ is the number of bits per M-PSK symbol and $E_b$ is the energy per information bit. The channel outputs are
\[ Y = (y_1, y_2, \ldots, y_N), \]

where the \( n \)th received PSK symbol is

\[ y_n = \rho_n x_n + \eta_n. \]

The parameter \( \rho \) is a measure of the in-phase and quadrature components of the signal. Assuming adequate interleaving and deinterleaving, so that each channel use is independent, then the channel transition probability may be expressed as a product [5]

\[ P_N(y / x) = \prod_{n=1}^{N} p(y_n / x_n), \]

where \( N \) is the number of code vectors along the path in question. The commonly used metric for coded systems is employed [7]

\[ m(y, x) = \ln(P_n(y/x)) = -\sum_{p=1}^{N} \| y_p - x_p \|^2, \]

and the general form of the Chernoff bound on the pairwise error probability between codewords can be expressed as

\[ P(X \to \hat{x}) \leq \prod_{n=1}^{N} E\left[ \exp\left( \lambda [m(y_n, \hat{x}_n) - m(y_n, x_n)] \right) \right], \]

where \( \lambda \) is the parameter over which the pairwise error probability is minimized. From [5 and 8], the general form of the pairwise bound above can be refined to the more useful form

\[ P(X \to \hat{x}) \leq \min_{\lambda} \prod \exp\left( 2\lambda^2 \rho^2 \sigma^2 \| x_n - \hat{x}_n \|^2 \right) \cdot E\left[ \exp\left( 2\lambda \rho \Re(x_n (\hat{x}_n - x_n)) \right) \right], \quad (3.1) \]

where \( \rho^2 = 1 \) and \( \sigma^2 = N_0/2 \). This form of the bound applies even if the channel symbols are not independent.
If the channel symbols are independent, the pairwise error bound between the correct symbol sequence and the received symbol sequence $P(X \rightarrow \hat{X})$ can be expressed as a product of the bounds on the probability of transition between individual symbols in the sequence. The pairwise error bound on transitions between individual symbols is a function of the Euclidean distance between the symbol pairs,

$$P(X \rightarrow \hat{X}) = \prod_{n=1}^{N} P(x_n \rightarrow \hat{x}_n) \leq \prod_{n=1}^{N} D(\alpha_n),$$

where $D(\alpha_n)$ is the Chernoff bound on the pairwise symbol error probability and $\alpha_n$ is the Euclidean distances between the symbols. From Ref. [5 and 8], the Euclidean distance between MPSK symbol pairs may be expressed as a function of $\varepsilon_n$, which is the phase difference between the transmitted symbol $x_n$ and the received symbol $\hat{x}_n$,

$$\alpha_n = \|x_n - \hat{x}_n\|^2 = 2E_c(1 - \cos(\varepsilon_n)).$$

A bound on the pairwise error probability per channel use in AWGN, Rician or Rayleigh fading channels can be obtained by averaging (3.2) over the applicable distribution. A summary of the bounds, on the pairwise error probability per channel use, is for each of the channels considered in this chapter, as Table 3.1 [5 and 8]. Slow flat fading is assumed. In the event of frequency selective fading, equation (3.1) would still apply; however, simplification to equation (3.2) would not be possible.

The expressions $D(\alpha_n)$ and $P(x_n \rightarrow \hat{x}_n)$ are used interchangeably in this chapter. They may differ under certain conditions, for example, in the presence of phase noise, which has not been considered. The accuracy of the bound is dependent on the modulation, code and channel; however, it is typically within a few tenths of a dB in AWGN and within about 1 dB in Rayleigh fading.

From (3.2) and Table 3.1, the probability of bit error ($P_b$) can be calculated using series expansions for the various codes, as shown for the case of BPSK in [9], and it
can be shown that QPSK is essentially equivalent to the case BPSK \((M = 2)\) \([10]\), since the contribution of the double error term in QPSK \((M = 4)\) is minimal, relative to the contribution to \(P_e\) from transitions between nearest neighbors in signal space (i.e., given ideal phase and timing).

Calculation of \(P_e\) is more complicated in the case of higher order PSK; especially, where TCM is concerned. However, \(P_e\) can be evaluated using a modified generating function, given knowledge of the code, as shown in \([2, 3 \text{ and } 7]\) and this is the approach taken.

Starting with the generating function, the probability of error caused by any one of \(a(d)\) incorrect paths at Euclidean distance \(\alpha\) from the correct or all 0's path through the trellis is \([9]\)

\[
P(e) \leq \sum_{d=0}^{\infty} a(d) P(X \rightarrow \hat{X}), \text{ where } P(X \rightarrow \hat{X}) \text{ is given by equation (3.2).}
\]

A union bound on the bit error rate can be obtained by weighting each term in (3.1) by the number of bit errors caused by an incorrect path diverging at node \(j\) in the trellis. This is equivalent to taking the partial derivative of the generating function \(T(P(X \rightarrow \hat{X}), I)\) with respect to \(I\)

Table 3.1: The pairwise error probability per channel use

<table>
<thead>
<tr>
<th>Channel</th>
<th>(D(\alpha_n))</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWGN</td>
<td>(\exp(-\alpha_n/4N_0))</td>
</tr>
<tr>
<td>Rician</td>
<td>(\frac{1 + K}{1 + K + \alpha_n/4N_0} \exp\left(-\frac{K\alpha_n/N_0}{1 + K + \alpha_n/4N_0}\right))</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>(\frac{1}{1 + \alpha_n/4N_0})</td>
</tr>
</tbody>
</table>
\[ P_b \leq \frac{1}{b} \left. \frac{\partial}{\partial a} \right|_{a=1} \mathcal{T}(P(X \rightarrow \hat{X})) = \frac{1}{b} \sum_{i=1}^{a} \sum_{d=d_f} \lambda(d,i) P(X \rightarrow \hat{X}), \]  

(3.3)

where \( i \) denotes the number of bit errors along the path and \( b \) is the number of information bits into the encoder (i.e., \( b = k - 1 \)). For the case of the common rate \( \frac{1}{2} \) constraint length 7 convolutional code, with generating polynomials \((133,171)\), the sequence of symbols along paths with the same Hamming distance from the all 0s path can be found [5 and 10]. From [3 and 7], an expression for the performance of TCM may be obtained by modifying (3.3). The discussion that follows describes how this is done and highlights the differences between the model employed herein and that of [3 and 7]. The case of 8-PSK TCM is taken as an example for ease of discussion.

3.1.2 Pragmatic MPSK TCM performance model and modifications

For Pragmatic MPSK TCM implemented using the standard rate \( \frac{1}{2} \) 64-state convolutional code, only one information bit is actually input into the rate \( \frac{1}{2} \) encoder, all other bits are unencoded [1]. In the case of 8-PSK TCM, there are two channel bits out of the encoder, plus one unencoded bit for a total of three channel bits. The resultant code rate is effectively \( \frac{3}{2} \), where the three output channel bits from the Pragmatic TCM encoder determine the MPSK channel symbol to be transmitted.

Continuing with the 8-PSK TCM example, the unencoded bit may be either a 1 or a 0. Assuming that the unencoded bit is the most significant bit (MSB), then the channel symbols out of the Pragmatic MPSK TCM encoder may be grouped into two types: those with a most significant bit (MSB) of 1 and those with an MSB of 0. For example, if the output of the encoder is 11, then the channel symbol would be either 011 or 111, where the digit to the far left is the MSB.

At the receiver, a hard decision is made on the MSB based on the detected phase change relative to the carrier phase reference. In the case of 8-PSK, this becomes one of two information bits output from the Pragmatic TCM decoder. Soft decision decoding proceeds on the basis of the relative phase change, without knowledge of the MSB.
The second information bit output from the Pragmatic TCM decoder is then the output of the Viterbi decoder, for the underlying rate $\frac{1}{2}$ 64-state convolutional code, from which the Pragmatic MPSK TCM was constructed. In effect, there are two possible channel symbols associated with each state transition in the trellis, which may be viewed as parallel paths through the trellis (i.e., one symbol if the unencoded bit is a 0 and another if the unencoded bit is a 1) [3 and 10]. As a result, one might expect that the pairwise error bound per channel use for 8-PSK may be viewed as having two components, $D_{n0}(\alpha_{n0})$ and $D_{n1}(\alpha_{n1})$. Similarly, 16-PSK TCM would have 4 components and 32-PSK TCM would have 8 components.

The contributions to the pairwise error bound per channel use are averaged over the possibility that the MSB was either a 1 or a 0, with equal probability. The Viterbi decoder is not considered to have knowledge of the MSB, as such, a relative phase change of the same magnitude and sign yields the same output from the Viterbi decoder, regardless of the MSB.

The probability that a 180° phase error occurs between antipodal signal pairs is assumed to be very small relative to transitions between neighbouring signal pairs. This is generally a good assumption, especially at reasonable values of $E_b/N_0$. However, if necessary, it may be possible to provide additional distance between signals. One way to increasing symbol separation would be to increase the transmit power. A more sophisticated approach would be to integrate the Pragmatic MPSK TCM transmitter with the antenna, which is a possible subject for future work.

An average error weight profile may be obtained

$$F(B, E_n, D(\alpha)) = \sum_a a_a D(\alpha / MSB = 0) \cdot p(MSB = 0) + a_a D(\alpha / MSB = 1) p(MSB = 1)$$

where $E_n$ is the “error frame”, or error pattern, at time $n$ [3], $a_a$ is the number of channel signals in the set that have a squared Euclidean error weight $a_a$. From [3], the number of branches emerging from a node in the trellis is $L = 2^{1-r}$, where $r = \frac{1}{4}$. For a rate $\frac{3}{4}$ code, this means that $L = \frac{1}{4}$. However, in this paper, the convolutional
code upon which the Pragmatic TCM is constructed remains modeled as the rate $\frac{1}{2}$ code that it is, such that, $L = \frac{1}{2}$.

Using the modified generating function approach to performance modeling of Pragmatic MPSK TCM of [3 and 7], with the above changes, gives rise to the error weight profiles of Table 3.2, which have been normalized by the number of branches $L$. The probability of bit error may now be evaluated by combining equation (3.5), with the weight profiles of Table 3.2, based on the modified generating function approach, as described in [3 and 7]

\[
P_b \leq \frac{1}{b} \left. \frac{\partial \bar{\tau}(P(X \to \bar{X}), I, L)}{\partial I} \right|_{I=1} = \frac{1}{b} \sum_{i=1}^{\infty} \sum_{d=d_f} \sum_{\alpha} i a(d, i) \prod_{n=1}^{N} F(B, E_n, D(\alpha_n)) L.
\]

(3.4)

In summary, the manner in which the basic weight profiles are computed is detailed in [3]; however, there are differences between the weight profiles in this paper and those computed based on [3]. The differences arise due to the handling of the unencoded bit(s) associated with Pragmatic TCM and averaging of the weight profiles over the unencoded MSB. For example, consider the case of the path at the minimum Hamming distance from the correct path. The sequence of codewords along the path of minimum Hamming distance is 110 → 001 → 110 → 110 → 000 → 010 → 110.

Table 3.2: Normalized error weight profiles for Pragmatic 8-PSK TCM

<table>
<thead>
<tr>
<th>Signals</th>
<th>Error pattern</th>
<th>$F(B, E_n, D(\alpha_n)) L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(000 or 100)</td>
<td>000</td>
<td>$2D(\alpha_0)$</td>
</tr>
<tr>
<td>(001 or 101)</td>
<td>001</td>
<td>$D(\alpha_1) + D(\alpha_0)$</td>
</tr>
<tr>
<td>(010 or 110)</td>
<td>010</td>
<td>$D(\alpha_3) + D(\alpha_4)$</td>
</tr>
<tr>
<td>(011 or 111)</td>
<td>011</td>
<td>$D(\alpha_3) + D(\alpha_7)$</td>
</tr>
</tbody>
</table>
A single bit error results in a departure from the all 0s path. The minimum Hamming distance along this path is 10 and seven codewords are transmitted (i.e., \( N = 7 \)) before reconvergence with the all 0s path. The contribution to the probability of bit error may be computed from (3.4) as

\[
P_b \leq \frac{1}{2} \cdot \left( D(\alpha_3) + D(\alpha_7) \right)^4 \cdot \left( D(\alpha_1) + D(\alpha_2) \right)^4 \cdot \left( 2D(\alpha_0) \right)^4 \cdot \left( D(\alpha_3) + D(\alpha_7) \right)^4 \cdot \left( D(\alpha_1) + D(\alpha_2) \right)^4.
\]

Relative to [3], the \( 2D(\alpha_0) \) term becomes \( D(\alpha_0) \) and the \( D(\alpha_3) + D(\alpha_7) \) becomes \( D(\alpha_7) \). Further, note that there are two factors of \( \frac{1}{2} \) in the equation above. One is due to the \( \frac{1}{2} \) term from equations (3.3) and (3.4). The second factor of \( \frac{1}{2} \) is valid for ML decoding [11].

The results that follow were calculated using the contributions to the \( P_b \) from paths at Hamming distance 10 and 12 in the unmodified generating function. The contributions to \( P_b \) from paths at distance 10 are provided in Table 3.3. (see Appendix 3A for the contributions to \( P_b \) from paths at Hamming distance 12). Note that the dependence on \( n \) in \( D_n(\alpha_n) \) has been dropped. This is possible because slow fading with interleaving is assumed, such that, the received channel symbols are independent from one another.

A slightly tighter version of the bound on \( P_b \), per equation (3.4), applies in the AWGN channel. This bound is described in [3 and 9] and is employed to evaluate the performance results for AWGN in section 3.2.

While the analytical approach described in this section may at first seem complicated, it is actually rather simple to implement and computationally fast, once the paths are known. In fact, knowledge of only a few paths is required when \( E_b/N_0 \) is large, since the higher order terms tend to drop off quickly, with increased power (i.e., increased Euclidean distance).
Table 3.3: The contribution to $P_s$ from terms at Hamming distance 10

\[
\begin{align*}
1 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
3 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
4 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
2 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
3 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
5 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
6 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
2 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
3 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
3 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
4 &: \left[(D(\alpha_3) + D(\alpha_5))^2 \cdot (D(\alpha_1) + D(\alpha_3))^4 \cdot (2 \cdot D(\alpha_0))^5 \cdot (D(\alpha_2) + D(\alpha_6))^4\right] \\
\end{align*}
\]

3.3. Pragmatic MPSK TCM performance

3.3.1. Performance in AWGN

A summary of the results for Pragmatic 8-PSK, 16-PSK and 32-PSK TCM in AWGN is provided as Fig. 3.2. The results for 8-PSK and 16-PSK TCM are different from those of [4]. It can be shown that the curves in this chapter have the same "waterfall" shape as the measured data published in [1], as well as, the simulated data in [4]. In fact, the results are within a few tenths of a dB of the measured and simulated results over a broad range, from a $P_s$ of $10^{-2}$ to $10^{-9}$. 
Figure 3.2: Pragmatic MPSK TCM and MPSK in AWGN

A comparison of results for Pragmatic 8-PSK TCM in AWGN is provided in Table 3.4a and a comparison of results for Pragmatic 16-PSK TCM is provided in Table 3.4b. Data points obtained per the analysis in this chapter are tabulated and presented alongside with sample data points for the analytic results of [2], the measurements of [1 and 4] and the simulated results of [4].

Table 3.4a: A comparison of sample results for Pragmatic 8-PSK TCM

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>1.6·10^{-2}</td>
<td>1.2·10^{-2}</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5.0</td>
<td>8.5·10^{-4}</td>
<td>7.0·10^{-4}</td>
<td>6.0·10^{-4}</td>
<td>5.0·10^{-3}</td>
</tr>
<tr>
<td>6.0</td>
<td>3.8·10^{-5}</td>
<td>3.6·10^{-5}</td>
<td>2.1·10^{-5}</td>
<td>1.5·10^{-4}</td>
</tr>
<tr>
<td>7.0</td>
<td>8.5·10^{-7}</td>
<td>9.0·10^{-7}</td>
<td>1.9·10^{-6}</td>
<td>2.8·10^{-6}</td>
</tr>
<tr>
<td>8.0</td>
<td>—</td>
<td>—</td>
<td>1.1·10^{-7}</td>
<td>4.1·10^{-8}</td>
</tr>
</tbody>
</table>
Table 3.4b: A comparison of sample results for Pragmatic 16-PSK TCM

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8.0</td>
<td>$7.0 \cdot 10^{-3}$</td>
<td>$0.1 \cdot 10^{-2}$</td>
<td>$-$</td>
<td>$2.6 \cdot 10^{-2}$</td>
</tr>
<tr>
<td>9.0</td>
<td>$3.0 \cdot 10^{-4}$</td>
<td>$5.0 \cdot 10^{-4}$</td>
<td>$5 \cdot 10^{-4}$</td>
<td>$0.1 \cdot 10^{-3}$</td>
</tr>
<tr>
<td>10.0</td>
<td>$0.1 \cdot 10^{-5}$</td>
<td>$6 \cdot 10^{-5}$</td>
<td>$6 \cdot 10^{-6}$</td>
<td>$4.0 \cdot 10^{-5}$</td>
</tr>
<tr>
<td>11.0</td>
<td>$2.1 \cdot 10^{-7}$</td>
<td>$-$</td>
<td>$2 \cdot 10^{-8}$</td>
<td>$7.7 \cdot 10^{-7}$</td>
</tr>
<tr>
<td>12.0</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$2 \cdot 10^{-8}$</td>
</tr>
</tbody>
</table>

From Table 4a, it can be seen that the analytical results for Pragmatic 8-PSK TCM presented in this chapter consistently upper bound the measured and simulated performance of [1 and 4], while the analytic results presented in [2] do not. Again, for the case of Pragmatic 16-PSK TCM in AWGN, the results presented in this chapter upper bound the measurements of [1 and 4], as well as, the simulated results of [4]. The analytic results of [2] for Pragmatic 16-PSK TCM under bound the data.

The difference between the results in this paper and those of [2], and the fact that the results presented in this chapter provide a consistently tight upper bound of the measurements of [2 and 4], can be readily seen when the sample data points are plotted and a plot is provided for the sample cases specified in Table 3.4a and Table 3.4b, as Fig. 3.3. The results for 32-PSK TCM are not found in [1 or 4], so a comparison is not possible. However, the results were derived in the same way as those for 8-PSK TCM and 16-PSK TCM.

Based on the results presented in this chapter, Pragmatic 8-PSK, 16-PSK and 32-PSK TCM perform about 2 to 3 dB in AWGN than uncoded MPSK, at a $P_b$ of $10^{-5}$, for the same bandwidth occupancy. Additionally, it can be seen that the performance of 16-PSK TCM is superior to uncoded QPSK in terms of both bandwidth and power at a $P_b$ of about $10^{-7}$. 
Figure 3.3: A comparison of results for Pragmatic 8-PSK TCM and Pragmatic 16-PSK TCM in AWGN

3.3.2. Performance in Rician fading

A summary of the performance of Pragmatic MPSK TCM in Rician fading is provided as Fig. 3.4. The performance degradation in Rician fading, relative to AWGN, depends on "K", which is the ratio of the direct component of the signal, to the scattered component of the signal. When there is no scattering, \( K \rightarrow \infty \) and the results approximate those of Fig. 3.2 for the AWGN channel. By comparison, \( K = 0 \) corresponds to the case of Rayleigh fading.

Rician fading is applies when \( 0 < K < \infty \). As a point of reference, the mobile satellite communication channel may be approximated as the case where \( K = 10 \) [5]. The performance for \( K = 25 \) approaches that in AWGN; however, there is a difference of approximately 0.5 dB, relative to the curves of Fig. 3.2. The difference appears to be due to the looseness of the bound, which is not as tight as the one employed to achieve the results of Fig. 3.2, per [2 and 9]. The tighter bound of [2 and 9] may be useful in Rician fading, but the improvement is less than in AWGN and the tighter bound does not apply in Rayleigh fading.
3.3.3. Performance in Rayleigh fading

A summary of the performance of Pragmatic MPSK TCM in Rayleigh fading is provided as Fig. 3.5. It is interesting to note that the shape of the $P_b$ curves in Rayleigh fading have a slight curvature, or flattening, at low error rates. This means that the relationship between $E_b/N_0$ and the $P_b$ is not quite inverse linear at high values of $E_b/N_0$ and suggests that the decoding metric may be "running out of steam".

This is believed to be due to sub-optimality of the maximum likelihood (ML) decoding metric in the Rayleigh fading channel (i.e., it is optimum in AWGN). The ML decoding metric appears to be best for smaller values of $E_b/N_0$, corresponding to a $P_b$ greater than or equal to about $10^{-6}$.

The results for pragmatic MPSK TCM in this section will now be applied to the analysis of Pragmatic MPSK TCM with symbol diversity and Pragmatic MPSK TCM concatenated with a Reed-Solomon outer code.
Figure 3.5: Pragmatic MPSK TCM in Rayleigh fading

3.4. MPSK TCM with symbol diversity

The performance of Pragmatic 8-PSK TCM with diversity may be evaluated by substituting the relationship for the Chernoff bound per diversity chip channel use [9 and 11] into eq. (3.2)

$$D\left(\frac{E_c}{N_0}, L\right) = D\left(\frac{E_c}{N_0 L}\right)^L \Rightarrow D\left(\frac{\alpha_n}{L}\right)^L,$$

where $E_c/N_0 = rE_b/L$ and $L$ is the level of diversity. Soft chip combining is assumed.

Fig. 3.6 illustrates the performance improvement with diversity. An improvement of about 2 to 3 dB was found relative to TCM without diversity. The improvement is dependent on the $E_b/N_0$ and tends to be greater at low values of $P_b$. The performance is consistent with that reported for 8-PSK TCM by [1] and others. The optimum diversity level was found to be $L = 2$ for $P_b = 10^{-5}$ and $L = 3$ for a $P_b$ in the range of $10^{-6}$ to $10^{-9}$. The results for 16-PSK TCM are similar [13].
Despite the improvement relative to TCM without diversity, the performance was still not as good as that of coded QPSK, for the same bandwidth occupancy. Symbol diversity with side information may be more useful. As can be seen from the results in this chapter, $P_e$ is heavily dependent on the contributions of a few key paths through the trellis, at least at low error rates. Side information could be used to “weight” the relative contributions of the symbols to the branch metrics to improve performance in a Rayleigh fading channel.

3.5. Concatenated Reed-Solomon MPSK TCM

Fig. 3.7 illustrates the concatenated Reed-Solomon (RS) TCM transmitter. The concatenated RS-TCM code is considered for use when the channel is less than ideal. The potential performance gain is examined in a Rayleigh fading channel.

![Figure 3.6: Pragmatic MPSK TCM with diversity in Rayleigh fading](image)

![Figure 3.7: Transmitter for concatenated MPSK RS-TCM](image)
Computation of the bit error rate is based on [11]. A (255,223) outer RS code with an inner MPSK TCM code is examined. The (255,223) RS code can correct up to 16 Q-ary symbols out of 256, with $Q = 8$ channel bits per symbol. From [11], the probability of a Q-ary symbol error out of the RS decoder is

$$P_{\text{symbol}} = 1 - (1 - P_{\text{inner}})^Q,$$

where $P_{\text{inner}}$ is the probability of a bit error out of the inner MPSK TCM decoder. The probability of bit error out of the RS decoder can be calculated using the equation below [11]

$$P_e = \sum_{i=1}^{N-1} \frac{i}{2(N-i)} \binom{N}{i} P_{\text{symbol}}^i (1 - P_{\text{symbol}}^{N-i}), \text{ where } N = 255 \text{ and } i = 16.$$

The results for concatenated RS-TCM are provided as Fig. 3.8. At low values of $E_b/N_0$, MPSK TCM outperforms concatenated RS-TCM. The concatenated code does not perform well when $E_b/N_0$ is too low because the input error rate to the RS decoder becomes unacceptably high.

Figure 3.8: Concatenated RS-TCM in Rayleigh fading
For example, an examination of the performance of RS-8PSK-TCM in Fig. 3.8 clearly shows that the concatenated code does not perform well relative to TCM alone when the channel bit error rate exceeds about $10^{-3}$, which corresponds to an $E_b/N_0$ of about 12.5 dB and translates to an $E_b/N_0$ of about 10.2 dB, as seen by the inner decoder. However, as expected, the performance of the RS-TCM concatenated code is superior at higher values of $E_b/N_0$. As can be seen from Fig. 3.8, the performance improvement is about 6 dB for 8-PSK, 16-PSK and 32-PSK, at $P_b = 10^{-8}$, which may be necessary to support certain types of communications in the future [14]. The performance advantage of the concatenated code is particularly interesting compared to the low bandwidth expansion of the (255,223) RS code (i.e., a factor of $2^{255}/2^{223}$).

3.6. Summary and concluding remarks

The performance of bandwidth and power efficient Pragmatic MPSK TCM was examined in AWGN, Rician fading and Rayleigh fading. The results for 8-PSK TCM and 16-PSK TCM are new and the results for 32-PSK are novel. Higher order PSK (i.e., larger M) offers greater bandwidth efficiency. For example, 8-PSK is 50% more bandwidth efficient than QPSK and 16-PSK is twice as bandwidth efficient as QPSK. However, the power efficiency of higher order MPSK decreases and the sensitivity to phase noise increases.

For typical data communication systems with a QoS requirement allowing operation at a bit error rate between $10^{-5}$ and $10^{-6}$, TCM is probably superior to the concatenated RS-TCM code examined in this chapter, from the perspective of performance and complexity. However, the concatenated RS-TCM code is superior when performance is at a premium, such as may be the case in future systems employing the Internet protocol suite (TCP/IP) over Asynchronous Transfer Mode (ATM) via satellite link, where the ability to achieve and maintain error rates of $10^{-8}$ or less is required [14].

Additionally, although TCM with diversity in the form of symbol repetition was found to be provide 2 to 3 dB of gain relative to TCM alone, it was neither as bandwidth efficient nor as power efficient as the RS-TCM code and is not considered further. A different and more promising form of diversity will be considered in Chapter 5.
References

14. J. Farserotu and A. Tu, "TCP/IP Over Low rate ATM-SATCOM Links", MILCOM '96, McLean Va., USA.
APPENDIX 3A: TERMS AT HAMMING DISTANCE 12

\[ a \left[ (D(\alpha_3) + D(\alpha_7))^b \cdot (D(\alpha_1) + D(\alpha_5))^c \cdot (2 \cdot D(\alpha_0))^d \cdot (D(\alpha_2) + D(\alpha_6))^e \right] \]

<table>
<thead>
<tr>
<th>term</th>
<th>A</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>term</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>19</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>20</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>21</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>22</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>23</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>24</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>25</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>26</td>
<td>7</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>27</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>28</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td>5</td>
<td>29</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>5</td>
<td>30</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>31</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>32</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>33</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>5</td>
<td>34</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>35</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>7</td>
<td>36</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>37</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>


Chapter 4

ATM-CDMA transceiver

4.1. Introduction

In this chapter, key issues and design considerations, for a CDMA transceiver capable of supporting the ATM-CDMA-SATCOM concept described in Chapter 2, are examined. The design combines the flexibility of ATM with that of CDMA by assigning a code per ATM connection or service class, as opposed to users. A user may require multiple codes to support a multimedia multicast service, but each user would be free to receive any and all components of the service, within the limits of their terminals capabilities.

The receiver is based on a simple generic DS-CDMA design, which is modified to support the ATM-CDMA-SATCOM concept. The ability to accommodate multiple simultaneous connections requires the reception of multiple signals at once. There are various options for accommodating the reception of multiple simultaneous signals, but the simplest is probably to have a bank of receivers in hardware [1], where each receiver tracks a different code. The limit on the number of CDMA receivers in the bank is a cost vs. design trade. A relatively large terminal, such as a major SATCOM injection site, or a wireless base station [1], can more readily accommodate the additional complexity than a small terminal, where it imposes a greater overhead cost on the design. In either case though, the number of available codes is not limitless and this must also be taken into consideration.

The number of codes, receivers and their complexity are important design considerations. Another key issue involves network flexibility. In a conventional CDMA system, a point-to-point or link would require only one transceiver per ground terminal and the allocation of a single code. By comparison, multipoint-to-multipoint communication requires multiple receivers and multiple codes at each terminal. In the proposed concept, the number of codes and receivers would be increased since
additional codes and receivers would be required to support the individual component services of a multimedia communication. However, in this concept, the focus is on point-to-multipoint broadcast or multicast communication, which requires fewer codes overall than full mesh multipoint-to-multipoint, and allows for simpler receivers at small terminals.

The proposed design also attempts to minimize the number of codes and receivers required to implement the concept of Chapter 2, by exploiting the commonality that exists between transmitter-receiver pairs in a multimedia communication. Unlike multipoint-to-multipoint communication from multiple transmitters, which experience different propagation losses and delays, as viewed at a common receive terminal in the network, the transmission of component multimedia services, between a single transmitter-receiver pair, occurs over the same satellite link, at the same time. The losses and delays experienced by the components of a multimedia service that are simultaneously transmitted over the same satellite link are identical, by definition, with the possible exception of transmission over a processed satellite, which will be discussed later in this chapter.

Assuming that the components of a multimedia service are transmitted separately and given that identical delays and losses are experienced when the components are transmitted simultaneously over the identical satellite link, it is proposed to employ time shifted versions of a single parent CDMA code, in order to spread and transmit the components of a given multimedia service request. The time shifts would be in multiples of chip intervals [2, p.357-358]. This is a form of synchronous CDMA [2]. A very tight time accuracy is required (e.g., within about one chip interval) over multiple channels.

Although synchronous CDMA may not be practical between different transmitter-receiver pairs [2], which experience independent and variable delays and may be especially difficult to implement for broadband CDMA, where the chip duration is very short, it should be possible to implement for simultaneous transmissions over identical transmitter-receiver pairs. If so, then time shifted versions of a single parent code may be employed to implement multiple connections between user pairs in support of multimedia communication, as required to achieve the full flexibility of the
ATM-CDMA-SATCOM concept described in Chapter 2. This, at least, is the goal. A
description of the transceiver follows. Acquisition and synchronization issues are also
examined.

4.2. ATM-CDMA transmitter

A high level illustration of the transmitter is provided as Fig. 4.1. Separate
transmissions by ATM service class are depicted. GPS is assumed as the source of
precise time and frequency, at least at satellite injection terminals, or base stations.

The PN codes employed to support the component services of a multimedia service
are time shifted versions of the same code, where the time shift (D) is nominally one
chip interval [2]. The code generator is illustrated as a tapped delay line. The
transmitter has two up conversion stages (U/C). The first up U/C stage converts the
individual spread signals (e.g., 200 MHz) from basedband to 700 MHz, where they
are summed together and then up converted to 30 GHz for transmission. Note that a
single user at a small terminal would require one module as illustrated in Fig. 4.1. A
large satellite injection terminal supporting multiple users would require multiple
modules.

Figure 4.1: Block diagram of the proposed ATM-CDMA transmitter
4.3. ATM-CDMA receiver

A block diagram of a possible DS-CDMA receiver is provided as Fig. 4.2., from which it can be seen that the key elements following down conversion of the signal are: code acquisition and tracking, carrier phase acquisition and tracking, data detection, decoding and gain control. As with the transmitter, multiple modules may be needed dependent on the degree of inter-networking between SATCOM terminals to be supported.

4.3.1. Code acquisition

Following PN correlation (de-spreading), detection of code lock is based on whether the total power exceeds a threshold dependent on the signal-to-noise ratio (S/N), the probability of acquisition and filter bandwidth. Energy combining of chips may be necessary, but initially simple single dwell non-coherent detection is assumed [2].

If the threshold condition is not satisfied, then the code will be shifted either forwards or backwards in time. Nominally, the shift will be a half chip interval. As envisioned, code acquisition would be performed on a control channel that is always present and common to all users. Code acquisition, also known as network entry and will be described further in section 4.4. To the greatest degree possible, common functions and components are shared within the receiver. For example, coarse acquisition of the "parent" code need only be performed once per family of codes and acquisition of the "child" codes amounts to tracking.

4.3.2. Direct sequence (DS) code tracking

Once code lock is obtained, code tracking must be performed. This could be accomplished using a simple non-coherent early-late gate delay lock loop [2], where the difference in the power between the early correlation and the late correlation determines whether the chip rate is advanced, delayed or unchanged.

Relative to coherent code tracking (e.g., via a matched filter), it has been shown that non-coherent code tracking offers the advantages of lower loss of lock rate under fading conditions [3], as may occur in mobile wireless communications. The reason is that the non-coherent tracking loop does not depend on carrier phase lock. Loss of lock can readily lead to loss of bit count integrity (BCI) and degradation of the QoS.
Figure 4.2: Block diagram of the proposed ATM-CDMA receiver

In a typical CDMA system, each code corresponds to a unique signal to be de-spread and received and each code must be tracked separately. In this concept, a family of codes derived from time delayed versions of a parent code may be employed to support multimedia communication, in which case, only the parent code(s) would need to be actively tracked. This is because the time delay associated with each child code relative to the parent is known apriori, since it is established at call setup and it is valid as long as the delay variations between the parent and child codes remains minimal.

4.3.3. Carrier acquisition and tracking (frequency)

Carrier lock is detected by subtracting the power in the in-phase (I) and quadrature (Q) channels and comparing the result to a threshold. A voltage controlled oscillator (VCO) is stepped until the carrier is detected. The VCO drives the DS spread time correlator used for de-spreading. This is a typical implementation. Carrier recovery is more novel.
4.3.4. Code aided carrier recovery

TCM, like any PSK modulation, is sensitive to phase and timing errors. This is especially true in a fading channel, and all the more so when frequency selective fading is considered. In some cases, adaptive equalisation may be required, but this is not expected to be the case for geosynchronous satellite links, where the coherence bandwidth is very large \[4\] and not addressed herein.

The combination of MPSK or MFSK TCM and CDMA may be complimentary. Intuitively, knowledge of time to within a fraction of a chip duration (e.g., an accuracy of a nanosecond or less) is better than knowledge to within the same fraction of a bit or symbol (e.g., microsecond accuracy). A clock aided carrier recovery system has been proposed for TCM \[5\]. From \[5\], there is a significant potential performance benefit provided that the time jitter is within certain limits; nominally, \(0.3 \cdot T\) for 8-PSK, where \(T\) is the symbol duration. A similar concept for carrier recovery and tracking is proposed using a clock extracted from the PN code. The "code" clock is then applied to an \(M^{th}\) power law device (\(M = 8\) PSK), as illustrated by Fig. 4.2 and described in \[5\]. The difference is that the clock would be derived from the early-late gate code tracking loop, hence the name code-aided carrier recovery.

![Image of code-aided carrier recovery system]

Figure 4.3: Method for extracting a PN code-aided clock
A possible method for deriving a clock from the PN code is illustrated by Fig. 4.3. The method employs a code clock pulse counter with AND gate. In the example shown, each clock pulse is 4 code clock pulses in duration. With clock-aided carrier recovery, jitter becomes the primary source of phase noise. This is different from the general model for the phase noise when M-law carrier recovery is employed. Both cases are of interest, and analysis is planned. Based on [5], jitter should be maintained within about 0.3 of a channel symbol. However, the $S/N$ ratio within the carrier recovery circuit will be reduced by about 5 dB, which could adversely affect PN code acquisition and tracking.

The basic concept for code-aided carrier recovery has been described. Code-aided carrier recovery may facilitate the use of higher order MPSK based modulations in a DS-CDMA system by limiting phase variations. Broader questions concerning timing accuracy, acquisition and synchronization of the proposed CDMA receiver must be understood, in order to determine the potential for trading $S/N$ for improved phase accuracy. Further, the benefit of code-aided carrier recovery may be lost, if the timing jitter is excessive. Time and frequency considerations will be discussed shortly. Code-aided carrier recovery is a subject for future analysis.

4.3.5. **Automatic Gain Control (AGC)**

The AGC loop is projected as an in-phase (I) and quadrature (Q) power measurement relative to the in-band noise $(S + N)/N$. The measurement can be made over an extended period of time, over a large number of chips, in order to improve the accuracy of the measurement for the purposes of acquisition and tracking.

AGC power measurements can provide side information to assist in the data detection and decoding process, for example, to estimate the noise power for the purposes of obtaining side information on the interference power from other users. Perhaps more importantly, with respect to ATM, AGC power measurements can be used to facilitate congestion control and connection admission control as described in Chapter 2. This is a subject for further analysis. Additionally, weighting of groups of chips may be employed to limit the impact of interference and may be especially interesting in a hybrid FH/DS-CDMA concept. This is a possible subject for further analysis.
4.3.6. **Summary of key features**

Key features of the proposed ATM-SATCOM-CDMA transceiver include:

- CDMA on both the forward and return links,
- reception of multiple simultaneous CDMA signals,
- PN code per ATM connection or service class and
- code sharing via time shifting for multimedia component services.

The last two features, described in Chapter 2, are unique to the ATM-SATCOM via CDMA proposal in this thesis. Other potentially novel elements of the CDMA transceiver remain to be evaluated, such as, the use of side information in the form of \((S + N)/N\) (especially in a hybrid FH/DS-CDMA concept) to facilitate decoding in the presence of interference or imperfect power control. Similar measurements could also be employed to facilitate ATM congestion control and connection admission control. In the case of a processed satellite, the integration of ATM connections, CDMA codes and MBA beams could enhance the efficiency of SATCOM resource allocation by allowing greater flexibility to match bandwidth and power allocations with service and QoS requirements. Additionally, code-aided carrier recovery, time estimation or time-tagging for delay estimation may be integrated into the design.

4.4. **Time and frequency uncertainty over satellite links**

Precise time and frequency (PT&F) play an increasingly important role in digital communication systems and networks. Precise time is required for entry into the network, network synchronization, and link acquisition. Precise frequency is required to maintain synchronization, and to ensure that the accumulation of frequency offset and time interval error (TIE), or time uncertainty, which may degrade performance in various ways, is maintained within acceptable limits. In this concept, time and timing accuracy are especially important since they are the basis for sharing a common code, as described earlier in this chapter and in Chapter 2.

The specific requirements for precise time and frequency can vary significantly from one type of communication system to another. In the case of digital satellite networks, the requirements are complicated by the additional delay, delay variation and Doppler
shift due to satellite motion. Satellites are never perfectly stationary. The degree of satellite motion is dependent on the operational station-keeping tolerances, which may be relaxed, and in some cases eliminated, in order to conserve fuel and prolong the life of the satellite. In the case of low earth orbit satellites, station-keeping is not applicable, and the orbital inclination may be large. Higher orbital inclination means greater variations in total path length relative to the mean path length, which tends to increase the delay, delay variation and Doppler shift.

Delay and delay variation are most relevant to geosynchronous satellite links. Doppler is most significant at higher operating frequencies and for constellations of rapidly moving low earth orbit satellites, as consistent with various concepts for future PCS. In either case though, time and frequency variations can be introduced at various points throughout the system; including, translation through the satellite transponder. The integration of SATCOM and terrestrial networks, the use of mobile terminals and the operation of CDMA networks require that precise time and frequency be kept within certain tolerances and that variations be accommodated.

In addition to increased delay, delay variation and Doppler shift, frequency drift due to the translation over the satellite transponder can accumulate. At SHF, the translation frequency on-board the satellite is usually either 650 MHz or 725 MHz. A translation frequency of 650 MHz is assumed in the analysis of Ka-band that follows. An oscillator stability consistent with that of good quartz, with an aging rate of about 1 part in $10^7$ per month, is considered, along with an initial frequency offset of not more than 100 Hz.

Transportable and mobile terminals, especially highly mobile terminals, tend to have small antenna apertures. The available link margin is generally less than that which might be achievable with larger aperture antennas. Typically, lower margin means lower $S/N$. For a DS-CDMA system, lower $S/N$ translates into longer acquisition time, which increases the importance of minimizing time uncertainty. In this system, it could also affect the ability to implement Pragmatic MPSK TCM and reduce the overall efficiency by more than merely the loss of $S/N$, as shown in Chapter 2.

For the all of the above reasons, an understanding of the issues associated with time,
frequency and synchronization is important and may be especially so in this concept. In the remainder of this section, the sources of PT&F uncertainty will be identified. CDMA acquisition performance will be evaluated based on a simple non-coherent early-late gate code synchronizer. The potential for trading $S/N$ for better phase accuracy in order to facilitate the use of code-aided carrier recovery will considered. The use of GPS may be necessary to provide a common high accuracy time source and to reduce the acquisition time sufficiently such that $S/N$ could be sacrificed to facilitate the implementation of more bandwidth efficient modulations, but this remains to be determined. An analysis follows.

4.4.1. The sources of time and frequency uncertainty

The requirements for time and frequency within a communication network can be categorized in terms of the distribution of timing, or clock, and the provision of a high stability frequency reference. In general, as time and frequency uncertainty increase, performance degrades. Acquisition takes longer. Loss of network synchronization may occur and reacquisition time increases. Loss of communications and down time are the result. Although the requirements for accurate time and frequency are interrelated and may be obtained from a common source, such as GPS, the specific requirements vary with communication subsystem, which may be susceptible to the affects of different sources of time and frequency uncertainty. For example, long term frequency variations that result in the accumulation of time interval error ($TIE$) at baseband, leading to bit slips, may result in a carrier offset at the 70 MHz input to the demodulator, which increases the acquisition time and contributes to carrier tracking slips.

In this section, the sources of time and frequency uncertainty are examined. The sources are categorized in terms of:

- propagation effects over the satellite,
- long term stability between oscillators and
- short term stability or jitter.

Transmission over the satellite introduces various forms of time and frequency
uncertainty. Delay, delay variation, Doppler shift and frequency variations are incurred, which may increase the total time and frequency uncertainty and the acquisition time.

Long term and short term oscillator stability are key sources of time and frequency uncertainty. Wander is the result of long term stability variations. Wander may lead to the accumulation of TIE between the SGTs and terrestrial networks. TIE is a form of time uncertainty.

The resolution of time uncertainty is an important factor with regards to PN acquisition, as well as, the integration of SATCOM into digital networks. The accumulation of TIE may lead to clock slips or deletes and the resultant loss of bit count integrity may adversely affect throughput. At the physical level, loss of carrier, phase lock and bit synchronization may lead to temporary outages. At higher levels, bursts of errors may occur if buffers overflow.

Jitter is a short term stability variation. It can adversely affect tracking and carrier recovery. It also contributes to both the total time and frequency uncertainty.

4.4.3. Propagation affects over the satellite
Satellites are never completely stationary. They move relative to the earth and their motion can give rise to a Doppler shift of the carrier frequency. The magnitude of the Doppler shift is related to the change in range between the satellite and the receive terminal, which depends on the type of orbit the satellite is in and the station-keeping capabilities of the spacecraft, as well as, whether the receive terminal is stationary, or mobile.

The relatively simple case of a geosynchronous satellite without north-south station-keeping, but with ideal east-west station-keeping, is considered in the analysis that follows. This is consistent with the operation of commercial satellites [6]. Geosynchronous satellites without north-south station-keeping move in a figure eight pattern about the sub-satellite point on the earth over the course of the day as described in [7]. The orbital plane may be viewed as inclined and an inclination of not more than ±5° is considered [6].
**Frequency uncertainty due to Satellite motion: Doppler shift** - The Doppler shift and related effects are the result of range variation over time. The range variation per unit time interval is known as the range rate \( \frac{dp}{dt} \). In computing the total Doppler shift experienced by a carrier over a satellite link, the contributions of both the uplink and downlink must be considered.

Doppler shift is frequency dependent. The fractional Doppler shift is the frequency deviation per unit frequency \( \frac{\Delta f}{f} \), which is a unitless quantity related to the range rate per equation (4.1)

\[
\frac{\Delta f}{f} = \left( \frac{f_{UL} - f_{DL}}{f_{UL}} \right) = -\frac{\left( \frac{dp_{UL}}{dt} + \frac{dp_{DL}}{dt} \right)}{c},
\]

(4.1)

where \( f_{UL} \) is the uplink frequency (Hz), \( f_{DL} \) is the downlink frequency (Hz), \( p_{UL} \) is the uplink path length (m), \( p_{DL} \) is the downlink path length (m) and \( c \) is the speed of light, which equals \( 3.0 \cdot 10^8 \) (m/s).

Generally, the vector sum of the uplink path variation \( \frac{dp_{UL}}{dt} \) and the downlink path variation \( \frac{dp_{DL}}{dt} \) must be considered, as shown in equation (4.1); however, simple addition is appropriate when a) the north-south motion of the satellite is much greater than the east-west motion and b) the link is between satellite ground terminals located in the same hemisphere, such that the satellite moves either towards or away from both terminals at the same time. Factors other than north-south station keeping, such as orbital eccentricity and east-west drift, can also have an impact on path variations and the Doppler shift; however, (1) is a reasonable approximation for orbital inclinations greater than about 3\(^\circ\). For a more detailed analysis refer to ref. [7].
Based on an orbital inclination of 5°, the maximum range rate would be about 50 m/s on the uplink and 50 m/s on the downlink, for a total of 100 m/s [8]. This results in a fractional Doppler shift of $3.33 \times 10^{-7}$. The Doppler shift seen by the receiver would be only about 667 Hz, based on the 2 GHz carrier frequency ($f_c$) (i.e., WARC-92 frequency allocations for UMTS satellite links), but it would increase to 8333 kHz for a typical 30/20 GHz EHF satellite link:

$$f_c\left(\frac{\Delta f}{f}\right) \equiv 8333 \text{ Hz (30/20 GHz carrier)}.$$  

The Doppler shift can substantially contribute to the frequency departure experienced by a carrier over a satellite link and to the total frequency uncertainty, as seen at the receiver. Oscillator drift is also a factor and will be considered shortly. Furthermore, the Doppler shift due to satellite motion is an instantaneous frequency deviation. The cumulative effect of frequency departure is also important. As will be shown next, the cumulative affect of satellite motion results in delay variations and time uncertainty, which can also have a significant impact on network T&S, as well as, network entry.

**Time uncertainty due to satellite motion** - In addition to Doppler shift, satellite motion can lead to the accumulation of time uncertainty. The problem may be viewed as the accumulation of delay due to the range changes, resulting from the motion of the satellite over the course of each day. The resultant contribution to the total time uncertainty due to satellite motion can be determined from

$$T_u = \frac{\text{Rate}}{c} \int_0^T (\Delta \tilde{p}_{UL} - \Delta \tilde{p}_{DL}) dt .$$  \hspace{1cm} (4.2)

$T_i$ is integration period, which is the time required for the satellite to move from the position of minimum path length to maximum path length, from the perspective of the transmit-receive terminal pair. In the case of a low earth orbit satellite, this may be interpreted to mean in and out of the ground terminals field-of-view. *Rate* is the transmission rate of the link in bits per second.
In the analysis that follows, as with the evaluation of Doppler above, the ground terminals are assumed to be located in the same hemisphere, such that the effects of delay variation are additive. This is the worst case, since the effects of terminals in different hemispheres may cancel. The most distant terminal pair, in terms of the total path length, is considered. This also bounds the variation, or uncertainty. The affects of east-west drift and orbital eccentricity are considered minimal, relative to north-south station keeping, which is a reasonable approximation for high orbital inclinations.

Given the above assumptions, the maximum time uncertainty in seconds due to delay variations resulting from the orbital inclination of equation (4.2), may be approximated using [9]

\[ T_u = \left( \frac{R_{\text{max},\ell} - R_{\text{min},\ell}}{c} + \frac{R_{\text{max},\ell} - R_{\text{min},\ell}}{c} \right) \]

Most commercial geosynchronous satellites have at least some degree of north-south station-keeping; however, there are exceptions and, for satellites without north-south station-keeping, variations of up to ±8 ms, or more, can occur due to satellite motion [6]. There are other factors to consider in estimating a reasonable bound on the time uncertainty to be resolved. Long term frequency stability is examined next.

4.4.4. Long term frequency stability: drift and offset
In addition to the motion of the satellite, oscillator drift and offset throughout the system can have a significant impact on frequency, as well as, time uncertainty. Both the long term and short term frequency stability are important. The effects of long term oscillator stability is examined in this section. The effects of short term are examined in the next section.

Frequency uncertainty due to wander and drift - One aspect of long term stability that is unique to SATCOM is drift of the oscillator used for frequency translation onboard the satellite. The uplink signal is not demodulated on the satellite, rather, it is translated to the downlink frequency and amplified at radio frequency (RF). In order
to examine the problem, the standard straight line approximation for drift due to aging will be assumed [10].

At time $t = 0$, the model assumes an initial frequency ($f_0$). Over time, $f_0$ is assumed to increase, or decrease, linearly with respect to a reference frequency ($f_r$) dependent on an oscillator aging rate ($a$). Aging is relevant primarily to quartz and rubidium oscillators. It is negligible with respect to cesium references [10]. GPS is based on cesium, although there may be a few GPS satellites that still rely on rubidium standards. Rubidium and good quartz oscillators will be considered as a reasonable worst case, for the purposes of the analysis that follows.

Based on the straight line model, assuming an upper bound of 1 part in $10^7$ per month for the aging rate of a high quality quartz oscillator on-board the satellite, and a translation frequency of 650 MHz for the transponder, the expected frequency departure is estimated to be not more than about 65 Hz, in one month's time. Given an initial frequency offset of 100 Hz, the total frequency departure after one month would be 165 Hz. When coupled with the value of 8333 Hz previously calculated for Doppler shift over a 30/20 GHz geosynchronous satellite link, the result is a total frequency uncertainty of about 8498 Hz, before consideration of short term stability.

Total frequency uncertainty affects entry into the CDMA network, as well as, the stability of the clock derived from the incoming signal. Slow variation of the carrier frequency due to oscillator drift on the satellite was not included, as it does not affect the accumulation of time uncertainty.

**Time uncertainty due to wander and drift** - Long term oscillator stability can also have an impact on the accumulation of time uncertainty. From [10], and given that $f = \frac{l}{t}$, where $f$ equals frequency in hertz and $t$ is the observation time in seconds

$$df = -t^2 \frac{df}{dt} = -\frac{dt}{t} f.$$ 

This implies that the magnitude of the fractional frequency uncertainty is equal to the magnitude of the fractional time uncertainty.
\[
\frac{\Delta f}{f} = \frac{\Delta t}{t},
\]  

(4.2)

The relationship between frequency and time (4.2) leads to the following expression for the resultant time uncertainty, which may be thought of as analogous to time interval error (TIE) between clocks

\[
|\text{TIE}| = \int \frac{\Delta t}{t} dt = \int \frac{f_0 + a f_0 t - f_r}{f_r} dt.
\]

Integration yields the familiar expression for TIE

\[
|\text{TIE}| = \frac{a t^2}{2} + y_o t + e(t), \text{ where } y_o = \frac{f_0}{f_r} - 1.
\]  

(4.3)

Intuitively speaking, the time squared dependency of the TIE arises due to the linear frequency variation with time due to aging. The linear component of the TIE with time is related to frequency inaccuracy, which is due to frequency offset and setability. The final term \(e(t)\) is a random factor that is due to a combination of initial time uncertainty and random short term variations, such as may result from jitter. TIE must be resolved during the process of PN code acquisition. To the extent that it is known or limited, acquisition is faster.

Consider the time uncertainty that must be resolved between a user terminal and a SATCOM injection terminal, or master terminal, when the user joins the network after having been out for an extended period of time; nominally, 24 hours. At the SATCOM injection terminal with GPS, the aging rate is effectively zero and \(y_o = 10^{-11}\). At the user terminal with a common quartz clock, the aging rate is about 1 part in \(10^5\) per month, \(y_o = 10^{-11}\) and \(e(t)\) is negligible (i.e., relative to the affects of aging and initial frequency offset over 24 hours). From (4.3), the contribution of the long term clock stability to the total time uncertainty would be 1.4 ms.
4.4.5. Short term stability: timing jitter between clocks

Short term clock variations are generally random and may be thought of as a form of timing jitter. Clock jitter is related to phase noise within the oscillator. Total phase noise from all sources is a very important factor in terms of overall system performance, but it is too broad a topic to be treated in detail this report. However, measurements suggest that the contribution to the total jitter due to phase noise at a typical satellite terminal, exclusive of the satellite transponder and the modems, may be characterized by a root normalized Allan variance of less than 1 part in $10^{10}$, in a 1 second duration [9]. On the ground, with respect to a mobile communication system with GPS at the hub and quartz at the mobile terminal, the short term stability is expected to be better than 1 part in $10^{10}$ (Allan variance over 1 second).

Short term stability and, to a lesser degree, long term stability may be viewed as the standard deviation of a zero mean Gaussian random variable. As such, it can be shown that the mean time between timing variations that exceed a threshold magnitude referred to herein as the "slip" size may be estimated from

$$MTBS^{-1} = \frac{I}{S \sigma_T \sqrt{2\pi}} \left| \frac{\Delta f}{f} \right| \exp \left( -\frac{\left| \frac{\Delta f}{f} \right|}{2 \sigma_T} \right) d\left| \frac{\Delta f}{f} \right|,$$

(4.4)

where the total variance ($\sigma_T^2$) is the sum of the short term ($\sigma_{sf}^2$) and long term ($\sigma_{lh}^2$) contributions. This expression can be simplified to

$$MTBS^{-1} = \frac{0.798 \sigma_T}{S},$$

where $S$ is the slip size in seconds.

From equation (4.4), assuming a short term stability consistent with that of a good quartz oscillator, a short term time variation or uncertainty, in the range of 1 $\mu$s to 10 $\mu$s, could occur on a regular basis (e.g., multiple times per day before consideration of other sources of time uncertainty). In fact, variations of several microseconds could occur every few seconds without Doppler correction.
Microsecond variations are not large and can be accommodated, but they can place
limitations on acquisition and reacquisition. Furthermore, the potential impact may be
greater if code-aided carrier recovery is considered. If the jitter on the code clock is
too large, then code-aided carrier recovery may not be useful.

It is proposed to model and evaluate the timing jitter associated with the symbol clock
derived from PN code clock, as a multiple of advanced or delayed fractions of a PN
chip. For example, assuming that code tracking is maintained to within 1/8 of a unit
interval (UI), or PN chip, where the figure of 1/8 of a UI may be viewed as the
granularity of the early-late gate synchronizer, then the jitter on the symbol clock
could be modeled using a binomial distribution for the probability of a sequence of
1/8 chip advances and delays. This is a possible subject for future work.

4.4.6. Summary of time and frequency uncertainty relative to SATCOM
A summary of the key time and frequency uncertainties identified in the preceding
paragraphs is provided as Table 4.1. These values will be applied in the analysis in the
remainder of this chapter.

4.5. The impact on CDMA acquisition and synchronization
The requirements for PT&F in a CDMA network are driven by chip rate and the need
for rapid network entry and re-entry. In it's simplest form, the problem is one of PN
code acquisition or reacquisition, which is complicated by time and frequency
uncertainty accumulated from various sources, as described in the preceding sections.
Long PN code sequences with no a priori knowledge of the true code phase position is
assumed.

The time uncertainty may be viewed as a window to be searched step-by-step. The
step size is in half chip increments, and the search rate (\( S_r \)) is dependent on the \( S/N \),
as well as the desired probability of PN sequence acquisition (\( P_{acq} \)) and probability of
false alarm (\( P_{fa} \)). Essentially, the time uncertainty, expressed in PN chips divided by
the applicable search rate in chips per second, determines the acquisition or
reacquisition time; however, the process is probabilistic and dependent on the \( S/N \).
The cost in $S/N$ for use of small terminals, bandwidth efficient modulation, such as MPSK TCM, or code-aided carrier recovery to limit phase variations that can adversely affect the performance of TCM and must be balanced against the requirements for acquisition and tracking.

A simple PN acquisition model is presented as Fig. 4.3, it is based on non-coherent detection [2]. The intent is to provide insight into the PT&F requirements, network entry (acquisition) and re-entry (reacquisition).

The mean acquisition time in seconds may be evaluated using the methodology of [2]

$$
\bar{T}_{acq} = \frac{(2 - P_a)(1 + K P_e)(q t_d)}{(2 P_a) \left[ \frac{N_u}{q} + \Delta f_c \tau_d (1 + K P_{fr}) \right]} \quad (4.5)
$$

Table 4.1: Summary of key sources of SATCOM time and frequency uncertainty

<table>
<thead>
<tr>
<th>Category</th>
<th>Nominal Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satellite propagation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Roundtrip delay</td>
<td>267 ms</td>
<td>40000 Km, GEO space-ground link</td>
</tr>
<tr>
<td>• Delay jitter</td>
<td>± 8 ms</td>
<td>north-south motion, ± 5 deg. orbital inclination</td>
</tr>
<tr>
<td>• Doppler shift</td>
<td>± 8333 KHz</td>
<td>± 5 degree orbital inclination, ± 50 m/s</td>
</tr>
<tr>
<td><strong>Frequency stability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Frequency offset</td>
<td>± 100 Hz</td>
<td>high stability quartz oscillator</td>
</tr>
<tr>
<td>• <strong>Frequency drift</strong></td>
<td>± 65 Hz</td>
<td>1 part in $10^7$/month, 30/20 GHz</td>
</tr>
<tr>
<td><strong>Long term clock stability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Quartz – Quartz</td>
<td>2.9 ms</td>
<td>Operation between quartz clocks, 1 part $10^6$</td>
</tr>
<tr>
<td>• Quartz – GPS</td>
<td>1.4 ms</td>
<td>quartz (1 part $10^6$) and GPS UTC (USNO)</td>
</tr>
<tr>
<td>• GPS - UTC (BIPM)</td>
<td>10 µs</td>
<td>cross border operation, GPS and UTC (BIPM)</td>
</tr>
<tr>
<td>• GPS - GPS</td>
<td>N/A</td>
<td>Common GPS reference</td>
</tr>
<tr>
<td><strong>Short term clock stability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Clock jitter</td>
<td>± 1-10 µs</td>
<td>short term quartz stability</td>
</tr>
<tr>
<td>• $\sigma$</td>
<td>± $10^{-10}$ s/s</td>
<td>root normalized Allan variance (oscillator)</td>
</tr>
</tbody>
</table>
Figure 4.3: The network entry process

Given $P_{fa}$, then the probability of single PN chip detection can be found

$$P_d = Q\left(\frac{Q^{-1}(P_{fa}) - \sqrt{B_{\tau_d}(S/N)}}{\sqrt{1 + 2S/N}}\right),$$

where $K$ is a penalty for false alarm (i.e., typically $K > 1$). No penalty (i.e., $K = 1$) is assessed in this analysis. The desired $P_{acq}$ must be chosen and the relationship between $P_{acq}$ and $P_d$ is $P_{acq} = 2P_d - P_d^2$.

The search is commonly performed in $\frac{1}{2}$ chip increments. The search rate is

$$S_r = \frac{0.5}{t_d} \text{ chips/second},$$

where $t_d$ is the dwell time of the detector. The number of code alignments to be searched ($q$) is $q = 2N_a$ code alignments/second. The number of chips to be searched ($N_a$) is given by
\[ N_u = \frac{T_u}{T_c} \]  

(4.6)

\( T_u \) is the total time uncertainty and \( T_c \) is the PN chip rate, which is equal to \( (1/W) \), where \( W \) is the spread bandwidth.

As can be seen by substituting (4.6) into (4.5), there is a linear relationship between time uncertainty and the average acquisition time. However, there are other variables and issues to consider. For example, a reduction in acquisition time is meaningless if the acquisition time is already trivial. Furthermore, a fairly realistic scenario must be considered in order to gain insight into the potential benefits of improved time and frequency accuracy and, ultimately, the potential tradeoff between more bandwidth efficient modulation and the cost in \( S/N \).

For the purposes of the analysis that follows, network entry is accomplished via a separate control channel. Control is performed out-of-band using a separate code common to all users. Network entry is assumed to be performed without prior knowledge of time uncertainty. This is open loop entry. Although open loop network entry may not be necessary for major nodes in a SATCOM network, which have access to PT&F, it may be of interest with respect to small terminals that lack access to PT&F.

The initial time uncertainty to be resolved is expected to be 100 \( \mu \)s, or less [11], but may be as high as several milliseconds, based on the analysis in section 4.2. Once the time uncertainty has been resolved on the control channel, it is also largely resolved for the purposes of acquisition on the data channel(s).

Consider a spread bandwidth be 204.8 MHz, as consistent with the example in Chapter 2. To start, QPSK modulation with \( r = \frac{1}{2} \) FEC is assumed on the control channel. The use of FEC on the control channel can improve the reliability of the control data, but it tends to increase the acquisition time, since coding reduces the effective \( S/N \) over the channel symbols.

A minimal 75 bps data rate (\( R \)) is assumed for control and signaling. A bandpass filter
with a one-sided bandwidth \( B \) of about 138 Hz is required to accommodate the
coded QPSK signal, assuming about \( \pm 100 \) Hz of residual frequency uncertainty (e.g.,
after Doppler resolution). A comparison of the performance given a simple 75 bps
control signal to the case of a 16 Kbps control signal, as consistent with ISDN
signaling rates, is a subject for future analysis.

The first step in CDMA network entry is acquisition of PN code lock. Based on the
methodology given in [2], the mean \( T_{av} \) can be evaluated for selected \( P_{fa} \) and \( P_{d} \),
based on the available post-detection \( S/N \) at the receiver, after losses due to
transmission, filtering, timing and frequency offset. The relationship between the pre
and post-detection \( S/N \) is

\[
\frac{S'}{N} = L \frac{S}{N}, \quad \text{where} \quad \frac{S}{N} = \frac{Rrke_k}{BN_0}.
\]

A loss of 2.5 dB is taken for worst case \( 1/4 \) chip timing misalignment, and 1 dB for
other pre-detection implementation losses. Details of these and other potential losses
are described in [2]. A plot of the mean PN acquisition time versus time uncertainty is
provided as Fig. 4.5. A pre-detection \( E_b/N_0 \) of 6.4 dB (i.e., \( S/N = 3.9 \) dB) is assumed,
as consistent with the use of QPSK with \( r = \frac{1}{2} \) FEC and operation at a BER of \( 10^{-8} \).

In a CDMA system, network entry may be divided into coarse acquisition, or code
lock, and fine acquisition or tracking [2]. In this case, coarse acquisition would be
accomplished via the control channel. Fine acquisition would be maintained
separately on each channel, which, in this concept, means each ATM VC. The total
time uncertainty to be resolved is dependent on uncertainties in range and system
time. The latter is typically the difference between "satellite" time and "terminal"
time. As can be seen from Fig. 4.4, the impact of range uncertainty on the mean
acquisition time tends to dominate over that of system time.

**Interpreting the results** - In order to help interpret the results, three example cases
will be considered. In case 1 both the hub and the remote or mobile terminal employ
good quartz oscillators.
**Fig. 4.4: Mean acquisition time vs. time uncertainty**

**Case 1 (quartz clocks)** - In the absence of GPS, position and range estimates may be derived from ephemeris data, which is updated on the control channel each time the user joins the network. Time-tagging of messages in the control channel could facilitate range estimation, resulting in improved accuracy and reduced time uncertainty. Dependent on the capabilities of the system, the position uncertainty could be anywhere from about 100 m, which is consistent with GPS, to not more than about 100 km (i.e., a rough estimate based on knowledge of the location of the base station). This places the contribution to the total time uncertainty, due to position inaccuracy, in the range of 33 ns to 333 μs.

Based on the concept described in Chapter 2, scheduled access would be a key mode of operation. Users may enter, or re-enter, the network after having been out for extended periods of time, during which, the user's receiver may be "off" to save power; in which case, the control channel would not be continuously monitored. In the event of re-entry 24 hours after having left the network and turned off the receiver, the total time uncertainty to be resolved may be as large as 3 to 10 ms (i.e., range, position and residual time uncertainty from all sources including clocks). In this case,
time interval error between clocks is the dominant source of time uncertainty to be resolved and, from Fig. 4.5, the acquisition time is expected to be very large.

Case 2 (GPS and UTC) - In this case, GPS is available at either the hub, or remote terminal, but not at both ends. A range accuracy of 100 m is assumed, along with a residual time uncertainty of 1 to 10 µs, from all sources (e.g., clocks), such as may occur given operation between a mobile terminal using GPS, which is based on UTC(USNO) and a hub using UTC (BIPM). From Fig. 4.5, the time required for coarse acquisition is estimated to be only a few seconds. Note that it would be necessary to multiply $T_{acq}$ by an additional factor of two, unless both the advance and delay uncertainties can be simultaneously searched [11]. From Fig. 4.5, it can also be seen that acquisition times of a minute, or less, are expected as long as the range uncertainty is less than 1 km and the residual time uncertainty is less than about 50 µs.

A mean acquisition time of a minute or less is probably satisfactory for scheduled use of the network, but certainly not for users wishing to re-join the network on an unscheduled basis. The results support the conclusion that scheduled users must go through coarse acquisition. When these users are finished, they leave the network and may turn their receivers "off". This may be viewed as analogous to the case of a scheduled corporate multicast transmission. If rapid re-entry is required, within seconds, then it may be necessary to leave the receiver "on" and maintain PN tracking on the control channel. These users do not leave the network. This may be viewed as analogous to a mobile user expecting to make, or receive, an unscheduled call.

Case 3 (GPS) - For small terminals that require fast acquisition, within seconds or less, the analysis indicates that GPS is needed at both the mobile and the hub (i.e. a range accuracy of 100 meters, and time accuracy to within a few nanoseconds [12]). Otherwise, a more sophisticated timing and synchronization plan would be required in order to implement the SATCOM CDMA network [11 and 13].

4.6. ATM-SATCOM delay variance analysis for multimedia component services

Delay variation, or jitter, may be divided into two types, which may affect the performance of the ATM-SATCOM system proposed in Chapter 2 in different ways.
The first type is delay variation that common to all of the ATM VC's over the SATCOM link and the second type is delay variation relative to components of a multimedia information service carried on separate ATM VC's. The focus is on the latter, which is unique to the proposal in Chapter 2.

**Delay variation common to all ATM VC's on an ATM-SATCOM link** - The results in section 4.2 indicate that a delay variation of several milliseconds is readily possible over a satellite link, without Doppler correction or pre-correction. In a typical SATCOM system, cyclical Doppler related delay variations would be accommodated in a Doppler buffer within the receiver [6]. In terms of QoS, the accumulated time uncertainty may contribute to the end-to-end delay and delay variation, but need not result in slips, deletions or loss of synchronization, as long as the buffer is adequately sized. With respect to the concept in Chapter 2, this form of delay variation is common to all ATM VC's over a given link and is not expected to pose any problems unique to the concept.

**Delay variation relative to the ATM VC's on an ATM-SATCOM link** - With respect to the concept described in Chapter 2, where the components of a multimedia service may be transmitted an ATM-SATCOM link using time shifted versions of a common parent PN code, relative delay variations experienced by different the individual PN codes could adversely affect performance. This is not a concern as long as the entire family of time shifted codes experience the same delay variation, at the same time, which they would, when transmitted simultaneously, over the identical path. However, any delay variation that occurs relative to one component service, but not another, could lead to loss of application synchronization. This is in addition to any QoS degradation due to delay and jitter common to all of the component services.

As an example, consider a multimedia information service that is divided into voice, data and video component services. If each component is transmitted separately, then delay and delay variation can occur between component services, as well as, end-to-end, as illustrated by Fig. 4.5. Relative delay variations that may be of little consequence between data and video component services, may be unacceptable between voice and video component services, where they may lead to a mismatch between the spoken word and the image.
Delay variation between multimedia component services would not be expected unless the multimedia service is devolved into its component parts and each part is transmitted separately. Nor would delay variation between multimedia component services be expected over a “bent pipe” satellite relay, where each component is subjected to the same delay and delay variation. However, relative delay variation could occur between multimedia component services, which have been transmitted separately, if the satellite employs on-board processing in a future ATM “switch-in-the-sky”. This is the result of queuing related delay variations inherent with ATM, which employs a separate queue(s) per service class. Note that this issue is not unique to the ATM “switch-in-the-sky”. In fact, it applies to any ATM switch, where the components of a multimedia service are passed separately; however, the “switch-in-the-sky” is assumed in the analysis that follows.
In lieu of a more detailed model of the switching elements on-board the satellite, the delay variation can be estimated based on standard techniques for modeling an M/D/1 queue [14-16]. The delay variance will be computed under different loading conditions and the probability that the variation exceeds a given threshold will be evaluated.

The delay variance is load dependent. Locally, the load is assumed to be from a single user, but in a satellite switched ATM system, the load may be from many simultaneous users and, therefore, more likely to be heavy and variable. From [14], the Laplace transform of the probability density function (PDF) for the delay of an M/G/1 queue is

\[ D(s) = \frac{s(1-p)M(s)}{[s - \lambda + \lambda M(s)]}. \]

For an M/D/1 queue, \( M(s) = e^{-sm} \), where \( m \) is the fixed packet length. Substituting \( s = \lambda (1 - z) \) into the equation above yields

\[ D(\lambda (1 - z)) = \frac{\lambda (1 - z)(1 - p)e^{-z(1 - z)m}}{[\lambda (1 - z) - \lambda + \lambda e^{-z(1 - z)m}].} \] (4.7)

From [14 and 15], the moments of the delay can be calculated by taking the derivative of equation (4.7) evaluated at \( z = 1 \) (i.e., or the Laplace transform of the delay evaluated \( s = 0 \))

\[ D(0)^{(n)} = (-1)^n D(z)^n. \]

From equation (4.7), given \( p = m\lambda \), the first moment of the delay is

\[ D(0)' = \bar{D} = m + \frac{mp}{2(1 - p)} \]
and the second moment of the delay is

\[
D(0)^2 = \bar{D}^2 = m^2 + \frac{m^2 p}{3(1 - p)} - \frac{mpD(0)}{(1 - p)}.
\]

The variance of the delay, or delay jitter, is given by [15]

\[
\sigma_d^2 = \bar{D}^2 - (\bar{D})^2.
\]

A plot of the delay variance under different loading conditions is provided as Fig. 4.6. From this figure, it can be seen that the variance at a given loading is larger for the case of low rate user access than for higher rate user access. This is viewed as analogous to the impact of phase noise at low data rates. Note that the delay variance curves shown in Fig. 4.6 apply to the case of a single multimedia SATCOM user. The case of multiple simultaneous users will be discussed next.

Using the variance, assuming that there may be many users and that the queue length is infinite, the probability of exceeding a relative delay threshold can be evaluated based on a Gaussian distribution as follows. Short term variations may be viewed as the standard deviation of a zero mean Gaussian random variable. As such, the mean time between timing variations exceeding a threshold magnitude \(S\) may be estimated by using the methodology for evaluating the mean time between "slips" of [17] and replacing \(\frac{\Delta t}{f}\) with \(\frac{\Delta t}{t}\), per equation (4.4).

A measure of performance is defined: the mean time between QoS "slips" (MTBQS), where a QoS slip is a delay variation that exceeds the acceptable threshold. The \(MTBQS\) can then be found by evaluating equation (4.8)

\[
MTBQS = \int_{-\infty}^{\infty} \frac{1}{S \sigma_D \sqrt{2\pi}} \frac{\Delta t}{t} \exp \left( -\frac{\Delta t^2}{2 \sigma_D^2} \right) \frac{\Delta t}{t} d\frac{\Delta t}{t},
\]  

(4.8)
Figure 4.6: Delay variance vs. traffic load for 64 Kbps, 512 Kbps and 2048 Kbps user access rates

where the total delay variance ($\sigma_D^2$) is the sum of the contributions of 1 to N simultaneous users accessing the satellite. If all of the users are independent, but identically distributed, then $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 \ldots \sigma_N^2$ [16], which reduces to $\sqrt{N}\sigma_D$. After integration, equation (4.8) becomes

$$MTBQS = \frac{S}{0.798\sigma_D}.$$  \hspace{1cm} (4.9)

Consider a delay variance of $\sigma_D^2 = 10^{-6}$ (seconds squared) and a QoS threshold slip size of $S = 100$ ms. From equation (4.9), the $MTBQS$ would be about once every 125 seconds. Based on Fig. 4.7 and equation (4.9), this could occur even at low traffic loads ($p \leq 0.1$), given 64 Kbps access rates. By comparison, a QoS slip of the same magnitude and average frequency of occurrence would only be expected at moderate to high traffic loads ($p \leq 0.7$), given 512 Kbps user access rates and only at very high
traffic loads \((p \leq 0.9)\), given 2048 Kbps user access rates.

The implications of the delay variation analysis may affect the design and performance of the system. However, the impact is subjective; it depends on the users perspective and may not be readily noticeable relative to the performance of low data rate video in general. Furthermore, recall that the lowest data rates in this concept are not intended to support video. They are primarily intended to support a limited subset of component services to small terminals.

Thus, the question of delay variation is most relevant to higher data rate access (i.e., at least a few hundred Kbps); in which case, the potential impact is much less. Additionally, the number of users accessing the satellite can affect the load and the delay variation. In which case, the load could be coordinated, at least where scheduled use is concerned.

The analysis of delay variation in this section focused on the potential impact of a processed satellite relative to a "bent pipe" relay; however, it also applies to the ground based switches. In Chapter 5, a similar issue will be examined with respect to the use of an ARQ protocol over terrestrial links for error recovery in a multicast diversity concept.

4.7. Summary and concluding remarks
A novel CDMA transceiver structure was presented in this chapter, as consistent with the implementation of the concept proposed in Chapter 2. The transceiver structure is designed to support communication of multimedia information services per ATM service class, where each service class is simultaneously transmitted via a separate CDMA code, in order to match the flexibility of ATM with the soft capacity limit of CDMA.

In principle, the design can be readily extended to the more general case where each ATM connection is matched to a separate CDMA code for transmission over the satellite link, as well as, to the case where each component of a multimedia service is also carried via a separate ATM VC – code, with the goal of enabling small terminals
to engage in subsets of a larger multimedia conference.

The proposed ATM-CDMA transceiver design requires a potentially large number of CDMA receivers and codes, and operation of CDMA to small terminals can impact acquisition times. The use of GPS was proposed to reduce the acquisition times given small terminals (low S/N), as well as, to facilitate the implementation of synchronous CDMA, such that the number of codes (and receivers) may be reduced by generating a family of time delayed codes from a parent.

Finally, transmission of a multimedia information service decomposed into it’s component services may lead to additional delay variations, for example, queuing given an ATM switch-in-the-sky. Based on the analysis in this chapter, such variations are expected to be within acceptable limits.
References

8. V. Slabinski, "Variation in range, range-rate and propagation delay and Doppler Shift a nearly geostationary satellite", Communications Satellite Corporation (Comsat), Washington D.C.
10. Application Note 52-2: Time keeping and frequency calibration, Hewlett Packard.
Chapter 5

Error correction and control for ATM-SATCOM links

5.1. Introduction

Error correction and control is very limited within the ATM protocol, but then ATM was not designed for operation over SATCOM links. The need for error correction and control is typically greater than over a satellite link than over a terrestrial F-O link. It is also dependent on the service, or at least the service class and, as already shown, there is more to efficiency than protocol overhead.

Three error correction and control techniques will be examined for improving the performance of ATM over SATCOM links: forward error correction (FEC), a selective repeat protocol (SRQ) and multicast automatic repeat request (ARQ) with diversity combining. The relative merit, shortfalls and potential role of each will be discussed. The error correction and control techniques considered in this paper are intended to be transparent to the ATM switch and augment the existing capabilities provided by commercial-off-the-shelf (COTS) satellite modems.

From a practical perspective, the use of a separate device between the ATM switch and the SATCOM modem is considered, in order to facilitate integration with existing equipment, with minimal change. Special error correction and control techniques for operation over satellite links would reside primarily within the ATM layer of an "ATM-SATCOM" integration device. However, the error control and correction capabilities could also be implemented within the ATM switch or the SATCOM modem. Additional performance may be possible by integrating specific techniques with the ATM switch and SATCOM modem, but this implies modification or replacement of existing equipment and not considered in this paper.

The purpose of error correction and control is to maintain the error rate within acceptable limits in terms of proper operation of the system, as well as, the end-to-end
QoS. Before examining potential error correction and control techniques for ATM over SATCOM, it is useful to consider the requirements for error correction and control imposed by operation of ATM over a SATCOM link. How sensitive is the operation of ATM over SATCOM to errors? To delay?

5.2. TCP/IP over ATM-SATCOM and the importance of error control
The performance of TCP/IP over ATM-SATCOM links is TCP limited. Specifically, the throughput per TCP connection it is limited by the acknowledgement window size and mechanism [1 and 2].

**Performance dependency on TCP window size** - Fig. 5.1 illustrates the dependency of performance on the size of the acknowledgement window. The results were obtained based on the methodology of [3] for evaluating the performance of TCP, which was modified for the case of an E1 2.048 Mbps ATM-SATCOM carrier per [1 and 2]. The bit error rate is assumed to be negligible.

The results clearly show the dependency on the transmit window size (W) in bytes. The required window size is determined by the product of the transmission rate and the roundtrip delay (i.e., satellite, plus end-to-end delay from any other sources). The transmit window must be large enough to store all of the transmitted data until it is acknowledged by the receiver; otherwise, the throughput may be degraded. The maximum available is TCP implementation dependent.

Given a roundtrip delay over the satellite link of about 0.5 seconds and a transmission rate of 2.048 Mbps, the required window size for continuous transmission without stopping to wait for an acknowledgement would be 1.024 Mbits. Thus, between the overhead of 20% and the fact that the window size is only 1/2 of the required 1.024 Mbits, the peak throughput would be only about 0.8 Mbps, per TCP connection.

Test and analysis indicate that the throughput of a TCP connection over an E1 2.048 Mbps ATM-SATCOM carrier is actually to about to about 0.8 Mbps due to a combination of protocol overhead (i.e., approximately 20%) and a maximum window size of 64 Kbytes, which is not sufficient to support a TCP connection at rates greater than 512 Kbits over a SATCOM link [1 and 2]. This is before consideration of errors.
Figure 5.1: Dependency of throughput on TCP window size E1 2.048 Mbps carrier, 9188 byte MTU, 0.25 second one-way delay

The combined affects of errors, delay and window size - Fig. 5.2 illustrates the combined affects of delay, errors and window size. Controlled measurements were made via a SATCOM delay and error simulator and modelled analytically [2]. As can be seen from Fig. 5.2, the analytical results are a close match with the measurements obtained via the simulator. In either case, the throughput, given a one-way delay of 0.25 seconds (i.e., roundtrip delay of 0.5 seconds) and a typical TCP implementation, with a window size limitation of 51 Kbytes per connection, would be about 750 Kbps. For additional details about simulated and on-air performance of ATM-SATCOM links refer to [1 and 2].

There are two key points to be discerned from Fig. 5.2. First, the throughput in Mbps drops off very rapidly when the probability of bit error exceeds a threshold of about $10^{-8}$. Second, as long as the window size is large enough and the error rate is low enough, the performance is limited by the capacity of the ATM-SATCOM carrier, less overhead, as opposed to the TCP window. The second point should be clear. It is much less clear as to what is sensitive to the error rate and why.
Figure 5.2: The combined effects of errors, delay and window size over an E1 2.048 Mbps ATM-SATCOM carrier

TCP acknowledgement mechanism - One would not ordinarily expect the 53 byte (424 bit cell) to be catastrophically sensitive to an error rate of 1 bit in $10^6$. Nor would one expect the 8192 byte (65536 bit) maximum transmission unit (MTU), which is roughly analogous to an IP packet, to be. However, the 51 Kbyte TCP transmit window (over 400 Kbits) might be. For example, on average, one out of every 2 to 3 windows would be negatively acknowledged, necessitating retransmission, at an error rate of $10^{-6}$. Furthermore, it would not matter whether it was a single bit error, or a burst of 51 Kbytes, the entire window's worth of data would need to be retransmitted.

Thus, the performance of TCP/IP over the long delay satellite link is TCP limited and sensitive to delay, errors and the combined effects of delay and errors. If operation of TCP/IP over ATM-SATCOM is to be effective, especially at rates higher than 2.048 Mbps, then additional measures of error control and correction may be required, if only to ease the burden on TCP. By comparison, UDP is much less sensitive to errors over the SATCOM channel and insensitive to the delay. It is also unreliable by definition.
Figure 5.3: Error control concept for ATM over SATCOM

5.3. Concept for error control tailored to the ATM service class

An illustration of a concept for error control and correction over ATM-SATCOM links is provided as Fig. 5.3. The illustration depicts the use of an E1 carrier with the G.804 PLCP. Other rates, interfaces and PLCPs are possible. Separate queues per service class are required. The concept allows for a different error correction scheme per service type. For the purposes of the discussion in this chapter, the four ATM service classes have been associated with the four AALs: class 1 = AAL 1, class 2 = AAL 2, class 3 = AAL 3/4 and class 4 = AAL 5. This is only an approximation made to simplify the discussion.

Tailoring of the error control and correction technique to the service type, as opposed to applying the same technique to all data, allows the overhead associated with error control to be reduced. This is because the overhead associated with error control and correction is effectively averaged, in accordance with the traffic profile, over component services on a virtual circuit, the services per SATCOM link and the links
over the satellite. The implication is that services which require more power efficient FEC may do so without necessitating additional overhead for all services.

5.4. A CPCs based truncated SRQ protocol for ATM AAL5

The use of a SRQ protocol is intended to support data communication. This includes broadcast data, file transfer and signaling services running over TCP/IP through ATM AAL 5. TCP/IP over ATM AAL 5 may also be considered to support video. The proposed SRQ protocol is not intended to guarantee error free communication. End-to-end reliability remains the responsibility of TCP. Rather, the SRQ is intended to reduce errors seen by the higher level TCP protocol. In this way, it eases the burden on TCP.

TCP is a "sliding window" protocol. It more sensitive to the combined effects of delay and errors than a SRQ protocol because it repeats an entire "windows" worth of data in the event of an error. The amount of data to be repeated in the event of an error can quickly become large as the combination of data rate and delay increases. Although there may be an SRQ capability in a future version of TCP, none exists at present.

The primary benefit of a SRQ protocol is that only the protocol data unit (PDU) in error is actually retransmitted. This saves bandwidth over the satellite link when errors occur as the channel begins to degrade. Additionally, SRQ protocols are less sensitive to delay than sliding window protocols [3]. The disadvantage of SRQ protocols is that they can be complicated to implement and memory intensive.

An ideal SRQ protocol would continue to retry transmitting an erroneous PDU until it was finally successful. It can be shown that the number of attempts required before success is geometrically distributed, with a mean dependent on the mean PDU error rate [3]. At high error rates, the memory required to implement an SRQ protocol can easily become very large and the delay may be increasingly long and variable, as the protocol repeatedly tries to send the PDU. The delay variability makes SRQ protocols unsuitable for constant bit rate (CBR) services.

The SRQ protocol examined in this paper is limited to a maximum of 3 retransmissions. The memory requirement is 3 times the round-trip delay, in the absence
of delay from other sources. Three attempts will be shown to be sufficient to significantly improve the overall performance.

The proposed SRQ protocol would be implemented in the CS portion of the ATM AAL sublayer. Error correction is consistent with the intended use of the CS sublayer of AAL 5 [4]. An illustration of the proposed CPCS-SRQ protocol frame is provided as Fig. 5.4, from which it can be seen that the protocol would be implemented using only 2 bytes out of the CPCS-PDU payload. The CPCS-PDU payload would be from 1 to a maximum of 65353 bytes instead of 1 to 65355 bytes. Otherwise, the CPCS-PDU remains unchanged.

The proposed SRQ protocol would rely on the existing CRC of the CPCS-PDU to determine if retransmission is necessary. Padding is employed to round off the CPCS-PDU to a multiple of 48 bytes, which is the ATM cell payload. Ideally the size of the CPCS-PDU payload is matched to the size of the higher layer TCP/IP packets, in order to minimize segmentation.

Testing suggests that an IP packet of 1024 bytes (8192 bits) is a reasonable value for operation over a satellite link [1]. Regardless of the exact value, the point is that the CPCS-PDU is expected to be very large compared to the 2 bytes required to implement the SRQ protocol. Consequently, the SRQ protocol has a minimal impact on the total overhead.

![Figure 5.4: ATM CPCS-SRQ Protocol Frame](image-url)
The 2 byte SRQ header may be allocated as follows: 2 bits for control, 2 bits for the number of retries \((M)\), 8 bits for CPCS-PDU numbering and 4 spare bits, which may also be used for numbering. Note that a retry is initiated when the 32-bit CPCS CRC detects an error. The performance of the SRQ protocol can be evaluated as illustrated by Fig. 5.5.

Let the underlying bit error rate be given as \(P_b\). The CPCS-PDU error rate per independent trial \(P_e\) is

\[
P_e = 1 - (1 - P_b)^N,
\]

where \(N\) is the size in bits of the CPCS-PDU. The probability that the CPCS-PDU is delivered free of error is then \(P_c = 1 - P_e\). The efficiency of the SRQ protocol may be evaluated from [3]

\[
SRQ_{\text{efficiency}} = \frac{1}{L}
\]

where \(L\) = Average number of CPCS - PDU transmissions

\[
L = \sum_{n=0}^{M-1} nP_c P_e^{n-1} + MP_e^{M-1},
\]

For \(M = 4\),

\[
L = 1P_c + 2P_c P_e + 3P_c P_e^2 + 4P_e^3
\]

Figure 5.5: ATM CPCS-SRQ protocol
The average probability of successful delivery of a CPCS-PDU may be found from

\[ P_{\text{success}} = \sum_{n=0}^{M-1} P_e P_n, \]

and the probability that the CPCS-PDU is not correctly received after \( M \) tries is

\[ P_{\text{fail}} = 1 - P_{\text{success}}. \]

This is the uncorrected CPCS-PDU error rate as viewed by higher layers (e.g., TCP/IP). The probability of failure to correctly deliver the CPCS-PDU can be translated into an average residual bit error rate; however, the performance of higher layer protocols is dependent on \( P_{\text{fail}} \) regardless of whether the failure is due to a single bit error within the CPCS-PDU, or a burst of errors. For example, the failure to deliver any CPCS-PDU correctly to TCP will necessitate the retransmission over the satellite of a TCP transmit window worth of data.

As can be seen from the preceding analysis, the performance of the SRQ protocol is not dependent on the round-trip delay, as it was for TCP. This is an important feature with respect to operation over SATCOM links. However, delay does factor into the memory required to implement the protocol.

The memory must be sufficient to save the CPCS-PDU's for a period of time equal to \( M-1 \) times the round-trip delay. For the case of \( M_{\text{max}} = 4 \), this allows for the possibility that a CPCS-PDU is not received correctly by the receiver on the first attempt, or either the first or second retransmission. Following the third and final retransmission (i.e., \( M_{\text{max}}-1 \)), a CPCS-PDU would be dropped by the SRQ, in which case, recovery is via TCP.

From Fig. 5.6, it can be seen that the larger the CPCS-SRQ-PDU, the more efficient it is in terms of the fraction of the data passed to higher layers when the error rate is low. However, the opposite is true when the error rate is high, as larger PDUs are more sensitive to bit errors.
Figure 5.6: CPCS-SRQ Protocol efficiency

For the case of an 8192 bit CPCS-SRQ-PDU, the greatest benefits occur when the input bit error rate is greater than about $10^{-5}$ and there is little additional benefit when the input bit error rate is less than about $10^{-6}$.

Fig. 5.7 illustrates the performance benefit of the truncated CPCS-SRQ protocol for the case of a CPCS-PDU of 8192 bits. The maximum number of attempts to deliver the data is $M_{\text{max}}$, which means that $M \in \{1,2,3,4\}$. The benefit as expressed in terms of the residual probability of bit error is nearly 4 orders of magnitude, at an input bit error rate of $10^{-5}$. As a practical reference, consider the case of TCP/IP over ATM AAL 5, operating via a 2.048 Mbps E1 carrier over a satellite link.

As shown in [1], the throughput of TCP/IP over ATM AAL 5 degrades from over 750 Kbps (TCP window limited) to less than 10 Kbps as the bit error rate increases from $10^{-8}$ to $10^{-6}$ and becomes virtually negligible at a bit error rate of $10^{-5}$. With the SRQ protocol ($M_{\text{max}} = 4$), the residual probability of bit error for an input error rate of $10^{-5}$ would be $4.7 \cdot 10^{-9}$. In which case, as can be seen from Fig. 5.6, there would be no degradation in the throughput at the ATM AAL 5 CPCS layer, apart from 0.2% overhead to implement the CPCS SRQ header.
5.5. **FEC options for ATM over SATCOM**

FEC provides another means of improving the performance of ATM over SATCOM or wireless links. The improvement can be substantial. A possible disadvantage is that the overhead required to implement FEC is typically greater than required to implement an SRQ protocol, and the overhead to implement the ATM protocol is already large. An advantage relative to the SRQ protocol is that the delay tends to be less and, at least in the case of block codes, the delay is fixed in duration. Low delay and delay variation are particularly desirable with respect to CBR services, which is why FEC is more suitable for use with voice or other real-time services than a SRQ protocol.

In this section, high rate block codes are considered as a means of improving communication of ATM over SATCOM links. A block code could be implemented within the ATM switch, the SATCOM modem or in a separate external box, which is the preferred approach. The block code is effectively concatenated with existing FEC within the SATCOM modem. Block codes operate on hard decisions about the
channel bits. Soft decision decoding is more efficient than hard decision decoding; however, hard decision decoding is the most practical, in this case, since decisions have already been made within the SATCOM modem on the inner channel bits and the outer code may be implemented externally.

The block code may be implemented within the physical layer or the ATM layer. Physical layer FEC can provide additional protection for the ATM header not afforded by the codes implemented at the ATM layer, or by CPCS-SRQ protocol described earlier in this paper. On the other hand, ATM layer FEC can be tailored to the service class, or even to component services. This is potentially more efficient. BCH codes and shortened block codes, matched to the ATM protocol, will be examined.

As bit errors over a SATCOM link may come in bursts either due to channel effects or the operation of the SATCOM modem (e.g., codec or differential encoding), burst error correction through the use of interleaved BCH codes and Reed-Solomon (RS) codes will be considered. Burst error correction is expected to be most applicable for codes implemented at either the physical layer, or the ATM layer, in support of information services other than those running over TCP, where the impact of error bursts tend to be masked by the affects of the sliding window retransmission mechanism.

The special case of a high rate BCH code mapped into spare bits in the popular E1 2.048 Mbps G.704 frame will also be examined. Although not as effective as lower rate codes, any improvement in performance is "free" from the perspective of the cost in terms of overhead. This technique applies to the E1 and higher order framing structures constructed from the E1 carrier. As such, it may be useful with respect to the integration of ISDN and circuit switched voice with ATM. Additionally, it may be applicable to the Intelsat IBS carrier and Drop-Insert interface, due to the similarity in frame structure.

A cross section of BCH and RS block codes are considered. A summary is provided in Table 5.1. For the \((n,k)\) BCH codes, \(n\) is the number of bits in a block and \(k\) is the number of information bits per block. The code rate is defined as \(\text{rate} = k / n\).
Table 5.1: Summary of Codes

<table>
<thead>
<tr>
<th>Code</th>
<th>(n,k)</th>
<th>(\lambda)</th>
<th>OH (%)</th>
<th>t</th>
<th>burst (bits)</th>
<th>Delay (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (BCH)</td>
<td>(1023,1013)</td>
<td>1</td>
<td>0.98</td>
<td>1</td>
<td>1</td>
<td>1024</td>
</tr>
<tr>
<td>2 (BCH)</td>
<td>(63,53)</td>
<td>1</td>
<td>18.9</td>
<td>1</td>
<td>1</td>
<td>64</td>
</tr>
<tr>
<td>3 (short BCH)</td>
<td>(443,423)</td>
<td>1</td>
<td>4.7</td>
<td>2</td>
<td>1</td>
<td>581</td>
</tr>
<tr>
<td>4 (BCH)</td>
<td>(511,457)</td>
<td>1</td>
<td>11.8</td>
<td>6</td>
<td>1</td>
<td>512</td>
</tr>
<tr>
<td>5 (BCH)</td>
<td>(255,223)</td>
<td>1</td>
<td>14.3</td>
<td>4</td>
<td>1</td>
<td>256</td>
</tr>
<tr>
<td>6 (BCH)</td>
<td>(2046,2026)</td>
<td>2</td>
<td>0.98</td>
<td>1</td>
<td>2</td>
<td>2048</td>
</tr>
<tr>
<td>7 (BCH)</td>
<td>(4092,4052)</td>
<td>4</td>
<td>0.98</td>
<td>1</td>
<td>4</td>
<td>4096</td>
</tr>
<tr>
<td>8 (short BCH)</td>
<td>(886,846)</td>
<td>2</td>
<td>4.7</td>
<td>2</td>
<td>2</td>
<td>2048</td>
</tr>
<tr>
<td>9 (RS)</td>
<td>(255,223)</td>
<td>N/A</td>
<td>14.3</td>
<td>16</td>
<td>8</td>
<td>2048</td>
</tr>
<tr>
<td>10 (short BCH)</td>
<td>(61,53)</td>
<td>N/A</td>
<td>15.1</td>
<td>1</td>
<td>8</td>
<td>3600</td>
</tr>
</tbody>
</table>

In the case of the RS codes, \(n\) represents the number of symbols per block and \(k\) represents the number of information symbols per block. Each symbol contains \(Q = 2^m\) bits, where \(m\) represents the burst error correcting capability of the RS code, which is related to the block size by the equation \(m = \log_2(n)\). Ordinary BCH codes do not correct burst errors (\(\lambda = 1\)); however, interleaved BCH codes [5] can correct bursts of errors (\(\lambda > 1\)). The BCH codes will be examined first.

5.5.1. Non burst error correcting block codes

The performance of a \(t\) error correcting BCH code can be evaluated from [5]

\[
P_{h_{aw}} = \sum_{i=t+1}^{n} (1 - P_h)^{m-i} P_h^i.
\]

A summary of the performance of the non burst error correcting block codes is provided by Fig. 5.8. The performance is expressed in terms of an I/O characteristic as it was for the SRQ protocol. The baseline depicts the performance without FEC (i.e., uncoded).

Testing [1 and 2] indicates that the performance of TCP/IP over ATM based systems deteriorates very quickly at error rates above \(10^{-8}\). The (1023,1013) BCH code begins to be effective at an input error rate of about \(10^{-6}\). Although not nearly as effective as
the lower rate BCH codes, even this high rate code can reduce the bit error rate seen by higher layers by more than an order of magnitude, at an input error rate of $10^{-7}$, at a cost of only 1% overhead.

The best of the BCH codes of Fig. 5.8 is code 4, which is the 6 error correcting (511,451) code. An improvement of more than 4 orders of magnitude is achieved, at an input bit error rate of $7 \cdot 10^{-4}$, relative to the desired output error rate of $10^{-8}$, at a cost of 11.7% overhead.

5.5.2. Burst error correcting block codes
Burst error correction is possible with BCH codes by constructing an interleaved code [5]. An interleaved BCH code may be constructed from a BCH code as illustrated by Fig. 5.9.

The burst error correction capability is equal to the interleaving degree $\lambda$. Alternatively, RS codes are powerful, burst error correcting codes, but they do not perform as well at low values of $E_b/N_0$.

![Figure 5.8: Non burst error correcting block code performance](image-url)
Figure 5.9: Illustration of a (λ, n, k) interleaved BCH code

Fig. 5.10 summarizes the performance of the burst error correcting codes defined in Table 5.1. Code 6 is an interleaved BCH code constructed from Code 1, the (1023,1013) BCH code, with λ = 2. Code 7 is also constructed from Code 1, but with an interleaving degree of λ = 4. Code 8 is a shortened BCH code constructed from Code 3. Code 9 is a 16 error correcting (255,223) RS code with a burst error correction capability of 8 bits. Code 8 is a shortened RS code constructed from the (255,223) RS code, which is also capable of correcting 16 errors of burst length 8 bits.

The best code, from the perspective of the lowest output bit error rate for a given input bit error rate, was code 9. But he overhead with this code is 12.6%, which comes off of the peak throughput achievable at higher layers. Dependent on the actual burst error correction requirements, code 8 may be of interest. This interleaved double error correcting BCH code is less effective than code 8 or code 10, but it is matched to the cell size and the overhead is only 4.5%. Code 10 is a short RS code, which is also matched to the cell size; however, the overhead is 13.1%.

**FEC matched to the spare bits in the E1 carrier** - The (1023,1013) BCH code and the interleaved codes constructed from it were the least effective of the codes examined from the perspective of the I/O performance.
Figure 5.10: Burst error correcting block code performance

However, this class of codes has the special quality that they can be implemented using only the spare bits of the E1 G.704 frame, or others framing structures based on the E1 (e.g., an 8.448 Mbps carrier with G.704 framing or Intelsat IBS framing).

This means that any coding gain, however limited, is essentially free, with respect to the cost in overhead. An illustration of how the (4092,4052) interleaved BCH code (i.e. $\lambda = 4$) can be mapped into the E1 CRC-4 Extended Super Frame (ESF) is provided by Fig. 5.11.

The mapping occurs over a 16 frame multiframe [6]. The 40 parity check bits of the code are mapped 1-to-1 into the 40 spare bits of the ESF, protecting 4052 information bits, which is sufficient to accommodate the 3840 bit (16 x 30 byte) payload in the ESF. Like other physical layer coding options, it applies to all traffic on the link. An illustration of the encoder is provided as Fig. 5.12.
<table>
<thead>
<tr>
<th>Submultiframe</th>
<th>Frame</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>M I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>P0</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>P4</td>
<td>P5</td>
<td>P6</td>
<td>P7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>P8</td>
<td>P9</td>
<td>P10</td>
<td>P11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>P12</td>
<td>P13</td>
<td>P14</td>
<td>P15</td>
</tr>
<tr>
<td>F II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P16</td>
<td>P17</td>
<td>P18</td>
<td>P19</td>
<td>P20</td>
<td>P21</td>
<td>P22</td>
<td>P23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P24</td>
<td>P25</td>
<td>P26</td>
<td>P27</td>
<td>P28</td>
<td>P29</td>
<td>P30</td>
<td>P31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P32</td>
<td>P33</td>
<td>P34</td>
<td>P35</td>
<td>P36</td>
<td>P37</td>
<td>P38</td>
<td>P39</td>
</tr>
</tbody>
</table>

E = CRC-4 error indication bits, C1 = CRC-4 bits, A = remote alarm indication bits, 
P = one of 40 parity check bits in the (λ, 1023, λ, 1013) code, where λ = 4.

**Figure 5.11:** Mapping of the tailored BCH code into the E1 CRC-4 ESF

![Diagram](attachment://diagram.png)

\[ g(x) = x^{10} + x^3 + 1 \]

**Figure 5.12:** (4095, 4055) BCH code encoder for use with the E1 CRC-4 ESF

5.6. Multicast with ARQ diversity for local error recovery

From a SATCOM perspective, diversity is most commonly employed to improve link availability and reduce power requirements, for example, in terms of rain margin. However, under the appropriate channel conditions, diversity is also a powerful error correction and control technique.

Broadcast or multicast of data via ATM over SATCOM may be employed to provide a diversity gain. There may be gain due to spatial separation where outages due to rain
are concerned over EHF, but there is also the potential for diversity gain where the information is transmitted to multiple ground terminals. A key assumption is that the receiver is noise limited on the downlink. This is typical of small terminals and, to the extent that the AWGN process is independent from one terminal to another, diversity gain is possible.

An examination of the link budget in Table 5.2 reveals that the forward link from the large terminal to the small terminal is limited by the receiver noise in the small terminal. The limitation is expected to be more pronounced given smaller terminals on the receive end and implementation may not be simple, additional processing may be required, and this is subject for follow-on work.

In order to realize a diversity gain in this context, multiple ground terminals must be interconnected. An illustration is provided as Fig. 5.13. Interconnection via a combination of the Internet and phone lines is shown. Other possibilities exist. Since the primary purpose of the interconnection is to support local error recovery, as opposed to recovery over the satellite, operation at very low data rates over the terrestrial interconnections is conceivable, as long as the need for retransmission remains infrequent.

The aforementioned diversity technique applies most directly to multicast data transmissions over ATM (i.e., TCP/IP over ATM AAL 5), but may have broader application dependent on the roundtrip delay required for local error recovery between ground terminals receiving the multicast transmission. The ability to rely on a local ground based neighbor for limited error recovery relaxes the need for error recovery over the satellite. As already shown, this is important because of TCP limitations over long delay-high data rate links. It is also important since error recovery via local terrestrial links may be more desirable for some information services than recovery over the satellite link in terms of the delay.

Local error recovery could be implemented using an automatic repeat request (ARQ) protocol similar to the SRQ protocol previously described. Upon detection of an error in the CPCS-PDU CRC, a one time ARQ for error recovery would be multicast to "terrestrial neighbors" in the "SATCOM multicast". In the special case of fixed rate
access using standard E1 2.048 Mbps carriers, the ARQ protocol could be implemented in the spare bits of the E1 carrier, like the (4095, 4055) block code. Although inflexible in terms of access rate, this option adds no additional overhead and could help protect the cell header, as well as, the CPCS-PDU.

The concept is probably best suited to regular scheduled use of SATCOM, which is why virtual paths (VP) are shown in Fig. 5.13. However, it is neither constrained to pre-scheduled transmissions, nor to VPs, and could be dynamic as long as terrestrial interconnectivity can be counted on between the SGTs.

The diversity gain may be evaluated as follows. Given N links, where each is independently limited by noise in the receiver, diversity gain may be estimated using the simple relationship

\[ P_e = \left[ 1 - (1 - P_s)^{\text{MTU}} \right]^N. \]

A summary of the potential performance benefit is provided as Fig. 5.14. The benefit was evaluated assuming a nominal MTU of 8192 bits (1024 bytes), where the MTU may be viewed as essentially an IP packet and it is assumed that the IP packet and the CPCS-PDU are matched.

Figure 5.13: ATM-SATCOM multicast with ARQ diversity ARQ for local low rate recovery of broadcast/multicast data
A reduction of the CPCS-PDU error rate of several orders of magnitude is possible with relatively few interconnected SGTs. Reducing the CPCS-PDU error rate has end-to-end significance, since a single erroneous CPCS-PDU passed to TCP/IP will necessitate retransmission over the satellite of a potentially much larger block of data. Note that the multicast gain is actually twofold: a) an $N$ level diversity gain dependent on the number of sites participating in the multicast and b) a reduction of the data rate over the satellite by up to a factor of $N$.

5.7. ARQ protocol delay analysis

In the previous section, the potential performance benefit of multicast diversity was examined and shown to be substantial. The concept relies on a terrestrial ARQ protocol for error recovery. In this section, the issue of queuing delay is examined.

The analysis is similar to the one performed in Chapter 4, with respect to queuing delay and delay variation over the ATM-SATCOM link, except that the focus in this section is on evaluation of the mean queuing delay, under different traffic loads and error rates, in order to verify the suitability of the terrestrial ARQ protocol, from the perspective of the end-to-end delay.

![Figure 5.14: The performance benefit with diversity](image-url)
As should be clear from the discussion in the preceding chapters, the transmit queuing delay can quickly become large under heavy loads and may exceed the nominal one-way propagation delay of 250 ms over a satellite link. Furthermore, the issue of queuing delay tends to worsen as the extent of the network increases, since delay may be experienced at multiple nodes. This is one of the reasons why ATM cell relay is advocated throughout the network, as opposed to conversion and re-conversion from one protocol to another, which typically necessitates additional buffering, leading to greater delay, which can adversely affect the performance of the ATM protocol.

As in Chapter 4, the analysis is based on asynchronous time division multiplexing (ATDM). In ATDM, all of the traffic from all of the users is multiplexed onto the same line; however, the full bandwidth is available to any given user, when the line is not busy. In the simplest case, traffic generated by multimedia information services, from one or more users, could be viewed as being multiplexed onto a single line for transmission. For example, data from various sources could be queued by service type for broadcast over the same VP, using a single PN code. Alternatively, traffic from various multimedia information services and component services could be queued by service type for transmission using a unique VC and PN code, as discussed in Chapter 2. In either case though, queuing is assumed to be by service type, as consistent with the ATM protocol.

If a protocol data unit (PDU) contains one or more errors, its contribution to the throughput is negated and retransmission of the information may be necessary, dependent on the information service or application. Retransmission of data reduces the throughput. Bit errors are the most common source of errors in a PDU, but errors can also occur due to clock slips, loss of synchronization and other problems. Only bit errors are considered.

Delay, such as incurred over a SATCOM link, can also have a significant impact on throughput. The longer the delay, the more information must be stored before acknowledgement and, if retransmission of data is required, the reduction in throughput is larger. A terrestrial ARQ is favored over SATCOM, in this regard, due to shorter delay. The impacts of errors and delay on throughput are protocol dependent. In the analysis that follows, the continuous ARQ scheme of [7] is
assumed, and only negative acknowledgements (NACKs) are considered.

The mean queue length in PDUs was selected as the basis for the performance comparison. A round-trip delay of 250 ms was assumed, which is typical of satellite communication links. A simple Poisson arrival process is considered, with one message per PDU. From [12], the generating function for the number of newly arrived messages in an interdeparture interval \((T)\) is

\[
A(z) = \frac{(1-p_e)\exp(-\lambda T (1-z))}{1-p_e \exp(-\lambda T (R+I)(1-z))}
\]

where \(M(z) = z\) for a simple Poisson arrival process, \(T\) is the interdeparture interval in seconds, or PDU duration adjusted for overhead (e.g., the ATM header), \(R\) is the round trip delay in PDUs, \(\lambda\) is the PDU arrival rate in PDUs per second and \(p_e\) is the probability that a PDU is in error.

The number of retransmissions is modeled using a geometric distribution, with a mean and variance dependent on \(p_e\). For an ATDM system, the mean number of PDUs waiting to be transmitted \((n_q)\) may be found from [12]

\[
n_q = \frac{(1-p)A(z, T)'}{1-A(z, T)'} + \frac{A(z, T)''}{2 \left[ 1-A(z, T)'ight]}.
\]

where \(p\) is the traffic intensity, which is the number of PDUs arriving in \(T\) seconds.

The mean number of PDUs arriving in \(T\) seconds is found by taking the derivative of \(A(z)\), with \(z\) set equal to 1 [12]. From Appendix 5A, the resulting expression is

\[
A'(z, T)_{z=1} = \frac{\lambda T [1-p_e R]}{(1-p_e)}
\]

Similarly, the mean squared number of arrivals is found by taking the second
derivative of $A(z)$, with $z = 1$

$$A''(z, T) = \left(\frac{T^2}{2 + R^2 + p_e R(1 + p_e R)}\right)$$

When both the round-trip delay and the error rate are zero, $A''(z, T)$ reduces to

$$A''(1, T) = 3(\lambda T)^2.$$ 

Clearly, the delay and delay variance are dependent on the traffic load (arrivals), the round-trip delay and the error rate. When the traffic load increases, so will the delay variance. Note that this is different than the analysis in Chapter 4, which was dependent on the load by service class, but independent of propagation delay and error rate.

The results of the ARQ protocol delay analysis are provided by Fig. 5.15. Three sample cases, corresponding to different operational scenario’s are considered: case 1 ($p_e = 10^{-5}$, roundtrip delay = 0.5 s), case 2 ($p_e = 10^{-3}$, roundtrip delay = 0.01 s) and case 3 ($p_e = 10^{-5}$, roundtrip delay = 0.01 s). A data rate of 2048 Kbps was assumed along with an average CPCS-PDU of 48 bytes. A small CPCS-PDUs was selected in order to upper bound the ARQ delay, which is more sensitive to the number of PDUs, than to their size.

With a roundtrip delay of 0.5 s, case 1 represents recovery over the satellite. From case 1, it can be seen that the queuing delay can quickly become large, possibly several seconds even under relatively light traffic loads, such as $p = 0.2$. By comparison, the mean queuing delay is generally much less than 100 ms, given terrestrial recovery with a roundtrip propagation delay of not more than 10 ms, even under high traffic loads (i.e., $p = 0.9$) and error rates, as can be seen from case 2 and case 3. A mean queuing delay of 100 ms, or less, is commonly acceptable for data and may also be acceptable for voice, while a delay of several seconds is certainly not. The results substantiate the viability of terrestrial recovery, and eliminate recovery over the satellite link from further consideration.
Figure 5.15: ARQ/multicast-diversity delay

An ARQ-multicast/diversity concept was proposed to improve performance when multicasting is considered to multiple, small, interconnected terminals. Diversity may be used in combination with the SRQ protocol and the FEC alternatives examined. As with the SRQ protocol, this technique is most directly applicable to data and possibly video services. Since error recovery is "local" and delay is much less than over the satellite link, this technique has potentially broader application to information services than the SRQ protocol.

5.8. Summary and concluding remarks
In this chapter, error control for ATM-SATCOM links was examined. The need for error control and correction is typically greater over SATCOM links, than over the terrestrial fiber-optic links for which ATM was conceived. A novel concept for matching the error control to the ATM service class was proposed., with the possibility of tailoring error control per ATM VC – CDMA code. The goal is improve overall SATCOM efficiency by using only the error control necessary to support each ATM service class, as a minimum, and ideally to be able to tailor the error control to
the individual ATM connections corresponding to a multimedia service, or even components of a multimedia service.

The concept calls for a layered approach to error control and three basic techniques were examined: a truncated SRQ protocol for data services, FEC for voice, data and video services and a multicast/ARQ protocol for local error recovery for data and possibly video, as opposed to recovery over the satellite.

The truncated SRQ protocol is novel from the perspective of it's proposed implementation in the CPCS sublayer of AAL 5. The FEC techniques are novel, where tailored to either the ATM protocol, or ATM-SATCOM carrier, through the use of shortened codes, interleaved codes and mapping to the spare bits of the standard 2.048 Mbps E1 carrier. The multicast/ARQ protocol is a novel concept in itself that attempts to treat the problem of ATM-SATCOM error control at a higher level, as part of an integrated solution tailored to, for example, a niche application.

The proposed multicast/ARQ diversity technique combines the multicast capability of ATM, with the wide area coverage of SATCOM and the tendency for the forward link from larger to smaller SATCOM terminals to be limited by noise in the receiver of the smaller terminal, in order to achieve improved power efficiency through diversity, improved bandwidth efficiency, as seen by the satellite, by reducing the need for retransmission over the satellite link, and improved QoS end-to-end, in terms of lower error rates and reduced delay, by virtue of local error recovery.

An important issue concerning the viability of the ARQ diversity protocol, especially with regards to the separate transmission of individual multimedia component services concerns delay due to queuing. Analysis in this chapter supports the conclusion that local error recovery is viable and preferable to recovery over the satellite.

Each of the techniques examined in this chapter have been shown to individually improve performance. In the next chapter, various combinations of techniques will be overlayed and the overall efficiency from a SATCOM perspective will be examined in conjunction with the modulation and codes considered in Chapter 3.
References


2. J. Farserotu and A. Tu, "TCP/IP over Low Rate ATM-SATCOM Links", IEEE MILCOM '96, McLean Va., USA.


APPENDIX 5A: DERIVATIVE OF THE GENERATING FUNCTION

Let $A(z)$ be the generating function for the number of newly arrived data units in an interdeparture interval

$$A(z) = \frac{(1 - P_e) e^{-\lambda T[1-M(z)]}}{1 - P_e e^{-\lambda T[R+1][1-M(z)]}}, \text{ where } M(z) = z.$$ 

The derivative of the generating function may be found from

$$A'(z) = \frac{f'(z)g(z) - g'(z)f(z)}{g(z)^2}, \text{ where }$$

$$f(z) = (1 - P_e) e^{-\lambda T[1-z]},$$

$$f'(z) = (1 - P_e) \lambda Te^{-\lambda T[1-z]},$$

$$g(z) = 1 - P_e e^{-\lambda T(R+1)[1-z]} \text{ and }$$

$$g'(z) = -P_e (\lambda T(R+1))e^{-\lambda T(R+1)[1-z]}.$$ 

$$A'(z) = \frac{f'(z)g(z) - g'(z)f(z)}{g(z)^2}$$

Taking the derivative

$$A'(z) = \frac{\left((1 - P_e)(\lambda T)e^{-\lambda T[1-z]}\right)[1 - P_e e^{-\lambda T(R+1)[1-z]}] + P_e (\lambda T)(R+1)e^{-\lambda T(R+1)[1-z]}(1 - P_e) e^{-\lambda T[1-z]}}{\left[1 - P_e e^{-\lambda T(R+1)[1-z]}\right]^2}$$

simplifying,

$$A'(z) = \frac{(1 - P_e)(\lambda T)e^{-\lambda T[1-z]} - P_e e^{-\lambda T(R+2)[1-z]} + P_e (R+1)e^{-\lambda T(R+2)[1-z]}}{\left[1 - P_e e^{-\lambda T(R+1)[1-z]}\right]^2}$$

and evaluating at $z = 1$, results in

$$A(Z)_{z=1} = \frac{(1 - P_e)(\lambda T)[1 - P_e + P_e (R+1)]}{(1 - P_e)^2} = \frac{\lambda T[1 - P_e R]}{1 - P_e}.$$
Chapter 6

SATCOM efficiency

6.1 Introduction
The primary measures of efficiency considered thus far are capacity in Erlangs and throughput in bits per second. In Chapter 2, the overall system capacity was evaluated in terms of the Erlangs of traffic that could be supported. The end-to-end throughput per connection was discussed relative to TCP/IP over ATM in Chapter 5. The potential bandwidth and power efficiency of several of the more promising error control and correction techniques from Chapter 5 are examined in this chapter.

6.2. Analysis of the results
The performance is evaluated in terms of the throughput in error free bits delivered to protocols above the ATM layer, which is normalized by the bandwidth occupied by the signal over the satellite. The results are expressed in terms of bits per hertz (bits/Hz) of carrier bandwidth over the satellite link. A “snap-shot” of the results is provided as Fig. 6.1, in order to illustrate the relative potential of the techniques. For the purposes of discussion, high speed file transfer via TCP/IP over ATM AAL 5 is considered, with a QoS requirement that $P_e \leq 10^{-8}$ [2]. A nominal 8192 byte CPCS-PDU is assumed. The performance may vary somewhat with CPCS-PDU size.

QPSK modulation, with rate 1/2 FEC, on a FDMA carrier, with a carrier spacing equal to the channel symbol rate, is considered as the baseline. A carrier spacing of 1.0 times the channel symbol rate is optimistic for FDMA, where a carrier spacing of 1.4 times the channel symbol rate is typical [2]; however, it is reasonable for single carrier CDMA. 8-PSK TCM on a CDMA carrier is considered for all other cases.

The baseline provides the best performance when $E_b/N_0$ is less than about 5.0 dB. The peak efficiency is 0.91 bits/Hz, as seen at the boundary between the ATM layer and TCP/IP. Although full performance is achieved from the perspective of the ATM
layer at about an $E_b/N_0$ of 5.5 dB, as evidenced by the flattening of the curve, an $E_b/N_0$ of 6.4 dB is actually needed to meet the QoS requirement that $P_b \leq 10^{-8}$. The reason for the discrepancy is that the link, in this example case, is TCP limited at the higher layer, not by the ATM layer.

8-PSK TCM without additional error correction and control outperforms the baseline in terms of the bandwidth efficiency starting at about 6.0 dB $E_b/N_0$, achieving a maximum of 1.81 bits/Hz at 7.2 dB $E_b/N_0$. However, 8.2 dB $E_b/N_0$ is needed to meet the QoS requirement.

The maximum efficiency of the SRQ protocol with 8-PSK TCM is just under 1.81 bits/Hz, due to the overhead associated with the SRQ protocol. At first glance the performance improvement would appear to be very small. However, this technique is more bandwidth efficient than the baseline beginning 5.8 dB $E_b/N_0$. The QoS requirement is satisfied at 6.5 dB $E_b/N_0$, which means that it is only 0.1 dB less power efficient than the baseline.

![Graph showing 8-PSK TCM with different options](image)

**Figure 6.1: ATM-SATCOM protocol stack efficiency 8192 bit AAL 5 CPCS-PDU**
The four error correcting (255,223) BCH code is 0.5 dB more power efficient than the baseline meeting the QoS requirement at $E_b/N_0$ of 5.9 dB and much more effective overall. A maximum bandwidth efficiency of 1.58 bits/Hz is achieved at an $E_b/N_0$ of 5.4 dB; however, this option is 12.6% less bandwidth efficient than SRQ technique for 6.5 dB $E_b/N_0$.

The ARQ-multicast/diversity technique with two interconnected ground terminals (2 level diversity) is as bandwidth efficient as the SRQ protocol, but 0.1 dB less power efficient based on the need for 6.6 dB $E_b/N_0$ to meet the QoS requirement. Note that the diversity here is in the availability of a good copy of the CPCS-PDU from a neighbor also receiving the multicast transmission. Additional improvement may be possible if diversity combining is performed on a bit or symbol basis, but this has implications in terms of additional system complexity, as well as, the rates over the terrestrial links.

8-PSK TCM with the SRQ protocol, the (255,223) BCH code and the ARQ-multicast/diversity (2 level diversity), is the most power efficient in this example, meeting the QoS requirement at 5.2 dB $E_b/N_0$. The bandwidth efficiency is essentially the same as for the (255,223) code alone. Alternatively, the (443,423) BCH code provides slightly better bandwidth efficiency at a cost of about 1 dB in power efficiency.

6.3. Summary and concluding remarks

The results presented in this chapter show that substantial improvement is possible in terms of overall power and bandwidth efficiency through the use of a layered approach to error control for ATM over SATCOM. Such an approach is particularly germane to the case of ATM. Only by tailoring error control to the ATM service and, preferably to the individual ATM connections, can the full benefit and efficiency of ATM over SATCOM be realized.

Implementation of a single common error control technique means that all services are treated alike, regardless of their individual requirements, or of the mix of services on the SATCOM link. This is inherently inefficient from a SATCOM perspective and
contrary to the purpose of the ATM protocol, which was designed to support multimedia information services through a combination of features including: flexible bandwidth allocation (only use what you need), bandwidth on-demand (only when you need it) and QoS negotiation with the network (only at the required QoS).

The use of a layered approach to error control tailored to the service can enhance efficiency by more realistically matching SATCOM resources to the mix of services and QoS requirements over the satellite link. In this context, SATCOM bandwidth and power may be viewed as part of a pool of resources from which to draw, where the actual usage at any given time is a function of the prevailing service mix, QoS requirements and the essential error control employed.
References

Chapter 7

Discussion, conclusions and recommendations

7.1. Discussion and conclusions

Techniques for broadband wide area networking based on IP over ATM-SATCOM links at Ka-band were examined in this thesis. The key technical issues with respect to operation over SATCOM are: delay, delay variation and error rate over the satellite link, as well as, outages due to rain and atmospheric affects. The later takes on particular significance where operation of high speed ATM networks is concerned. Current TCP limitations can also affect performance at high speeds; especially, over long delay satellite links.

From a SATCOM perspective, the efficiency of ATM lies in its ability to flexibly allocate capacity, it’s support for multimedia services and, potentially, on the manner in which multicasting, priority queuing and other features are exploited. ATM is very sensitive to traffic intensity, type and mix. Previous generations of SATCOM systems were designed in terms of voice circuit capacity. This approach does not make sense for multimedia, mixed rate services running over ATM, where unused capacity is intended to be available to any other user on a priority basis.

In a broader context, the key question is cost, which may be viewed in terms of bandwidth, power and ultimately economics. Efficiency is critical and can only be achieved with a system level solution tailored to meet the users specific information exchange requirements, which include service type, QoS requirements, traffic intensity and mix, as well as, user geographic considerations. Providing communication to anyone, anywhere, anytime time is an admirable goal. However, there may be a niche for a lower cost system that only attempts to satisfy the "80% of the users that impose only 20% of the system design cost".

The potential benefits and limitations of combining ATM, SATCOM and CDMA
were examined in Chapter 1. A novel concept was proposed for integrating ATM, SATCOM and CDMA in Chapter 2, which calls for the allocation of a CDMA code per ATM service class and, ideally, per ATM connection. With ATM, capacity is allocated based on the service type and QoS requirements. The soft capacity limit of CDMA is a function of QoS. The purpose is to match the flexibility of ATM with that of CDMA, in order to improve overall efficiency.

To this end, the performance of bandwidth and power efficient TCM was considered in Chapter 3. The analysis is an enhancement of existing models, which facilitates the evaluation of overall efficiency in Chapter 6.

CDMA transceiver issues were addressed in Chapter 4. The transceiver is designed to be capable of supporting the integration of ATM and CDMA via the allocation of a separate code per ATM service class, ATM VC, or even component of a multimedia service. Operation of multiple simultaneous VCs via separate CDMA codes may be required, which would necessitate the use of a bank of receivers. Based on the results of Chapter 4, the relative delay variation given separate transmission of multimedia component services over a satellite link is not expected to be an issue. The use of GPS is envisioned to facilitate the implementation of synchronous CDMA and the use of time delayed versions of a parent code, at least for multimedia component services, as well as, to reduce acquisition times to small terminals.

Special error control techniques tailored to service class and the ATM-SATCOM protocol stack were presented in Chapter 5. These include a truncated SRQ protocol implemented in the ATM CPCS sublayer for data services, high rate FEC matched to the ATM-SATCOM protocol stack for all service classes and a more comprehensive approach to error control for ATM-SATCOM, based on multicasting and diversity combining for small terminals that are downlink receiver noise limited, for data and possibly video services.

ATM is designed to support QoS negotiation with the network. Queuing for the various services classes is separate and prioritized. It is only natural to consider tailoring error control to the service class. By tailoring FEC (and error control in general) to the service class, and preferably to the individual ATM VCs, only the
essential FEC overhead is employed over the satellite. From a SATCOM perspective, this is inherently more efficient than a purely physical layer approach, where all services are treated alike and the overhead is dictated by the service with the most stringent QoS requirements.

In Chapter 6, the error control techniques of Chapter 5 were combined with the modulation and coding options of Chapter 3, in order to evaluate the potential improvement in the overall efficiency of SATCOM resource utilization. Substantial improvement is possible, dependent on the mix of information services to be supported and their QoS requirements.

There is a common perception that broad acceptance of ATM in the market is dependent on the need for information services that require the flexibility, performance and speed of ATM. It is difficult to know what these services will be, but perhaps broadband wide area networking via ATM over SATCOM is such a service.

7.2. Recommendations for future work
During the course of this work a number of related issues were found to be of interest and are the potential subjects of future work. A summary of these issues is provided.

7.2.1. ATM “switch in the sky” vs. the “bent pipe” relay
The concept presented in Chapter 2 applies to both the case of an unprocessed “bent pipe” satellite relay, as well as, an ATM “switch-in-the-sky”. Their are advantages in either case. There are also significant differences in capability and complexity. A technique for allocating capacity in Erlangs from a pool of resources was proposed in Chapter 2, based on the combination of ATM and CDMA. The effectiveness of this technique remains to be proven and the efficiency may vary dependent on whether a portion of the capacity is allocated to individual ground stations for autonomous local capacity control, as in the case of the “bent pipe” satellite relay, or whether the satellite is the master controller of all capacity allocations, as consistent with the ATM “switch-in-the-sky”. A study of satellite capacity allocation and related ATM CAC for ATM over SATCOM via CDMA is the possible subject of future work. Other possible subjects related to on-board processing include the use of a phased array antenna or MBA and the impact and design of satellite switching and crosslinks.
7.2.2. ATM-SATCOM congestion control

Decisions concerning congestion control may be most relevant from the perspective of the satellite, for example, on a future processed satellite with an on-board ATM switch; however, information can also be obtained for the purposes of congestion control at ground terminals that are similarly noise limited. An assumption in this case is that the noise is user limited not thermal noise limited. This is typical of CDMA systems. By comparison, it is much more complicated to discern similar information about a multi-carrier FDMA based system.

7.2.3. Phased array or MBA

The concept for integration of ATM and SATCOM via CDMA presented in Chapter 2, with additional detail in Chapter 4 and Chapter 5, relies on the association of a PN code per ATM VC. Multimedia communication is decomposed into its component parts and carried on multiple VC-codes, allowing for flexible treatment of the individual services. Small disadvantaged users could readily participate in a subset of the services within a given multimedia conference.

The use of a phased array antenna, or multiple beam antenna, on the satellite could further enhance the flexibility of the concept by allowing an association to be made between information services, PN codes and ATM VC's and SATCOM beams. In other words, services could be more closely associated with the requisite EIRP and bandwidth required, leading to greater efficiency; however, such an antenna could be very complex and costly to implement (i.e., sophisticated beam forming network to support a large number of VC’s). The design and cost of a phased array or MBA antenna capable of allocating beams to VC’s or groups of VC’s in a multicast is proposed as the subject of further research.

7.2.4. Capacity allocation and CAC algorithm development

The ability to allocate SATCOM capacity in Erlangs, from a pool of resources, is an important element of the proposal in Chapter 2. In an ATM-SATCOM system, the allocation of satellite capacity and the ATM CAC algorithm are closely related (i.e., or at least they should be). Satellite capacity has traditionally been allocated in terms
of a frequency band in a single channel per carrier FDMA system, a timeslot in a TDM on FDMA system, or a time and burst rate in a TDMA system.

The concept for allocating capacity based on traffic in Erlangs described in Chapter 2 is well suited to an ATM-CDMA system, but it is novel from a SATCOM perspective. Important issues need to be resolved. For example, knowledge of the available capacity is only as good as the ability to maintain power control. System simulation is proposed a next step in the process of concept validation.

7.2.5. ATM vs. IPv6
Multicasting and priority queuing are important features of ATM and of the concept proposed in Chapter 2. In the future, multicasting and priority queuing may be supported by IPv6, as well as, ATM. It can be argued that only limited capabilities for multicasting and priority queuing will be available with IPv6, but even a limited capability may be enough to obtain a substantial portion of the benefit and to implement the ATM-SATCOM concept of Chapter 2. As an analogy, consider the difference between hard decision decoding, soft decision decoding using a 3-bit quantizer and full soft decision decoding based on ideal quantization. Given the dominance of the Internet protocol suite today, a possible subject of future work is to compare the concept of Chapter 2 based on ATM, to a revised concept based on IPv6.

7.2.6. Enhanced error correction and control
At present, there is little in the way of error correction and control within the ATM layer. Various options were presented in Chapter 5 and the efficiency of select options was evaluated in Chapter 6.

Higher rates and T1 frames - Several of the techniques examined in Chapter 5 were intended to be implemented within the E1 2.048 Mbps frame format. Thus, they are applicable at higher rates (e.g., E2 8.448 Mbps), but this case has not been specifically examined. Further, similar techniques could conceivably be employed with respect to the T1 frame structure, but the frame format is different and any codes tailored to the E1 frame structure would have to be revised.
Integration of the Pragmatic TCM transmitter with the antenna - The probability that a 180° phase error occurs between antipodal signal pairs is typically very small relative to transitions between neighboring signal pairs, for reasonably high values of $E_b/N_0$. If necessary, it may be possible to provide additional distance between signals by combining the Pragmatic MPSK TCM transmitter with the transmit antenna. For example, an association could be made between the MSB and the antenna polarization (e.g., right polarization if the MSB is a 1 and left polarization if it is a 0). This may require an adaptive antenna, such as a phased array, or multiple beam antenna, as well as, changes to the receiver and is a possible subject for follow-on work.

Turbo codes - As noted in Chapter 3, Turbo Codes may be an alternative to MPSK TCM. Turbo Codes offer potentially superior performance for high rate transmission, but they are not yet as mature as "Pragmatic MPSK TCM", which can be implemented using a COTS codec. Further, the results of Chapter 2 indicate that the capacity in Erlangs, is more sensitive to bandwidth than to power. Revisiting the ATM-SATCOM efficiency analysis in Chapter 6 for the case of Turbo codes is the possible subject of future work, in combination with an assessment of whether the additional performance gain is worth the cost in complexity and additional delay due to interleaving.

7.2.7. Review of emerging ATM-SATCOM systems

At present, a number of emerging satellite systems either plan to employ ATM over SATCOM, or are seriously considering the possibility. Examples include: Teledesic, Celestri, Spaceway and KaStar. Recently, Comsat began offering the first commercial ATM service over a satellite. This service is intended to support Internet access from Puerto Rico to points in South America. A survey of existing and planned ATM-SATCOM services is proposed for comparison with the concept presented herein.

7.2.8. Comparative performance of access techniques

The focus in this thesis has been on ATM over SATCOM via CDMA. This combination was selected because the capability to flexibly allocate bandwidth with ATM and the soft capacity limit of CDMA are complimentary. Other techniques are may be of interest, dependent on the application and operational requirements, such as
ATM over SATCOM via TDMA or FDMA, and a comparative performance analysis is proposed.

7.2.9. FH/DS-CDMA
The concept proposed in Chapter 2 relies on the integration of ATM and SATCOM via spread spectrum. Specifically, it is based on broadband DS-CDMA. Other spread spectrum techniques may be of interest. Hybrid FH/DS-CDMA is considered to be a particularly promising technique. From Chapter 1, hybrid FH/DS-CDMA can be employed to increase the overall spread bandwidth and therefore the capacity, it is useful where the available spectrum is not contiguous and it may be possible to reduce the delay due to interleaving, in some cases. Perhaps most significantly though, it may be simpler to obtain side channel information necessary to evaluate the available capacity, at any given instant, and to support resource allocation and ATM CAC. With CDMA, side information may require the transmission of a separate pilot tone. By comparison, each block of frequency hopped data can provide side information, for example, for power control relative to an average, or expected value.

7.2.10. MPSK TCM in the presence of partial band interference
The objective of this analysis is to model the performance of MPSK TCM in the combined AWGN and mutual interference environment using bounds similar to those of section 3.1. This work would require substantial modification to the existing model, but may be especially interesting in conjunction with the aforementioned hybrid FH/CDMA approach, where unexpectedly strong signals are detected as interference and their contributions are de-weighted or normalized.

7.2.11. The impact of imperfect phase estimation on MPSK TCM
Carrier phase estimation is never perfect. Phase error can have a significant impact on MPSK performance; especially, as M increases. In systems employing M-law carrier recovery, typical with MPSK modulation, the phase variation can be modeled using a Tikhonov distribution. When clock aided carrier recovery is considered, the phase error or phase uncertainty of the recovered signal is limited by the clock signal and it’s time uncertainty. Short term clock timing variations may be modeled using a Gaussian distribution and it is proposed to model the phase variations with clock aided carrier recovery as a Gaussian process, rather than a Tikhonov process.
7.2.12. Code-aided carrier recovery
The possibility of using a form of clock-aided carrier recovery based on the PN code clock in a CDMA system was raised in Chapter 1 and Chapter 4. The concept, referred to as code-aided carrier recovery could be employed to facilitate the use of higher order MPSK TCM and may also have applicability where a phased array antenna is implemented, in terms of synchronization. While the concept shows promise, there is a cost in terms of reduced S/N and the benefits, if any, remained to be determined. Clock aided carrier recovery may be useful dependent on the size of the satellite ground terminals and the quality of service to be supported. Initial modeling has been performed; however, the analysis remains largely a subject for follow-on work.

7.2.13. Implementation using COTS H/W and S/W
Given the fact that ATM technology is still not mature, applications are scarce and hardware continues to evolve, a practical question is to determine to whether the concept of Chapter 2 can be implemented using COTS hardware and software or, if not, to what degree modification or enhancement would be required. This is potentially very important question, well within the scope of potential future work, but it is considered necessary to develop an ATM-SATCOM testbed, perhaps as part of a larger ATM-mobile wireless testbed, in order to obtain a reliable answer.

7.2.14. System cost estimation study
The concept proposed in Chapter 2, is geared towards business use. In order to increase cost effectiveness, service would be focused almost exclusively on the regions where business activity is greatest (e.g., 50 degrees north to 50 degrees south). A limited satellite constellation is considered and the initial concept calls for a "bent pipe" satellite. Estimation of the cost of the proposed system and comparison with other systems offering comparable services is a possible topic for future work.
JOHN R. FARSEROTU
ELECTRICAL ENGINEER

EXPERTISE:

Satellite communications, wireless communications, ATM-SATCOM networks, telecommunications, digital communications, modulation and coding

EDUCATION:

Delft Technical University, 1993 - present
Delft, The Netherlands
Telecommunication and Traffic Control Systems Group
Doctoral Student, candidate for graduation 1 December 1998
Broadband networking via IP/ATM over Ka-band SATCOM

George Washington University, 1983 - 1986
Washington D.C., U.S.A.
M.S.E.E. - Communications, 1986

University of Maryland, 1978 - 1982
College Park, Md., U.S.A.
B.S.E.E., 1982

EXPERIENCE:

Centre Suisse d'Electronique et de Microtechnique (CSEM): Neuchatel, Switzerland (9/98 - present)
Mr. Farserotu is currently the Head of the Wireless Communication section at CSEM, where he is responsible for leading a team of 12 to 14 engineers. He is also Program Manager of the Multimedia Access and Distribution (MAD) project at CSEM. The MAD project involves implementation of a Hiperlan type 2 wireless LAN and CSEM is responsible for leading the baseband subsystem development effort.

NATO C3 Agency (NC3A): The Hague, The Netherlands (9/90 - 9/98)
As a Principal Scientist in the Satellite Communication (SATCOM) Branch at NC3A, Mr. Farserotu was a member of the NATO SATCOM Post-2000 team, evaluating SATCOM requirements, options and technologies for the future. From 1994 through 1998, he has also been a member of the ATM project team. The focus of his work was on the integration of ATM and SATCOM, networking and performance analysis and characterization. He participated in the conduct, organization and planning of international tests and technology demonstrations with the US, the UK and France. He performed test and analysis of TCP/IP over ATM AAL5 via satellite links contributing to all aspects of the evaluation of TCP/IP over low data-rate ATM over satellite links. The performance was evaluated in terms of the throughput via a combination of on-air satellite test, simulation and analysis.

Previously, Mr. Farserotu was a task leader on the Communications System Network Interoperability (CSNI) program. CSNI is a multi-national technology demonstration
involving the development of a Wide Area Network (WAN) based on Open Systems Interconnection (OSI) protocols. Local area networks (LAN) within participating organizations were interconnected via satellite and other transmission media. He was responsible for implementation of the SATCOM nodes at STC, which required integration of a UNIX workstation with an access controller running a Time Division Multiple Access (TDMA) protocol developed using the VxWorks realtime software development environment.

As a Senior Scientist, Mr. Farserotu performed test and analysis on commercial Intelsat Business Systems (IBS) modems and provided advice to NATO working groups. He performed systems engineering test and analysis for video teleconferencing (VTC), LAN-SATCOM integration and Public Area Branch Exchange (PABX) SATCOM experiments, with emphasis on timing and synchronization (T&S) and he is regarded as an expert on T&S within NATO.

Stanford Telecommunications, Inc.: Mclean, Virginia, USA (10/85 – 9/90)
At Stanford Telecommunications (STel), Mr. Farserotu was a Senior Engineer and Program Manager on the Battle Management Communications Development program for RADC. This was a satellite communication system design and analysis project for the Strategic Defense Initiative (SDI). He was responsible for technical leadership of up to five junior engineers.

As a Senior Engineer and Task Leader, Mr. Farserotu was responsible for design and analysis of the communication waveform, system and subsystem in a stressed environment. Specifically, he led junior engineers in the quantitative evaluation of the link margins required to close communication links under stressed conditions. He also led the analysis of degradation due to imperfect jammer state information. Our work resulted in waveform recommendations that were presented to the Strategic Defence Systems (SDS) Communications Working Group, and he received an outstanding performance award from STel for my contribution. He also received an outstanding performance award for the optical communications conceptual design and analysis that I performed. Additionally, he contributed to the Laser vs. Radio Frequency (RF) technology tradeoff studies evaluating performance, weight and power.

The BDM Corporation: McLean, Virginia, USA (1/83 – 10/85)
As an Associate Staff Member at BDM, Mr. Farserotu performed Defence Satellite Communication System (DSCS) Super High Frequency (SHF) analysis in support of the USPACOM communications architecture. Specifically, I was responsible for a quantitative link assessment of the advantages of SHF over UHF SATCOM under intentional interference.

Mr. Farserotu contributed to the network simulation and modelling studies of a tactical communication system. His specific responsibilities included: derivation of all traffic loading data for the baseline model, definition of the network topology, correction of the call alternate routing tables in network model on the VAX-11/780, and evaluation of GOS and SOS. Previously, he was responsible for a statistical analysis of frequency hopping, which identified flaws in the initial frequency management concept. Additionally, he supported the Automatic Test Equipment (ATE) program performing digital and analog circuit analysis.

COMSAT Laboratories: Clarksburg, Maryland, USA (9/80 – 12/82)
As a Co-op intern in the optical communication laboratory at COMSAT, Mr. Farserotu
performed test, characterization and sensitivity analysis of semiconductor devices. He also
designed a transimpedance amplifier and performed mathematical and computer modelling
to evaluate the potential as an optical receiver front-end, he performed spectral analysis of
light sources using a monochromometer and electro-optic switch isolation measurements.

PROFESSIONAL SOCIETY ACTIVITIES:

• Member of the IEEE (1983 - present) and the IEEE COMSOC (1986 - present)
• Secretary of the IEEE Joint Benelux Chapter, Communications and Vehicular Technology
  (1993 - present)
• Secretary of the IEEE Vehicular Technology Conference 1999 in Europe (VTC'99 Europe)
  Organizing Committee (1997 - present)
• Review of papers submitted to the "Wireless Personal Communications An International
  Journal", as well as PIMRC '94 and local IEEE chapter symposia
• Chairman of the session on networks, IEEE SCVT '97 symposium at Eindhoven TU in The
  Netherlands

COURSES AND TUTORIALS:

• Developed a 1 day course on Satellite Communications for the Third Joint European
  Compact Course on Mobile Personal Communications Systems (MPCS) 1998, presented at
• Developed a tutorial on MILSATCOM and presented it to members of industry seeking the
  Spacetech Master of Technology degree at The Technical University of Delft, Delft, The
  Netherlands in May 1997 and April 1998.

JOURNAL PUBLICATIONS:

1. J. Farserotu and R. Prasad, Broadband Wide Area Networking via IP/ATM over
   SATCOM, IEEE Communications Magazine (accepted for publication).
2. J. Farserotu, A Performance Analysis of MPSK TCM, Wireless Personal Communication
   An International Journal, Kluwer publishers (accepted for publication).
3. J. Farserotu and R. Prasad, Wide Area Networking via ATM over Ka-band SATCOM with
   CDMA, IEEE Journal on Selected Areas in Communications (JSAC) (accepted for
   publication).

CONFERENCES, SYMPOSIA AND WORKSHOPS:

1. J. Farserotu and R. Prasad, ATM over SATCOM via EHF, to be presented at the IEEE
   Global Communications Conference (GLOBECOM) '98, Sydney, Australia, 8 - 12,
   November 1998.
2. R. Prasad and J. Farserotu, Tutorial on Air Interface and Access Schemes for Land
   Mobile Satellite Communications, invited paper, to be presented at the 6th Digital Signal
   Processing (DSP) Workshop DSP '98, at ESTEC, Noordwijk, The Netherlands, 24-26
   September, 1998.
3. J. Farserotu and R. Prasad, Error Control and Correction Techniques for ATM over
   SATCOM, to be presented at the 6th Digital Signal Processing (DSP) Workshop DSP '98,
4. J. Farserotu and A. Tu, TCP/IP over Low Rate ATM-SATCOM Links, IEEE MILCOM
OTHER PUBLICATIONS:
