KTH Electrical Engineering

Mobile Satellite Broadcast and Multichannel Communications — Analysis and Design

CRISTOFF MARTIN

TRITA–S3–SB–0528
ISSN 1103–8039
ISRN KTH/SB/R - - 05/28 - - SE

Doctoral Thesis in Signal Processing
Stockholm, Sweden 2005

Signal Processing
KTH Royal Institute of Technology
SE-100 44 Stockholm, Sweden
Tel. +46 8 790 6000, Fax. +46 8 790 7260
http://www.s3.kth.se

The work presented in this thesis is partly funded by

SES GLOBAL S.A.
L-6815 Château de Betzdorf
Luxembourg
http://www.ses-global.com

and

LIASIT
Luxembourg Advanced Studies in Information Technologies
University of Luxembourg
211, route d'Esch
L-1471 Luxembourg
Luxembourg
http://www.liasit.lu

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen fredagen den 29 april 2005 klockan 13.00 i Sal D2, Kungl Tekniska högskolan, Lindstedtsvägen 5, Stockholm.

Tryck: Universitetsservice US AB

Abstract

In this thesis, analytical analysis and design techniques for wireless communications with diversity are studied. The impact of impairments such as correlated fading is analyzed using statistical models. Countermeasures designed to overcome, or even exploit, such effects are proposed and examined. In particular two applications are considered, satellite broadcast to vehicular terminals and communication using transmitters and receivers equipped with multiple antennas.

Mobile satellite broadcast systems offer the possibility of high data rate services with reliability and ubiquitous coverage. The design of system architectures providing such services requires complex trade-offs involving technical, economical, and regulatory aspects. A satisfactory availability can be ensured using space, terrestrial, and time diversity techniques. The amount of applied diversity affects the spectral efficiency and system performance. Also, dedicated satellite and terrestrial networks represent significant investments and regulatory limitations may further complicate system design.

The work presented in this thesis provides insights to the technical aspects of the trade-offs above. This is done by deriving an efficient method for estimating what resources in terms of spectrum and delay are required for a broadcast service to reach a satisfactory number of end users using a well designed system. The results are based on statistical models of the mobile satellite channel for which efficient analytical design and error rate estimation methods are derived. We also provide insight to the achievable spectral efficiency using different transmitter and receiver configurations.

Multiple-element antenna communication is a promising technology for future high speed wireless infrastructures. By adding a spatial dimension, radio resources in terms of transmission power and spectrum can be used more efficiently. Much of the design and analysis work has focused

on cases where the transmitter either has access to perfect channel state information or it is blind and the spatial channels are uncorrelated.

Herein, systems where the fading of the spatial channels is correlated and/or the transmitter has access to partial channel state information are considered. While maintaining perfect channel knowledge at the transmitter may prove difficult, updating parameters that change on a slower time scale could be realistic. Here we formulate analysis and design techniques based on statistical models of the multichannel propagation. Fundamental properties of the multi-element antenna channel and limitations given by information theory are investigated under an asymptotic assumption on the number of antennas on either side of the system. For example, limiting normal distributions are derived for the squared singular values of the channel matrix and the mutual information. We also propose and examine a practical scheme capable of exploiting partial channel state information.

In both applications outlined above, by using statistical models of the channel characteristics in the system design, performance can be improved. The main contribution of this thesis is the development of efficient techniques for estimating the system performance in different scenarios. Such techniques are vital to obtain insights to the impact of different impairments and how countermeasures against these should be designed.

Acknowledgments

Five years ago, Professor Björn Ottersten, KTH supervisor extraordinaire, provided me the opportunity to start as a doctoral student. Through his continuous support, this work is nearing completion. Dr Alexander Geurtz, my supervisor at SES, ensured that I stayed focused on the work which outcome now is in your hands. Björn, Alexander, I am grateful that you shared some of your valuable time and your insights into technical detail and level of professionalism never cease to impress.

As part of the LIASIT project in the framework of the University of Luxembourg I have been privileged with financial support from SES GLOBAL S.A. and a BFR grant. The time spent at SES was rewarding, providing industrial insight to satellite broadcast and improving the relevance of this thesis. None of this would have been possible without the efforts of Dr Gérard Hoffman, LIASIT director, as well as Edward Ashford, Robert Bednarek, Gerhard Bethscheider, Christophe Loeillet, John Lothian and others from SES. Technical discussions with Guy Harles and Ray Sperber at SES as well as interaction with Harald Ernst and the other participants of the ESA "Mobile Ku-band Receiver Demonstrator Project" were most rewarding.

My colleagues at the signal processing lab at KTH have always been a great source of inspiration. Collaboration with Svante Bergman and Joakim Jaldén led to joint publications. The first class administrative and computer support is not forgotten either.

I would also like to thank all new friends who made me feel at home in Luxembourg. Without you the last years would have been an uphill struggle. The continuous support of family and friends back home even though I might have seen absent at times was highly appreciated. Very special thanks to Sandra for her patience.

A number of people directly contributed to this dissertation. I am particularly grateful to Dr Jack Winters for acting faculty opponent and

to the grading committee. Finally I am indebted to Svante Bergman who proofread and commented this manuscript.

Cristoff Martin
Luxembourg, March 2005

Contents

Chapter 1

Introduction

In this thesis we consider analysis and design techniques for wireless communications with diversity. Based on statistical channel models, methods are proposed for two different applications; satellite broadcast to moving vehicles and multiple-input multiple-output point to point communication. Digital mobile broadcast allows for traditional services such as audio and video but other, mobile specific, services like travel information, map updates and other data services could also be envisioned. Using a satellite system potentially offers the possibility of reliability and ubiquitous coverage in an efficient manner. Multiple-input, multiple-output, MIMO, communication systems employ multiple antennas on both the transmitter and the receiver sides of the communication link. This way, the limited spectral and power resources available can be used more efficiently creating a foundation for tomorrow's high speed wireless connections.

The purpose of this chapter is to provide a general introduction and a road-map to the thesis. Section 1.1 provides high level problem formulations and main contributions. A brief introduction to digital wireless communication systems with emphasis on topics considered in this thesis is found in Section 1.2. Section 1.3 provides a background to the considered mobile satellite broadcast system. An overview of systems currently deployed, in deployment and in design is provided. We outline how the overall design process of a commercial mobile satellite broadcast system could be envisioned and discuss where our results may contribute. In Section 1.4 we motivate the interest in MIMO communications. To that end we attempt to provide some intuition to the gain in capacity that multi-antenna systems can provide. We also provide a background

to the considered problems. Someone already familiar with these topics may wish to skip ahead to Section 1.5 which outlines the rest of the thesis. Appendix 1.A briefly presents contributions outside this thesis and extensions of the work presented herein. Finally, Appendix 1.B and Appendix 1.C provide an overview of the notation and abbreviations used throughout.

1.1 Analysis and Design Based on Channel Statistics

While the two considered applications are diverse there are common denominators. In both cases, system performance could be improved if the transmitter was aware of the current propagation conditions. However, perfect information about the instantaneous channel state is in most scenarios unrealistic. In the broadcast case, even if a feedback channel is available, collecting information about the current channel conditions for all receivers would be impossible also for a moderate number of users. To characterize the channel of a multi-input, multi-output system, estimates of the channels between all the different transmit and receive antenna elements are required. If the channel changes rapidly, providing this information via a feedback channel is costly. There are also a number of sources of error in the estimates including delay and sampling. Thus even if the estimates are available they cannot be completely trusted.

Instead of basing the design of the transmitted signal on estimates of the instantaneous channel response, it is in many cases more favorable to design the system based on its statistical properties. That is, instead of adapting to the current quality of the channel, the system is designed after its usual behavior. For example, the fluctuations in attenuation of the mobile satellite channel are mainly the result of long term events due to shadowing and blockages from obstacles surrounding the mobile terminal. By selecting the parameters of the broadcast system after the typical distribution of duration and frequency of such events, a robust and efficient system can be designed. Similarly, maintaining channel estimates at the transmitter may be prohibitively expensive in a multi-input, multi-output scenario. As an alternative, feeding back parameters that change more slowly such as statistical properties may be more attractive. Thus system design techniques based on the statistical properties of the system are of interest. For these types of techniques to be effective, efficient analysis techniques are required. The derivation of such methods is the

main contribution of this thesis.

With access to efficient analysis techniques, the impact of various impairments can be studied and elaborate designs derived. The behavior of communication channels is frequently characterized by correlation. For example, mobile satellite broadcast and MIMO channels are often modeled as correlated in time and space respectively. Unless properly taken cared of, this type of effect may have a severe impact on the system performance. Using the derived analysis techniques we propose system designs that exploit knowledge of channel statistics. This way, the impact of correlation can be reduced and we show that well designed correlated systems in some scenarios outperform their ideal uncorrelated counterparts.

1.2 Wireless Communication Systems

In this section a brief overview of a general wireless communication system with emphasis on the parts considered in the latter part of this thesis is provided. In particular we try to provide intuition to concepts such as coding, diversity, and channel capacity.

1.2.1 System Overview

Wireless communication systems such as those considered in this thesis can be seen as consisting of a number of steps. Below we present an overview of these steps which is illustrated in Figure 1.1. Standard books such as [Pro01] describe this process in much more detail.

> *Source coding:* A digital communication system may carry many different kinds of content. For example, streaming and file delivery one-way services might be provided by a broadcast system, while a point-to-point system also may provide data communication and two-way voice services. In all cases the content should be encoded in a manner suitable for the data rate available while providing sufficient quality for the end user. To meet this goal the data is compressed and redundancy removed. The source coding approach taken depends largely on which type of content is to be provided. Two-way voice communication has tight delay constraints, but the requirements on sound quality are relatively relaxed. File delivery services on the other hand normally require that files are not dis-

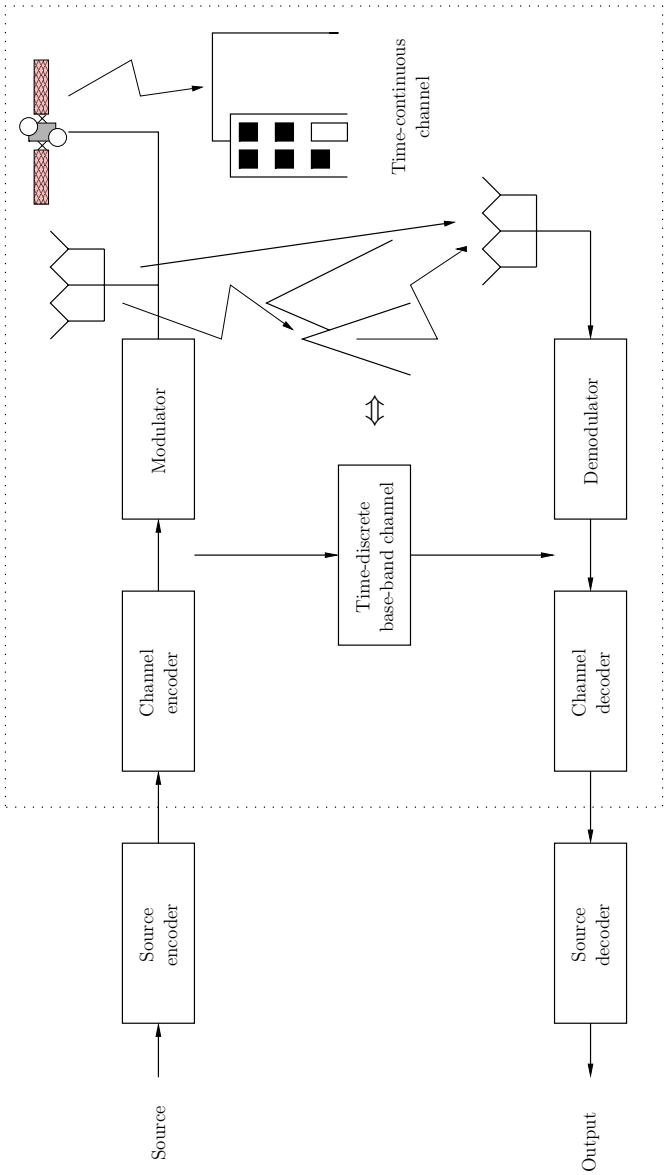


Figure 1.1: Schematic of a wireless communications system.

torted by the source encoding process and delivered error free, but may on the other hand have more relaxed delay requirements.

To some extent, the source coding used in the system also defines requirements on the rest of the system. For example, error mitigation schemes used in modern audio coding for digital radio applications might tolerate losses up to a few percent of the frames in the content stream [FJK+02]. Similarly, voice encoders used for point-to-point communication can often tolerate some errors before the distortion becomes noticeable [Nat88]. For data services, such as the delivery of files, requirements on the error rate performance are in general much more stringent. Depending on what services are considered and what source coders are used, quality of service requirements on the content delivery can be established, see for example [Rap95].

Channel coding: Wireless communication is challenging as the channel impairments create errors in the transmitted data stream. One efficient technique to overcome these effects is to use channel coding. A channel code works by adding redundancy to the data in a structured way, such that when the received data is corrupted by noise and other error sources, a satisfactory quality in the decoded data can still be ensured.

Modulation: Transmission over a wireless channel is inherently an analog time continuous process. Therefore the encoded digital data symbols are modulated on a time-continuous waveform and up-converted to a carrier frequency. The resulting signal or signals are then transmitted over one or more antennas. At reception this process is reversed; the received data is down-converted, filtered and sampled resulting in a decision metric suitable for the channel decoder.

The mobile wireless channel: The mobile wireless communications channel suffers from a number of impairments. In this thesis, mainly two effects are considered. First, thermal noise disturbs the received signal. Second, the power of the desired signal may vary considerably over time, the channel is *fading*. Fading can be caused by a number of different effects. Most obviously, long term effects such a shadowing of the line of sight component will affect the received signal strength. The signal may also arrive at the receiver along multiple paths. Depending on the lengths of the different paths, different components might add constructively or

destructively. The signal level may therefore fluctuate significantly making reception impossible unless proper countermeasures are applied.

The combined effect of the modulation, the wireless communications channel and the demodulation steps can be modeled using a time-discrete base-band model, see Figure 1.1 and standard references such as [Pro01]. The channel impairments are then modeled using an additive term for the noise and a multiplicative term for the fading. This will be the approach taken throughout this thesis.

Designing and analyzing communication systems for all possible combinations of noise and fading would not be practical. Instead the channel can be modeled in statistical terms and the system designed after the behavior defined by the channel statistics. The channel statistics used depend on what type of environment and what type of services are considered. For example, the mobile satellite broadcast channel is characterized by long term events in term of shadowing due to obstacles in the terminal environment. As the satellite power is limited the system must be designed to overcome this using other means. For indoor WLAN type of applications the situation is radically different. In most cases there is no line of sight propagation path but the system must rely on multi-path propagation to communicate.

In this thesis we are in general considering analysis and design techniques for the communication process from channel coding to channel decoding, i.e. the steps within the dashed box of Figure 1.1.

1.2.2 Channel Coding and Diversity Techniques

To overcome the effects of fading and noise on the channel, coding and diversity are efficient countermeasures. By adding redundancy in a structured manner, a channel code can be used to recover data damaged in the transmission process. An efficient mean to overcome fading is to use diversity techniques. The idea is to transmit redundant data over independent channel realizations such that if some of the transmitted data is lost due to fading the original data can still be recovered. In many cases it is favorable to combine diversity and channel coding techniques, such approaches have for example resulted in space-time coding techniques for terrestrial [TSC98] and satellite [FKKK01] systems. In this thesis this

type of coded diversity is considered and we will mainly consider two means of providing it.

Time Diversity. Due to changes in the relative positions of the transmitter receiver pair or due to changes in the environment the communication channel will change in time. Thus, by spreading the transmission in time, diversity can be provided. The required duration to achieve reasonably independent channel realizations depends on a number of factors including terminal velocity and carrier frequency. It also depends on which effect causes the channel to fade; fading due to shadowing normally lasts much longer than fading due to multi-path propagation. Unfortunately, introducing time diversity means introducing a delay as the data cannot be recovered until a sufficient part of a code word has been received.

From a reliability viewpoint, the longer the time period the data can be spread over the better. However, most services are associated with a delay requirement which limits the amount of redundancy that can be provided. For example, a two-way voice connection must introduce a minimum of delay to be attractive. Audio and video streaming services can be attractive even with a considerable delay, but even here there are limitations. While it may be acceptable that an audio or video broadcast service is played out many seconds or even minutes after the transmission started, a delay when starting the terminal or switching content channel is more frustrating.

Spatial Diversity. By transmitting and/or receiving the data over several antennas the correlation in the fading can be reduced and the performance of the communication system improved. This is illustrated in Figure 1.1. The antenna separation required to achieve satisfactory results largely depends on the carrier frequency and the environment. If the fading is mainly caused by shadowing, as for a mobile satellite channel, the different antennas must be sufficiently separated to make it unlikely that a single object in the environment will shadow more than one transmit antenna [VCFS02]. When the fading is caused by multi-path propagation the antennas can be very close provided there is sufficient near field scattering [SFGK00].

Applying diversity techniques will not render the individual channel realizations completely independent. For example, as there are limits on the length of an interleaver, there will be some correlation when time

diversity techniques are applied. This will hold true also for spatial diversity techniques. In general, this will mean that the performance of the diversity scheme is reduced. One of the main contributions of this thesis is to analyze these types of effects and design countermeasures against the fading correlation. If properly dealt with, in some situations correlation may even improve the system performance [MBD89, IUN03].

1.2.3 Channel Capacity

In his landmark work [Sha48], Shannon introduced the channel coding theorem, perhaps the most famous and important result of information theory. Roughly speaking, this theorem states that communication over a channel that introduces errors can take place with arbitrarily low probability of error as long as the information rate in the communication is below a threshold called the channel capacity. Conversely, error free communication at rates above this threshold is not possible. Unfortunately, for most channels, communication at the limit given by capacity is impractical as it requires infinitely complex receivers and infinitely long delays. For a more precise definition of channel capacity, see [CT91].

In this thesis we will use the channel capacity as an indicator of the maximal performance that can be achieved using more practical coding and modulation schemes. This is of relevance since modern coding and modulation schemes [BG96, Gal62] can operate near this limit.

1.3 Mobile Satellite Broadcast

In the future a wide range of entertainment and information data services will be available for drivers and passengers of moving vehicles. These could include audio, video and back-seat entertainment as well as traffic, travel and other information. Not all services are necessarily directly noticed by the end user, for example software and databases in the vehicle could be updated silently. Most of the services mentioned here could be well served by an advanced broadcast or multicast wireless system providing sufficient data rates and coverage. Such a system must also be capable of supporting multiple service classes as the requirements on for example file delivery and audio entertainment are vastly different.

In this section terrestrial and satellite systems that could provide some of the services above are described. We briefly discuss mobile satellite channel impairments and suitable countermeasures. Finally, we present

a vision for an overall design of a commercial mobile broadcast system and how our work fits in this framework.

1.3.1 Terrestrial Communication Systems

Broadcast systems today mainly provide information and entertainment services to mobile vehicles using analog FM radio. While FM radio can provide a reasonable sound quality, at least given the noisy car environment, these systems still suffer from shortcomings motivating new and improved broadcast infrastructures. To begin with, the variation in terms of the number of simultaneous radio channels is lacking. While additions to the analog networks, such as the European Radio Data System, RDS [Mar89], can provide some data services, for example limited traffic and travel information, for more advanced data services the data rates offered are insufficient.

To allow for more advanced services, digital terrestrial broadcast systems have been designed and spectrum released allowing for a larger number of audio channels and more advanced data services. An example of a standard which has been successful in some markets is the Digital Audio Broadcasting, DAB Eureka-147 [ETS01]. Using digital techniques, for national coverage, DAB can utilize the available spectrum about one order of magnitude more efficient than analog FM [RZ93]. With the advancements in source and channel coding currently designed systems could provide even greater gains in resource efficiency. Better spectrum utilization allows for more advanced services like higher audio quality and a greater selection of content. Digital transmission also allows for data services such as advanced traffic and travel information [Mar03]. Terrestrial digital video broadcast standards such as DVB-T [ETS04a] are currently in deployment in many parts of the world. However these systems have primarily been designed for fixed services making them less suitable for mobility and wide area mobile coverage. The more recent DVB-H standard [ETS04b] will allow for terrestrial data broadcast that is more suitable for reception in mobile and handheld terminals.

While third generation and beyond mobile phone system have been designed with multimedia services in mind, they are in general not well suited for broadcast applications. Apart from the obvious inefficiency in using individual communication links for broadcasting services, data rates tend to be reduced when the receiver is not near a base station. Like envisioned in [HM00], these types of systems and mobile broadcast systems may in the future complement each other. The broadcast system

can provide high data rate services with extensive coverage requirements, while the mobile phone system handles return channels and services that require individual links.

In general, coverage is a problem for terrestrial systems. While providing coverage for densely populated areas may be attractive, the resource requirements to provide extensive coverage also in rural areas may be overwhelming. For mobile services this might become even more apparent as the mobility itself for some designs means that the coverage is reduced [DGMD04]. Terrestrial based services are in general also confined within national limits. For services targeting moving vehicles ubiquitous coverage is highly desirable. This issue is addressed with the Digital Radio Mondiale, DRM, standard [ETS04c]. By reusing spectrum earlier allocated for analog AM radio (< 30 MHz), extensive coverage can be provided. Unfortunately the amount of suitable spectrum is extremely limited meaning that the supported data rates will be small. In fact, the projected data rates are so small that even though state of the art source coding algorithms are utilized, the music sound quality will only be "near analog FM."

The above discussion illustrates some of the inherent limitations of terrestrial broadcast. Many of these can be addressed by using hybrid satellite terrestrial networks. For example, it is possible to provide ubiquitous coverage with attractive data rates.

1.3.2 Mobile Satellite Broadcast Systems

Providing broadcast services via a satellite system is attractive as tremendous coverage can be provided using a single transmitter. For services to moving vehicles this advantage becomes even more apparent as they can be received also by travelers in remote areas or even by receivers in different countries. A number of commercial mobile satellite broadcast networks have recently been launched and some projects are running for the design of next-generation systems.

Commercially Deployed Networks

Over the continental US, two mobile hybrid satellite-terrestrial broadcast networks are in commercial operation today, Sirius Satellite Radio and XM Satellite Radio [Sir05a, XMR05]. Another network, run by MBCO and TU Media Corporation [MBC05, TUM05], is currently being launched over Japan and South Korea.

The two American systems are similar, both mainly targeting vehicular users even if handheld and fixed terminals are available. Both competitors use licensed spectrum consisting of 12.5 MHz in S-band ($\sim$ 2300 MHz) which in these systems is used to provide mobile end users with approximately 4 Mbits per second of data [FJK+02]. The main service of both these systems is audio. Using advanced source coding techniques, more than 100 audio channels can be received out of which 60-70 are music channels and the remaining 50-60 are used for talk. Some additional data services exist, like stock tickers, and in the future the intention is to also provide some video. The services are commercial free, instead subscriptions are required which cost USD 10–13 ($\approx$ EUR[1] 8-10) per month. By the end of 2004 XM had attracted about three million subscribers expecting to add another two in 2005 [XM 05]. Sirius reached one million subscribers late 2004, predicting more than 2 million before 2006 [Sir05b].

The system currently being deployed over Japan and Korea differs from its American counterparts in that it primarily targets hand-held terminals. Both dedicated terminals and terminals integrated in mobile phones are foreseen. The service offering is also slightly different with more focus on video and data information. For this system 25 MHz of S-band capacity has been allocated which is used to provide a total of 7 Mbits per second in audio, video and data services [MSS99, ITU01]. The system will provide about 10-15 video, 20-30 audio and a number of data information channels for each of the two target countries [MOB05, Lee04]. Revenue will be generated from advertisement, end user subscriptions and on-line shopping. For the Korean part, a monthly fee of 13 000 won ($\approx$ EUR 10) is required for basic services with additional fees for premium and pay per view content. In their Korean network Tu Media Corporation predicts 8 million subscribers by 2010 [Lee04]. Access to the all non-premium audio and video channels in the Japanese network requires a monthly subscription fee of about 2 500 yen ($\approx$ EUR 18), [MOB05].

While mobile satellite broadcast is taking off in America and Asia, there is currently no system deployed over Europe. With many different countries and languages, mobile broadcast services with pan-European coverage would be attractive. This would for example allow international travelers to have access to their national content. While there may be commercial potential there are regulatory and technical challenges. To provide traditional, near real-time, streaming services, terrestrial diver-

[1] Exchange rates as of end January 2005

sity is a necessity. For terrestrial repeaters, licenses are required, meaning that the owner of the network may have to require licenses for all countries which are to be served. Europe is also located far north meaning that the satellite elevation angle will be small if geostationary satellites are used, decreasing the quality of the mobile channel. More elaborate satellite constellations may reduce this problem; however, other European characteristics such as narrow streets and many small cities may further complicate system design.

Networks Currently Being Designed

A number of system designs have been proposed in the technical literature, see e.g.[GG94, WLS99, LGW00]. Here we will just mention two ambitious system designs that may provide interesting insights to the future of mobile satellite broadcast.

The S-DMB system considered in the EU MAESTRO-project aims at providing multimedia broadcast services to mobile phones [CCF+04]. By using a transmission scheme nearly identical to that already used in third generation UMTS phones, the additional complexity required in the end user terminals is minimized. To reduce the bandwidth requirements of the system, apart from traditional broadcast, forward and store services are foreseen where the handset automatically stores content based on user preferences for later play-out.

Another interesting design is developed in the ESA "Mobile Ku-band Receiver Demonstrator Project," [KuM04]. In this project innovative techniques, including advanced channel coding and a novel service concept are designed [EHS+04]. The service concept allows the system to provide a seamless end user experience even if a large part of the data is lost. This way mobile broadcast services can be provided without the need of terrestrial transmitters. Thus infrastructure complexity and cost can be reduced and regulatory issues alleviated. The system is also designed to use Ku-band ($\sim$ 10-13 GHz) capacity normally utilized for direct to home broadcast and other fixed applications. Compared with the cramped L-band ($\sim$ 1.5-1.6 GHz) and S-band ($\sim$ 2.5-2.6 GHz) normally considered for mobile satellite services, Ku-band potentially offers an abundance of spectrum allowing for high data rate services. This would also allow existing satellites to be utilized in the initial system roll-out allowing for cost and risk reduction. Using such high frequencies for a mobile service is not without problems though. The propagation conditions are harsh as even light foliage will attenuate the transmitted signal

severely. High gain directive receive antennas are also needed to suppress interference and to provide sufficient signal strength. Even though lately there has been significant progress in this area, see e.g. [WW04, SV04], this type of antennas will always be bulkier and more expansive then those typically used with dedicated satellite systems in licensed frequency bands.

1.3.3 The Mobile Satellite Broadcast Channel

In mobile broadcast systems line of sight conditions between the transmitter and the receiver cannot be guaranteed without imposing constraints on the mobility of the terminal. For terrestrial broadcast sufficient power levels can still be ensured by designing the system with a link margin that allows for severe attenuations. In the satellite case though, due to the vast distances involved and power limitations in the satellite, providing such margins is not practical. Thus mobile satellite broadcasting is challenging as obstacles in the terminal environment will cause severe, frequent, and long lasting attenuations in the received signal. One attractive solution to overcome these channel impairments is to use diversity techniques [Bri95]. The idea is to provide the same data from different sources such that if one source is temporarily unavailable the original content can still be recovered. This type of approach is taken in the commercial mobile satellite broadcast systems deployed or in deployment today [XMR05, Sir05a, MBC05]. These systems overcome the land mobile satellite channel impairments using combinations of time, space, and terrestrial diversity techniques [FJK$^+$02, MSS99]. An overview of techniques used in other portable and mobile satellite broadcast networks including WorldSpace and Eureka-147 DAB may be found in [ITU01].

The two systems deployed over the continental US use similar designs to provide diversity. In both cases, two continuously transmitting satellites are providing space diversity. XM is using two satellites in geostationary orbits. Sirius is using a more elaborate space setup with three satellites in geosynchronous, highly elliptical orbits. With this choice of orbit, two satellites are always north of the equator, providing considerably higher elevation angles than what would have been the case with geostationary satellites. To ensure service availability in scenarios where space diversity is not sufficient the signal from one of the satellites is delayed a few seconds. This is useful when for example driving through a short tunnel where otherwise both satellite channels would be blocked. Using space and time diversity, the systems can ensure satisfac-

tory service availability in rural and suburban areas, both XM and Sirius targets an availability of better than 99%, see [MN02, Dav02]. For urban areas however, the above techniques are insufficient. To ensure acceptable conditions also in these areas both systems have deployed terrestrial repeater stations which provide coverage in cities and other attractive areas where satellite coverage is difficult. Mainly because of its more elaborate satellite infrastructure, the Sirius system requires significantly less repeaters [Lay01].

In Japan and Korea a slightly different approach has been taken. Here only a single satellite is used meaning that space diversity cannot be exploited. Some time diversity is provided by using low rate codes and interleavers spanning a few seconds in time [MSS99]. Without space diversity, more terrestrial repeaters are necessary. As this system primarily targets hand held terminals, low power repeaters are used to provide the services indoors at popular locations such as shopping centers and train stations.

While diversity techniques are necessary to provide an acceptable service quality, providing diversity is not without drawbacks.

Time diversity is provided by spreading the transmitted data in time and adding redundancy. As the data is spread in time a delay is introduced and the added redundancy reduces the effective data rate. Intuitively, using a short delay means that the effect of a channel blockage becomes more severe, requiring additional redundancy to overcome the impairment. Thus there is a trade-off between the amount of delay and the amount of redundancy applied. Adding delay also increases the memory requirements in the receiver raising the terminal cost.

Spatial diversity, provided by multiple transmitting satellites or a network of terrestrial repeaters, may be used to ensure that a satisfactory number of end users receives the service also in difficult terminal environments. Such techniques require large investments in infrastructure and may also pose regulatory challenges. In this type of system the available bandwidth must be shared between the different transmitters which in general lowers the achievable data rates.

The above illustrates some of the aspects that should be considered in an overall design, taking commercial, service and cost aspects into account. This is further elaborated below.

1.3.4 Designing a Commercial Mobile Satellite Broadcast System

The design of a commercial mobile satellite broadcast system is a complicated trade-off between commercial, technical, and regulatory aspects. The intention behind this section is to present a vision for how such a design could be performed, how commercial requirements could be taken into account in the technical system design and how the work presented in this thesis is useful in this process.

Our vision for the design of a commercial mobile satellite broadcast system is illustrated in Figure 1.2. For the designer of a commercial broadcast system the goal of the overall system design, including services, business model, and broadcast network, is to maximize the profit of the owner of the network. That is, the difference between the revenue and the cost for generating those proceeds should be maximized. In this overall design, the designer would iterate between commercial aspects, e.g. how much revenue the delivery of a selection of audio, video and data services could generate, and technical aspects like the cost of the infrastructure necessary for the content delivery.

In Figure 1.2 broadcast services are delivered to mobile end users. By complying with certain commercial requirements on the content delivery, revenue is generated. Requirements on the content delivery include amount of content delivered, content quality parameters, service and tuning delays, coverage, availability etc. Depending on what requirements are satisfied different amounts of revenue are generated, e.g. real-time services are more attractive than non real-time services and would tend to generate a larger income. Also, for the end user, the cost is not limited to content and service provisioning, but does also include equipment and installation. There might also be non-monetary costs involved, for example bulky antennas may affect the appearance of private cars reducing the perceived value of the service. Note that the figure does not intend to imply any particular business model which would be well beyond the scope of this thesis. For example revenue is not necessarily generated through end user subscriptions but could come from advertisement or other sources. Also, the content provider and the deliverer might be different entities etc.

The goal of the overall technical design is to design a broadcast system, including transmission scheme, infrastructure, and end user equipment that satisfies the commercial requirements on the content delivery at the smallest possible cost. Note that the cost of complex systems like

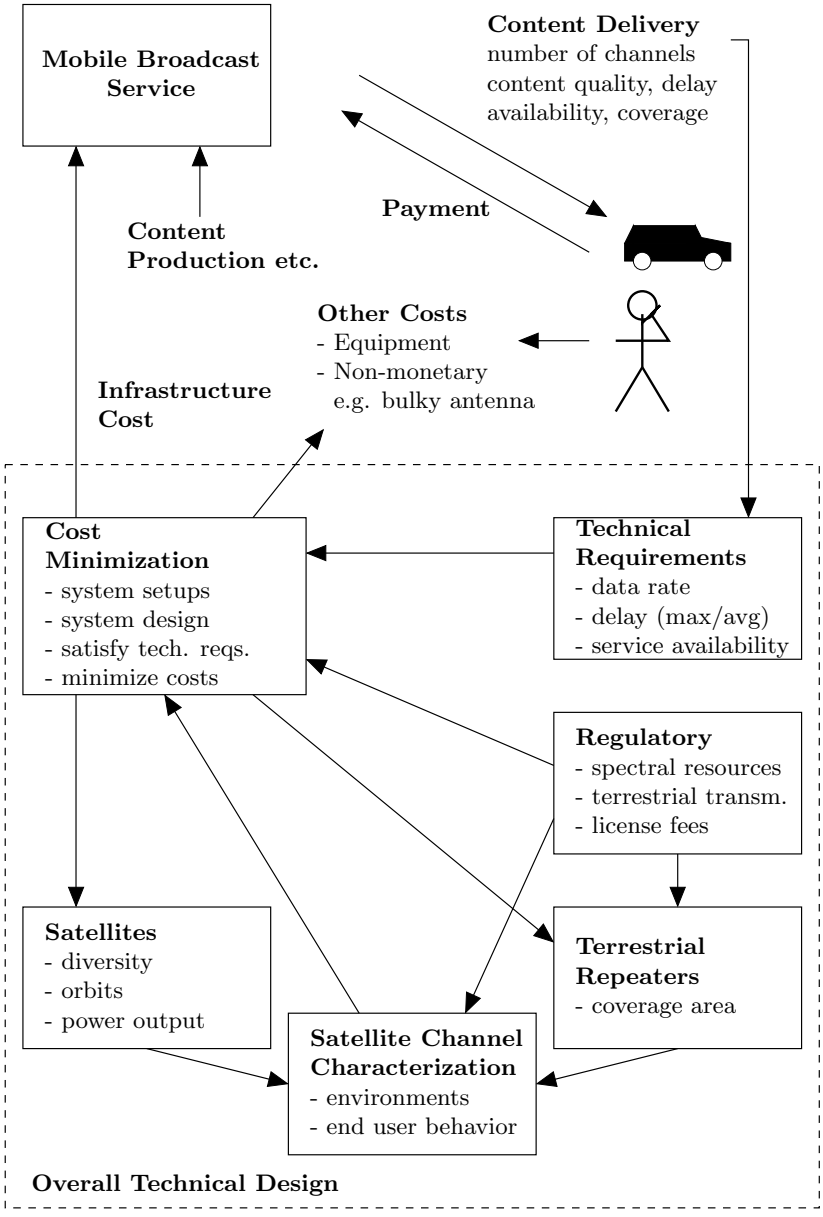


Figure 1.2: Design of a commercial mobile satellite broadcast system.

hybrid satellite terrestrial broadcast networks is not limited to the cost of the infrastructure. Normal running expenses may also include license fees, rental of sites used for terrestrial repeaters etc. In Figure 1.2 we outline how an overall technical design could be performed.

The commercial requirements on the content delivery system need to be translated into technical requirements on the system. That is, the required number of audio and video channels, content quality, coverage etc. should to be translated into technical requirements like data rates, acceptable delay and service availability. While this might seem straightforward, this translation implies choices of certain parameters in the design. For example the choice of source coder will be implied by the above process. This choice does not only affect the content quality but also the sensitivity to errors in the transmission.

Regulatory aspects are also important for the system design. They define which spectral resources are available. The amount of bandwidth available defines how much overhead can be tolerated for coding and diversity techniques. The carrier frequency used largely defines the propagation conditions; high carrier frequencies may have a severe impact on the propagation conditions making the design of the used transmission scheme more challenging. Licensed frequency bands may also come with other restrictions, including limitation on terrestrial repeater deployment, transmission power output etc. Spectrum licenses could also include fees and restrictions on coverage that also need to be taken into account in the system optimization.

Given the constraints from the technical requirements and the regulator, the goal of an overall technical design would be to find the system infrastructure that minimizes the cost. This optimization process could be performed by choosing a system infrastructure, including satellite and terrestrial transmitter setups as well as end user equipment characteristics, model the behavior of the various communication channels, and try to find a transmission scheme that satisfies the given constraints, see Figure 1.2. Depending on if the requirements can or cannot be met, more relaxed or more elaborate system configurations can be considered until a well optimized design has been found.

Producing an overall technical design according to the task above is a challenging task from a communication viewpoint. While techniques for estimation of terrestrial propagation conditions and design for mobile terrestrial broadcast networks are well established [RZ93], corresponding methods for mobile satellite communication are less developed. First, for the optimization of a transmission scheme employing time and space

diversity techniques, having a detailed understanding of the long term propagation statistics is essential. Considerable efforts have during the last decades been put into improving our understanding of the mobile satellite channel, see e.g. [LCD+91, LB98, ITU03]. Even though some interesting work exists on predicting the mobile channel long term statistical properties without explicit channel measurements [STE01], efficient methods for predicting these properties based on different satellite orbits and geographical terminal locations appear to still need more investigation. Second, detailed and time-consuming simulations are needed to fine tune and verify the system design. However, for an iterative design process as that outlined in this section, efficient techniques for performance prediction and optimization of transmission schemes suitable for mobile satellite broadcast are essential. This problem is addressed in the first part of this thesis.

1.4 Multi-Input, Multi-Output Communications

As terrestrial wireless communication systems become more popular and more demanding in terms of data rates, their design becomes increasingly challenging. Traditionally, there are only a few options to allow for more users and higher throughput. One way is to allocate a larger frequency band, however the amount of useful spectrum is scarce. By increasing the transmit power, the number of bits that can be reliably transmitted over a certain bandwidth can be increased. Unfortunately, increasing the output power is not attractive for battery operated devices and for multi-user systems this will in general also lead to increased interference levels and no improvement in system capacity. Thus, to improve the situation without adding additional access points, the actual communication channel needs to be improved.

One way to improve the channel is to employ multiple element antennas, antenna arrays, at both transmitter and receiver. This way a multi-input multi-output, MIMO, system is formed. By adding this additional infrastructure it is possible to create parallel spatial channels allowing for tremendous increase in spectral efficiency [Win87, PK94, FG98, Tel99]. In this thesis we investigate methods for analysis and design of MIMO systems. The work is based on statistical models describing channel imperfections and channel estimate uncertainties. We limit ourselves to the analysis of a single communication link, interference and other aspects of

multi-user systems are not considered.

In this section some intuition into the gain from using MIMO is provided and we discuss why the work in this thesis is of relevance.

1.4.1 The MIMO Link Capacity Gain – Intuition

By being able to exploit spatial properties of the communication channel MIMO communication systems have the potential to provide several times the capacity of single antenna systems. The idea behind this gain is illustrated in Figure 1.3. Here, communication between an n_{t} antenna transmitter and an n_{r} antenna receiver is considered. For simplicity in this example we will assume that $n_{\mathrm{r}} \geq n_{\mathrm{t}}$. Note however that such a constraint does not have to be enforced in general. In all cases the different signals transmitted over the transmitter antenna elements share the same radio resource and hence these will interfere at the receiver antenna elements.

First, consider system (a). In this case the antenna elements of the transmitter are well separated and there is line of sight propagation between the transmitting and the receiving antenna elements. Such a system could for example correspond to a number of satellites communicating with a multi-element antenna terminal. This will be further discussed in the satellite broadcast part of this thesis. It could also correspond to a number of single antenna terminals communicating with a multi-antenna base station. By combining the signals impinging on the different elements of the receiving array in an elaborate manner, the reception of the receiving array can be made directive. This means that signals transmitted from the different elements of the transmitting array can be separated based on their spatial properties. Thus, in this scenario n_{t} parallel channels can be formed, each using the same spectral resources. To compare this MIMO system with its single element antenna counterpart, the propagation conditions must also be taken into account. If the total power output should be the same, each transmitting antenna in the MIMO case would only be allowed to transmit with $1/n_{\mathrm{t}}$ of the transmit power of a single antenna system. As n_{r} copies of the signal can be combined coherently at reception, the signal to noise ratio of each received transmission will stay at least as good as in the single antenna scenario. This means that for system (a), n_{t} parallel channels in space can be formed, each with identical bandwidth and at least as good reception quality as in the single antenna scenario. Thus the total amount of data that can be transmitted using a system of type (a) is approximately n_{t} times higher

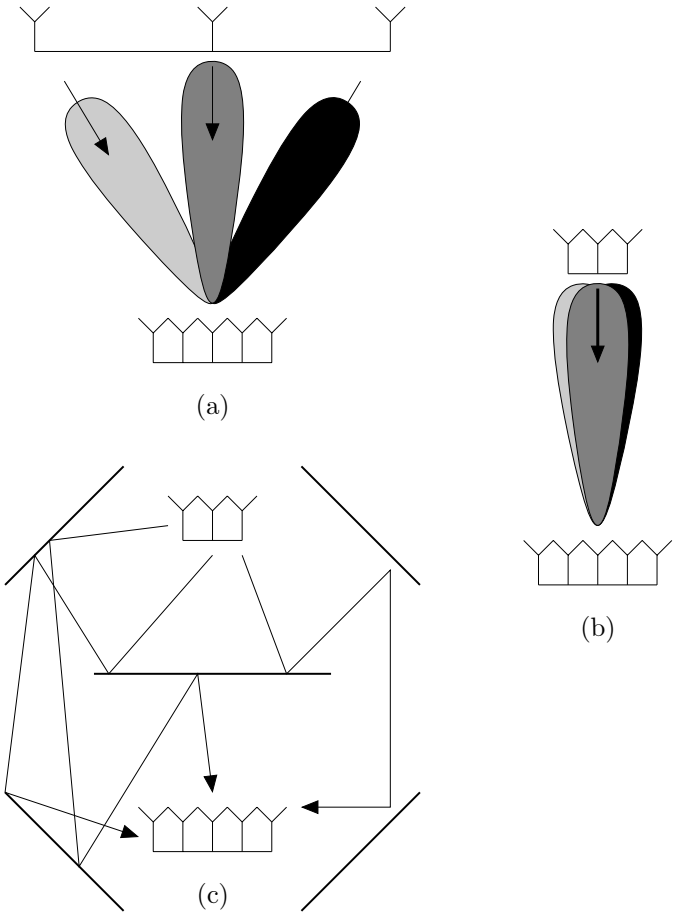


Figure 1.3: Three multi-element antenna communication systems. System (a), well separated antennas on the transmitter site – spatial channels can be formed using directional properties of the receiving array. System (b), the aperture limits the resolution of the array – spatial channels cannot be formed and the capacity gain from using the arrays is limited. System (c), no line of sight propagation – multi-path propagation can be used to resolve spatial channels. Note that provided that the transmitter has access to estimates of the communication channels, identical spatial channels may be created also if the direction of communication is reversed.

than in the single antenna scenario.

Second, consider system (b). In this case the antenna elements of transmitting array are not well separated. As the resolution of the antenna array is limited by the aperture size, for reasonably sized arrays separating the transmitted signals based on their spatial properties is in general not feasible. Thus in this case techniques applicable for system (a) cannot be applied and in general the gain is limited. Thus, systems of type (a) would provide an enormous capacity gain, but require arrays with physical dimensions which are not practical for all scenarios. Systems with realistic systems dimensions, here exemplified by (b), do not provide the same gain in performance. Fortunately, if multi-path propagation is taken into account, useful system configurations exist also for reasonable array dimensions.

Finally, consider system (c). In this case the antenna apertures are small, and there is no line of sight propagation between the transmitting and receiving array. Instead the transmitted signal will be scattered by the environment surrounding the antennas and the signal will propagate along multiple paths. An example of a system typically corresponding to (c) would be an indoor wireless local area network. As the different signals arrive at the receiving antenna elements, due to small differences in propagation paths, they will superposition incoherently. Provided there is sufficient scattering around the communicating devices it will again be possible to resolve n_{t} spatial channels between the transmitter and the receiver. Again, with a total output power constraint, each spatial channel will offer similar characteristics as the single spatial channel formed by a single element antenna system. Thus by parallelizing the transmission the data rate can be increased on the order of n_{t} times.

1.4.2 The MIMO Communications Channel

The MIMO communication channel can be characterized by the phase and attenuation properties of the individual spatial channels between the different transmit and receive elements. Depending on the system setup different models are applicable. For examples consider the systems of Figure 1.3.

For (a) and (b), given that line of sight propagation dominates, each spatial channel should provide a similar attenuation and only differ in phase due to the differences in path length. For (a) the difference in received phase characteristics between the transmitting signals is sufficient for the receiving array to separate the signals based on their spatial

properties while for (b) this is not the case.

A signal transmitted from one of the transmit antennas arrives at one of the receiving elements of system (c) along a large number of paths. Such channels are commonly modeled as Rayleigh fading, i.e. the attenuation and phase realizations of the spatial channels are drawn from Rayleigh and uniform random distributions respectively. Under ideal conditions, with richly scattering environments surrounding both arrays, the fading process of the channels between the different transmit and receive antennas can be considered independent. This way, the channel responses corresponding to the different transmitter antennas will likely be sufficiently different to allow efficient signal separation.

1.4.3 System Imperfections

In general the data rate performance of a MIMO communication system can be improved if the transmitter has access to estimates of the channel [Tel99]. However, maintaining transmitter channel estimates may be difficult if the channel is changing for example due to terminal or environment mobility. Also, for systems such as (c) in Figure 1.3, the assumption of uncorrelated fading is not realistic. In practice, depending on the amount of scattering surrounding the antenna sites, there will always be some correlation in the fading. This illustrates two of the imperfections that need to be taken into account in the analysis and design of this type of system.

Different countermeasures can be designed to overcome system imperfections such as those discussed above. For example, while it might not be practical to maintain perfect channel estimates at the transmitter, updating parameters that changes on a slower time scale may be feasible. Such parameters could for example include statistic properties of the channel which change on a slower pace than the individual channel realizations. If the effects of different statistical properties can be analyzed in an efficient manner, design schemes based on these properties can be devised, improving the system reliability and performance. This problem is addressed in the second part of this thesis.

1.5 Thesis Outline

This thesis treats analysis and design techniques for wireless communication. In particular diversity techniques are studied in conjunction

with statistical channel models. Two specific applications are studied, mobile satellite broadcast and terrestrial multiple-input, multiple-output (MIMO) communication. The remainder of the thesis is organized in three parts.

1.5.1 Part I, Mobile Satellite Broadcast

A mobile broadcast infrastructure, using satellite, terrestrial, and time diversity techniques, is considered. As terrestrial mobile broadcast is relatively well understood, emphasis is placed on areas where no terrestrial coverage is available. Based on statistical models, efficient techniques for estimating the service availability and optimizing system resources are derived. The models considered are applicable to vehicular users and characterize the temporal behavior of the channel by the average time and frequency the user spends in different reception quality conditions. Another contribution in this part of the thesis is a discussion on how the transmitters of the system can be multiplexed to provide spatial diversity in a spectrally efficient manner.

This part of the thesis is organized as follows. Chapter 2 presents a vision for the considered end-to-end system, outlines the part in detail and states contributions which are also related to other work in the field. Fundamental limitations of different methods of organizing the transmit diversity are discussed in Chapter 3. This organization affects the achievable data rates of the system. We argue that by using a more sophisticated receiver setup the spectral efficiency can be radically improved compared with techniques in use today. In Chapter 4 two transmission schemes with corresponding system models are proposed. Based on these models, an efficient technique for estimating the service availability is introduced in Chapter 5. Having an efficient method of estimating the system performance, elaborate design schemes can be devised. This is illustrated in Chapter 6. Here several design schemes are discussed, for example we show how trade-offs between delay and coding overhead can be analyzed and how different transmission schemes compare. Finally, Chapter 7 provides some concluding remarks and technical insights to an end-to-end system design. Areas of future work are pointed out and we discuss issues not covered by our analysis.

1.5.2 Part II, MIMO Communications with Partial Channel State Information

In this part a communication link between a transmitter and a receiver both employing multiple antennas is studied. In particular we are interested in system imperfections such as correlation in the fading of the different spatial channels and access to non-perfect channel state information at the transmitter. These issues are here modeled in statistical terms. Based on an asymptotic assumption we derive techniques for analyzing the impact of spatial correlation on the system performance. A practical scheme for exploiting non-perfect transmitter channel state information is also proposed and evaluated.

Chapter 8 provides a detailed outline of this part including contributions and relationship with earlier work. An overview of the considered system and models used in the later chapters are presented in Chapter 9. Based on an asymptotic assumption on the number of antennas on either side of the system, Chapter 10 derives statistical properties of the MIMO channel eigenvalues. These eigenvalues can be seen as the channel gain in different directions and largely determines the achievable system performance. The usefulness of the results in Chapter 10 is illustrated in Chapter 11 which derives approximations of the mutual information and the outage capacity of the correlated MIMO channel. The derived approximations are on a simple form which is useful for further analysis and design. This is shown in Chapter 12, where the earlier results are used to optimize the transmission scheme to approach capacity. Inspired by the results in Chapter 12, in Chapter 13 an efficient method for adapting a more realistic transmission scheme to non-perfect transmitter channel state information is proposed. Finally, Chapter 14 concludes this part and provides some ideas for further work.

1.5.3 Part III Epilogue

The thesis is summarized and common points between the two parts discussed.

Appendix 1.A Contributions Outside and Extensions Beyond This Thesis

In this section contributions outside the thesis and extensions of the work herein are presented.

1.A.1 Joint Channel Estimation and Data Detection for Interference Cancellation

During the last decades, several methods for suppressing interfering signals using antenna arrays have been proposed, [PN98, PP97]. Most interference suppression methods can be divided into two steps. First, the wireless channel of the desired user is estimated along with the channels or some statistics of the interfering signals. Second, the desired data is detected while the interfering signals are suppressed. While such methods have computational advantages, joint channel estimation and data estimation methods, where data and channel characteristics are estimated simultaneously, can show superior performance [CP96]. Examples of algorithms for joint channel estimation and detection include [TVP96, vdVTP97, AdCS98]. However, there are issues that still require further attention.

For example, most current communication standards provide training information, i.e. predetermined symbols included in the transmissions. From [dCS97] it is clear that a semi-blind receiver that considers both the known training information and the unknown data simultaneously achieves better performance than a receiver that only considers the training information or only considers the unknown data. Also, the training information used in the wireless systems of today is designed to make the channel estimation at the receiver as simple as possible. However, by designing this redundancy more elaborately a system could achieve better performance. In [SGP02], a scheme for combined channel estimation, equalizing and coding is proposed, this includes a joint design of training information and error correction coding. Unfortunately, the proposed scheme is highly computationally complex. If receivers and training schemes could be invented that allow for efficient and robust implementations while providing better performance and smaller training information requirements the wireless networks of today could be improved.

A scenario with a one antenna terminal communicating with a multi-antenna base station is considered. To evaluate the possible performance of such a system when generalized training information and non-synchronized interference is present, a simple semi-blind iterative joint estimation and detection algorithm and a Cramér-Rao bound are proposed. The estimation and detection algorithm is an extension of the ILSP-algorithm [TVP96] that allows for inter-symbol and non-synchronized co-channel interference and that also incorporates generalized training information in a natural fashion. The derived Cramér Rao bound is a modification of the bound derived in [dCS97] that allows for non-synchronized interference as well as the generalized training scheme. This work was presented in,

> Cristoff Martin and Björn Ottersten. Joint channel estimation and detection for interference cancellation in multi-channel systems. In *Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing*, 2000.
>
> Cristoff Martin and Björn Ottersten. On robustness against burst unsynchronized co-channel interference in semi-blind detection. In *Asilomar Conference on Signals, Systems and Computers*, 2000.

and also appeared in the licentiate thesis [Mar02].

1.A.2 Semidefinite Programming for Detection in Linear Systems – Optimality Conditions and Space-Time Decoding

Optimal maximum likelihood detection of finite alphabet symbols in general requires time consuming exhaustive search methods. The computational complexity of such techniques is exponential in the size of the problem and for large problems sub-optimal algorithms are required. In this work, to find a solution in polynomial time, a semidefinite programming approach is taken to estimate binary symbols in a general linear system. A condition under which the proposed method provides optimal solutions is derived. As an application, the proposed algorithm is used as a decoder for a linear space-time block coding system and the results are illustrated with numerical examples. This work was presented in

> Joakim Jaldén, Cristoff Martin, and Björn Ottersten. Semidefinite programming for detection in linear systems – optimality condi-

tions and space-time decoding. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2003.

1.A.3 Performance Analysis for Maximum Likelihood Detection of Linear Space Time Block Codes

To adapt the transmission to the channel state information currently available it is important to be able to efficiently estimate the performance of the transmission scheme. This is illustrated in Chapter 13 where the data rate is maximized by adapting the transmitted data to the channel knowledge available at the transmitter. In the design process a simple and efficient method to predict the error rate performance is required. The estimator used in Chapter 13 has its limitations however. For example, the channel statistics must follow a certain structure and by ignoring self interference the performance of the estimator is sometimes limited. These aspects have been addressed in

Svante Bergman, Cristoff Martin, and Björn Ottersten. Performance analysis for maximum likelihood detection of high rate space-time codes. *IEEE Trans. Signal Processing*, 2005. Submitted.

and

Svante Bergman, Cristoff Martin, and Björn Ottersten. Bit and power loading for spatial multiplexing using partial channel information. In *ITG Workshop on Smart Antennas*, April 2004.

This work allows for more general transmitter channel state information models than what was considered in Chapter 13. By including the effects of crosstalk between different spatial carriers the performance of the estimator is also improved. Also more general linear dispersive space time block coding schemes are allowed.

Appendix 1.B Notation

$\boldsymbol{X}$, $\boldsymbol{x}$	$\boldsymbol{X}$ is a matrix and $\boldsymbol{x}$ is a column vector.
$\boldsymbol{X}^{\mathrm{T}}$, $\boldsymbol{X}^{*}$	The matrix transpose, the conjugate transpose of the matrix $\boldsymbol{X}$.
$\mathbb{C}$	The set of complex numbers.
$\boldsymbol{I}$	$\boldsymbol{I}$ is an identity matrix.
$\boldsymbol{0}$	$\boldsymbol{0}$ is a zero matrix.
$\boldsymbol{X} \otimes \boldsymbol{Y}$	The Kronecker product of $\boldsymbol{X}$ and $\boldsymbol{Y}$ [Gra81].
$\operatorname{vec} \boldsymbol{X}$	The columns of $\boldsymbol{X}$ stacked in a vector.
$(\boldsymbol{X})_{ij}$	The ith, jth element of the matrix $\boldsymbol{X}$.
$\lvert\boldsymbol{x}\rvert$	The norm of the vector $\boldsymbol{x}$, $\lvert\boldsymbol{x}\rvert^2 = \boldsymbol{x}^*\boldsymbol{x}$.
$\lVert\boldsymbol{X}\rVert_{\mathrm{F}}$	The Frobenius norm of the matrix $\boldsymbol{X}$, $\lVert\boldsymbol{X}\rVert_{\mathrm{F}}^2 = \sum_i \sum_j \lvert(\boldsymbol{X})_{ij}\rvert^2$.
$\operatorname{Tr} \boldsymbol{X}$	Trace of $\boldsymbol{X}$, $\operatorname{Tr} \boldsymbol{X} = \sum_k (\boldsymbol{X})_{kk}$.
δ_{kl}	Kronecker delta function, $\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ otherwise.
$\mathrm{E}\{X\}$	The expected value of X.
$\mathcal{N}(m, \sigma^2)$	Normal distribution with mean m and variance σ^2.
$\mathcal{CN}(m, \sigma^2)$	Circularly symmetric complex normal distribution with mean m and variance σ^2.
$\xrightarrow{\mathcal{L}}$	Convergence in law (distribution).
$\Pr(X < x)$	The probability that X is smaller than x.

Appendix 1.C Abbreviations

AM	Amplitude Modulation
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase-Shift Keying
CDMA	Code Division Multiple Access
DAB	Digital Audio Broadcasting
DVB	Digital Video Broadcasting
DRM	Digital Radio Mondiale
FDMA	Frequency Division Multiple Access
FM	Frequency Modulation
GSM	Global System for Mobile Communications (originally Groupe Speciale Mobile)
IID	Independent Identically Distributed
ILSP	Iterative Least-Square with Projection
LDPC	Low-Density Parity-Check
LOS	Line Of Sight
MIMO	Multiple-Input, Multiple-Output
NLOS	Non-Line of Sight
OFDM	Orthogonal Frequency-Division Multiplexing
PHS	Personal Handyphone System
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
RDS	Radio Data System
SDMA	Spatial Division Multiple Access
SFN	Single Frequency Network
SNR	Signal-to-Noise Ratio
STBC	Space-Time Block Code
TDMA	Time Division Multiple Access
ULA	Uniform Linear Array
UMTS	Univeral Mobile Telecommunications System
WLAN	Wireless Local Area Network

Part I

Mobile Satellite Broadcast

Chapter 2

System Vision and Contributions

In this part of the thesis a mobile satellite broadcast system is considered. To overcome the channel impairments that characterize this type of system, space, terrestrial, and time diversity techniques are suitable. The main contributions are efficient methods for the analysis and design of suitable transmission schemes. For transmit diversity broadcast systems we also discuss the fundamental limitations for different multiplexing options.

This chapter is organized in two sections. In the first an overview of the envisioned mobile satellite broadcast system is given. The second outlines the rest of this part of the thesis and also relates our work to earlier results.

2.1 End-to-end System Vision

The purpose of the broadcast system considered in this part of the thesis is to deliver content data to mobile receivers. The delivery of the content should satisfy certain quality of service requirements. Examples of such requirements include service availability and maximal delay. The design of the system should optimally satisfy this purpose at the smallest possible cost. Here, a hybrid satellite-terrestrial broadcast system is considered. The system consists of a number of components as illustrated in Figure 2.1. In this section these are presented in some detail to set the

stage for the remaining chapters.

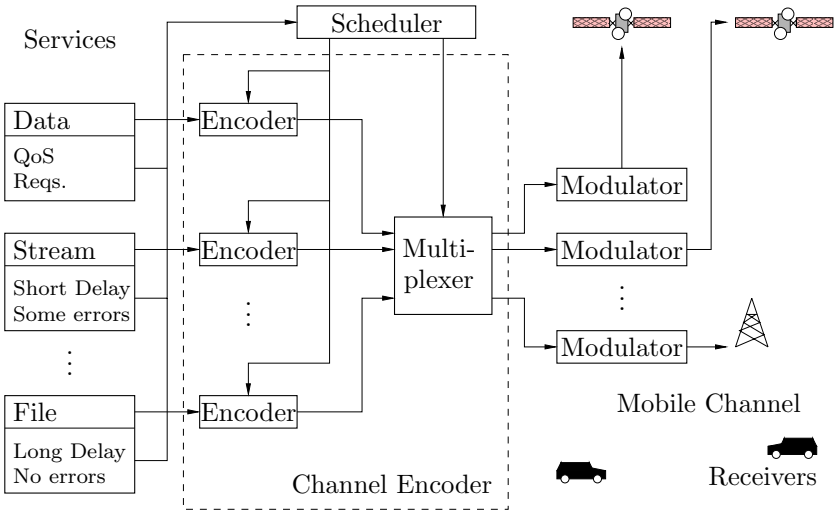


Figure 2.1: An end-to-end mobile satellite broadcast system with transmit diversity.

2.1.1 Services

The system may be used to transfer several different types of data content. Two natural classes of data are streams and files. A stream would provide a continuous service like traditional audio and video. A file distribution service could provide traffic and other travel information [Mar03], updates for vehicle software, maps etc. Files can also be used to distribute audio and video content which can be played out according to the preferences of the end user or according to pre-made play-lists [EHS+04].

Every piece of content is associated with quality of service requirements as discussed above. These might be fixed for different classes of content, or dynamically updated. For example, it might make sense to give priority to important emergency messages which may require shorter delay and higher reliability than other content.

It should be noted that in Figure 2.1 the content part is simplified. In the figure the content is assumed to already have been source coded. For

streaming services, to exploit the available data transportation capabilities efficiently, variable rate source coders are normally used [FJK+02]. As many content streams are multiplexed together this means that on average the transmission resource can be used more efficiently. However, given that this resource is finite the different source encoders must be coordinated.

2.1.2 Channel Encoder

The channel encoder can be seen as consisting of two parts. The first part is an encoder which applies an error correcting code to protect the content data from noise and channel fades. The amount of coding overhead should correspond to the quality of service requirement of the content. The second part is the multiplexer, the purpose of which is three-fold. First, it divides the communication channel into virtual sub-channels, one for each input data source. Second, it divides the encoded data stream into multiple data streams, one for each transmitter in the system. Last, by allowing multiple codewords to be transmitted simultaneously, the transmission time for each codeword is prolonged. This way it is possible to create an *outer interleaver* of a length that is sufficient to handle also the long term channel impairments of the mobile satellite channel. Again the exact settings of the multiplexer should be set according to the service requirements. The principle functionality of the multiplexer is illustrated in Figure 2.2.

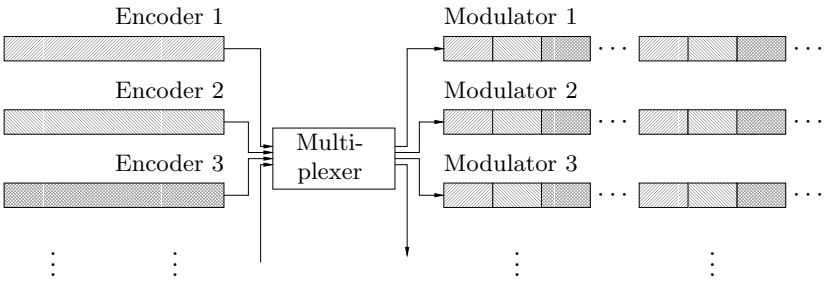


Figure 2.2: Principle operation of the multiplexer.

For the encoder, the idea is to protect the smallest possible data unit useful for the receiver applications. For file delivery type of services a data unit would correspond to a complete file while for streaming type of services it could correspond to one content frame. This design choice

is motivated by the fact that most files are useless if they contain errors and error concealment techniques for streaming services in general discard partly erroneous frames anyway [FJK+02].

Several multiplexing schemes can be considered to allow multiple services to use the same physical channel. In the systems deployed over the continental US the content data is multiplexed in time. The multiplexing used in the system in deployment over Japan and Korea is based on multichannel direct sequence spread spectrum techniques. Such techniques have also been considered in the literature, see e.g. [GG94, WLS99]. In this thesis, a time multiplexing scheme will primarily be considered. The multiplexer divides the incoming data units into segments which are multiplexed in time and sent to the modulators corresponding to the different transmitters. If only the reception of a single data unit is considered the transmission of this data unit will therefore appear bursty, see Figure 2.2. We believe that this is more favorable than for example code division multiplexing for the following reasons:

1. Using time division multiplexing allows for modulation schemes with low peak to average power ratios. Thus the power amplifier in the satellites can be operated near saturation.

2. Because of the burstyness of the transmission, assuming the receiver is only decoding a fraction of the data units, parts of the receiver can be turned off while waiting for data of interest. This would be similar to the time-slicing schemes used in e.g. DVB-H [VP04, ETS04b] and allows battery operated receivers to reduce their power consumption.

3. Last but not least, as we show in Chapter 5 and Chapter 6, by carefully designing the time multiplexing, exploiting the correlated nature of the mobile satellite channel, the service availability can be improved.

It should be noted that many of the techniques presented in this part of the thesis still are applicable to other types of multiplexers. In Chapter 4 transmission schemes suitable for mobile satellite broadcast will be considered in more detail.

2.1.3 Modulation and the Mobile Channel

The data streams at the output of the channel encoder are modulated and transmitted over the different transmitters. Thus, the different data

streams are multiplexed on the wireless channel. The modulation used should be chosen according to the channel and transmitter characteristics. For example, in the North American systems, QPSK modulation is used for the satellite transmission and OFDM for the terrestrial component [FJK+02]. This way, the power amplifiers in the satellite can operate efficiently while the terrestrial repeaters can be organized in an efficient single frequency network and frequency-diversity can be easily exploited at the receiver. The multiplexing of the different transmitters must also allow for efficient recombination in the receiver. Several options are available, in the North American system the different transmitters are separated in frequency. In the Japanese/Korean system, the satellite and the terrestrial transmitters share the same radio resource. The transmitter multiplexing scheme used has implications for the achievable system performance. This will be discussed in more detail in Chapter 3 and Chapter 7.

The mobile satellite channel is characterized by long term events caused by shadowing from objects in the terminal environment. How severe these events are depends on a number of factors including carrier frequency and the type of obstacle. The elevation angle to the satellite is also important. A large elevation angle means that channel blockages are less likely and of shorter duration. Unfortunately, if geo-stationary satellites are used the elevation angle will be small unless the coverage area is close to the equator. This has prompted an interest in alternative satellite orbits such as the highly elliptical orbit used by the Sirius system over the continental US. In the literature large constellations of low-orbit satellites have also been considered. Different orbit alternatives and their impact on mobile satellite broadcast are discussed in [Pro00]. No matter which satellite orbits are chosen, for efficient system analysis and design, channel models that are manageable yet accurate are essential. The statistical channel models considered in this thesis are described in Chapter 4.

2.1.4 Receivers

In the receiver, the signals from the different transmitters are received and recombined, providing the end user with the desired service. In this thesis we will primarily analyze systems targeting moving vehicles. Stationary and slow moving receivers are assumed cooperative, ensuring their antennas are located in acceptable receive conditions.

2.2 Outline and Contributions

Below this part of the thesis is outlined in some detail. Main contributions are stated and related to earlier work. Where applicable we also point out where our results have been published.

In the thesis we mainly address the channel encoder, the transmit diversity multiplexing and the effects of the mobile channel on the reception, see Figure 2.1. Thus we will not address issues regarding source coding and transmission scheduling. The remainder of this part consists of five chapters. In the first, different methods of multiplexing the signals from the different transmitters to allow efficient recombination at the receiver are discussed. We propose the use of a small receiver antenna array to separate the transmitted signals in space. Such a solution could potentially provide large gains in spectral efficiency. In the following three chapters, transmission schemes suitable for the mobile satellite broadcast channel are analyzed and designed. Our results allow for estimation of the impacts of time and transmit diversity techniques and we believe these are useful in a design process as that outlined in Section 1.3.4. The final chapter provides conclusions, insights to system design, ideas for further work and a critical discussion.

2.2.1 Chapter 3, Transmit Diversity Multiplexing: Fundamental Limitations

To exploit transmit diversity the signals emitted by the different transmitters must be designed such that they can be easily recombined at reception. By using frequency multiplexing as in the systems currently deployed over the continental US [FJK+02], exploiting this spatial diversity at the receiver is straightforward. Transmitting the same information at different frequencies also reduces the limited and expensive spectrum available for this type of service. By configuring the different transmissions as a single frequency network, in principle a more efficient usage of the spectrum can be achieved. Such a configuration is considered for the Japanese/Korean system currently in commercial deployment [MSS99]. Due to the vast distances involved in satellite diversity systems, practical design of a single frequency network is challenging. Here we discuss the fundamental limitations of various multiplexing schemes and we propose the use of a small adaptive antenna at the receiver to separate the different transmissions and improve the spectral efficiency. System architectures are outlined, an example illustrates the potential gain and

implementation issues are discussed. This discussion has previously appeared in,

> Cristoff Martin, Alexander Geurtz, and Björn Ottersten. Spectrally efficient mobile satellite real-time broadcast with transmit diversity. In *Vehicular Technology Conference*, September 2004.

2.2.2 Chapter 4, System Model

This chapter presents system models and transmission schemes suitable for mobile satellite broadcast systems employing time and transmit diversity.

Two transmission schemes suitable for transmission over the mobile satellite broadcast channel with diversity are presented. The first scheme, *symbol level coding*, provides time and space diversity by adding an error correcting code to the data and spreading the transmission over space and time. If the applied code is sufficient, the original data can be recovered even if parts of the data have been lost during the transmission. If extensive time diversity is provided, this scheme will require large amounts of memory in the receiver as for optimal performance soft values representing the likelihoods of the received symbols are required. That motivates the second scheme, *packet level coding*, where the transmitted data is divided into packets and two layers of coding are used to protect against channel impairments. An inner channel code protects individual packets and an outer erasure code protects if some packets are lost. This way decisions from the channel code decoding can be taken early relaxing the receiver memory requirements. Transmission schemes similar to the latter one have earlier been presented in e.g. [ESS04].

In [WM99] it is shown that if the interleaving used to communicate over a correlated fading channel is not perfect sometimes system performance can be improved by reducing the block size of the code. Analogously, in Chapter 6 we will show that for mobile satellite broadcast systems, depending on the acceptable delay in the system and the temporal channel characteristics, the segmentation of the transmitted data can be optimized to maximize service availability. Therefore we will allow for different segmentation strategies in both the transmission schemes outlined above.

To be efficient, mobile satellite communication system designs should be based on the characteristics of the mobile satellite channel. During the last decades considerable efforts have been made in measuring and

modeling such channels [LB98]. In this work diversity techniques are considered and as long term events, such as signal blockages due to obstacles surrounding the terminal, may last many seconds, models that include these temporal aspects need to be used. Here, these types of effects are modeled as a random process that defines transitions between different states where each state has a physical meaning such as line of sight, lightly shadowed, or completely blocked signal. Thus the channel is characterized by the statistical properties of the random process and the fast fading behavior due to multi-path propagation within each state. These properties depend on several factors including terminal environment, vehicle speed, carrier frequency and satellite elevation. In this chapter, models where the state transitions are modeled as a Markov chain are considered [LCD+91, VD92]. Such models have been shown to be reasonably accurate and, as we will show, allow efficient analysis.

2.2.3 Chapter 5, Service Availability Estimation

In this chapter we propose an efficient technique for estimating the satellite broadcast service availability for moving vehicular terminals. Based on the statistical channel models and the transmission schemes outlined in Chapter 4, we propose techniques that allow for quick evaluation of different transmission schemes, parameters, and system configurations. Thus the results presented in this chapter are useful in iterative design processes such as those envisioned in Section 1.3.4.

Our results are partially inspired by earlier work on correlated terrestrial wired and wireless channels [Cup69, YW95, WM99]. In these papers the performance of communication systems operating on channels with time correlation are studied. To model the correlation a two state Markov model is considered and the performance is estimated by computing the probability of being a certain number of times in each state of the chain. In this work, we extend these results to time-varying Markov chains with an arbitrary number of states and we also allow for more general channel models and coding concepts.

As a side result, in Appendix 5.A we derive an efficient recursive technique for enumerating Markov chains. Such techniques have previously been considered in for example [PB99], however we believe our derivation and results are significantly simpler and we also allow for time-varying transition probabilities.

2.2.4 Chapter 6, Design Examples

The design of mobile satellite broadcast networks has been the topic of several papers in the literature. In [GG94] the authors propose and analyze a system using code division multiplexing techniques to broadcast to mobile terminals using one transmitting satellite in a single frequency configuration with a network of terrestrial repeaters. The design does not foresee time diversity explicitly, but assumes that terrestrial diversity and a sufficient link margin can be used to provide satisfactory quality of service. Time diversity aspects are taken into account in [WLS99, LGW00] which consider mobile satellite broadcast using code and time division multiplexing designs. In these papers the time diversity is not designed using a statistical model describing the long term behavior of the channel but rather to handle signal blockages up to a certain length in time. Other interesting concepts such as using multi-layer audio source codes to reduce the tuning delay introduced by long interleavers are also proposed. In [ESS04] a file delivery system designed to overcome signal blockages is outlined. Files that are to be transmitted are divided into packets and two layers of coding are used to protect against channel impairments. An inner channel code protects individual packets and an outer erasure code protects the file if some packets are lost. By multiplexing the packets with packets from other files an extremely long *outer interleaver* can be created capable of protecting the data even from long term outages. This corresponds to the packet level coding scheme in this thesis.

In this chapter we illustrate how the analysis techniques derived in Chapter 5 can be useful in a system design process. Based on statistical properties of the mobile satellite channel different designs are proposed. For example we show how the coding overhead can be minimized while guaranteeing acceptable service availability and the trade-off between coding overhead and delay is illustrated.

Most of the results of Chapter 5 and Chapter 6 have been submitted as

> Cristoff Martin, Alexander Geurtz, and Björn Ottersten. Statistical analysis and optimal design of mobile satellite broadcast with diversity. *IEEE Trans. Veh. Technol.*, 2005. Submitted.

and parts of the results have also appeared in,

> Cristoff Martin, Alexander Geurtz, and Björn Ottersten. File based mobile satellite broadcast systems: Error rate computation and

QoS based design. In *Vehicular Technology Conference*, September 2004.

Cristoff Martin, Alexander Geurtz, and Björn Ottersten. Packet coded mobile satellite broadcast systems: Error rate computations and quality of service based design. In *European Workshop on Mobile/Personal Satcoms and Advanced Satellite Mobile Systems Conference*, September 2004.

2.2.5 Chapter 7, Concluding Remarks and Future Work

The contributions of the first part of the thesis are summarized, the results are critically discussed and opportunities for further work are outlined.

Chapter 3

Transmit Diversity Multiplexing: Fundamental Limitations

To avoid outages in real-time services provided by mobile satellite broadcast systems their design must overcome radio signal blockages from obstacles in the environment. In practice this can be accomplished using a combination of time and space diversity. While providing diversity is necessary to ensure a satisfactory service availability, providing it in general results in lower data rates as redundant information is transmitted. In this chapter the fundamental limitations of a mobile satellite broadcast system with space diversity are discussed. We argue that, similar to terrestrial MIMO systems, deploying a multiple element receiver allows for significantly improved spectral efficiency. While this discussion does not take time diversity and more detailed channel models into account as it is done in Chapter 5 and Chapter 6 the same relative comparison should still hold. We will also briefly touch upon this subject in Chapter 7.

This chapter is organized as follows. Section 3.1 provides an overview of different methods of multiplexing the signals from the different transmitters to facilitate efficient signal recombination at the receiver. Where applicable references to related systems are given. In Section 3.2 example multiplexing configurations are outlined. Fundamental limitations on the data rates the different systems can provide are discussed in Section 3.3. The discussion is concluded with a numerical example in Section 3.4.

3.1 General

The systems deployed today over the continental US [FJK+02, XMR05, Sir05a] use a combination of two transmitting satellites and a network of terrestrial repeaters to provide the necessary spatial diversity and to ensure coverage also in difficult environments. To separate the various transmitted signals at the receiver, the available frequency band is split in three approximately equal parts which are used by the two transmitting satellites and the terrestrial repeaters respectively, see Figure 3.1. In Japan and Korea a system employing a single satellite and an extensive network of gap-fillers is currently in commercial deployment [MBC05, MSS99]. As only a single satellite transmitter is used it is possible to organize the network in an efficient network configuration where all users share the same radio resource using direct-sequence spread spectrum techniques.

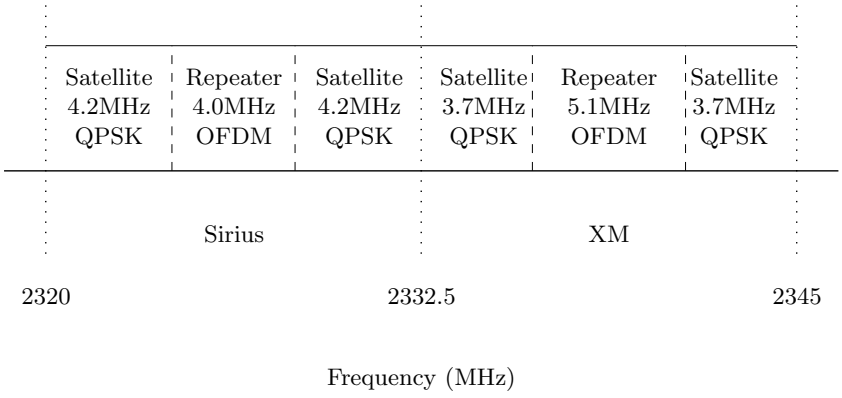


Figure 3.1: Frequency band allocation and use for the US mobile satellite broadcast systems [FJK+02].

To be able to provide mobile satellite broadcast services to affordable and small mobile terminals and to get reasonable propagation conditions usage of expensive and scarce licensed radio spectrum is necessary. Hence, providing such services via a highly efficient air interface is essential. Unfortunately, in current systems, providing satellite diversity means transmitting identical information over different frequency bands wasting this precious resource.

One way of improving the spectral efficiency of broadcast networks is

to let all transmitters share the same radio resource in what is called a single frequency network, SFN. This is for example supported by the terrestrial DAB and DVB standards for broadcast of audio and video [ETS01, ETS04a]. In principle, the spectral efficiency of mobile satellite broadcast networks could be improved by using similar techniques even though the vast distances between the transmitting satellites will make the design of such a system more complex and less efficient.

During the last decades, adaptive antenna solutions, where multiple antennas and transceiver chains are employed at the base stations and/or terminals, have received considerable attention in terrestrial wireless communications for their potential to allow different users to share the available radio resources more efficiently [PP97]. Commercial mobile phone base stations which exploit adaptive antenna solutions exist for the purpose of interference suppression and SDMA[1] in at least the GSM and the (East Asian) PHS systems [Eri00, Arr]. Recently, research has focused on systems where, by employing adaptive antennas on both sides of a terrestrial communication link, the data rate can be significantly increased [Tel99]. Adaptive antenna terminals have also been suggested as means of increasing data rates in broadcast systems allowing multiple broadcast stations to share the same radio resource [PK94].

In this chapter, we propose the use of small adaptive antennas at the receivers of a mobile satellite broadcast system as a means to exploit the satellite and terrestrial spatial transmit diversity and to improve the spectral efficiency. By employing multiple antennas and receiver chains at each terminal it is possible to separate the different impinging signals based on their spatial characteristics instead of their frequency, time or code. In principle, such a scheme is capable of providing several times the spectral efficiency of the techniques used today using identical satellite transmission power, link margins, time and spatial diversity. For comparison we outline other possible system configurations and the potential gain in spectral efficiency is illustrated with a simple example.

3.2 System Configurations

In this section a few different system configurations are discussed in the framework of a mobile satellite broadcast infrastructure consisting of s simultaneously transmitting satellites for space diversity and a single fre-

[1] SDMA – Spatial Division Multiple Access, in analog with FDMA, TDMA and CDMA.

quency network of terrestrial repeaters. For the service, spectrum of bandwidth $(s+1)B$ is available. Emitting the same information over all transmitters means that the broadcasted information can be recovered even if one or more signals are blocked but it does not necessarily imply that the same signal is transmitted. For example, the modulation formats over the terrestrial and satellite links may be adapted to the different channel conditions and for optimal performance the data transmitted over the different channels can be jointly coded to provide a coding gain on top of the diversity, see e.g. [FKKK01] and the transmission schemes outlined in Chapter 4. In all system configurations it is assumed that a common power output is available from each satellite and for the different setups a common total bandwidth is considered.

The various system setups are visualized with an example system consisting of $s = 2$ satellites and a single frequency network of terrestrial repeaters similar to the systems in commercial use over the continental US today, see Figure 3.2. In the examples, the total bandwidth available for the system is thus $3B$.

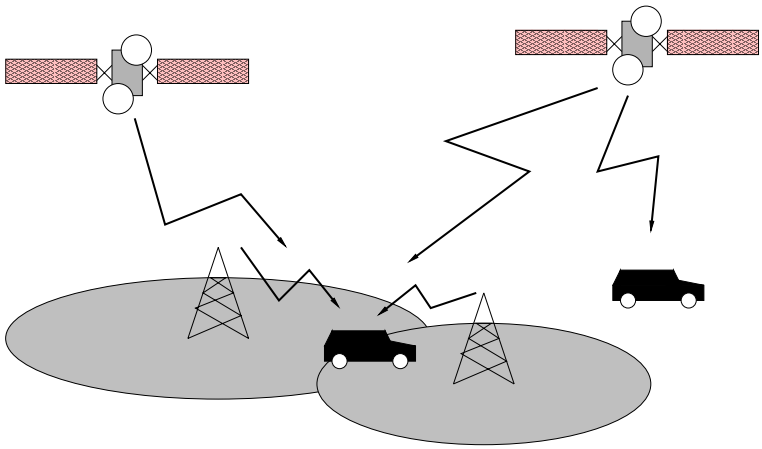


Figure 3.2: Example system setup consisting of $s = 2$ simultaneously transmitting satellites and a network of terrestrial repeaters in a single frequency network configuration. The system is designed so that the transmitted data can be recovered even if only a single satellite signal is received.

3.2.1 Multiple Frequency Network Setup

For reference a conventional system with the transmitted signals separated in frequency is considered. To simplify comparisons with more elaborate transmission setups, the satellite and terrestrial transmitters use equal bandwidths, i.e. the bandwidth available for each transmitter is B. The frequency band allocation for our example is illustrated in Figure 3.3.

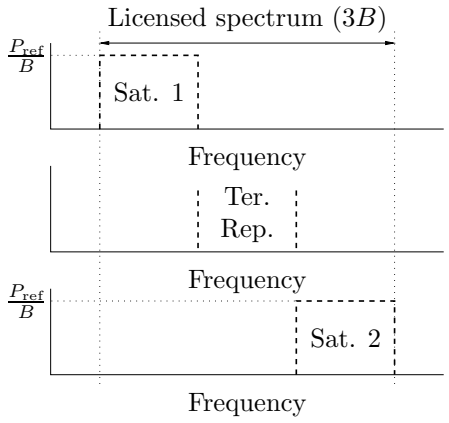


Figure 3.3: Reference single antenna, multiple frequency mobile satellite broadcast network. All transmitters transmit the same information (albeit possibly using different coding and modulation) for diversity. The y-axis illustrates the received signal power spectral density.

3.2.2 Single Frequency Network Setups

By using different frequency bands to transmit the same information, multiple frequency broadcast networks are wasting valuable bandwidth. Therefore, in general it is desirable for digital broadcast systems to operate in a single frequency network configuration where the available radio resource is shared amongst all transmitters.

In principle, one way to construct a single frequency network would be to let all transmitters transmit the same signal in the same frequency band. As different receivers are in different locations, the transmissions cannot be synchronized but the various signals will impinge with different delays. For the receiver, the different arriving signals appears as

inter-symbol interference and optimally, the originally transmitted information is recovered using maximum likelihood sequence estimation. In practice, due to the time dispersion of the channel, for high data rates such schemes would be extremely computationally demanding and therefore specially suited modulation schemes need to be employed. Examples of such schemes include orthogonal frequency division multiplexing, OFDM, and direct-sequence code division multiplexing.

In OFDM, the chosen modulation scheme for the terrestrial repeater networks in the XM and Sirius satellite broadcast systems; the data is parallelized and transmitted on a large number of frequency carriers resulting in a data rate per carrier much lower than the total data rate. Similarly, in direct-sequence spread spectrum single frequency networks, such as that chosen for the terrestrial repeaters and single satellite in the MBCO system [MSS99], the transmitted data is parallelized and transmitted using orthogonal spreading codes similar to the forward link of a DS-CDMA system. For these schemes to be effective the resulting symbol period on the parallel channels must be much larger than the time delay spread of the channel. Clearly, if the channel has considerable length this requirement may be difficult to fulfill. For example, in a mobile system the coherence time of the channel also needs to be taken into account.

As discussed above, the design of a single frequency network is simplified if the different signals can be synchronized on a symbol period level at reception. For this reason two single frequency network configurations are considered; one where all transmitters use and share the entire spectrum and one where the satellites are separated in frequency while the repeaters share their bandwidth with the satellites. This would mean that each satellite uses a bandwidth of either $(s+1)B$ or $(s+1)B/s$. In the second configuration the repeaters can synchronize with the signals transmitted from the satellites and an OFDM or direct-sequence code division multiplexing approach could be taken to implement the network in practice, see e.g. [GG94, MSS99]. For the example the bandwidth configurations are illustrated in Figure 3.4.

3.2.3 Proposed Adaptive Antenna Receiver Setup

Consider an adaptive antenna receiver consisting of n_{r} antennas each connected to demodulation and analog to digital converters. By combining the digital outputs from the different antennas in an elaborate manner, various goals can be achieved. For example, by using a maximum ratio combiner the signal to noise ratio can be improved and the receive diver-

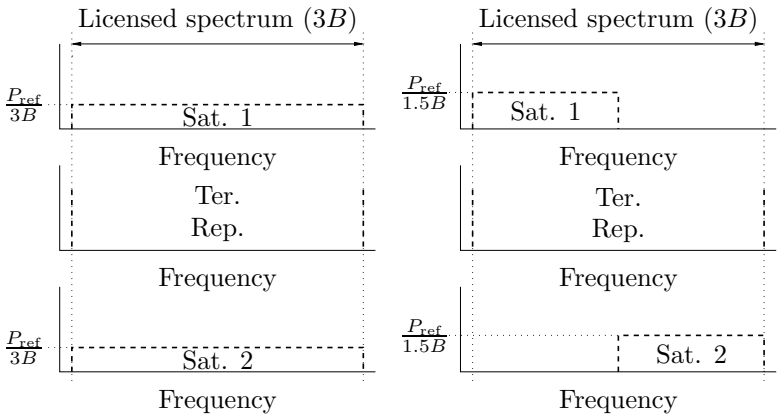


Figure 3.4: Bandwidth configuration in the two single frequency network configurations. The y-axis illustrates the received signal power spectral density.

sity exploited to overcome deep fades due to multi-path effects. Such a use of receive diversity has previously been considered for the land mobile satellite channel in e.g. [BSB+89, MTHK99] and similar ideas are implemented in the terminals of the MBCO system [MSS99]. In terrestrial cellular communications, adaptive antennas is an established technique on the base station side. Here the spatial properties of the antenna array are used to lower the interference level in the system, to reduce the frequency reuse distance and in some systems even to increase the network capacity using SDMA schemes. In SDMA systems several user terminals are allowed to share the same radio resource, e.g. the same time/frequency slot in a time division multiple access system. Instead the different terminals are separated at the base station based on spatial properties of the received and transmitted signals. Examples of commercial use of adaptive antennas today include interference suppression in GSM [Eri00] and SDMA in the East Asian PHS system [Arr].

Here the use of a small adaptive antenna at the receiver to improve the spectral efficiency of a mobile satellite broadcast system is considered. Two types of system configurations are considered.

First, a multiple-frequency configuration where the adaptive antenna is used to provide a diversity gain is considered. Here an increase in signal to noise ratio can be provided in stationary conditions and spatial diversity can also provide robustness against fast fading due to multi-

path. An identical bandwidth allocation as in the single antenna, multiple frequency network setup is assumed. For the configuration in our example see Figure 3.5.

Second, a receiver structure as that depicted in Figure 3.6 is considered. Here all signals are transmitted using the same radio resource and thus all transmitters are allocated a bandwidth of $(s+1)B$. In this single frequency setup the signals arriving from the different transmitters are separated at the receiver based on their spatial properties. If multiple signals are received they can then be recombined using maximum ratio combining or a space-time code for improved performance. In practice, the number of receive antenna elements should at least match the number of transmitted signals, i.e. $n_{\mathrm{r}} \geq s+1$. Notice the similarity of this approach with the return link in SDMA, Figure 3.7, and that well established techniques exist for the signal separation, [PP97]. Also, as the channels between the transmitters and receive antennas can be estimated using the received signals, no physical information regarding the position of the array, transmitters nor relative antenna element positions is necessary for the signal separation.

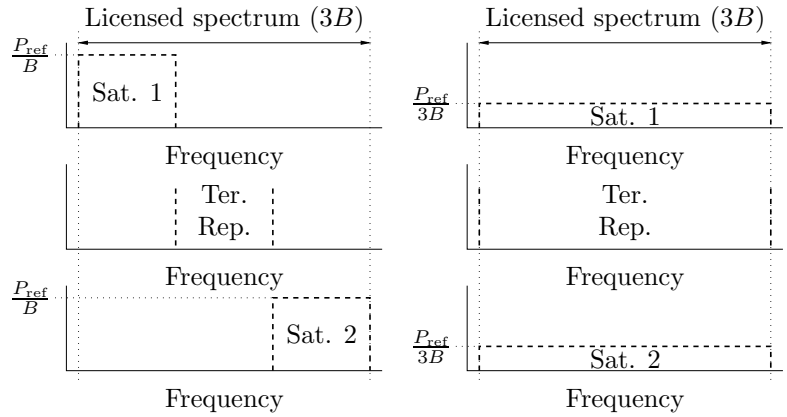


Figure 3.5: The example system in the two configurations considered for use with an adaptive antenna receiver. The y-axis illustrates the received signal power spectral density per receive antenna.

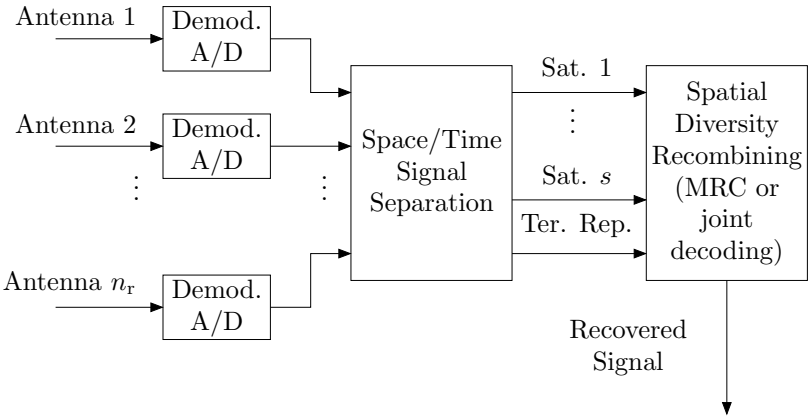


Figure 3.6: Principle functionality of an adaptive antenna single frequency satellite/terrestrial receiver.

3.3 Spectral Efficiency of the Different Configurations

The Shannon capacity of a channel is the highest data rate where data can, in principle, be communicated at an arbitrary low probability of error, see Section 1.2.3. Using modern coding schemes, such as turbo or LDPC coding, [BG96, Gal62, MN97], one can achieve system performance near the bounds of capacity and this measure is therefore of relevance. To estimate the spectral efficiency that can be achieved using the different multiplex configurations of Section 3.2, we assume that these broadcast systems are designed for a worst case scenario where only a single satellite signal is received. To keep the results simple a non-fading flat additive white Gaussian noise channel is used to model the propagation between this single satellite and the receiver. Thus, to estimate the achievable spectral efficiencies for the different system configurations, we compute the capacity of a flat AWGN channel, band-limited in the different system setups as indicated by the figures in Section 3.2. In the discussion below we will denote this channel the design channel. Note that as a mobile environment with non-directive antennas is considered in reality, the channel would fade due to multi-path. Here we assume that these kinds of effects are taken into account by providing sufficient link margin and interleaving. Either way, taking more realistic channel models into

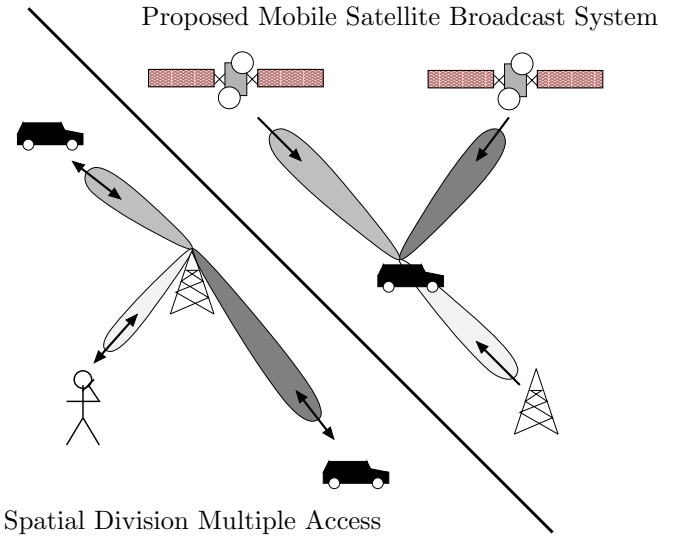


Figure 3.7: Comparison of spatial division multiple access and the adaptive antenna, single frequency broadcast network proposed in this chapter. Note that the broadcast network only requires directive reception and since the different signals contain identical information they can be recombined to improve performance.

account would only obscure the results as the relative performance never the less should be similar.

The capacity of an AWGN channel, band-limited to W with noise spectral density of $N_0/2$ and received average power limited to P is well known to be [CT91],

$$C = W \log_2 \left(1 + \frac{P}{N_0 W}\right) \quad \text{[bits/s]}.$$

Based on this result the capacity of the design channel can easily be derived for the different multiplexing configurations.

3.3.1 Design Channel Capacity Computation

In this section the Shannon capacity for the design channel is computed for the different system configurations. To be fair, the received signal

power per receive antenna in all configurations is P_{ref} and the noise spectral density is $N_0/2$.

First the reference multiple frequency system of Section 3.2.1 is considered. Here the bandwidth of the received satellite signal is B and the capacity of the design channel may be found as

$$C_{\text{ref}} = B \log_2 \left(1 + \frac{P_{\text{ref}}}{N_0 B}\right). \tag{3.1}$$

Now consider the single frequency network configurations, see Section 3.2.2. Here the bandwidths for the reference channels would be $(s+1)B$ and $(s+1)B/s$ respectively and the corresponding capacities are,

$$C_{\text{SFN1}} = (s+1)B \log_2 \left(1 + \frac{P_{\text{ref}}}{(s+1)N_0 B}\right) \tag{3.2}$$

and

$$C_{\text{SFN2}} = \frac{s+1}{s} B \log_2 \left(1 + \frac{sP_{\text{ref}}}{(s+1)N_0 B}\right). \tag{3.3}$$

Finally, the adaptive antenna system configurations from Section 3.2.3 are considered. For the system with the different signals separated in frequency, the bandwidth of the design channel is B. By combining the antenna outputs such that the satellite signal is combined coherently and the noise incoherently the received signal to noise ratio is increased n_{r} times. Thus the capacity for our design channel would be,

$$C_{\text{Adaptive}} = B \log_2 \left(1 + \frac{n_{\text{r}} P_{\text{ref}}}{N_0 B}\right). \tag{3.4}$$

When the transmitters are organized in a single frequency network fashion not only is n_{r} times the signal power available, but also $s+1$ times the bandwidth. We have

$$\begin{aligned} C_{\text{SFN,Adaptive}} &= (s+1)B \log_2 \left(1 + \frac{n_{\text{r}} P_{\text{ref}}}{(s+1)N_0 B}\right) \\ &\geq (s+1) C_{\text{ref}}, \end{aligned} \tag{3.5}$$

where the last inequality follows from the assumption that $n_{\text{r}} \geq s+1$ in Section 3.2.3 and equality holds if $n_{\text{r}} = s+1$.

3.4 Numerical Example

In Figure 3.8 the resulting spectral efficiencies of the different multiplexing configurations are illustrated as functions of the design signal to noise ratio for the single receive antenna, multiple frequency network reference system. For the adaptive antennas setups an $n_{\mathrm{r}} = 3$ element antenna array is used. Using more antennas would further improve performance and could also simplify the implementation of necessary signal processing. As this simple example illustrates, by changing the frequency configuration, significant gains in terms of data rate can potentially be realized. Especially for the two single frequency network configurations the gain can be huge. At the same time it should be noted that it is unclear how a single antenna single frequency network can be implemented in a multiple satellite infrastructure and that the gain from a multiple antenna system comes at the cost of a more bulky and expensive receiver.

In the above discussion the impact of time diversity has not been taken into account. The next couple of chapters are intended to provide some additional insights when such effects are included. In Section 7.1 we will again discuss the impact on the spectral efficiency when both time and space diversity is employed.

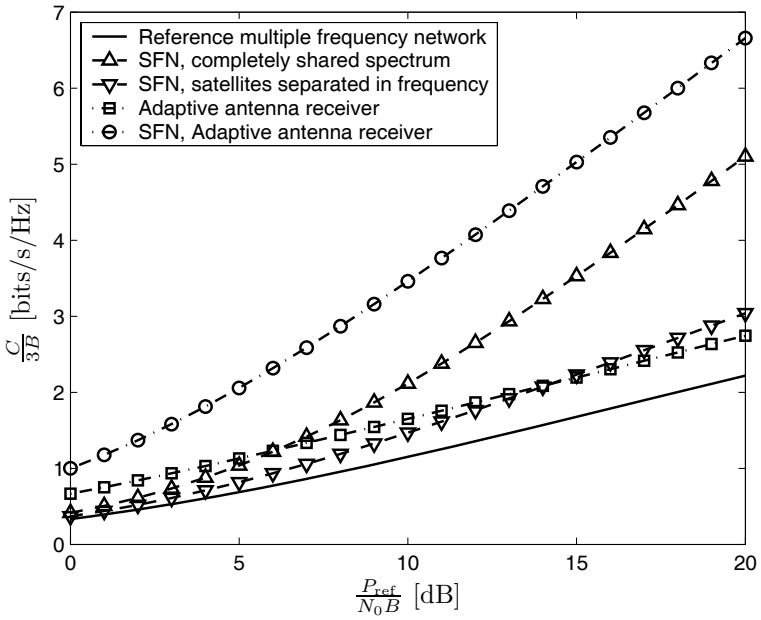


Figure 3.8: Normalized capacities, "spectral efficiencies," for the design channel of the various system configurations as functions of the design SNR for the single antenna multiple frequency reference system. In the example $s = 2$ satellites are simultaneously transmitting and $n_{\mathrm{r}} = 3$ antenna elements are employed in the adaptive antenna receivers.

Chapter 4

System Model

The mobile satellite channel is characterized by long term events in terms of severe and frequent blockages of the channel due to obstacles in the terminal environment. Thus the transmission scheme of an efficient broadcast system must be designed taking these long term correlation characteristics into account. This chapter presents two such transmission schemes with corresponding statistical channel models. The two transmission schemes both allow for time and transmit diversity techniques, but differ in performance and computational complexity. In Chapter 5 and Chapter 6 analysis and design techniques for the system model presented in this chapter are proposed.

Section 4.1 describes two transmission schemes suitable mobile satellite broadcast. To overcome channel impairments coded time and transmit diversity techniques are employed. In an example we illustrate how the segmentation of the transmitted data affects the system error rate performance. We will therefore consider the data segmentation as one design parameter in the transmission scheme. In Section 4.2 land mobile satellite channel models suitable for the transmission schemes are presented.

4.1 Transmission Schemes

In this thesis we consider transmission schemes designed to protect the smallest data unit useful for the receiver applications. For file delivery type of services a data unit would normally correspond to a complete file

while for streaming type of services it would correspond to one content frame. This design choice is motivated by the fact that most files are useless if they contain errors and error concealment techniques for streaming services in general discard partly erroneous frames anyway [FJK+02].

To overcome the channel impairments of the mobile satellite channel, diversity techniques are considered. Spatial diversity can be provided using multiple transmitting satellites and terrestrial repeaters. To provide time diversity, error correcting codes of rates lower than a line of sight link budget implies are used and the data transmission is spread in time. Two methods are studied, either time diversity is provided using low rate direct channel codes or by dividing the data unit in packets each protected by an inner channel code sufficient to overcome noise in near line of sight conditions and an outer erasure code to recover lost packets. In this paper these schemes are referred to as *symbol level coding* and *packet level coding* respectively. In both cases the transmission of a data unit consisting of u symbols multiplexed in time with data from other data units is considered.

4.1.1 Symbol Level Coding

Symbol level coding exploits time diversity by employing a modulation scheme and a channel code that provides an information rate r_{s} (bits per channel use) lower than the rate implied by a line of sight link budget and spreading the transmitted data over a time t_{delay}. If multiple satellites are used, the coded data is divided into multiple streams and transmitted over the different satellites. This scheme is illustrated in Figure 4.1. As we will argue in Section 4.1.4, in some scenarios it is favorable to not spread the transmitted symbols evenly in time. This is here addressed with a scheme where the data from a data unit is divided into k *segments* each consisting of c symbols before multiplexing with other data units. Using this multiplexing scheme, an outer interleaver can be created that is sufficiently long to cope with long term effects on the channel. Note that the case of symbol interleaving is included as a special case ($c = 1$) and that, if needed, the techniques presented in Chapter 5 can handle more general, non-uniform, distributions of transmitted symbols in time.

4.1.2 Packet Level Coding

A data unit transmitted using packet level coding is divided into $k - l$ packets, each consisting of p bits. To overcome blockages, l packets are

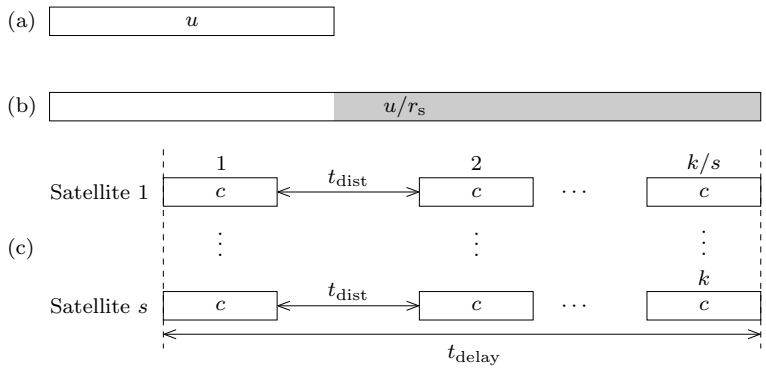


Figure 4.1: Symbol level coding scheme and transmission of a jointly coded data unit over s satellites. (a) Original data unit consisting of u bits. (b) Channel code providing an information rate r_s (bits per channel use) added, r_s is chosen smaller than a line of sight link budget would imply. (c) The coded data unit is divided into k segments, each consisting of c symbols. The segments are divided into one stream for each satellite, multiplexed with segments from other data units and transmitted, resulting in a time distance t_dist between each segment. This way a delay, t_delay, is introduced from the first to the last symbol transmitted by one satellite.

added with erasure coding, i.e. an erasure code of rate $r_{\mathrm{p}} = (k-l)/k$ is used. Each packet is modulated and individually channel coded resulting in an information rate r (bits per channel use) on the symbol level. The modulation and coding is assumed sufficient for a transmitted packet to be quasi error free provided that there are no signal interruptions for the duration of the reception of the packet. Time diversity is provided by spreading the packets evenly over a time t_{delay} between the first symbol of the first packet to the last symbol of the last packet. If multiple satellites are used, the packets are divided into one stream for every satellite and t_{delay} is the delay per satellite. The adaptation of the data for the channel is illustrated in Figure 4.2. Similar techniques for overcoming the impairments of the satellite mobile broadcast channel have previously been introduced in e.g. [ESS04].

4.1.3 Comparison of the Two Schemes

The symbol level coding scheme is optimal in the sense that it uses a single code to jointly overcome noise and blockages. Thus, if there are no blockages additional link margin is provided which may prove beneficial in cases where the terminal is in a shadowed state. This also avoids a trade-off problem present in the packet level coding scheme where one needs to decide how much of the total coding overhead should be used for direct channel coding and how much should be used to overcome blockages. On the other hand the packet level coding scheme has practical advantages. Decoding of the symbol level scheme requires soft values representing the likelihoods of the received symbols. If t_{delay} is large this represents significant amounts of memory in the receiver limiting the amount of time diversity that can be provided. If packet coding is used, decisions resulting from the channel code decoding can be taken earlier relaxing this requirement. The packet level coding scheme also makes it easier to handle data units of different sizes. The numerical examples in Section 6.6 illustrate how the techniques derived in this paper can be used to estimate the spectral efficiency resulting from choosing one of the transmission schemes. One could also imagine joint schemes where packet level coding is used as an outer code of a symbol level scheme. While such schemes are not pursued further herein, the techniques of Chapter 5 should be possible to extend to such a case provided that the transmissions of neighboring packets are not overlapping in time, this is somewhat elaborated on in Section 7.2.

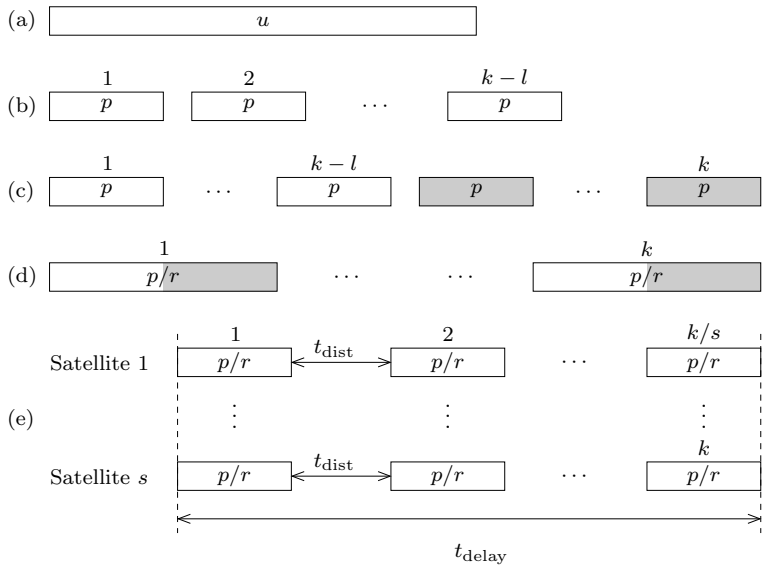


Figure 4.2: Packet level coding scheme and transmission over s satellites. (a) Original data unit consisting of u symbols. (b) Data unit divided into $k - l$ packets of p symbols. (c) l redundancy packets added for erasure coding, resulting in a total transmission size of n_t. (d) Packets modulated and coded resulting in an information rate r (bits per channel use) to protect against noise in non-blocked conditions. (e) Multiplexing with packets from other data units and transmission. Packets transmitted over s satellites, the distance in time between two packets transmitted over the same satellite is t_dist resulting in a total delay of t_delay in the transmission by one satellite.

4.1.4 Design Parameters

For both schemes there are a number of parameter values that need to be chosen to provide the desired system performance. Obviously, the coding overhead to overcome channel blockages and the total acceptable delay need to be optimized. Intuitively, short delays means that the impact of a channel blockage becomes more severe increasing the coding overhead necessary for satisfactory performance. Thus, requirements on a short delay introduce a cost in terms of additional bandwidth usage. A design parameter that is less obvious is the number of segments or packets that are transmitted, k. A small k means that a large part of the transmitted code word is lost if the signal is blocked. On the other hand, increasing k, while maintaining a fixed t_{delay}, reduces t_{dist} meaning that the probability of loosing consecutive packets or segments increases.

In Figure 4.3, the impact of the choice of k is illustrated through a simple example. Consider the transmission over a channel with two states, either the transmitted signal is received or completely lost. This could be seen as corresponding to the cases of line of sight and completely blocked propagation conditions for a mobile satellite link. We assume that an error correcting code sufficient for recovery of the original data unit provided that two thirds of the transmitted data is received has been chosen. Similarly to [WLS99, LGW00], we want to design the transmission scheme to be able to cope with channel blockages up to a certain length in time, t_{block}. First, consider a system where the coded data symbols are spread evenly over a time t_{delay}, see (a) in the figure. Under the assumption of the error correcting performance of the transmission scheme, to guarantee reception, at least a delay of $t_{\text{delay}} = 3t_{\text{block}}$ needs to be introduced. Second, consider transmitting the coded data in three bursts as in (b) of the figure. To satisfy the simple design criterion of this example, it is sufficient to choose t_{delay} such the separation in time between two bursts is at least t_{block}. Thus, using the second scheme the required delay for error free operation is reduced.

Clearly the design criterion used for Figure 4.3 is overly simplified. For example, in a scenario with several short blockages the design (b) might fail while the more conservative (a) will work as long as the total duration of blockage does not exceed t_{block}. On the other hand, if t_{delay} is reduced, due to the correlated nature of mobile satellite channels, the probability of another blockage is also reduced. This illustrates why it is important to take more realistic models for the temporal behavior of the mobile satellite channel into account in the design of robust transmission

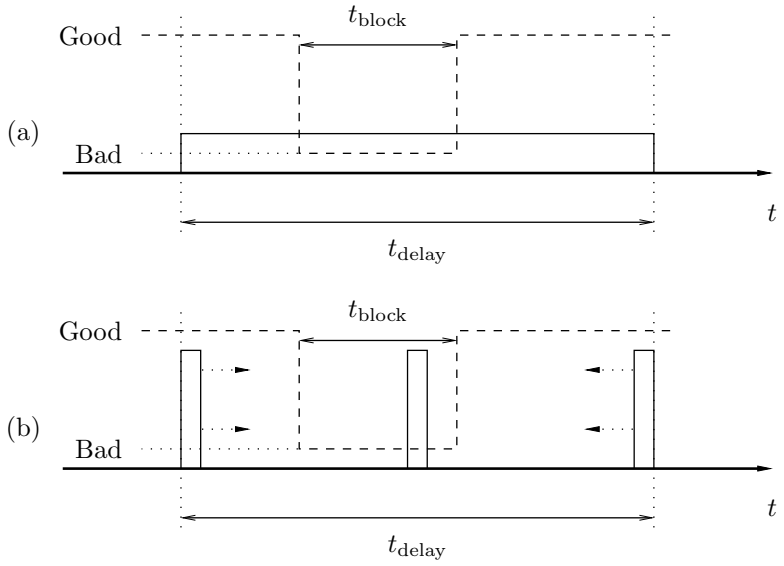


Figure 4.3: Example of outer interleaver designs. The dashed curve illustrates the current receive signal quality as the channel enters a bad state for the time t_{block}. The continuous curve shows the number of coded bits received per time unit. Two multiplexing configurations are considered creating different interleaving structures. In both cases it is assumed that the original data can be recovered if two thirds of the transmission is received in the good state. In (a) the transmitted data is spread evenly in time while in (b) the data is transmitted in three bursts. Assuming that there will only be a single blockage of the channel for the duration of the outer interleaver, the delay, t_{delay}, introduced to provide time diversity can be reduced by segmenting the data unit before transmission.

schemes exploiting coded time diversity.

4.2 Land Mobile Satellite Channel Model

From a time diversity perspective, the land mobile satellite channel is characterized by long term blockages and shadowing of the direct path between the satellite and the mobile terminal. A natural approach to model this phenomenon is to divide the channel into states where each state represents one type of event. For example, the long term temporal behavior of the channel could be modeled as a line of sight state, a shadowed state, a blocked state and the transitions between these states. Events acting on a shorter time scale such as fast fading due to multi-path propagation are modeled by drawing channel coefficients from a random distribution defined by the current state of the channel. Different random processes have been proposed to mimic the state durations and the transitions between different states. Examples include Markov processes, see e.g. [LCD$^+$91, VD92, FGF$^+$97] and "semi-Markov" processes where general probability distributions are used to model the duration of the different states, see e.g. [ITU03, HVG91, BT02].

Here we consider multi-state channel models where the probability of being in a certain state at some point in the future only depends on the current state and not on earlier states, i.e. Markov processes. These seem to prevail in the channel modeling literature, see e.g. [LCD$^+$91, FGF$^+$97, VD92], and we only consider discrete time Markov chain models. In this case, the state transitions of the Markov chain can be seen as sampling of the channel states.

4.2.1 Symbol Level Coding

To model the communication of a data unit using the symbol level scheme, a discrete time complex-valued model is considered where the received data, y_s, is the output of a frequency flat fading additive white Gaussian noise channel with input x_s. That is, one use of the channel is modeled as

$$y_s = h_s x_s + v_s \tag{4.1}$$

where h_s models the fading channel and v_s is the noise. This is a reasonable model for the sampled output of the matched filter in a synchronized flat fading system. If needed, most of the techniques presented in the following can be easily extended to more complex channels, for example

including frequency selective fading or multiple receive antennas. The signal to noise ratio, ρ, is defined as

$$\rho = \frac{\mathrm{E}\{|h_s x_s|^2\}}{\mathrm{E}\{|v_s|^2\}}. \tag{4.2}$$

To model fast fading, at each time instance, h_s is independently drawn from a distribution defined by the current state of a j_{tot}-state Markov chain with states $S_1 \ldots S_{j_{\mathrm{tot}}}$, for an example see Figure 4.4 where a 3-state model is illustrated. Each of the states are associated with a function, $f_1(h_s), \ldots, f_{j_{\mathrm{tot}}}(h_s)$ that defines the probability distribution from which h_s is drawn. The different states of the Markov chain have physical interpretations in the current channel conditions. Examples of interpretations include line of sight, light shadowing or complete signal blockage. The long term behavior of this type of channel is thus defined by the transition probabilities between the different states and the short term behavior by the probability density of h_s defined by the state. During the years, significant efforts have been made in estimating these parameters from channel measurements and physical modeling [LB98, FVCC+01, STE01].

Notice that the transmission time of a segment may be longer than the sample time between the transitions in the Markov chain. This means that for a detailed analysis the transition probabilities can be seen as time varying over the transmission of one code word with different transition probabilities depending on if the transition occurs between two transmitted segments or within one transmitted segment.

4.2.2 Packet Level Coding

To model the system behavior at packet level of the packet level scheme, an erasure channel where the transmitted packet, x_p, is either perfectly received or completely lost is considered. The probability of loosing a packet, e_p, is time varying according to the current state of a j_{tot}-state Markov chain. In one subset of the states the applied channel code allows for error free decoding of the received packets, $e_p = 0$, while for the remaining states the received signal is insufficient, $e_p = 1$. As an example, consider a two state model with one good state, G, and one bad state, B. Such a channel is completely characterized by the transition probabilities between the different states, p_g and p_b, see Figure 4.5. If the transmission time of a packet is significant compared to the average time the physical channel spends in non-blocked conditions, accurate estimation of the error probability requires that the probability of receiving partial packets is

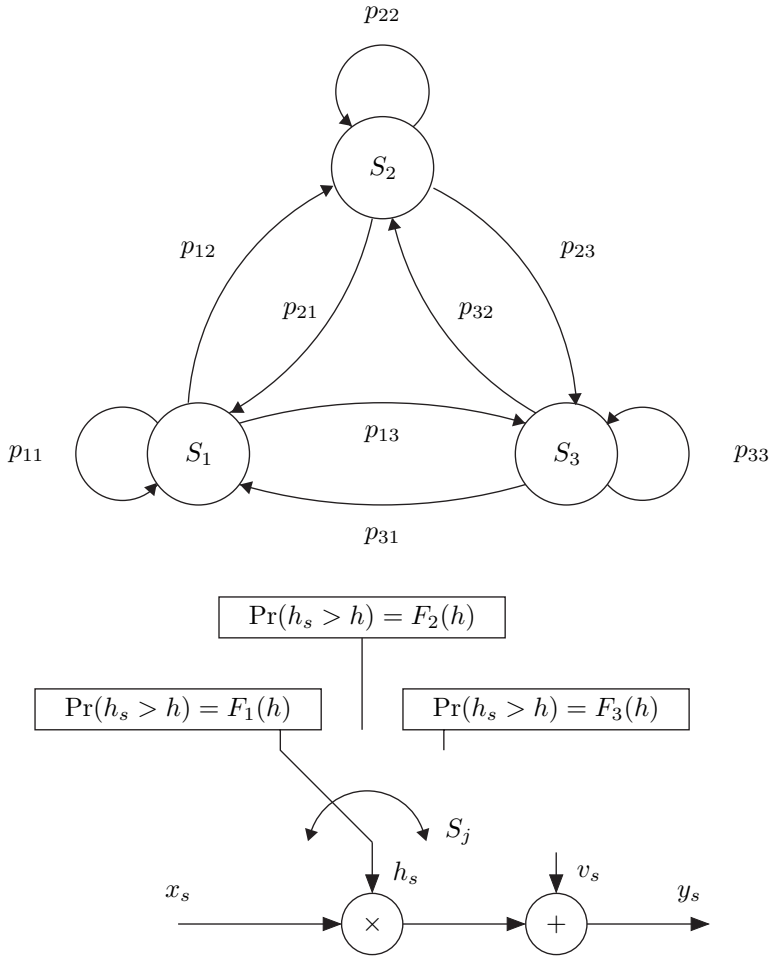


Figure 4.4: Example of a ($j_{\text{tot}} = 3$)-state Markov channel like that considered in the examples of Chapter 6. A flat fading additive white Gaussian noise channel is considered. The channel coefficient, h_s is independently drawn from a distribution defined by the current state of a Markov chain.

taken into account. If the physical channel can be modeled as a two state model this effect can be included directly in the packet level transition probabilities, see Appendix 4.A. Note that using techniques originally derived for Gilbert-Elliot channels, see e.g. [YW95] or Appendix 5.A.4, the results in Chapter 5 are straightforward to extend to include scenarios where the receiver has some probability of receiving packets partly received in the bad state or loosing packets received in the good state, i.e. for states with e_p such that $0 \leq e_p \leq 1$.

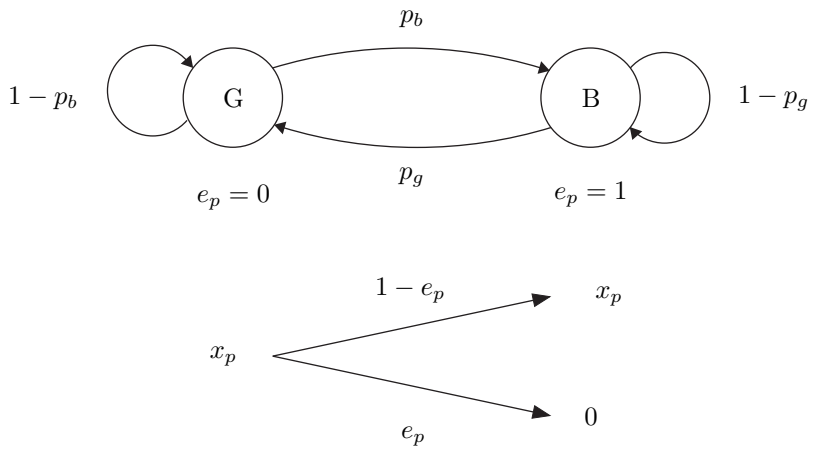


Figure 4.5: Example of a ($j_{\text{tot}} = 2$)-state Markov model used for the packet level coding scheme. A transmitted packet x_p is correctly received if the channel is in state G and lost if the channel is in state B.

Appendix 4.A Packet Level Markov Model

If large data units are transmitted using the packet level coding scheme defined in Section 4.1.2, depending on the number of packets the data unit is divided into there might be a considerable probability of a state change in the underlying physical channel during the transmission of a packet. In this appendix it is shown how this effect can be taken into account when deriving the transition probabilities, p_b and p_g for the packet level channel model shown in Figure 4.5.

For the derivation of these transition probabilities the following assumptions and notations are introduced:

- It is assumed that the temporal behavior of the underlying physical (symbol level) channel model is characterized by a two state Markov chain. The two states, denoted S_1 and S_2, are chosen such that the symbol level coding is sufficient to recover a packet entirely received in S_1 while if some part of the packet is received in S_2 the packet is lost. Note that Markov chains with a larger number of states cannot be reduced to a two state chain. If the symbol level code is powerful enough to recover packets partly received in S_2 the results in this thesis are straightforward to extend to include such effects.

- The transition probabilities that characterizes the two state Markov model modeling the physical channel, $p_{11} = 1-p_{12}$ and $p_{22} = 1-p_{21}$ are assumed available from measurements or other analysis.

- Transmission of one packet spans n_{p} samples in the symbol level Markov chain and there are n_{dist} samples between each packet.

- The stationary probabilities of being in either the good or the bad state in the symbol level Markov model are given by $\Pr(S_1^\infty)$ and $\Pr(S_2^\infty)$. These probabilities follow from the well known expressions,

$$\Pr(S_1^\infty) = \frac{p_{21}}{p_{12}+p_{21}} \text{ and } \Pr(S_2^\infty) = \frac{p_{12}}{p_{12}+p_{21}}.$$

  For the packet level Markov model the analog notation $\Pr(\mathrm{G}^\infty)$ and $\Pr(\mathrm{B}^\infty)$ is used.

First, p_b, the probability of loosing a packet if the previous packet was received successfully is derived. Note that $1 - p_b$ is the probability of successfully receiving a packet when the previous packet also has been

received. If a packet should be successfully received, the first sample and the remaining $n_{\mathrm{p}} - 1$ samples of the packet must all have been received in state S_1, we have

$$1-p_b = \underbrace{(\Pr(S_1^{\infty}) + \Pr(S_2^{\infty})(1 - p_{12} - p_{21})^{n_{\mathrm{dist}}+1})}_{(a)} \underbrace{p_{11}^{n_{\mathrm{p}}-1}}_{(\mathrm{b})}. \tag{4.3}$$

In (4.3) the part (a) is the probability that the first sample of the current packet is correctly received given that the last sample of the previous packet was correctly received, see e.g. [WM99], the part (b) gives the probability that the remaining $n_{\mathrm{p}} - 1$ samples of the transmission of the packet are in S_1.

Second, p_g, the probability of receiving a packet if the channel was in the bad state during the transmission of the previous packet, is computed. This is done by first deriving the stationary probabilities of the packet level Markov channel model. The stationary probability of being in the good packet state, $\Pr(\mathrm{G}^{\infty})$, can easily be found,

$$\Pr(\mathrm{G}^{\infty}) = \Pr(S_1^{\infty}) p_{11}^{n_{\mathrm{p}}-1}.$$

Using that $\Pr(\mathrm{B}^{\infty}) = 1 - \Pr(\mathrm{G}^{\infty})$ and the well known

$$\Pr(\mathrm{G}^{\infty}) = \frac{p_g}{p_b + p_g} \text{ and } \Pr(\mathrm{B}^{\infty}) = \frac{p_b}{p_b + p_g}$$

the transfer probability from the bad to the good state of the packet level model may be found as,

$$p_g = p_b \frac{\Pr(\mathrm{G}^{\infty})}{\Pr(\mathrm{B}^{\infty})}.$$

Chapter 5

Service Availability Estimation

In the design of a mobile satellite broadcast system there is a large number of design parameters to choose. Examples include coding rate, interleaving length and data segmentation but one may also consider different system configurations in terms of satellites and terrestrial coverage. While detailed simulations are necessary in a system design process, due to the large number of parameters, efficient performance estimation techniques are important. In this chapter we derive an efficient technique for estimating the data unit reception probability when the two transmission schemes defined in Section 4.1 are applied to overcome channel impairments as those defined in Section 4.2.

Based on some simplifying assumptions in Section 5.1, the problem can be divided into two parts, analyzing the temporal behavior of the channel and analyzing the code performance over the different channel states. In Sections 5.2 and 5.3 we show how such analysis can be performed while Section 5.4 combines these results to estimate the end user receive probability. Finally, Appendix 5.A derives efficient enumeration techniques for Markov chains.

5.1 Simplifying Assumptions

To perform an efficient error analysis a few assumptions regarding the satellite reception capability and spatial diversity are necessary.

5.1.1 Satellite Channel Reception

The error analysis techniques proposed in this section require two assumptions on the satellite channel reception:

- The probability of correct decoding only depends on how much of a transmitted codeword is received in either state.
- The receiver channel state information is always in a steady state. That is, the receiver channel knowledge does not improve nor deteriorate over time. Neither is there any delay in resynchronization if the synchronization have been lost due to some signal blockage.

Based on the first of these assumptions the problem of estimating the data unit error probability can solved by dividing it into two parts, a temporal and code performance part. That is, if one can estimate the receive probability given that the transmitted codeword has been received in a certain combination of channel states, the probability of error can be computed from the probability of that distribution of the states. The second assumption preserves the Markov property of the channel, meaning that the states can be efficiently enumerated, see Appendix 5.A.

5.1.2 Spatial Diversity

Two types of spatial diversity are considered, space and terrestrial.

For space diversity, with multiple transmitting satellites, it is assumed that the amount of data received in the different states from the various satellites is statistically independent. In practice, correlation between the different satellite channels is likely and the results achieved for satellite diversity could be seen as an upper bound on the achievable performance. It should be noted that as much larger time latencies can be accepted in broadcast applications than for two-way type of communications, the transmission of the individual segments or packets over the different satellite channels can be separated in time reducing the correlation. This also means that much of the measurements of satellite correlation and the analysis of its impact on the communication performance that has been performed for low latency systems, see e.g. [Lut96, VCFS02], are not directly applicable.

For terrestrial diversity, with terrestrial repeater stations covering environments too resource demanding for the satellites, we assume that all, or a certain percentage, of the users in those environments receive

the service via the repeaters. Thus, these end users are excluded in the satellite coverage analysis.

5.2 Temporal Analysis

To estimate the error probability according to the scheme outlined above the temporal behavior needs to be analyzed, i.e. the probability of receiving a code word in a certain distribution of states is required. Let n_{tot} be the total number of state changes, samples, of the Markov chain that the transmitted code word spans and let $n_1, \ldots, n_{j_{\text{tot}}}$ correspond to the number of samples of a transmitted code word that has been received in $S_1, \ldots, S_{j_{\text{tot}}}$, i.e. $n_{j_{\text{tot}}} = n_{\text{tot}} - \sum_{j=1}^{j_{\text{tot}}-1} n_j$. We are interested in $p_{\text{m}}(n_1, \ldots, n_{j_{\text{tot}}-1})$, the probability that of the n_{tot} samples of a transmitted data unit n_1 samples are received in S_1, n_2 in S_2 and so on. Given the transition probabilities of the Markov chain, p_{m} can be efficiently computed using the techniques in Appendix 5.A. For an illustration of how the probabilities, p_{m}, may appear, see Figure 5.1. If t_{delay} and k are large $p_{\text{m}}(n_1, \ldots, n_{j_{\text{tot}}-1})$ would converge such that the amount of data received in the different states would correspond to their stationary probabilities. However, as a finite delay is considered in this example, there is significant variance around this value.

If satellite diversity is considered, the amount of data received in the different states from the different satellites is added up. In Section 5.1.2 it was assumed that the amounts of data received in the different states via the different satellites is independent. Thus the probability of receiving a transmitted data unit in a certain combination of states can be found by convoluting the individual probability functions of each satellite channel. For example, in the case of two continuously transmitting satellites, $p_{\text{m}}(n_1, \ldots, n_{j_{\text{tot}}-1})$, the total probability function for the two satellite signals, can be found as the convolution of the individual probability functions of the two channels, $p_{\text{m}}^{\text{Sat1}}(n_1, \ldots, n_{j_{\text{tot}}-1})$ and $p_{\text{m}}^{\text{Sat2}}(n_1, \ldots, n_{j_{\text{tot}}-1})$, i.e.

$$p_{\text{m}}(n_1, \ldots, n_{j_{\text{tot}}-1}) = p_{\text{m}}^{\text{Sat1}} * p_{\text{m}}^{\text{Sat2}}(n_1, \ldots, n_{j_{\text{tot}}-1}), \tag{5.1}$$

where $*$ denotes $(j_{\text{tot}} - 1)$-dimensional convolution. Since the different satellites are assumed well separated in space, their fading statistics in the different states may vary considerably due to differences in elevation angles and satellite antenna design [MN02]. Such effects can be incorporated in the present analysis by increasing the number of states in the

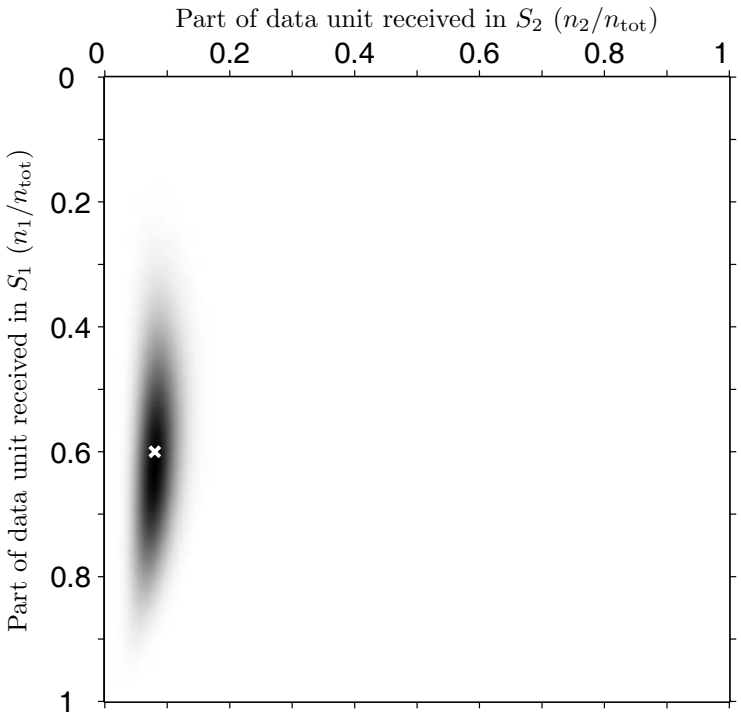


Figure 5.1: Example of $p_{\mathrm{m}}(n_1, \ldots, n_{j_{\mathrm{tot}}-1})$ for a ($j_{\mathrm{tot}} = 3$)-state Markov chain model, the darker the plot the higher the probability of that combination of states. The part of the transmitted data unit not received in either S_1 or S_2 is received in S_3. The transmission of 500 segments spread over 90 s using a single satellite in the urban environment defined in Chapter 6 is considered. The different states, S_1, S_2 and S_3 correspond to line of sight, shadowed and blocked conditions respectively. One segment corresponds to one sample in the Markov chain. Using an infinitely long interleaver p_{m} would be concentrated around the stationary probability of the channel, the white "x". Due to the finite delay however the variance around this value is substantial.

Markov chain, where some states may be unique to one of the satellite channels.

Note that as long as the Markov assumption on the channel holds, $p_{\mathrm{m}}(n_1, \ldots, n_{j_{\mathrm{tot}}-1})$ can be efficiently computed analytically using the techniques derived in Appendix 5.A. If more elaborate random processes are considered for the long term behavior of the channel, brute force enumeration or Monte-Carlo type computations are in general required, reducing the efficiency and precision of the temporal analysis.

5.3 Code Performance

To estimate the data unit error rate performance of the transmission schemes the performance of the channel code must be known or approximated. Thus, we must compute or estimate $p_{\mathrm{c}}(n_1, \ldots, n_{j_{\mathrm{tot}}-1})$, the probability of correctly decoding a coded data unit spanning a total of n_{tot} samples of the j_{tot}-state Markov chain out of which $n_1, \ldots, n_{j_{\mathrm{tot}}-1}$ were received in $S_1 \ldots S_{j_{\mathrm{tot}}-1}$ and the remaining $n_{j_{\mathrm{tot}}} = n_{\mathrm{tot}} - \sum_{j=1}^{j_{\mathrm{tot}}-1} n_j$ in $S_{j_{\mathrm{tot}}}$.

In many cases p_{c} cannot be computed exactly but must instead be estimated in some way. In the literature, different bounds and approximations that may be suitable for the symbol level coding scheme exist. For example [SS01] provides tight bounds on the performance of modern codes such as LDPC and turbo-like codes. An upper bound on the error-free code performance of the symbol level scheme can be found from the capacity of fading Gaussian channels similar to the capacity results in [ESH04]. Since modern coding schemes such as turbo or LDPC-codes are known to perform near capacity such a bound is of relevance [BG96, MN97] and will therefore be considered in the design examples in Chapter 6. If the receiver has access to perfect channel estimates and the process generating h_s is stationary ergodic, with the notation from Section 4.2.1, the channel capacity may be found as [BPS98] (in nats per channel use),

$$C = \mathrm{E}\{\log(1 + |h_s|^2)\}. \tag{5.2}$$

For an end user who receives a share n_j/n_{tot} of a transmitted data unit in S_j, the probability density function of h_s is,

$$f(h_s) = \sum_{j=1}^{j_{\mathrm{tot}}} \frac{n_j}{n_{\mathrm{tot}}} f_j(h_s) \tag{5.3}$$

where $f_j(h_s)$ is the probability density function of h_s in S_j. Thus, from (5.2), the channel capacity of such an end user may be found as

$$\begin{aligned} C(n_1,\ldots,n_{j_{\text{tot}}}) &= \mathrm{E}\{\log(1+|h_s|^2)\} \\ &= \int \log(1+|h_s|^2) \sum_{j=1}^{j_{\text{tot}}} \frac{n_j}{n_{\text{tot}}} f_j(h_s) dh_s \\ &= \sum_{j=1}^{j_{\text{tot}}} \frac{n_j}{n_{\text{tot}}} C_j, \end{aligned} \tag{5.4}$$

where C_j is the channel capacity of the channel in S_j. If an information rate r_{s} is used in the transmission, the set of end users that reliably can decode the received data can be upper bounded by the set of users with $n_1,\ldots,n_{j_{\text{tot}}}$ satisfying

$$C(n_1,\ldots,n_{j_{\text{tot}}}) > r_{\text{s}}. \tag{5.5}$$

Hence an approximation of p_{c} that is an upper bound on the achievable error-free code performance for a transmission scheme with information rate r_{s} can be found as

$$\begin{aligned} &p_{\text{c}}(n_1,\ldots,n_{j_{\text{tot}}-1}) \\ &\quad\approx \hat{p_{\text{c}}}(n_1,\ldots,n_{j_{\text{tot}}-1}) \\ &\quad= \begin{cases} 1 & \text{for } C(n_1,\ldots,n_{\text{tot}} - \sum_{j=1}^{j_{\text{tot}}-1} n_j) > r_{\text{s}} \\ 0 & \text{otherwise} \end{cases} . \end{aligned} \tag{5.6}$$

Note that the error and the error free volumes in $\hat{p_{\text{c}}}$ are separated by a plane defined by $C(n_1,\ldots,n_{\text{tot}} - \sum_{j=1}^{j_{\text{tot}}-1} n_j) = r_{\text{s}}$ in the $(j_{\text{tot}}-1)$-dimensional space spanned by $n_1,\ldots,n_{j_{\text{tot}}-1}$.

This limit is illustrated by the continuous curve in Figure 5.2. The information rate in this example is 1/2 bit per channel use and the three state channel defined in Section 6.1 is considered. To demonstrate the relevance of the bound when compared with a more realistic estimate it can be compared with code performance derived through Monte-Carlo simulations. The gray-scale plot of Figure 5.2 shows $p_{\text{c}}(n_1,n_2)$ of a BPSK-modulated rate 1/2 LDPC-code. From this figure we may note that the considered LDPC code follows a similar behavior where the erroneous and error free areas are separated by a line indicating that expressions similar to (5.6) may be used to approximate the performance of modulation schemes and codes that are realistic to implement.

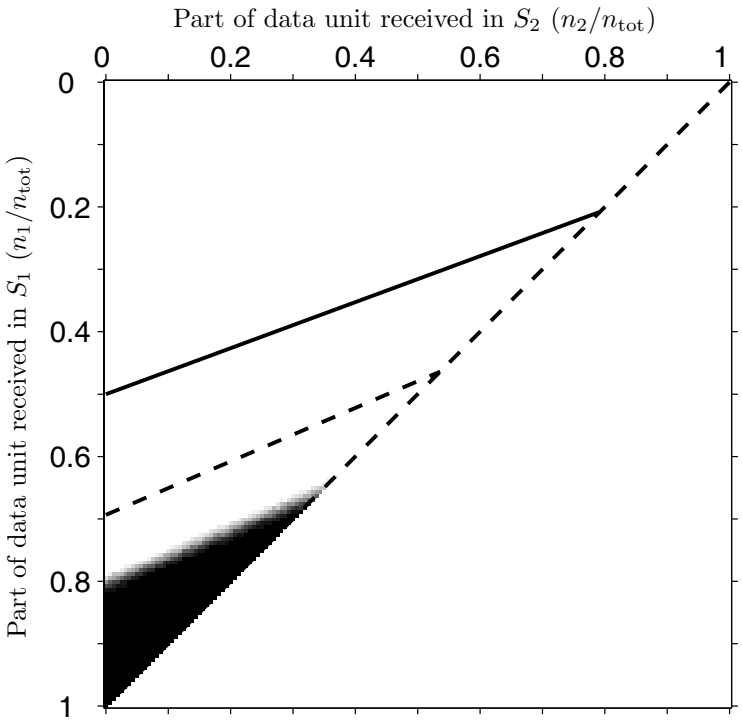


Figure 5.2: Example of $p_{\mathrm{c}}(n_1, \ldots, n_{j_{\mathrm{tot}}-1})$ when a $j_{\mathrm{tot}} = 3$-state Markov chain model is considered, the darker the plot the higher the probability of successful reception. Here the performance of an (16000, 8000) LDPC code transmitted using BPSK is considered. The code had a column weight three and a maximum of 30 iterations were used in the decoding. Code generation, encoding and decoding were performed using software available at [Mac04]. Fading statistics are defined in Section 6.1.1. The LDPC code performance can be compared with the upper bound (5.5) at transmission rate 1/2 bit per channel use, with continuous input symbols x_s for the superimposed continuous line and with BPSK for the dashed line. Note that the majority of the loss in code performance comes from using such a small signal constellation. In this particular case, the remaining loss corresponds to approximately 1.2 dB in signal power.

For the packet level scheme, only packets received in the set of "good" states can be used for recovering the data unit. An upper bound is the performance of the optimal erasure code for which the transmitted data unit can be recovered provided that any $(k-l)$ packets have been successfully received. For example, if the two state model in Figure 4.5 is considered,

$$p_{\mathrm{c}}(n_1) = \begin{cases} 1 & \text{for } n_1 \geq k-l \\ 0 & \text{for } n_1 < k-l \end{cases} \tag{5.7}$$

where n_1 in this case corresponds to the number of packets received in state G.

5.4 Data Unit Receive Probability

Based on the assumptions in Section 5.1, the probability of correctly receiving a code word, p_{rec} can be computed as,

$$p_{\mathrm{rec}} = \sum_{n_1} \cdots \sum_{n_{j_{\mathrm{tot}}-1}} p_{\mathrm{c}}(n_1, \ldots, n_{j_{\mathrm{tot}}-1}) p_{\mathrm{m}}(n_1, \ldots, n_{j_{\mathrm{tot}}-1}). \tag{5.8}$$

For a graphical intuition to this process, this may be viewed as combining the temporal analysis illustrated in Figure 5.1 with the estimated code performance illustrated by Figure 5.2. Clearly p_{rec} is a function of many different variables, an example of a subset could include the environment, ϵ, the number of segments or packets, k, the code used to protect against long term channel impairments, r_{s} or r_{p}, and the allowed delay, t_{delay}. For the symbol level coding case we write

$$p_{\mathrm{rec}} = p_{\mathrm{rec}}(\epsilon, k, r_{\mathrm{s}}, t_{\mathrm{delay}}), \tag{5.9}$$

and use the same notation when packet level coding is considered.

As illustrated in (5.8), by applying the technique for estimating the data unit error rate performance presented in this section the analysis of the temporal behavior of the channel and the performance of the code used can be separated. This allows for efficient analysis and elaborate design techniques.

For a broadcast system (5.8) has an interesting interpretation. Assume that end users in the same environment have similar channel statistics. Then $p_{\mathrm{m}}(n_1, \ldots, n_{j_{\mathrm{tot}}-1})$ can be seen as the share of moving terminals that have received a data unit in a combination of states given

by $n_1, \ldots, n_{j_{\mathrm{tot}}-1}$ and $p_{\mathrm{c}}(n_1, \ldots, n_{j_{\mathrm{tot}}-1})$ as the share of those users that can decode the unit correctly. Consequently, p_{rec} can be interpreted as the share of the moving terminals in the considered environment that can decode a particular data unit, i.e. the service availability.

In the next chapter we will explore this further by considering some specific design examples.

Appendix 5.A Efficient Enumeration of Finite State Markov Models

To estimate the probability of receiving a data unit using the methods presented in this chapter it is necessary to compute the probability of having been a certain number of times in each state of a Markov chain. For efficient usage of the presented technique it is vital that this probability is computed in an efficient manner. The references [Cup69, YW95, WM99] are examples of earlier work where the probability has been computed and analyzed for a two state Markov chain. In [PB99] techniques for Markov chains with any finite number were presented. Here a simple recursive technique to compute the necessary probability is proposed. While efficient, not only is the method simple to derive and implement, but it is also especially suited for cases when we are interested in computing the entire probability function and the results are applicable to Markov chains with non-stationary transition probabilities.

5.A.1 Proposed Technique

Consider a Markov chain with j_{tot} states denoted $S_1, \ldots, S_{j_{\text{tot}}}$. We are interested in the statistical properties of the $j_{\text{tot}}-1$ element random vector $\vec{N}_S(n_{\text{tot}})$, where $(\vec{N}_S(n))_j$ is the number of times the Markov chain has been in S_j when n consecutive states labeled according to their order in time as $1 \ldots n$ have been observed. Note that the chain at time n has been $n - \sum_{j=1}^{j_{tot}-1} (\vec{N}_S(n))_j$ times in $S_{j_{\text{tot}}}$. The state of the Markov chain at time n is $S(n)$. The chain is characterized by its transition probabilities, where the, possibly time-varying, transition probability from S_a at time $n-1$ to S_b at time n is $p_{ab}(n)$, and the probability of starting the Markov chain in S_j is $\Pr(S(1) = S_j)$. Here an efficient recursive technique for computing $p_{\text{m}}(n_1, \ldots, n_{j_{\text{tot}}-1}) = \Pr(\vec{N}_S(n_{\text{tot}}) = [n_1, \ldots, n_{j_{\text{tot}}-1}]^{\text{T}})$ is derived.

Let $p_{\text{m}}^{(j)}(n, n_1, \ldots, n_{j_{\text{tot}}-1})$ be the joint probability of the Markov chain being in S_j at time n and that the Markov chain has been n_1 times in S_1, n_2 times in S_2 etc., i.e.

$$\begin{aligned} &p_{\text{m}}^{(j)}(n,n_1, \ldots, n_{j_{\text{tot}}-1}) \\ &\qquad = \Pr\left(S(n) = S_j \cap \vec{N}_S(n) = [n_1, \ldots, n_{j_{\text{tot}}-1}]^{\text{T}}\right). \end{aligned} \tag{5.10}$$

Note that from $p_{\mathrm{m}}^{(j)}$ the sought for p_{m} can easily be computed as,

$$p_{\mathrm{m}}(n_1,\ldots,n_{j_{\mathrm{tot}}-1}) = \sum_{j=1}^{j_{\mathrm{tot}}} p_{\mathrm{m}}^{(j)}(n_{\mathrm{tot}},n_1,\ldots,n_{j_{\mathrm{tot}}-1}). \tag{5.11}$$

It is obvious that $p_{\mathrm{m}}^{(j)}$ can be computed recursively using the step,

$$p_{\mathrm{m}}^{(j)}(n,\ldots,n_j,\ldots) = \sum_{k=1}^{j_{\mathrm{tot}}} p_{\mathrm{m}}^{(k)}(n-1,\ldots,n_j-1,\ldots)p_{kj}(n) \tag{5.12}$$

where $p_{\mathrm{m}}^{(j)}(1,\ldots)$ have been initialized using the initial probabilities $\Pr(S(1)=S_j)$, $j=1,\ldots,j_{\mathrm{tot}}$. Thus, by combining (5.11) and (5.12) an efficient recursive relation for computing $p_{\mathrm{m}}(n,n_1,\ldots,n_{j_{\mathrm{tot}}-1})$ has been established.

5.A.2 Efficient Implementation

For the temporal analysis of Section 5.2 we are interested in $p_{\mathrm{m}}(n_1,\ldots,n_{j_{\mathrm{tot}}-1})$ evaluated in all or many different combinations of $n_1,\ldots,n_{j_{\mathrm{tot}}-1}$. Fortunately, the technique derived in Appendix 5.A.1 can be used to compute all values simultaneously. Here this is illustrated using a 3-state Markov chain. Note that the technique is applicable to Markov chains with any finite number of states.

To simplify the presentation, the following matrix notation is introduced. Let the $(n_{\mathrm{tot}}+1)\times(n_{\mathrm{tot}}+1)$ matrix $\boldsymbol{P}_{\mathrm{m}}$ be defined such that its (n_1+1)th, (n_2+1)th element corresponds to $p_{\mathrm{m}}(n_1,n_2)$ and let the $(n+1)\times(n+1)$ matrices $\boldsymbol{P}_{\mathrm{m}}^{(1)}(n)$, $\boldsymbol{P}_{\mathrm{m}}^{(2)}(n)$ and $\boldsymbol{P}_{\mathrm{m}}^{(3)}(n)$ be defined such that their $(n_1{+}1)$th, $(n_2{+}1)$th elements correspond to $p_{\mathrm{m}}^{(1)}(n,n_1,n_2)$, $p_{\mathrm{m}}^{(2)}(n,n_1,n_2)$ and $p_{\mathrm{m}}^{(3)}(n,n_1,n_2)$ respectively. Using this notation, the proposed method is initialized as,

$$\begin{aligned}
\boldsymbol{P}_{\mathrm{m}}^{(1)}(1) &= \begin{bmatrix} 0 & 0 \\ \Pr(S(1)=S_1) & 0 \end{bmatrix}, \\
\boldsymbol{P}_{\mathrm{m}}^{(2)}(1) &= \begin{bmatrix} 0 & \Pr(S(1)=S_2) \\ 0 & 0 \end{bmatrix}, \\
\boldsymbol{P}_{\mathrm{m}}^{(3)}(1) &= \begin{bmatrix} \Pr(S(1)=S_3) & 0 \\ 0 & 0 \end{bmatrix}.
\end{aligned} \tag{5.13}$$

The recursive step (5.12) can then be computed for all outcomes simultaneously as,

$$\begin{aligned}
\boldsymbol{P}_{\mathrm{m}}^{(1)}(n) &= \begin{bmatrix} 0 & 0 \\ \sum_{j=1}^{3} \boldsymbol{P}_{\mathrm{m}}^{(j)}(n-1)p_{j1}(n) & 0 \end{bmatrix}, \\
\boldsymbol{P}_{\mathrm{m}}^{(2)}(n) &= \begin{bmatrix} 0 & \sum_{j=1}^{3} \boldsymbol{P}_{\mathrm{m}}^{(j)}(n-1)p_{j2}(n) \\ 0 & 0 \end{bmatrix}, \\
\boldsymbol{P}_{\mathrm{m}}^{(j)}(n) &= \begin{bmatrix} \sum_{j=1}^{3} \boldsymbol{P}_{\mathrm{m}}^{(j)}(n-1)p_{j3}(n) & 0 \\ 0 & 0 \end{bmatrix}.
\end{aligned} \tag{5.14}$$

Using the recursion (5.14), $\boldsymbol{P}_{\mathrm{m}}$ may be found as,

$$\boldsymbol{P}_{\mathrm{m}} = \sum_{j=1}^{3} \boldsymbol{P}_{\mathrm{m}}^{(j)}(n_{\mathrm{tot}}). \tag{5.15}$$

5.A.3 Computational Complexity

A naive method to compute $p_{\mathrm{m}}(n_1, \ldots, n_{j_{\mathrm{tot}}-1})$ would be to simply add up the probability for all possible state sequences in the chain that fits each parameter setting. However, for n_{tot} states in a j_{tot}-state Markov chain $j_{\mathrm{tot}}^{n_{\mathrm{tot}}}$ sequences are possible and thus the computational time such an approach would increase exponentially in n_{tot} if all parameter settings need to be considered.

Here, the computational complexity of the technique discussed in Section 5.A.2 is analyzed. Again, the analysis is restricted to a 3-state Markov Chain. Generalization of the discussion to any finite number of states is straightforward. Each of the matrices $\boldsymbol{P}_{\mathrm{m}}^{(1)}(n)$, $\boldsymbol{P}_{\mathrm{m}}^{(2)}(n)$ and $\boldsymbol{P}_{\mathrm{m}}^{(3)}(n)$ contains $n(n+1)/2$ non-zero elements. Thus, for every recursive step (5.14), $9(n-1)n/2$ multiplications and an equal number of additions need to be made. The total number multiplications and additions required to reach the n_{tot}th step can be found as,

$$\frac{9}{2}\sum_{n=2}^{n_{\mathrm{tot}}}(n-1)n = \frac{3}{2}(n_{\mathrm{tot}}-1)n_{\mathrm{tot}}(n_{\mathrm{tot}}+1),$$

i.e. the computational complexity of the algorithm grows as n_{tot}^3. For a general finite number of states j_{tot} it is easy to show that the complexity using this algorithm grows in n_{tot} as $n_{\mathrm{tot}}^{j_{\mathrm{tot}}}$.

5.A.4 Example: Error Probability of a Generalized Gilbert-Elliot

A model for burst type errors in communication channels is the Gilbert-Elliot model [Ell63]. Here a binary symmetric channel is considered where the error probability of the bit transmitted at time t is $e(t)$. The error events are independent from one time to another, but the probability of error is given by the state of a two state Markov chain, i.e. $e(t) = e_1$ if the chain is in S_1 and $e(t) = e_2$ if the channel is in S_2, see Figure 5.3. Here, the transmission over n_{tot} state transitions is considered and we are interested in the number of errors of such a transmission N_e. For the probability of receiving a certain number of bits in error, $\Pr(N_e = n_{\text{tot}})$, efficient exact expressions and bounds have been derived, see e.g. [Cup69]. The usefulness of the earlier results in this appendix is here illustrated by deriving the error probability of a generalized Gilbert-Elliot model where the error probability of the binary symmetric channel is given by the state of a j_{tot}-state Markov chain, i.e. $e(t) = e_j$ where $j = 1 \dots j_{\text{tot}}$ is the current state. The presented result also allows for time-varying transition probabilities.

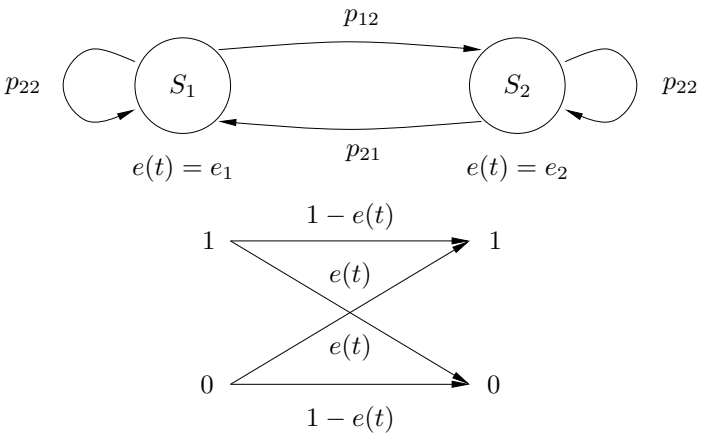


Figure 5.3: Gilbert-Elliot time-varying binary symmetric model for transmission burst errors. The error probability varies between two levels correlated according to a Markov chain.

We begin our derivation by letting the random variables $N_e^{(j)}$ be the

number of bits received in error in S_j. Note that $N_e^{(j)}|(\vec{N}_S)_j = m_j$, the number of bits received in error in S_j given that the chain is in that state m_{tot} times, is a random variable formed by m_j independent trials each with a probability e_j. Thus $N_e^{(j)}|(\vec{N}_S)_j = m_j$ is binomially distributed and

$$\begin{aligned}\Pr(N_e^{(j)} = n_j|(\vec{N}_S)_j = m_j) &= \binom{m_j}{n_j} e_j^{n_j}(1-e_j)^{m_j-n_j} \\ &= p^{(j)}_{N|\vec{N}_S}(n_j).\end{aligned} \tag{5.16}$$

Since the error events of the binary symmetric channel are independent the probability of receiving n_{tot} bits in error given that the channel has spent time in each state as given by $\vec{N}_S$ can be found from the convolution of the probability functions for the different states.

$$\begin{aligned}\Pr(N_e = n_{\text{tot}}|\vec{N}_S) &= p^{(1)}_{N|\vec{N}_S} * p^{(2)}_{N|\vec{N}_S} * \cdots * p^{(j_{\text{tot}})}_{N|\vec{N}_S}(n_{\text{tot}}) \\ &= p^{\text{tot}}_{N|\vec{N}_S}(n_{\text{tot}})\end{aligned} \tag{5.17}$$

The sought for $\Pr(N_e = n_{\text{tot}})$ may then be found by computing the average of (5.17) weighted over the probabilities of all $\vec{N}_S$ which may be found using the techniques in Section 5.A.1.

Chapter 6

Design Examples

Based on the receive probability computations of Chapter 5 we derive and present techniques for analysis and design of mobile satellite broadcast systems. The results are illustrated in a common scenario where the temporal behavior of the channel is based on measurements available in the literature. The intention behind the examples is to illustrate the usefulness of the proposed design techniques, not to present a system design.

The common system configurations considered in the design examples are detailed in Sections 6.1–6.2. The first example, described in Section 6.3, illustrates the implementation of the receive probability estimation technique proposed in Chapter 5. The importance of optimizing the segmentation of the data unit is demonstrated by different designs in Section 6.4. Section 6.5 shows how the technique can be applied to produce optimized designs satisfying constraints on the service availability and other quality of service parameters. In Section 6.6 the required resource usage using different transmission schemes and infrastructure setups is compared. Finally, Section 6.7 illustrates how the trade-off between delay and coding overhead can be analyzed.

6.1 Channel Parameters

While the purpose of the numerical examples below is not to propose realistic system designs, we use channel state transition parameters based on channel measurements available in the literature.

6.1.1 Symbol Level Coding

For the time behavior we consider results from [SKE+02, REKH02] where four different environments, urban, suburban, rural, and highway, have been defined and relevant parameters extracted from measurements in Ku-band ($\sim$ 10-12 GHz). The channel is modeled as having three states, line of sight, shadowed, and blocked, corresponding to S_1, S_2, S_3, and can be characterized by the average duration, $\bar{t}_1$, $\bar{t}_2$, $\bar{t}_3$, and stationary probabilities of each state, $\Pr(S_1^\infty)$, $\Pr(S_2^\infty)$, $\Pr(S_3^\infty)$. For reference these values are given in Table 6.1. From this table the transition probabilities of the Markov chain can be computed using techniques found in e.g. [KKM97].

	Avg duration (s)			Stationary probability		
Environment	$\bar{t}_1$	$\bar{t}_2$	$\bar{t}_3$	$\Pr(S_1^\infty)$	$\Pr(S_2^\infty)$	$\Pr(S_3^\infty)$
Urban	4.2	0.34	3.0	0.60	0.08	0.32
Suburban	2.0	0.29	0.22	0.78	0.17	0.05
Rural	2.0	0.24	0.23	0.78	0.16	0.06
Highway	3.0	0.26	1.0	0.89	0.08	0.02

Table 6.1: Average state durations and stationary probabilities of the four defined environments. The different states, S_1, S_2, and S_3, correspond to line of sight, shadowed and blocked channel conditions respectively. From the transition probabilities of [SKE+02].

To simplify the simulations and the interpretation of the results the fading statistics in the different states are identical for the different environments. In all cases the channel, h_s, is flat Ricean fading. Note that in practical systems it might be necessary to consider more elaborate fading statistics [LB98] and depending on carrier frequency, bandwidth and antenna characteristics the fading of the channel may not be frequency flat. The parameters for the different states may be found in Table 6.2. These parameters are primarily chosen to illustrate the proposed techniques and to facilitate an easy comparison with the packet level coding scheme, not to reflect realistic propagation conditions. For S_3, the blocked state, it is assumed that the received signal strength is too low to extract any useful information.

The presentation is simplified by the assumption that the transmission time of one segment is much smaller than the average duration of one state. Thus the probability of state changes within the transmission of a

segment can be ignored. This means that the results become independent of the actual size of the data unit and the data rate of the channel as long as the above assumption holds.

State	SNR, ρ	Ricean factor	C_j (bits/use)
S_1 (LOS)	0 dB	$-\infty$ dB	1.00
S_2 (Shadowed)	-5 dB	0 dB	0.37
S_3 (Blocked)	$-\infty$ dB	-	0.00

Table 6.2: Parameters of the Ricean fading channel for the different states. Signal to noise ratio has been defined in (4.2), the Ricean factor is the power of the multi-path, fading component relative to the direct, specular component, C_j is the channel capacity in S_j in bits per channel use.

6.1.2 Packet Level Coding

To allow for direct comparison between the packet and symbol level coding schemes, the symbol level channel on which the individual packets are transmitted is assumed to be identical to that in Section 6.1.1. The information rate, r, of the modulation and channel code that protects the individual packets is chosen such that a packet is correctly decoded if it is received in S_1 and lost if the packet is received in any of the other two states, i.e. the set of "good" states consists of S_1 only. Note that unlike the symbol level coding scheme this means that data received in S_2 cannot be exploited to increase the probability of successful decoding. Similar to the symbol level coding case it is assumed that the time it takes to transmit one packet is small compared with the average state duration so that the probability of a state change within a packet can be ignored.

6.2 Code Performance

6.2.1 Symbol Level Coding

The code performance is estimated using the approximation based on capacity (5.6) derived in Section 5.3. This approach has the benefits of being independent of the choice of modulation and coding scheme as

well as allowing for a straightforward and fair comparison with the packet level coding scheme. The capacities for channels stationary in either state may be found in Table 6.2. Figure 6.1 illustrates the code performance by plotting the border between the error-free and erroneous regions of $\hat{p_c}$ for transmission with inverse information rates between one and six in steps of one quarter. From this figure we may for example notice that to decode a segment of which half has been received in S_1 and half in S_2, transmission with information rate $r_s \approx 2/3$ bits per channel use would suffice.

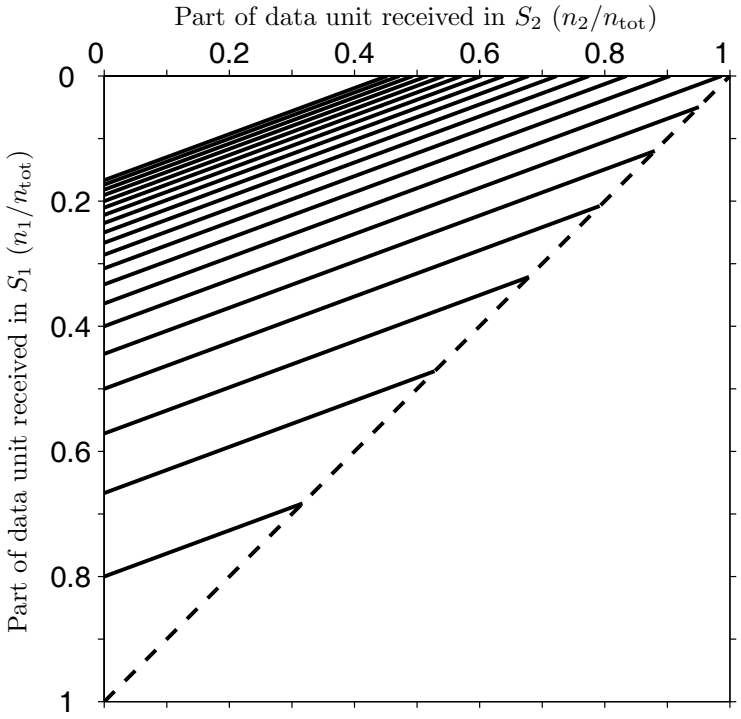


Figure 6.1: Estimates of the code performance for information rates, in bits per channel usage, ranging from 4/5, bottom line, to 1/6, top line.

6.2.2 Packet Level Coding

For fairness in comparison with the symbol level coding case, the usage of an optimal erasure code is considered for protection against lost packets, i.e. if any $k-l$ packets have been received the code word can be recovered, see (5.7). Note that while e.g. Reed-Solomon codes achieve this bound in practice they are only practical for certain rates and code sizes and more flexible but suboptimal codes may have to be considered. In Section 6.1.2 it was assumed that the rate of the channel code of the packet layer coding system was chosen such that only packets received in S_1 can be recovered. Using the same approximation on the code performance as in the symbol level coding case, an information rate $r = 1$ bit per channel use suffices to protect the packets making comparisons of the total coding overhead easy.

6.3 Example of Service Availability Estimation

To realize broadcast designs that satisfy constraints on the service availability, the receive probability for users in different environments must be estimated efficiently. This can be achieved by using the techniques presented in Chapter 5 which is illustrated in Figure 6.2. In the figure, the probability function of the temporal behavior, $p_{\mathrm{m}}(n_1, n_2)$ is plotted for the suburban environment. The curves superimposed on $p_{\mathrm{m}}(n_1, n_2)$ limit the areas where the approximations of $p_{\mathrm{c}}(n_1, n_2)$ are one for transmission with information rates 4/5, 2/3, 3/5, 1/2, and 1/3 bits per channel use. The continuous curves corresponds to the symbol level code performance approximation (5.6) and the dotted line to the packet level code expression (5.7). By summing $p_{\mathrm{m}}(n_1, n_2)$ over the area defined by $p_{\mathrm{c}}(n_1, n_2)$ the service availability, p_{rec} can be estimated, see Section 5.4.

Note that the figure graphically illustrates the gain in performance that the symbol level coding scheme can achieve by exploiting data received in S_2. In this particular example, the coded data unit was divided into $k = 30$ segments or packets before transmission. However, k affects the receive probability and should in a well designed system be chosen to maximize receive probability.

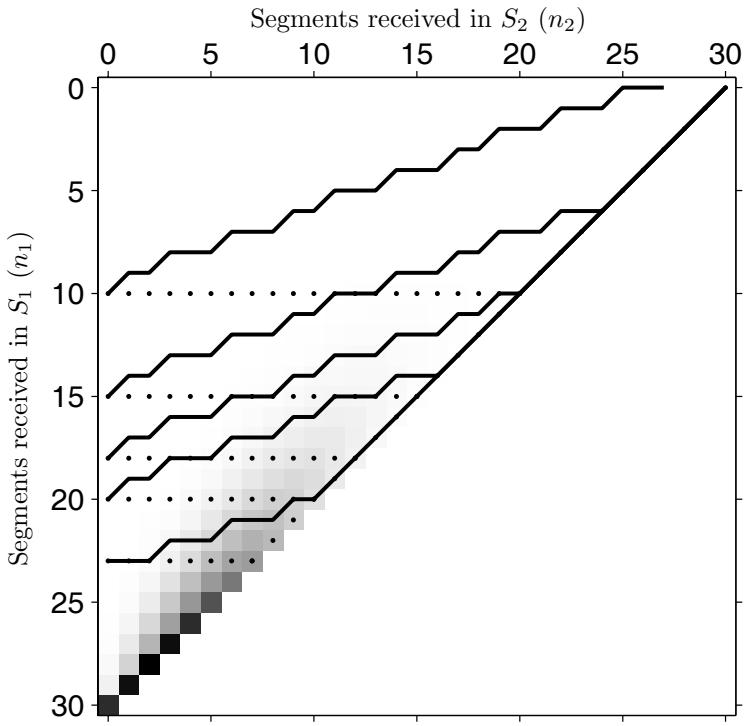


Figure 6.2: Service availability estimation. Analysis of the suburban environment where a data unit has been divided into 30 segments or packets and transmitted over 7 seconds. The probability for an end user to have received data in a certain combination of states, $p_{\mathrm{m}}(n_1, n_2)$, is gray-scaled such that high probability regions are dark, i.e. similar to Figure 5.1. Superimposed on the image are the limits of the code performance, $p_{\mathrm{c}}(n_1, n_2)$, with information rates 4/5, 2/3, 3/5, 1/2 and 1/3 bits per channel usage, i.e. similar to the lines plotted in Figure 6.1. The continuous and dotted lines limit the performance of the symbol and packet level schemes respectively. The corresponding receive probabilities, p_{rec}, were estimated to 0.70, 0.93, 0.97, 0.994 and 0.9998 for the symbol level coding scheme, 0.55, 0.77, 0.86, 0.95 and 0.993 for the packet level coding scheme.

6.4 Interleaver Optimization

As illustrated in Section 4.1.4, the number of packets or segments a data unit is divided into affects the data unit error rate performance. A small number means that only loosing a few packets or segments results in an error and the increase in size increases the probability of only receiving partial packets. Too many on the other hand means that the time, t_{dist}, between the transmitted packets or segments decreases and their error probabilities correlate. Using the receive probability computation derived in Chapter 5 the error rates using different packet sizes can be computed efficiently. The code performance also affects the exact performance and when the data unit is divided into a small number of packets or segments there may be significant threshold effects and therefore the exact number should be chosen carefully. Figure 6.3 shows the receive probability as a function of the number of packets when a data unit protected by a rate 1/2 optimal erasure code is transmitted using the packet level coding scheme. Due to the short time spreading of 7 s the urban environment has not been considered.

This example illustrates how the number of packets can be optimized to maximize receive probability. With the parameter choices used in the example the optimal k is almost identical for all environments, i.e. to maximize the receive probability $k = 8$ packet shall be transmitted out of which any $k - l = 4$ are sufficient for recovery of the original data unit. However, depending on the channel statistics, transmission delay, and coding overhead, this might not always hold and instead k must be optimized over all environments. As a broadcast system is considered one should use an optimization criterion that meets the goals of the broadcaster. In Figure 6.3 two possible criteria are illustrated:

- The average performance weighted by the population sizes in the different environments, i.e. use $k = k_{\text{avg}}$, where

$$k_{\text{avg}} = \arg\max_{k} \sum_{\epsilon} q_{\epsilon} p_{\text{rec}}(\epsilon, k, r_{\text{p}}, t_{\text{delay}}) \tag{6.1}$$

  and q_{ϵ} is the population size in environment ϵ. This criterion attempts to maximize the total number of end users that receive a transmitted data unit. In Figure 6.3 equal size populations are assumed for all environments. For examples, see [MGO04a, MGO04b].

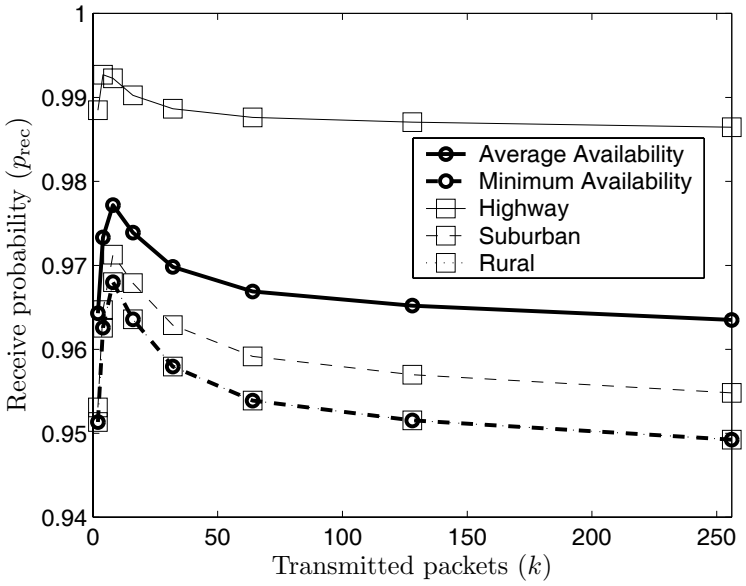


Figure 6.3: Service availability as a function of the number of transmitted packets. The data unit is spread over 7 s using packet level coding and the applied erasure coding is sufficient for decoding even if only 50% of the transmitted data is received. The curves corresponding to the minimum receive probability and the rural environment coincide.

- The performance in the worst environment, i.e. use $k = k_{\min}$, where

$$k_{\min} = \arg\max_{k} \min_{\epsilon} p_{\text{rec}}(\epsilon, k, r_{\text{p}}, t_{\text{delay}}). \tag{6.2}$$

  This ensures that users in all target environments receive the service with satisfactory quality of service. This criterion is used in the examples for the rest of this chapter. In Figure 6.3 the rural environment provides the worst performance for all packet sizes.

6.5 Quality of Service Based Design

For a broadcast system to be attractive it is important that it can provide a satisfactory quality of service for the end user. For example, it is vital that a sufficient number of end users are reached by the service, but there may also be other criteria like a maximal acceptable delay. By having efficient means of estimating the system receive probability performance, complicated system design criteria can be formulated. As an example, consider designing a system that uses the minimum coding overhead that provides at least a service availability, $p_{\text{rec}}^{\min}$, in the worst satellite channel environment given some acceptable delay. For a symbol level coded scheme this criteria can be formulated as finding the minimum information rate, r_{s}, that satisfies,

$$\max_{k} \min_{\epsilon} p_{\text{rec}}(\epsilon, k, r_{\text{s}}, t_{\text{delay}}) \geq p_{\text{rec}}^{\min}. \tag{6.3}$$

Figure 6.4 illustrates how this problem can be solved using a brute force approach. Here the service availabilities in the suburban, rural, and highway environments are plotted at different r_{s}. The data unit is transmitted using a single satellite over a time interval of 7 seconds using the symbol level coding scheme. At each r_{s}, $k = k_{\min}$ has been chosen according to (6.2).

From this figure the information rate can be minimized while ensuring quality of service constraints in terms of availability and acceptable delay. If we for example would like to reach 99% of the end users with the data unit in all considered environments an information rate of approximately 1/2 bit per channel use would be sufficient. Notice that end users in the highway environment are the easiest to satisfy as long as the requirement on service availability is not too severe. However, due to the relatively long average durations in the blocked state, obtaining very high service availability is more challenging. Obviously other similar design choices

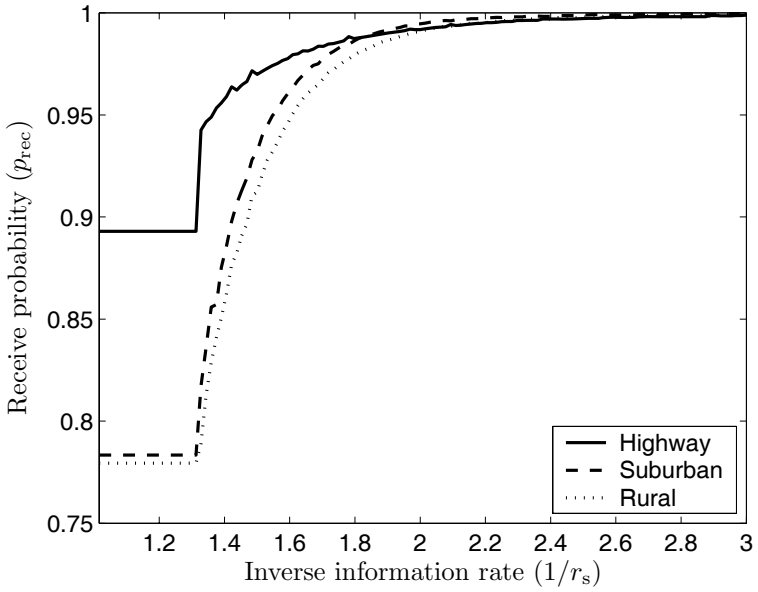


Figure 6.4: Optimized system performance as a function of the information rate for users in three different environments. Symbol level coding is used with a 7 s delay and for each information rate the optimal k has been found to maximize the service availability in the worst environment.

can be proposed, for example, the minimal transmission delay required to provide a design service availability at some acceptable coding overhead can be considered, see [MGO04b].

In Section 6.4 it was illustrated how the segmentation of the transmitted data affects the service availability at a fixed coding overhead. It is interesting to see how the optimal number of segments develops as the requirements on the service availability and the coding overhead increases. This is illustrated in Figure 6.5 where an identical scenario as that of Figure 6.4 is considered. In the figure, the number of segments leading to the maximum service availability is plotted as a function of the coding overhead. Both the k that maximizes the service availability in a single environment and $k_{\min}$, the k that maximizes the performance in the worst environment, are included.

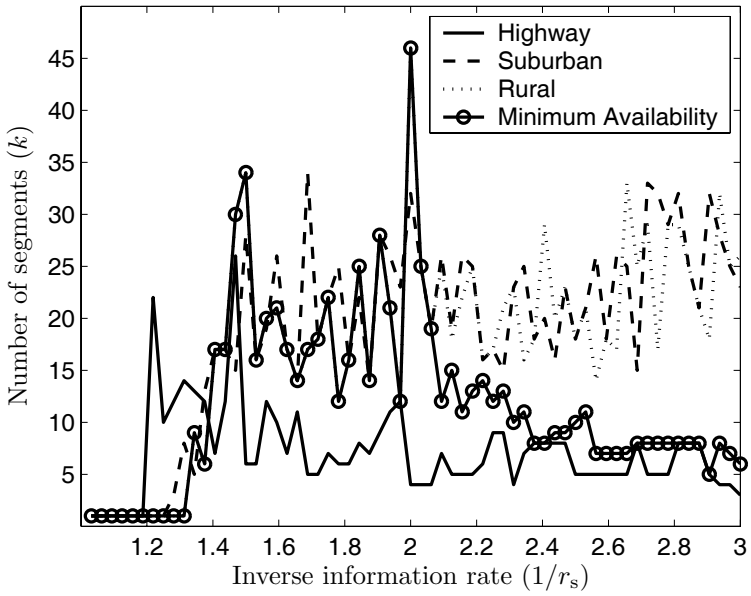


Figure 6.5: Optimized number of segments for end users in three different environments. Symbol level coding with a 7 s delay.

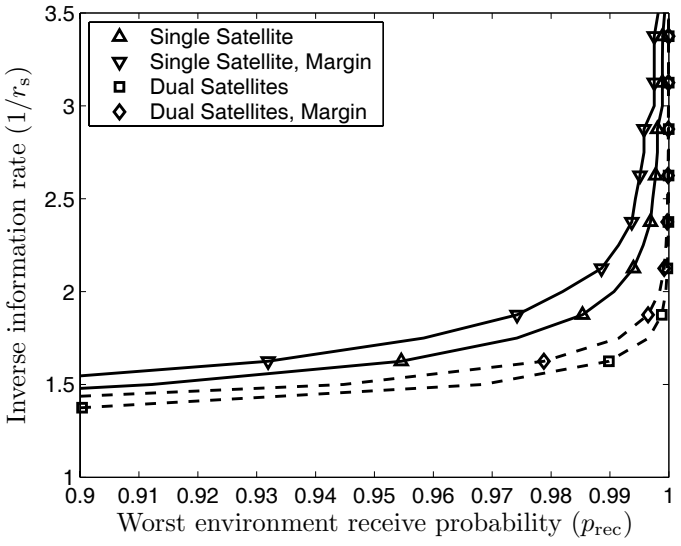
The horizontal axis of Figure 6.5 can be divided in three parts. For high information rates, the applied coding scheme is not powerful enough to recover the original data unit if parts of it have been received in the

shadowed and blocked states. Hence the best strategy is to transmit the entire coded data unit in a single segment. As more extensive coding is applied, transmit diversity can be exploited by spreading the transmission in time. For medium information rates, the rural environment is the most difficult and will dominate the choice in k. In Figure 6.4 we noticed that to achieve very high availability, the highway environment becomes the most difficult. This is also reflected in the segmentation chosen in Figure 6.5. Notice that as the information rate decreases, $k_{\min}$ approaches the k that maximizes the availability in the highway environment.

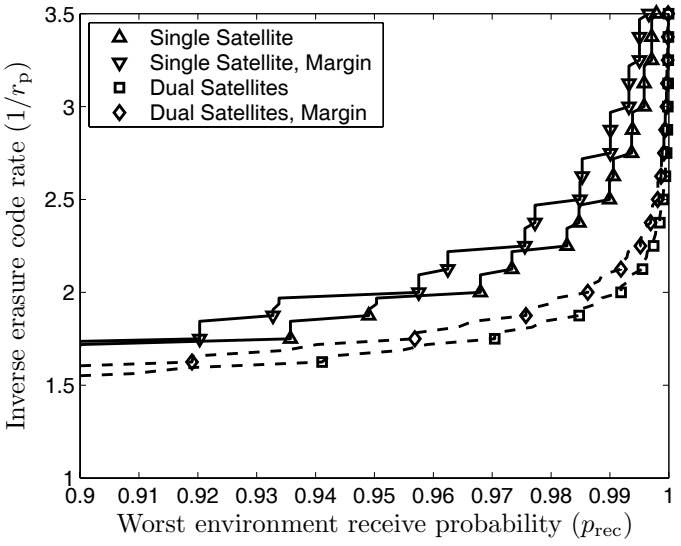
6.6 System Design and Infrastructure Considerations

One of the main goals with this work is to provide technical insights to how system resource requirements in terms of coding and interleaving overheads are affected by changes in the system infrastructure and channel statistics. Figure 6.6 and Figure 6.7 illustrate how the techniques proposed in this thesis can be used for such analysis. In these figures the required inverse information rates to reach a certain number of end users in the worst of an ensemble of environments are plotted. At every point of the curves the number of packets or segments, k, has been selected as $k = k_{\min}$ according to (6.2). Two different scenarios are considered for both the symbol level and the packet level coding schemes. In Figure 6.6 users in all environments except the urban are targeted while the systems in Figure 6.7 are designed to reach mobile end users in all defined environments. Note the difference in t_{delay} for the two figures.

The dashed curves are results computed when two satellites have been used to provide space diversity. The channels between the mobile terminals and the satellites were assumed statistically independent and identically distributed. By comparing these curves with the continuous curves representing a single transmitting satellite the gain from space diversity in terms of data rates can be quantized. These examples confirm the intuitive result that high requirements on the receive probability makes the transmit diversity system increasingly attractive. Note that the code is applied before the transmitted stream is divided over the various satellites and thus the comparison with the single satellite scenario is fair as far as usage of spectral resources goes. This also implies that with inverse information rates below 2 bits per channel use, segments or packets from both satellites are required to decode the data unit. Clearly this is

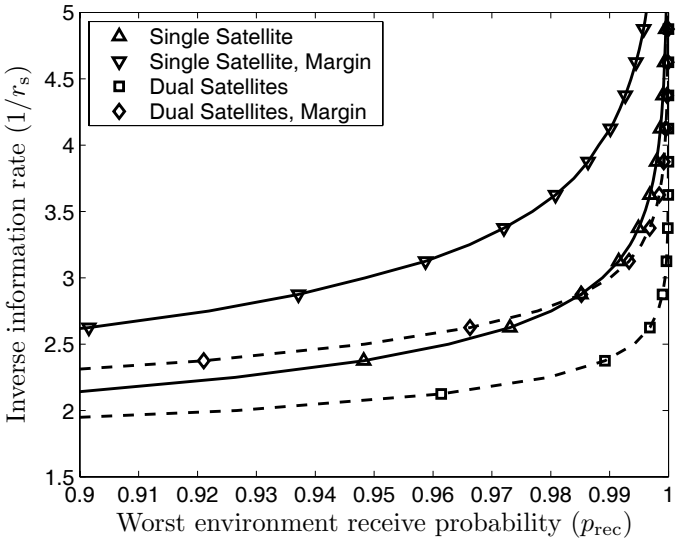


(a) Symbol level coding, 7 s delay

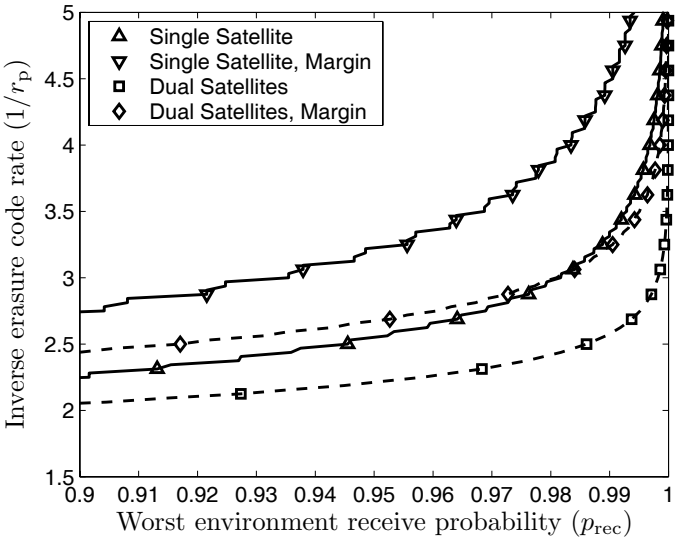


(b) Packet level coding, 7 s delay

Figure 6.6: Required inverse code rates to reach a certain part of the end users in the worst of the highway, suburban, and rural environments.



(a) Symbol level coding, 90 s delay



(b) Packet level coding, 90 s delay

Figure 6.7: Required inverse code rates to reach a certain part of the end users in the urban environment.

not attractive in practice as this would mean that a stationary receiver in perfect line of sight with one satellite will not be able to recover the transmitted data if the other satellite is blocked.

For most wireless communication systems a link margin is used to ensure satisfactory performance in the presence of, for example, rough weather conditions, shadowing, implementation losses etc. A mobile broadcast system relying on time diversity also requires such a margin. It may be meaningful to also consider a *time margin* to ensure satisfactory performance in scenarios where the temporal channel conditions are less favorable than in the channel model. One way of introducing such a time margin is to compute how much additional coding overhead is required to satisfy some quality of service requirement. This is illustrated in Figure 6.6 and Figure 6.7 where the required coding overheads for channels without margin are compared with designs that provide the same service availability over channels where the average duration and stationary probability of being in S_3 have been increased by 25%, i.e. $\bar{t}_3$ and $\Pr(S_3^\infty)$ of Table 6.1 have been increased by 25% and $\Pr(S_1^\infty)$ and $\Pr(S_2^\infty)$ decreased accordingly. From Figure 6.6 we may notice that introducing such a time margin would imply an increase in the inverse coding rate that usually is less than 10%. However, as can be seen in Figure 6.7, when the channel conditions are more severe, providing this extra margin can be quite costly.

Finally, by comparing Figure 6.6(a) and Figure 6.7(a) with Figure 6.6(b) and Figure 6.7(b) the relative performance of different transmission schemes can be examined. For a fair comparison the information rate, r, resulting from the modulation and coding used to transmit and protect the packets transmitted using in the packet level coding scheme should also be included. In these examples, the channel conditions have been chosen such that an information rate $r = 1$ bit per channel use allows for a fair comparison. As can be expected the gain from using the symbol level scheme is most pronounced in the system scenario considered in Figure 6.6 and where environments with relatively large amounts of shadowing are considered. When the urban environment is considered as in Figure 6.7, the time spent in the shadowed state is smaller and thus the gain from exploiting data received in that state is decreased. For an exact analysis of the relative performance of the different transmission schemes the exact performance of the different codes considered for implementation should be taken into account. The packet level coding scheme may turn out more efficient if the erasure code used performs well or if additional time diversity can be afforded in that scenario.

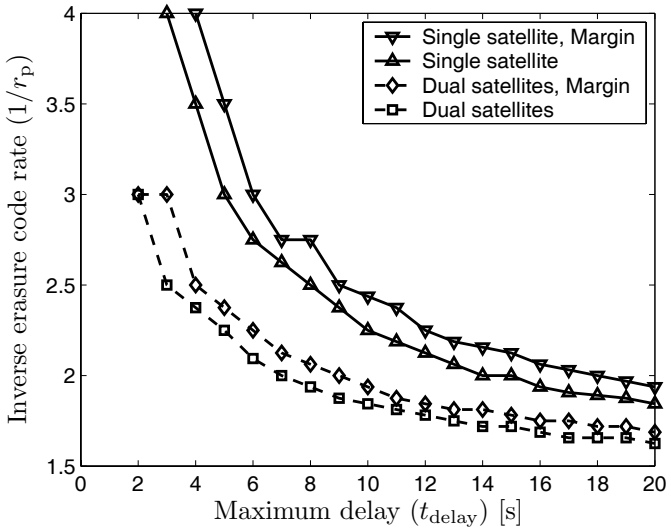
6.7 Delay vs. Coding Overhead Trade-Off

As previously discussed there is a trade-off between the acceptable delay and the coding overhead necessary to reach a certain percentage of the end users. Intuitively this is because with a short delay there is more variance in the amount of channel blockage forcing a larger margin in the code performance to compensate. This means that tight delay requirements impose a cost. That is, to provide the service with the desired availability using a shorter delay either the effective data rate needs to be reduced or more spectrum or power is necessary.

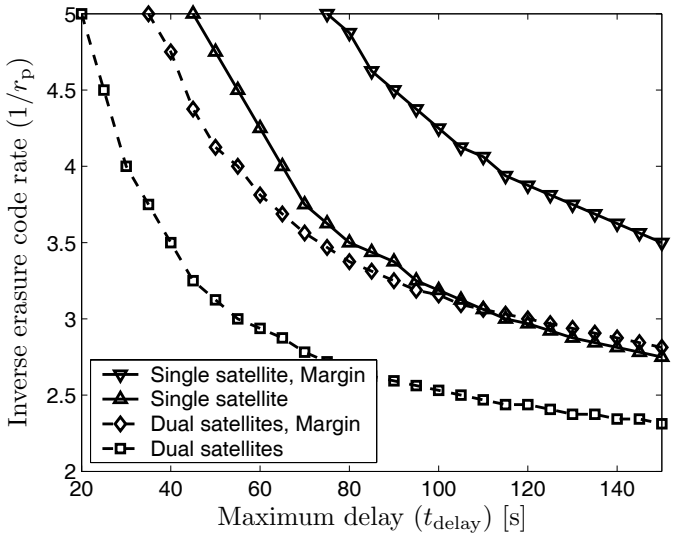
As the method derived in Chapter 5 provides an efficient mean to estimate the service availability, it is useful in evaluating this type of trade-off. In Figure 6.8 this is illustrated for a packet level coded system. Here the required inverse erasure code rate to reach 99% of the end users in the worst of an ensemble of environments is plotted as a function of the maximal acceptable delay. Similarly to the example of Section 6.6, scenarios with and without satellite diversity as well as the required time margin to provide the service if the blocked state is 25% more frequent are considered. On each point of the curves the minimum coding overhead to satisfy the availability constraint has been found using packet sizes optimized according to (6.2). As illustrated by Figure 6.4, the service availability is a non-decreasing function of the inverse information rate allowing for efficient search algorithms.

Figure 6.8 shows that providing services with strict delay requirements may indeed be quite costly in terms of bandwidth usage. A few other interesting observations can be made. To serve stationary users which may have access to only one of the satellites an inverse erasure code rate of at least two should be used. Thus, in the satellite diversity case of Figure 6.8(a) where all environments except the urban are considered, there is no point in allowing for more delay than 7–9 s. Comparing Figure 6.8(a) with Figure 6.8(b) one should also notice the difference in cost of providing a time margin. This implies that a system designed to operate under urban conditions will be more sensitive to model errors, i.e. for services that require a reliable forward link it is desirable to serve these users using, for example, terrestrial repeaters.

Together with the analysis of the spectral efficiency of different transmit diversity multiplexing schemes in Chapter 3, results as those presented in Figure 6.8 can provide important design insights. This is one of the topics discussed in the next chapter.



(a) Suburban, rural, highway environments



(b) Urban environment

Figure 6.8: Required inverse erasure code rate to reach 99 % of the end users in the worst environment at different delays.

Chapter 7

Concluding Remarks and Future Work

This chapter summarizes our results applicable to mobile satellite broadcast. Conclusions and main contributions are presented and we try to provide some general insights into air interface design. Finally some ideas for future work are outlined and we discuss effects not covered by the present work.

7.1 Conclusions

This part of the thesis treats mobile satellite broadcast with diversity. Main contributions include efficient analysis and design techniques based on statistical channel models as well as a discussion on the fundamental limitations of various transmit diversity multiplexing schemes. In the latter discussion we argued that by deploying an adaptive antenna at the receiver, gains in capacity comparable with those achieved by terrestrial MIMO systems can be provided. Below conclusions on these results are presented and some design aspects discussed.

7.1.1 Transmit Diversity Multiplexing Techniques

In this part of the thesis, different network configurations to provide transmit diversity in a mobile satellite broadcast system were outlined. The use a small adaptive antenna receiver to separate the different trans-

mitted signals in a single frequency network configuration was proposed. We argued that, for a system with transmit diversity such a solution could improve the spectral efficiency several times compared to the conventional systems in use today.

It should be noted that these results have been produced under the assumption of small implementation losses and the gain comes at the cost of more complicated and expensive receivers. On the other hand, the gain of this type of system configuration is so large that potentially one could reduce the requirements on the different components in the receiver and thus lower the additional cost. For example, while a larger number of antennas is required at the receiver, the requirements on the individual antenna elements could be relaxed and they could be made smaller and simpler. Although for slightly different reasons, adaptive antennas on the terminal side of a broadcast system are already implemented for commercial use in for example the Japanese/Korean MBCO system [MSS99].

Antenna arrays also have other benefits, especially in a mobile environment with fading due to multi-path propagation. As the different channels fade, with a spatial channel receiver the chance that all receiving antennas should be in a deep fade is small and in such scenarios the performance benefits of the array may be even larger than those indicated by the discussion in Chapter 3. Also, since the different transmitters can be separated in space it is possible to use different coding and modulation techniques for the satellite and terrestrial signals even in a single frequency network configuration. This would for example make it possible to avoid transmitting signals with high peak-to-average ratios such as OFDM or multi-channel direct sequence code division multiplexing over satellite transponders.

7.1.2 Mobile Satellite Broadcast with Diversity

Herein an efficient technique for estimating the receive probability of a transmitted data unit given transmission parameters and channel statistics is derived. The method is general in that it allows for analysis of different transmission schemes, channel models with arbitrary number of states and a selection of system infrastructures. Based on the technique several elaborate design schemes can be considered and the effect of various parameters analyzed. For example, different designs of the coded time diversity may be examined and schemes where transmission resources like coding overhead or delay are minimized while ensuring sat-

isfying service availability considered. The results also allow for efficient estimation of the impact of changes in the system infrastructure. We may for example estimate the relaxation of required radio resources that can be provided by employing space and terrestrial diversity. Design trade-offs, such as that between coding overhead and acceptable delay can also be considered. If adequate channel statistics are available even the effects of different satellite orbits can be taken into account. This way, the work presented in this thesis can provide insights into the technical aspects of an overall system design including commercial requirements, cost, and system performance.

7.1.3 Insights into Technical Design and Performance

While a proper overall system design as discussed in Section 1.3.4 requires that detailed commercial requirements are taken into account in the technical part of the design, some insight can be provided solely from technical aspects.

For mobile satellite broadcast systems with space diversity, there is in general a limit on how much time diversity it is meaningful to provide. This is because, to serve stationary and slow moving terminals, sufficient data should be transmitted even if only the signal of a single satellite is received. Otherwise acceptable receive conditions from more than one satellite needs to be ensured and the provided space diversity will actually reduce the service availability area for stationary receivers. For example, with the 99 % availability constraint considered in the example illustrated in Figure 6.8(a), for the dual satellite system providing more than 7-9 s of delay does not provide any further gain in coding overhead. Thus, with space diversity there is a "sweet spot" for the acceptable delay where minimum coding overhead is ensured while providing a certain service availability.

At this delay sweet spot it is interesting to compare the spectral efficiency of the mobile satellite diversity system with a single satellite system serving stationary users. To make the comparison simple, inefficiencies in the used channel code and transmission scheme are ignored. The comparison of the transmission diversity multiplexing efficiency provided in Chapter 3 will then hold provided that the multiplexing scheme allows the different transmitters to transmit individually coded data. Note that the last remark excludes traditional single frequency techniques where identical signals are transmitted from all transmitters. If the multiple

frequency configuration of Section 3.2 is considered for the diversity multiplexing, each transmitter of the mobile system will only have access to $1/(s+1)$ of the spectrum of the stationary system. Thus the achievable data rates of such a mobile system would be considerably lower compared with a single satellite system designed for a line of sight scenario. Next, consider the adaptive antenna single frequency configuration for the diversity multiplexing. This system would allow identical bandwidths for the stationary and the mobile transmit diversity systems. Assuming identical receiver antenna arrays and ignoring imperfections in the spatial signal separation, the mobile system should provide identical data rates as the stationary system. Thus, at the cost of providing sufficient time and transmit diversity, a mobile satellite broadcast system may offer comparable spectral efficiency as a line of sight system with identical infrastructure at the receivers.

7.2 Further Extensions and Discussion

7.2.1 Further Work

Below a few possible extensions of the work presented in this thesis are outlined. We discuss transmission, implementation, overall system and measurement verification aspects, but other issues like channel modeling and prediction should also be addressed to fulfill the goals outlined in Section 1.3.4.

More Elaborate Transmission Schemes

Several natural extensions of the techniques presented herein exist. To begin with, more general transmission schemes may be analyzed with similar techniques. For example one could allow for non-uniform distributions of the transmitted segments in time to reduce the delay for receivers in favorable conditions [MGO04b]. One could also combine the symbol and packet level coding schemes by adding additional channel coding on the individual packets and spreading the data in time on both packet and symbol level. For the analysis, if the individual packets are not overlapping, the Markov property could be preserved and similar analysis applied.

With an efficient estimation of the service availability, several elaborate design schemes beyond those in Chapter 6 can be constructed. For

example, if sufficient channel data can be provided the availability in different geographical locations and using different satellite orbits etc. may be included in the design.

Implementation Issues

In Chapter 3 it was shown that using an adaptive antenna at the receiver might be attractive to boost spectral efficiency. As compared with most terrestrial implementations of adaptive antenna communication systems this application has some implementation advantages. For example, the different transmitters will always be well separated and since the transmitted data is jointly coded, the more "interference" the more useful data is available for the decoding. However, there are other aspects that require more attention, for example it would be attractive to use different modulation schemes for the terrestrial and space transmissions to accommodate for the different channel conditions. This will complicate the use of non-linear signal separation techniques at the receiver. If linear signal separation techniques are used, where the various transmissions are separated solely based on their spatial properties, the receiver array may require additional elements to reduce the impact of interference. The above outlines some of the points that need to be addressed when realistically estimating the performance of this type of system and implementing it.

End-to-end System Aspects

To provide a complete analysis of an end to end system such as that envisioned in Chapter 2 the work contained herein needs to be expanded. For example, the present analysis does not take into account the explicit scheduling and multiplexing of files. If different service classes with different requirements are allowed, designing a scheduler as illustrated in Figure 2.1 is challenging. Here analysis and design techniques similar to those derived in this thesis may prove useful.

Section 7.1.3 provides some basic insights to the impact on spectral efficiency from using different system configurations in conjunction with transmit and time diversity. However, for a detailed understanding this topic needs to be addressed further. For example the impact of organizing the different transmitters in a single frequency configuration in conjunction with time diversity should to be examined.

Verification Against Measurements

The results derived in this part of the thesis are based on various assumptions and a relatively simple statistical channel model as outlined in Section 4.2 and Section 5.1. To validate the performance of the service availability estimation and insights provided from the different design examples, these results need to be compared against channel measurements or detailed simulations.

7.2.2 Discussion

At the same time as the methods derived in this part of the thesis can provide important insights, it must be emphasized that they are based on approximations which are valid under certain assumptions. This means that some issues that affect the system performance cannot easily be evaluated using the presented analysis. Important examples of such issues include:

Synchronization and Channel Estimation. To reacquire synchronization and estimate the channel coefficients after they have been lost in general requires some time under which the received signal cannot be decoded. If this delay is long, or if the quality of the channel estimates change over time, it needs to be taken into account in a performance analysis. To some extent this could be included in the present analysis by increasing the average time and probability of a blocked state. However exact analysis can not be performed while maintaining the assumption under which the techniques of Chapter 5 were derived. Thus, to analyze these types of effects Monte-Carlo simulations or other techniques may need to be applied.

Correlation in the fast fading. Channel realizations within a fast fading channel state are in practice always going to have some correlation which will degrade performance. How much depends on a number of factors including vehicle speed, the number of segments the coded data unit is divided into, the terminal environment etc. This effect could be included in the analysis by allowing for some correlation between different channel realizations but for a detailed analysis system simulations are likely needed.

Transmit diversity, self interference. There are a number of ways to organize the transmitters of a broadcast network to provide trans-

mit diversity. In the XM and Sirius radio networks the available frequency band is divided into three parts, with one part allocated for each of the transmitting satellites and one for the network of terrestrial repeaters. In the Japanese/Korean MBCO system the transmitting satellite and the terrestrial repeaters are organized in a single frequency network using code division multiplexing techniques. Either way, to provide high spectral efficiency, in schemes where the different transmitters share the same frequency band some interference between the different transmitted signals is likely. This would also hold if a small adaptive antenna is used at the receiver as proposed in Chapter 3. While some analysis for such effects exists in the literature [Mic02] it may be difficult to include in the results derived herein.

Part II

MIMO Communications with Partial Channel State Information

Chapter 8

System Vision and Contributions

We consider analysis and design techniques for flat fading MIMO communication systems. The receiver has access to perfect channel state information while the transmitter knowledge is limited to statistical properties of the fading. This type of system is normally characterized by a random complex valued channel matrix where each entry describes the attenuation and phase of the channel between two antenna elements. Various statistical properties of this matrix are analyzed and transmission schemes are proposed.

This chapter consists of two sections. First, a short general introduction is given and we outline problems that will be addressed in the later chapters. Second, the remaining chapters of this part of the thesis is outlined. To that end, contributions of this thesis and publications of our work are pointed out and related to earlier results.

8.1 System Vision

Future wireless communication systems are likely to require considerably higher data rates than systems of today. At the same time, the available radio spectrum is scarce and increasing transmit power is not an option in multi-user systems. To solve this problem, communication systems employing multiple antennas at both the transmitter and receiver sites, forming multiple-input, multiple-output, MIMO, architectures, have been

suggested [PGNB04]. The channel capacity, the maximum rate at which data can be communicated without errors, has been shown to increase proportionally to the minimum number of antennas at the transmitter and receiver sites under favorable propagation conditions, [Tel99, FG98, CTKV02]. Therefore, using multi-element antenna configurations is one promising technique for tomorrow's high speed wireless communication.

This thesis investigates analysis and design techniques based on statistical properties of a narrow-band MIMO channel. A flat fading MIMO channel is commonly characterized by a *channel matrix* $\boldsymbol{H}$, where the ith, jth complex element of $\boldsymbol{H}$ models the random attenuation and phase between the ith receiver and jth transmitter. The receiver is assumed to have perfect knowledge of the realizations of $\boldsymbol{H}$ while the transmitter knowledge is limited to the stochastic process generating $\boldsymbol{H}$. This could correspond to a communication system with a limited feedback link which does not allow transmission of the instantaneous channel responses but allows feedback of parameters that change on a slower time scale. The contributions to MIMO communications in this thesis can be divided into two parts, an analysis part and a design part.

In the analysis part, consisting of Chapter 10 and Chapter 11, we are interested in statistical properties of the eigenvalues of $\boldsymbol{H}\boldsymbol{H}^*$ and of fundamental system limitations set by information theory. The statistical properties of these eigenvalues are interesting as they are related to the channel attenuation in different spatial dimensions and they limit the achievable system performance. However, deriving the true distributions and characteristics of these quantities is complicated except for simple special cases. Instead, the properties are studied under the assumption that the number of antennas on either side of the MIMO system is large. As will be demonstrated, the results derived herein still provide useful approximations for limited array sizes. The increase of capacity in a MIMO system comes at the price of increased cost, space, and computational complexity. This price is more easily motivated at the access point where space and power is less limited and the cost is shared among several terminals. Therefore the assumption that more antennas are employed at the access point than at the terminals is reasonable. In this thesis a downlink scenario where the number of transmit antennas is larger than the number of receive antennas is studied. This scenario is often the most critical due to asymmetrical data rate service requirements. However, similar approaches could be applied to the uplink case.

In the design part, Chapter 12 and Chapter 13, we attempt to optimize a MIMO communication system with access to channel statistics at

the transmitter. Two scenarios are considered. In the first a near optimal design that attempts to maximize the mutual information is propose based on the analysis techniques derived in Chapter 10 and Chapter 11. While such a transmission scheme is not practical to implement, the resulting design can provide important insights. As an example our results show how spatial correlation, if properly dealt with, in some scenarios can improve performance. Second, in Chapter 13, inspired by the results of Chapter 12, we propose a more practical design. The goal of the design is to devise a more realistic transmission scheme that can adapt to the channel state information currently available at the transmitter.

8.2 Outline and Contributions

This section outlines this part of the thesis, states contributions and relate these to earlier work. Where applicable earlier publications of our results are pointed out.

8.2.1 Chapter 9, System Overview and Models

A model describing a narrow-band fading MIMO communication system is presented. The receiver is assumed to be perfectly synchronized with the transmitter and have perfect knowledge of the channels impulse response. Channel knowledge at the transmitter on the other hand is more uncertain as it only has access to fading statistics. In particular we consider channels where the fading can be modeled using complex Gaussian distributions. Two cases will be studied, either the spatial channels are Rayleigh fading and correlated or they are independently Ricean fading. In [VM01] such models were named covariance and mean feedback.

8.2.2 Chapter 10, Asymptotic Channel Eigenvalue Distribution

To understand the effects of random fading on the performance of a MIMO communication system, it is vital to understand the statistical behavior of the eigenvalues of $\boldsymbol{H}\boldsymbol{H}^*$. During the remainder of this thesis these are denoted *channel eigenvalues*. Amongst other things, their distribution determines the capacity of the system. Also, practical methods to increase transmit rate, such as spatial multiplexing with bit and power loading [RC98] depend on these properties.

Unfortunately, deriving the distribution of the channel eigenvalues is not straightforward. While expressions exist for the case when all channel elements are uncorrelated, [And84, Ede89], they are in many cases too complicated for practical use. For more efficient analysis and to study more complicated scenarios approximations are necessary. One way to derive approximations of the channel eigenvalues is to study the deterministic empirical distribution the eigenvalues converges to when the number of antennas on both sides of the system is large. This is done in for example [CTKV02, Mül02a, Mül02b]. In this thesis the limiting statistical distributions of the channel eigenvalues are derived for the case when the number of antennas on one side of the communication system is large. The main goal is to find a simple closed form approximation of the distribution that provides insight to, for example, how correlation between the channel elements affects system performance.

The asymptotic channel eigenvalue distributions have a direct analogy to the asymptotic eigenvalue distributions of sample covariance matrices. Our derivation is based on previously published results regarding asymptotic analysis of such eigenvalues. In [And63] a limiting distribution of the eigenvalues of sample covariance matrices in the real, time independent case was derived, in [Gup65] this result was extended to the complex case. Asymptotic second order moments of the eigenvalues for the complex, time dependent case were derived in [SS97].

In Chapter 10 limiting distributions of the channel eigenvalues for the case when the number of transmit antennas is large are derived. The derivation follows the earlier results in [And63, Gup65, SS97]. The derived distribution allows for spatial correlation, is closed form and simulations indicate that it provides a good approximation of the true distributions also at realistic system dimensions. Note that the non-zero eigenvalues of $\boldsymbol{H}\boldsymbol{H}^*$ and $\boldsymbol{H}^*\boldsymbol{H}$ are identical and thus it should be straightforward to derive similar results if the number of receive antennas is large instead. We also allow for correlated transmit data, i.e. we introduce a hermitian positive semidefinite matrix $\boldsymbol{Q}$ and study the eigenvalues of $\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*$. This is interesting since the correlation of the transmitted data streams can be optimized to maximize system performance. Such aspects are taken into account in the next two chapters.

8.2.3 Chapter 11, Asymptotic Mutual Information and Outage Capacity

Understanding the effects of correlated fading on the capacity of a MIMO communication system is complicated. Even computing the capacity for uncorrelated systems is difficult except for simple special cases [Tel99, FG98, Fos96]. The case of correlated fading among channel elements is an additional complication and already to analyze the mutual information one has to resort to time-consuming simulations or numerical integration providing limited insight to the problem.

As an alternative to simulations, bounds and approximations of the MIMO capacity have been derived. In [SFGK00] useful upper and lower bounds on the outage capacity are derived. The bounds are derived under the assumption that the fading is uncorrelated on either transmit or receive site. Although the derived bounds are not closed form they are significantly simpler than the original expressions and provide some intuition to the problem. Also, in [SFGK00], a closed form upper bound of the average mutual information of a MIMO communication system is derived under the same conditions. An upper bound based on the concave nature of the mutual information and Jensen's inequality is found in [LK02, Loy01]. The bound copes with correlation at both transmit and receive sites, and has the advantage of being simple to compute.

An approximation of the channel capacity was derived in [CTKV02]. This result is valid asymptotically as the number of antennas on both sides of the MIMO channel approaches infinity. Correlation is allowed at both transmit and receive sites according to the same model applied in this thesis. However the derived expression is not closed form and does not allow finite numbers of antennas at one of the sites. Also, to work well as an approximation of the capacity, the expression requires large arrays. A similar approach is also taken in [Mül02b] but with a different channel correlation model. In [MSS03] an approximation of the MIMO channel capacity is derived for cases when the number of transmit and receive antennas is large. The derived approximation is shown to be accurate also at reasonable sized arrays and the results also suggest that the outage capacity can be obtained from the cumulative distribution of a normal distribution.

In [WV01] a simple expression for the outage capacity is derived for the case where either the number of transmit or the number of receive antennas is large. Correlation is allowed in the system, but only on the site with a finite number of antennas. Under this one-sided correlation

it is shown that the outage capacity can be obtained from the cumulative distribution of a limiting normal distribution. The paper considers Rayleigh as well as Ricean fading and it also presents higher order asymptotic approximations.

To illustrate the usefulness of the eigenvalue statistics derived in Chapter 10, simple expressions of the mutual information and the outage capacity of a MIMO communication system are formulated in Chapter 11. The derived expressions are asymptotic in the number of antennas on the transmit side. They have the advantages of being closed form, allowing for spatial correlation on both transmit and receive sites according to the same model as in [CTKV02] as well as correlation between the transmitted data signals. We show that under certain conditions on the correlation between the spatial channels and transmitted data the outage capacity can still be obtained from a normal distribution as the number of transmit elements increases, thus extending this aspect of [WV01]. Simulations as well as channel measurements indicate that the derived expressions work well as approximations of the true properties also for limited array sizes.

Most of Chapters 10 and 11 have previously been published in

> Cristoff Martin and Björn Ottersten. Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading. *IEEE Trans. Wireless Commun.*, July 2004.
>
> Cristoff Martin and Björn Ottersten. Analytic approximations of eigenvalue moments and mean channel capacity for MIMO channels. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2002.
>
> Mats Bengtsson, Cristoff Martin, Björn Ottersten, Ben Slimane, and Per Zetterberg. Recent advances on MIMO processing in the SATURN project. In *Proc. IST Mobile Communications Summit*, June 2002.

8.2.4 Chapter 12, Approximate Transmit Covariance Optimization

This chapter considers the optimization of the transmit covariance given knowledge of the spatial correlation of the MIMO channel. Recently, several authors have addressed this topic. Relevant work includes [VM01] and [JG04] which consider this type of problem when the channel matrix

is only correlated on the transmission side. In [JB04] this requirement is relaxed and correlation also on the receiver side is taken into account. Conditions are also derived under which beamforming solutions, where only a single spatial channel is used, are optimal. The analytic results of these papers reveal some of the structure of the optimal covariance between the transmitted signals, but tedious Monte Carlo methods are required to find the exact solution. The computational complexity of such methods is high and the insight they provide is therefore limited. This issue was partly addressed in [SM03]. Here an analytical approach is derived for optimizing an architecture consisting of two transmit and many receive antennas. The channel is assumed to be correlated only at the transmitter side. While the work presented in this paper provides significant simplifications and insights, the exact optimization still requires numerical integration.

In the above work, the transmit covariance is chosen to optimize the exact average mutual information. Due to the complexity of the exact expression this is a difficult problem to solve and as an alternative approach approximations can be optimized. Such an approach is taken in this paper and has also been explored in the literature. In [IUN03] the optimal transmit covariance is approximated by the covariance found using the average channel variance. This allows for some interesting observations, for example the authors show that if properly dealt with, spatial covariance may well increase capacity. Other examples of where approximate approaches to the transmit covariance optimization has been taken include [MSS03].

Using the approximation of the MIMO channel average mutual information derived in Chapter 11 we find a transmit signal distribution that optimizes the system. Numerical examples are provided that, for example, illustrates that if compensated for, in some scenarios, a correlated channel can provide a higher channel capacity than an uncorrelated channel. Most of the results in this chapter have previously appeared in

> Cristoff Martin and Björn Ottersten. Approximate transmit covariance optimization of MIMO systems with covariance feedback. In *Asilomar Conference on Signals, Systems and Computers*, 2003.

8.2.5 Chapter 13, Bit and Power Loading With Mean or Covariance Feedback

From a capacity viewpoint, the optimal communication scheme when the channel is partially known at the transmitter is to transmit code words that result in Gaussian distributed symbols with power and correlation selected based on the available information [GJJV03]. This way, transmit power is allocated to the directions where it is the most useful, i.e. the less attenuation in a direction the more power is allocated and higher data rates can be supported.

For practical systems, using Gaussian distributed symbols is not an option. Instead the transmitted data symbols belong to some finite alphabet, resulting in an optimization problem where different constellation sizes and transmit powers are allocated to different directions to satisfy some constraint on the quality of the received data and to maximize data rates. Here, we term this type of technique spatial loading and in this chapter we propose a simple optimization scheme for optimizing the transmitted data when a maximum likelihood detector is used in the receiver. The proposed scheme allows for both mean and covariance feedback and we demonstrate gains compared with systems not exploiting the available information. While this type of system may appear simplistic, combined with a powerful error correcting code, similar designs without access to channel state information have been shown to perform near capacity on spatially uncorrelated channels [HtB03].

Most practical techniques for exploiting the gain promised by multi-element antenna systems are either based on the assumption of no channel knowledge at the transmitter and uncorrelated channels or perfect knowledge at the transmitter. Examples of techniques included in the first category are space-time coded and BLAST systems [Ala98, TSC98, Fos96], the latter can be exemplified by spatial loading schemes with perfect channel knowledge such as [RC98, ZO03]. Recently there has been significant interest in transmission schemes that can exploit partial transmitter channel state information. In [JSO02] a transmission scheme was devised where orthogonal space-time block coding was combined with transmit weights to minimize the bit error rate when partial channel state information is available at the transmitter. A special case of this transmission scheme is in [XZG04] extended to frequency selective channels. By adding bit and power loading in the frequency domain based on partial channel state information system performance is improved. Thanks to the structure of the underlying orthogonal space-time block

code the receiver complexity is limited and the design of the transmit weights straightforward. Unfortunately this structure is also a constraint that in some scenarios means that not all spatial dimensions can be exploited. For example, if the channel is perfectly known the scheme will converge to a one-dimensional beamforming solution leaving the remaining dimensions unused. From a single link data rate perspective it is well known that this is not an optimal solution [RC98]. Instead it might be attractive to consider a spatial multiplexing solution which offers more degrees of freedom at the cost of increased design and implementation complexity. The authors of [SS03] propose a scheme where a linear precoder is added to a BLAST system. Based on partial channel state information the precoder is chosen to minimize the error rate of the system. Other relevant work include [Nar03] where antenna selection based on covariance feedback is implemented. However, in both papers above, the data rates on the different spatial channels are constrained to be equal. In this chapter, we propose a design that attempts to maximize the data rate under an error rate constraint. Similar to spatial loading techniques using perfect channel knowledge [RC98] our scheme maximizes the data rates by exploiting favorable directions of transmission. Partial channel state information is allowed and perfectly known, completely unknown and correlated channels are included as special cases. For those extremes our technique converges to well known solutions.

Most of the results of this chapter have previously been appeared in

> Cristoff Martin, Svante Bergman, and Björn Ottersten. Simple spatial multiplexing based on imperfect channel estimates. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2004.
>
> Cristoff Martin, Svante Bergman, and Björn Ottersten. Spatial loading based on channel covariance feedback and channel estimates. In *Proceedings European Signal Processing Conference*, September 2004.

The results presented in this chapter are based on joint work with Svante Bergman.

8.2.6 Chapter 14, Concluding Remarks and Future Work

In Chapter 14.2 the contributions of the second part of the thesis are summarized, the results are critically discussed and opportunities for further

work outlined.

Chapter 9

System Model

This part of the thesis considers design and analysis techniques for terrestrial multi-input, multi-output communication systems. Such systems are often characterized by a channel matrix, $\boldsymbol{H}$, whose elements describe the attenuation and phase properties of the channels between the transmitter receiver pairs of the system. Here, $\boldsymbol{H}$ is modeled in statistical terms and we consider models that allow for realistic propagation conditions.

Section 9.1 provides a general system overview. In Section 9.2 we detail statistical channel models that capture some of the typical behavior of MIMO channels in different scenarios. In the next few chapters, analysis and design techniques are derived based on these properties.

9.1 General

In this chapter we present a channel model for a communication system consisting of a single transmitter receiver pair both employing multiple antennas. One possible application is wireless local area networks, WLAN, such systems require high speed communication and are less limited by space constraints than for example mobile phones making it possible to employ multiple antennas at both ends of the communication channel.

The channel is flat fading, i.e. the time-discrete channel between the input of one of the modulators at the transmitter and the output of the demodulators at the receiver is modeled by a multiplicative fading factor and an additive noise component. While, by nature, high

speed wireless communication usually results in inter-symbol interference, this might be combated using modulation schemes such as orthogonal frequency-division multiplexing, OFDM, [WE71, Cim85]. OFDM avoids inter-symbol interference by transforming the frequency selective channel to multiple orthogonal flat fading channels separated in frequency. While OFDM suffers from some loss in signal to noise ratio and transfer rate, avoiding inter-symbol interference significantly simplifies the receiver making OFDM the modulation scheme of choice for many high speed WLAN standards.

The received signals in a multi-element antenna communication system can be modeled as a superposition of the transmitted signals as illustrated in Figure 9.1. The signals propagate along multiple paths through the environment, resulting in a fading channel that here will be modeled in statistical terms. If there is sufficient scattering in the environment, the channels between different transmitter receiver pairs will provide different attenuation and phase. With less scattering in the near field, the different propagation paths will be more similar and the channel responses will be correlated in space [SFGK00, MBB+00]. There might also be a dominant path or a line of sight component affecting the channel statistics [YBOB02].

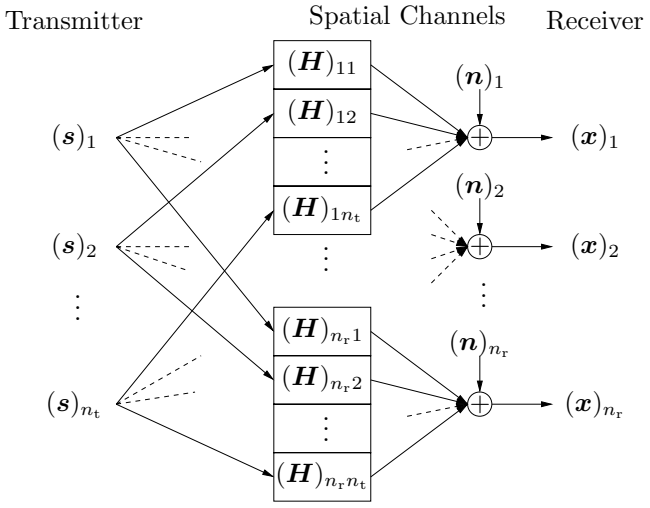


Figure 9.1: A MIMO Communication System.

In the following section we will present statistical models for the scenarios above. Various properties of these models are analyzed in the next few chapters. In all cases, the transmitter is assumed to have access to the parameters of the statistical process generating the MIMO channel responses. Based on the analysis, transmission schemes can be designed that adapt to the current channel statistics. Such schemes will be presented and analyzed in the last chapters of this part.

9.2 Data Model

We consider a MIMO system consisting of n_{t} transmitters communicating with n_{r} receivers over a flat fading channel. Transmission over the channel is modeled through a channel matrix, $\boldsymbol{H} \in \mathbb{C}^{(n_{\mathrm{r}} \times n_{\mathrm{t}})}$, the elements of which model the attenuations and phase differences between the various transmitter/receiver pairs. The output, $\boldsymbol{x} \in \mathbb{C}^{n_{\mathrm{r}}}$, resulting from a single usage of the channel to transmit the data $\boldsymbol{s} \in \mathbb{C}^{n_{\mathrm{t}}}$, is then found as,

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{n}, \tag{9.1}$$

where $\boldsymbol{n} \in \mathbb{C}^{n_{\mathrm{r}}}$ is additive white complex Gaussian noise, see Figure 9.1. To make the required output power of the different transmission schemes presented in the paper comparable, the SNR is defined as the ratio between the power of the received data signal, $\mathrm{E}\,|\boldsymbol{H}\boldsymbol{s}|^2$, and the received noise power, $\mathrm{E}\,|\boldsymbol{n}|^2$, when equally powered uncorrelated data is transmitted, i.e. when $\mathrm{E}\{\boldsymbol{s}\boldsymbol{s}^*\}$ is a scaled identity matrix. Without loss of generality, the received noise is normalized so that $\boldsymbol{n} \in \mathcal{CN}(0, \boldsymbol{I})$ and the channel matrix is normalized so that $\mathrm{E}\,\|\boldsymbol{H}\|_{\mathrm{F}}^2 = n_{\mathrm{t}} n_{\mathrm{r}}$. Applying the definition and normalizations above then gives the SNR as $\mathrm{E}\,|\boldsymbol{s}|^2 = P$.

9.2.1 MIMO Channel Models

In what follows, the receiver is assumed to have perfect knowledge of $\boldsymbol{H}$ when it tries to detect or decode the transmitted symbols and code words. However, the channel knowledge at the transmitter is not necessarily perfect or complete. There are several reasons for the transmitter channel estimate to be in error. For example, the channel may have changed between the estimation and the usage of the estimates due to delay or the channel estimate may have been obtained via a low rate feedback channel not able to provide accurate estimates due to heavy quantization. For a more detailed discussion see e.g. [JSO02]. In scenarios where the

channel changes too fast to keep the transmitter estimates updated it may be more attractive to estimate the channel statistics which change at a slower pace. This will reduce the bandwidth and delay requirements on the feedback channel.

To exploit the partial channel knowledge at the transmitter, $\boldsymbol{H}$ is modeled stochastically. The model for $\boldsymbol{H}$ depends on the type of information available. Here it is assumed that $\boldsymbol{H}$, given the available channel knowledge, is complex Gaussian,

$$\begin{aligned} \operatorname{vec} \boldsymbol{H} &\in \mathcal{CN}(\operatorname{vec} \hat{\boldsymbol{H}}, \boldsymbol{R}_{\boldsymbol{H}}), \\ \boldsymbol{R}_{\boldsymbol{H}} &= \mathrm{E}\{\operatorname{vec}(\boldsymbol{H} - \hat{\boldsymbol{H}}) \operatorname{vec}^*(\boldsymbol{H} - \hat{\boldsymbol{H}})\}, \end{aligned} \tag{9.2}$$

thus the transmitter channel state information consists of $\hat{\boldsymbol{H}}$ and $\boldsymbol{R}_{\boldsymbol{H}}$. In practice, such knowledge could be estimated at the receiver and obtained at the transmitter via a feedback link.

In this work two special cases of the model above are considered, mean and covariance feedback, see [VM01]. Covariance feedback is applicable for non-line of sight conditions in richly scattering environments, e.g. indoor, wireless local area networks. This scenario will be considered in all the remaining chapters. Mean feedback is useful in scenarios where a line of sight component exists, but it might also be seen as modeling of errors in a channel estimate. This type of model will be considered in Chapter 13.

Covariance Feedback

Consider the *Kronecker channel model* [MBB+00] which is realistic for indoor non-line of sight conditions and limited array apertures [YBO+01]. Here the channel mean, $\hat{\boldsymbol{H}}$, is zero and the covariance matrix $\boldsymbol{R}_{\boldsymbol{H}}$ is structured as $\boldsymbol{R}_{\boldsymbol{H}} = {\boldsymbol{R}_{\mathrm{t}}}^{\mathrm{T}} \otimes \boldsymbol{R}_{\mathrm{r}}$ (where $\otimes$ is the Kronecker product), i.e. $\boldsymbol{H}$ may be generated as

$$\boldsymbol{H} = {\boldsymbol{R}_{\mathrm{r}}}^{1/2} \boldsymbol{G} {\boldsymbol{R}_{\mathrm{t}}}^{1/2} \tag{9.3}$$

where $\boldsymbol{G}$ is an n_{r} by n_{t} zero mean complex Gaussian matrix with independent identically distributed elements distributed as $\mathcal{CN}(0, 1)$. The non-negative definite matrices $\boldsymbol{R}_{\mathrm{t}}$ and $\boldsymbol{R}_{\mathrm{r}}$ can be interpreted as the covariance between the transmitters, the rows of $\boldsymbol{H}$, and the receivers, the columns of $\boldsymbol{H}$, respectively. The normalization imposed in Section 9.2 implies that $\boldsymbol{R}_{\mathrm{r}}$ and $\boldsymbol{R}_{\mathrm{t}}$ should be chosen such that $\operatorname{Tr} \boldsymbol{R}_{\mathrm{t}} = n_{\mathrm{t}}$ and

$\mathrm{Tr}\,\boldsymbol{R}_{\mathrm{r}} = n_{\mathrm{r}}$. This type of model will be considered in all the remaining chapters of this part of the thesis.

As a simple model for the amount of spatial correlation in some of our examples, we introduce the following transmit covariance matrix,

$$\boldsymbol{R}_{\mathrm{t}} = \begin{bmatrix} 1 & {\rho_{\mathrm{t}}}^* & {\rho_{\mathrm{t}}^2}^* & \cdots & {\rho_{\mathrm{t}}^{n_{\mathrm{t}}-1}}^* \\ \rho_{\mathrm{t}} & 1 & {\rho_{\mathrm{t}}}^* & & {\rho_{\mathrm{t}}^{n_{\mathrm{t}}-2}}^* \\ \vdots & & & \ddots & \\ \rho_{\mathrm{t}}^{n_{\mathrm{t}}-1} & \rho_{\mathrm{t}}^{n_{\mathrm{t}}-2} & & \cdots & 1 \end{bmatrix}. \tag{9.4}$$

A similar model is used for $\boldsymbol{R}_{\mathrm{r}}$, but with ρ_{r} instead of ρ_{t}, $|\rho_{\mathrm{r}}| < 1$, $|\rho_{\mathrm{t}}| < 1$. This model may be reasonable in the case of a uniform linear array, ULA, in a richly scattering environment and will be used to illustrate the behavior as a function of the correlation parameters. It has previously been applied in the context of capacity computations in e.g. [Loy01], and will be referred to as the *exponential correlation* case. However, it remains to be shown how well this model fits reality. Note that unless stated otherwise, the results in this thesis apply to general covariance matrices $\boldsymbol{R}_{\mathrm{r}}$ and $\boldsymbol{R}_{\mathrm{t}}$.

Mean Feedback

The channel state information consists of the channel mean $\hat{\boldsymbol{H}}$ and the covariance matrix $\boldsymbol{R}_{\boldsymbol{H}} = \sigma_{\boldsymbol{H}}^2 \boldsymbol{I}$. Note that $\hat{\boldsymbol{H}}$ can be seen as an estimate of $\boldsymbol{H}$ and $\sigma_{\boldsymbol{H}}^2$ as a measure of the uncertainty in this estimate. The normalizations in the previous section imply that $\sigma_{\boldsymbol{H}}^2 = 1$ results in $\hat{\boldsymbol{H}} = 0$ or no channel knowledge while $\sigma_{\boldsymbol{H}}^2 = 0$ results in $\hat{\boldsymbol{H}} = \boldsymbol{H}$ or perfect channel knowledge. Another interpretation of the model is a channel resulting from a line of sight path and local scattering in which case $\hat{\boldsymbol{H}}$ is rank one [YBOB02]. This channel model will be considered in Chapter 13.

Chapter 10

Asymptotic Channel Eigenvalue Distribution

In this chapter we are interested in the statistical properties of the channel eigenvalues. These eigenvalues can be seen as defining the channel attenuation in different spatial dimensions and therefore plays a central role in the analysis and design of a MIMO communication system. For example, as will be shown in Chapter 11, they define the channel capacity. We also believe that they may play an important role in the analysis and design of more practical transmission schemes like the spatial loading scheme presented in Chapter 13. In general, deriving the eigenvalue distribution of a MIMO channel using a realistic statistical channel model is difficult. Here a distribution is derived that is limiting when the number of transmit antennas is large. A non-line of sight indoor channel is considered, i.e. the channel matrix elements are correlated according to the covariance feedback model in Section 9.2.1. It should be noted that our results are straightforward to extend to the case of a large receive array.

This chapter is organized as follows. In Section 10.1 we define various entities used in this and the following two chapters. Most importantly, the term channel eigenvalue is exactly defined. Section 10.2 presents a limiting (for $n_{\mathrm{t}} >> 1$) distribution of the channel eigenvalues. Using numerical examples, Section 10.3 verifies that the derived distribution may serve as an approximation in systems with finite dimensions. Finally, Appendix 10.A provides proof for the results of Section 10.2.

10.1 Notation and Definitions

In this and the following two chapters an indoor, non-line of sight scenario is considered. To that end, the elements of the channel matrix are assumed to be zero-mean circularly symmetric complex Gaussian, i.e. $\hat{\boldsymbol{H}} = \boldsymbol{0}$ see (9.2). The spatial correlation between the channel matrix elements is modeled according to

$$\begin{aligned} \boldsymbol{R}_{\boldsymbol{H}} &= \mathrm{E}\{\mathrm{vec}\,\boldsymbol{H}\,\mathrm{vec}^*\,\boldsymbol{H}\} \\ &= \boldsymbol{R}_{\mathrm{t}}{}^{\mathrm{T}} \otimes \boldsymbol{R}_{\mathrm{r}}, \end{aligned} \tag{10.1}$$

where $\otimes$ denotes the Kronecker product [Gra81], i.e. the correlation feedback model is considered, see Section 9.2.1. Note from (10.1) that the $(n_{\mathrm{t}} \times n_{\mathrm{t}})$ matrix $\boldsymbol{R}_{\mathrm{t}}$ and the $(n_{\mathrm{r}} \times n_{\mathrm{r}})$ matrix $\boldsymbol{R}_{\mathrm{r}}$ are scaled covariance matrices for the rows (transmit antennas) and columns (receive antennas) of $\boldsymbol{H}$. To normalize the system, $\boldsymbol{R}_{\mathrm{r}}$ and $\boldsymbol{R}_{\mathrm{t}}$ are scaled so that $\mathrm{Tr}\{\boldsymbol{R}_{\mathrm{r}}\} = n_{\mathrm{r}}$ and $\mathrm{Tr}\{\boldsymbol{R}_{\mathrm{t}}\} = n_{\mathrm{t}}$.

The covariance matrix of the transmitted data is denoted $\mathrm{E}\{\boldsymbol{s}\boldsymbol{s}^*\} = \boldsymbol{Q}$. By limiting the transmitted power to P, i.e. $\mathrm{E}\{\boldsymbol{s}^*\boldsymbol{s}\} = \mathrm{Tr}\,\boldsymbol{Q} = P$, the SNR of the system using the normalizations above is P. The SNR has then been defined as the ratio between the power of the received data signal, $\mathrm{E}\,|\boldsymbol{H}\boldsymbol{s}|^2$, and the received noise power, $\mathrm{E}\,|\boldsymbol{n}|^2$, when equally powered uncorrelated data are transmitted, i.e when $\boldsymbol{Q} = \boldsymbol{I}P/n_{\mathrm{t}}$.

Define the channel eigenvalues as the eigenvalues of $\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*$. Let these be denoted $\boldsymbol{\lambda} = [\lambda_1 \ldots \lambda_{n_{\mathrm{r}}}]$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n_{\mathrm{r}}}$, and notice that

$$\mathrm{E}\{\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*\} = \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\boldsymbol{R}_{\mathrm{r}}. \tag{10.2}$$

Furthermore, $\boldsymbol{\lambda}^{\mathrm{r}} = [\lambda_1^{\mathrm{r}} \ldots \lambda_{n_{\mathrm{r}}}^{\mathrm{r}}]^{\mathrm{T}}$, $\lambda_1^{\mathrm{r}} > \lambda_2^{\mathrm{r}} > \cdots > \lambda_{n_{\mathrm{r}}}^{\mathrm{r}}$ are used to denote the ordered eigenvalues of $\boldsymbol{R}_{\mathrm{r}}$. Note that this formulation requires that all eigenvalues of $\boldsymbol{R}_{\mathrm{r}}$ are different. This simplifies the derivation and the resulting distribution in the asymptotic analysis in the next section, see [Gup65]. Finally the eigenvalues of $\boldsymbol{R}_{\mathrm{t}}{}^{1/2}\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}{}^{1/2}$ are written $\boldsymbol{\lambda}^{\mathrm{tq}} = [\lambda_1^{\mathrm{tq}} \ldots \lambda_{n_{\mathrm{t}}}^{\mathrm{tq}}]^{\mathrm{T}}$ where $\boldsymbol{R}_{\mathrm{t}}{}^{1/2}$ is a Hermitian square root factor of $\boldsymbol{R}_{\mathrm{t}}$.

10.2 Main Result

This section presents a limiting (for $n_{\mathrm{t}} >> 1$) distribution of the eigenvalues of $\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*$. The result is based on previous results regarding eigenvalue distributions of sample covariance matrices [And63, Gup65, SS97]. The main result of this chapter is summarized in the following theorem.

Theorem 1 *Assume that* $\boldsymbol{H}$ *and* $\boldsymbol{Q}$ *satisfy the conditions of Section 10.1,* $\lambda_i^{tq} < \infty$ *for all* i*, as* $n_t \to \infty$*, and that*

$$\lim_{n_t \to \infty} \frac{\left(\sum_{n=1}^{n_t} (\lambda_n^{tq})^3\right)^{\frac{1}{3}}}{\left(\sum_{n=1}^{n_t} (\lambda_n^{tq})^2\right)^{\frac{1}{2}}} = 0. \tag{10.3}$$

Then the limiting distribution of $\sqrt{n_t}(\boldsymbol{\lambda} - \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_t\}\boldsymbol{\lambda}^r)$ *is* $\mathcal{N}(\mathbf{0}, \boldsymbol{R}_{\tilde{\lambda}})$*, where the* i, j*th element of the covariance matrix* $\boldsymbol{R}_{\tilde{\lambda}}$ *of the limiting distribution is given by*

$$(\boldsymbol{R}_{\tilde{\lambda}})_{ij} = \delta_{ij} (\lambda_i^r)^2 \left\| \boldsymbol{Q}\boldsymbol{R}_t \right\|_F^2 n_t. \tag{10.4}$$

Proof: See Appendix 10.A. ■

We have used δ_{ij} to denote the Kronecker delta function, $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. The condition (10.3), gives a sufficient condition on the correlation between the transmitted signals and transmit side channel correlation for convergence of the limiting distribution as the number of transmit antennas increases, see Appendix 10.A. In practice this means that as more transmitter chains are added to the system, the added channels must not be completely correlated with the previous ones but should add to the diversity of the system.

Notice that the eigenvalues are asymptotically independent. If $\boldsymbol{Q}$ is a scaled identity matrix, i.e. uncorrelated transmitters, from the covariance expression it is clear that under the condition $\mathrm{Tr}\{\boldsymbol{R}_\mathrm{t}\} = n_\mathrm{t}$ the asymptotic variance of the channel eigenvalues has a minimum for $\boldsymbol{R}_\mathrm{t} = \boldsymbol{I}$. This is expected since uncorrelated columns provide more averaging (and thereby less variations) than the correlated case.

10.3 Verification Against Simulation

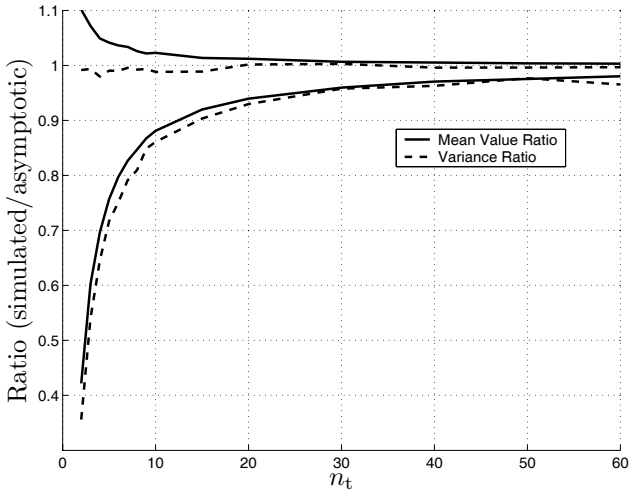
From Theorem 1 it is natural to approximate the distribution of $\boldsymbol{\lambda}$ as $\mathcal{N}(\mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_\mathrm{t}\}\boldsymbol{\lambda}^\mathrm{r}, \boldsymbol{R}_{\tilde{\lambda}}/n_\mathrm{t})$ at large but finite n_t. To test the accuracy of this asymptotic approximation, it has been compared with sample distributions from simulations. Note that when $n_\mathrm{t} < \infty$ there is a non-zero probability that the distribution may generate $\lambda_i < \lambda_j$ for some $i < j$, i.e. in a sense the approximation is unordered. To generate the sample distributions, a large number of channel matrices were generated according to the assumptions in Section 9.2 and from these, channel eigenvalues were

computed, sorted and inserted into histograms resulting in an ordered statistic. Thus, in the following figures the ordered statistics from the simulations are compared with the unordered asymptotic statistics. This comparison is valid as long as there is not too much overlap between the distributions, i.e. at high n_{t}. The performance with considerable overlap between the distributions of the individual eigenvalues might be better than that shown in the Figure 10.1 and Figure 10.2. For simplicity, in all the examples below $\boldsymbol{Q}$ is chosen as $\boldsymbol{Q} = \boldsymbol{I}/n_{\mathrm{t}}$.

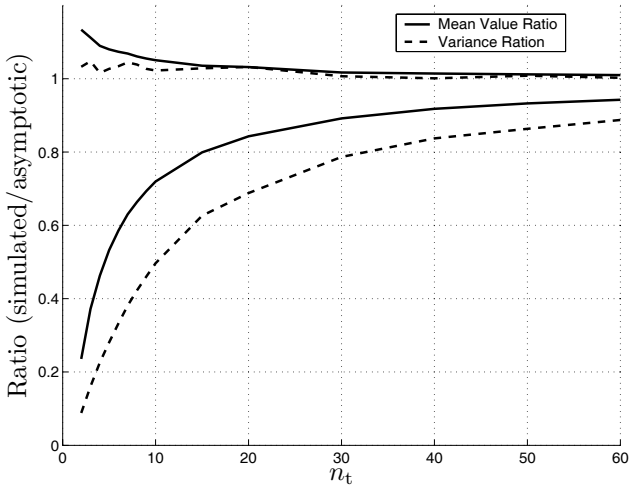
Figure 10.1(a) and Figure 10.1(b) investigate the convergence of the mean and variance of the asymptotic distribution as the number of antennas increases. The figures display the ratio between the simulated eigenvalue mean and variance and the asymptotic mean and variance as a function of the number of transmit antennas. In both figures a system with a two antenna receiver is studied. The receiver and transmitter correlation matrices, $\boldsymbol{R}_{\mathrm{r}}$ and $\boldsymbol{R}_{\mathrm{t}}$, are chosen according to (9.4) with $\rho_{\mathrm{r}} = 0.7$. Figure 10.1(a) displays a case when the transmit antennas are well separated, $\rho_{\mathrm{t}} = 0$, while in Figure 10.1(b) the transmit side is correlated and $\rho_{\mathrm{t}} = 0.7$.

From these figures we notice that convergence is slower in the correlated data case than in the uncorrelated case. This is expected since in this case there is less averaging. Also, the mean and variance of the larger eigenvalue converges faster than the mean and variance of the smaller eigenvalue. In many cases, such as capacity computations, see (11.4), the larger eigenvalues are more important to the application. Therefore this is an attractive feature of the derived distribution.

To investigate whether the limiting normal distribution of the eigenvalues might serve as an approximation of the true distribution also at finite array sizes, in Figure 10.2 we study simulated and asymptotic channel eigenvalue distributions for an $n_{\mathrm{t}} = 10$, $n_{\mathrm{r}} = 2$ system. Again, $\boldsymbol{R}_{\mathrm{r}}$ and $\boldsymbol{R}_{\mathrm{t}}$ were chosen according to (9.4) with $\rho_{\mathrm{r}} = 0.7$ and $\rho_{\mathrm{t}} = 0$. Even though the sample distributions are clearly not Gaussian (they do not take on negative values for example), there is still reasonable agreement with the limiting normal distribution. As the number of transmit antennas grows, this similarity increases.



(a) $\rho_t = 0$



(b) $\rho_t = 0.7$

Figure 10.1: Convergence of the mean and variance of the channel eigenvalues of an $n_r = 2$ system with and without correlation on the transmitter side. In both figures $\rho_r = 0.7$ and the top two curves correspond to the larger eigenvalue.

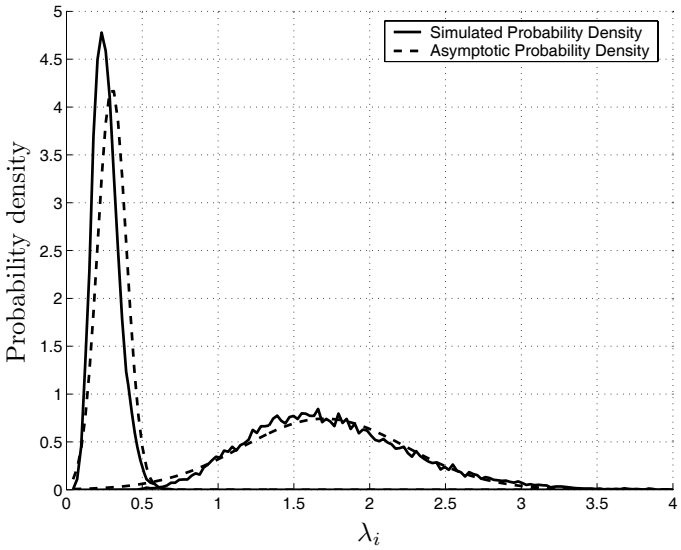


Figure 10.2: Simulated and asymptotic channel eigenvalue probability densities of an $n_\mathrm{t} = 10$, $n_\mathrm{r} = 2$, $\rho_\mathrm{r} = 0.7$ and $\rho_\mathrm{t} = 0$ system.

Appendix 10.A Proof of Theorem 1

The derivation of the limiting eigenvalue distribution of Theorem 1 is based on the observation that, if we let $\boldsymbol{h}(k)$ denote the kth column of $\boldsymbol{H}\boldsymbol{Q}^{1/2}$ we may write

$$\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^* = \sum_{k=1}^{n_\mathrm{t}} \boldsymbol{h}(k)\boldsymbol{h}^*(k) = \hat{\boldsymbol{R}}. \tag{10.5}$$

Therefore the derivation of an asymptotic eigenvalue distribution is practically identical to deriving the same property of sample covariance matrices. Note, from the normalization of $\boldsymbol{Q}$, $\mathrm{Tr}\,\boldsymbol{Q} = P$, the sum should converge to $\mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_\mathrm{t}\}\boldsymbol{R}_\mathrm{r}$ as n_t increases. An asymptotic eigenvalue distribution of sample covariance matrices was derived in [And63] for the case of time independent real samples, in [Gup65] this result was extended to the complex, time independent case. In [SS97] asymptotic second order moments of the sample covariance eigenvalues were derived for the complex, time dependent case.

We divide the proof in three parts, first notation is introduced, second the asymptotic second order moments are derived and third the limiting distribution is motivated.

10.A.1 Notation

Let $\boldsymbol{v}_1^\mathrm{r}, \ldots, \boldsymbol{v}_{n_\mathrm{r}}^\mathrm{r}$ be orthonormal eigenvectors corresponding to the eigenvalues of $\boldsymbol{R}_\mathrm{r}$, $\lambda_1^\mathrm{r}, \ldots, \lambda_{n_\mathrm{r}}^\mathrm{r}$. Furthermore,

$$\boldsymbol{S}_\mathrm{r} = \begin{bmatrix} \boldsymbol{v}_1^\mathrm{r} & \ldots & \boldsymbol{v}_{n_\mathrm{r}}^\mathrm{r} \end{bmatrix} \tag{10.6}$$

and

$$\boldsymbol{\Lambda}_\mathrm{r} = \begin{bmatrix} \lambda_1^\mathrm{r} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_\mathrm{r}}^\mathrm{r} \end{bmatrix}, \tag{10.7}$$

hence $\boldsymbol{R}_\mathrm{r} = \boldsymbol{S}_\mathrm{r}\boldsymbol{\Lambda}_\mathrm{r}\boldsymbol{S}_\mathrm{r}^*$, $\boldsymbol{S}_\mathrm{r}^*\boldsymbol{S}_\mathrm{r} = \boldsymbol{S}_\mathrm{r}\boldsymbol{S}_\mathrm{r}^* = \boldsymbol{I}$. Correspondingly, the random variables and vectors, λ_i and $\boldsymbol{v}_i$, $i = 1 \ldots n_\mathrm{r}$, and the random matrices $\boldsymbol{S}$ and $\boldsymbol{\Lambda}$ are defined such that $\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^* = \hat{\boldsymbol{R}} = \boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^*$. The diagonal of $\boldsymbol{\Lambda}_\mathrm{tq}$ consist of the eigenvalues of ${\boldsymbol{R}_\mathrm{t}}^{1/2}\boldsymbol{Q}{\boldsymbol{R}_\mathrm{t}}^{1/2}$, $\lambda_1^\mathrm{tq}, \lambda_2^\mathrm{tq}, \ldots, \lambda_{n_\mathrm{t}}^\mathrm{tq}$. For convenience, let $\bar{\boldsymbol{\Lambda}} = \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_\mathrm{t}\}\boldsymbol{\Lambda}_\mathrm{r}$, i.e. a diagonal matrix with the ordered eigenvalues of $\mathrm{E}\{\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*\}$, $\bar{\lambda}_i$ on the main diagonal. Note that $\bar{\lambda}_i = \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_\mathrm{t}\}\lambda_i^\mathrm{r}$, $i = 1 \ldots n_\mathrm{r}$.

10.A.2 Eigenvalue Moments

In this section asymptotic second order moments of the eigenvalues of $\boldsymbol{HQH}^*$ are derived. To derive them $\boldsymbol{T}$ and $\boldsymbol{Y}$ are introduced such that

$$\boldsymbol{T} = \boldsymbol{S}_{\mathrm{r}}^* \hat{\boldsymbol{R}} \boldsymbol{S}_{\mathrm{r}} = \boldsymbol{S}_{\mathrm{r}}^* \boldsymbol{S} \boldsymbol{\Lambda} \boldsymbol{S}^* \boldsymbol{S}_{\mathrm{r}} = \boldsymbol{Y} \boldsymbol{\Lambda} \boldsymbol{Y}^*. \tag{10.8}$$

Note that $\boldsymbol{Y}$ is unitary. It is well known that the eigenvectors of $\hat{\boldsymbol{R}}$ are not unique and in order for (10.8) to define $\boldsymbol{Y}$ uniquely, we require $\mathrm{Re}(\boldsymbol{Y})_{ii} \geq 0$, $\mathrm{Im}(\boldsymbol{Y})_{ii} = 0, \forall\ i$. Let $\sqrt{n_{\mathrm{t}}}(\boldsymbol{T} - \bar{\boldsymbol{\Lambda}}) = \boldsymbol{U}$, $\sqrt{n_{\mathrm{t}}}(\boldsymbol{Y} - \boldsymbol{I}) = \boldsymbol{W}$ and $\sqrt{n_{\mathrm{t}}}(\boldsymbol{\Lambda} - \bar{\boldsymbol{\Lambda}}) = \boldsymbol{D}$. Then (10.8) may be written

$$\bar{\boldsymbol{\Lambda}} + \frac{1}{\sqrt{n_{\mathrm{t}}}}\boldsymbol{U} = \left(\boldsymbol{I} + \frac{1}{\sqrt{n_{\mathrm{t}}}}\boldsymbol{W}\right)\left(\bar{\boldsymbol{\Lambda}} + \frac{1}{\sqrt{n_{\mathrm{t}}}}\boldsymbol{D}\right)\left(\boldsymbol{I} + \frac{1}{\sqrt{n_{\mathrm{t}}}}\boldsymbol{W}\right)^* \tag{10.9}$$

or equivalently

$$\begin{aligned} \boldsymbol{U} = {} & \boldsymbol{W}\bar{\boldsymbol{\Lambda}} + \boldsymbol{D} + \bar{\boldsymbol{\Lambda}}\boldsymbol{W}^* \\ & + \frac{1}{\sqrt{n_{\mathrm{t}}}}\left(\boldsymbol{W}\boldsymbol{D} + \boldsymbol{W}\bar{\boldsymbol{\Lambda}}\boldsymbol{W}^* + \boldsymbol{D}\boldsymbol{W}^*\right) \\ & + \frac{1}{n_{\mathrm{t}}}\boldsymbol{W}\boldsymbol{D}\boldsymbol{W}^*. \end{aligned} \tag{10.10}$$

Also, since $\boldsymbol{I} = \boldsymbol{Y}\boldsymbol{Y}^* = (\boldsymbol{I} + \boldsymbol{W}/\sqrt{n_{\mathrm{t}}})(\boldsymbol{I} + \boldsymbol{W}/\sqrt{n_{\mathrm{t}}})^*$, we have

$$\boldsymbol{0} = \boldsymbol{W} + \boldsymbol{W}^* + \frac{1}{\sqrt{n_{\mathrm{t}}}}\boldsymbol{W}\boldsymbol{W}^*. \tag{10.11}$$

For convenience we adopt on the convention used in [SS97] and use "$\simeq$" to denote a first-order approximation, that is an equality where terms of order $1/\sqrt{n_{\mathrm{t}}}$ and $1/n_{\mathrm{t}}$ that do not influence the asymptotic statistics of λ_i have been neglected. Using this convention,

$$\boldsymbol{U} \simeq \boldsymbol{W}\bar{\boldsymbol{\Lambda}} + \boldsymbol{D} + \bar{\boldsymbol{\Lambda}}\boldsymbol{W}^*, \tag{10.12}$$

$$\boldsymbol{0} \simeq \boldsymbol{W} + \boldsymbol{W}^* \tag{10.13}$$

are obtained from (10.10) and (10.11). Substituting $\boldsymbol{W}^* \simeq -\boldsymbol{W}$ from (10.13) into (10.12) we have from the main diagonal

$$(\boldsymbol{D})_{ii} \simeq (\boldsymbol{U})_{ii} \tag{10.14}$$

or

$$\lambda_i - \bar{\lambda}_i \simeq \boldsymbol{v}_i^{\mathrm{r}*}\hat{\boldsymbol{R}}\boldsymbol{v}_i^{\mathrm{r}} - \bar{\lambda}_i \tag{10.15}$$

and hence

$$\begin{aligned}\mathrm{E}\{(\lambda_i-\bar{\lambda}_i)(\lambda_j-\bar{\lambda}_j)\} \simeq \mathrm{E}\{&\boldsymbol{v}_i^{\mathrm{r}*}\hat{\boldsymbol{R}}\boldsymbol{v}_i^{\mathrm{r}}\boldsymbol{v}_j^{\mathrm{r}*}\hat{\boldsymbol{R}}\boldsymbol{v}_j^{\mathrm{r}}+\bar{\lambda}_i\bar{\lambda}_j \\ &-\bar{\lambda}_j\boldsymbol{v}_i^{\mathrm{r}*}\hat{\boldsymbol{R}}\boldsymbol{v}_i^{\mathrm{r}}-\bar{\lambda}_i\boldsymbol{v}_j^{\mathrm{r}*}\hat{\boldsymbol{R}}\boldsymbol{v}_j^{\mathrm{r}}\}.\end{aligned} \tag{10.16}$$

In the above expression all terms except the first are easy to evaluate. Using the formula for the expectation of four complex Gaussian matrices [JS88], and the circularity assumption of the complex Gaussian variables, we have,

$$\begin{aligned}\mathrm{E}\{\hat{\boldsymbol{R}}\boldsymbol{v}_i^{\mathrm{r}}\boldsymbol{v}_j^{\mathrm{r}}\hat{\boldsymbol{R}}\} &= \sum_{k=1}^{n_{\mathrm{t}}}\sum_{l=1}^{n_{\mathrm{t}}}\mathrm{E}\{\boldsymbol{h}(k)\boldsymbol{h}^*(k)\boldsymbol{v}_i^{\mathrm{r}}\boldsymbol{v}_j^{\mathrm{r}*}\boldsymbol{h}(l)\boldsymbol{h}^*(l)\} \\ &= \sum_{k=1}^{n_{\mathrm{t}}}\sum_{l=1}^{n_{\mathrm{t}}}\big(\mathrm{E}\{\boldsymbol{h}(k)\boldsymbol{h}^*(k)\}\boldsymbol{v}_i^{\mathrm{r}}\boldsymbol{v}_j^{\mathrm{r}*}\,\mathrm{E}\{\boldsymbol{h}(l)\boldsymbol{h}^*(l)\} \\ &\quad+\mathrm{E}\{\boldsymbol{h}(k)\boldsymbol{h}^*(l)\}\boldsymbol{v}_j^{\mathrm{r}*}\,\mathrm{E}\{\boldsymbol{h}(l)\boldsymbol{h}^*(k)\}\boldsymbol{v}_i^{\mathrm{r}}\big)\,.\end{aligned} \tag{10.17}$$

Remember that $\boldsymbol{h}(k)$ denotes the k column of $\boldsymbol{H}\boldsymbol{Q}^{1/2}$. Notice from the structure of the spatial correlation on the channel (9.3), $\mathrm{E}\{\boldsymbol{h}(k)\boldsymbol{h}^*(l)\} = (\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}})_{lk}\boldsymbol{R}_{\mathrm{r}}$. Also, since $\boldsymbol{v}_i^{\mathrm{r}}$ is the ith eigenvector of $\boldsymbol{R}_{\mathrm{r}}$, $\boldsymbol{R}_{\mathrm{r}}\boldsymbol{v}_i^{\mathrm{r}} = \lambda_i^{\mathrm{r}}\boldsymbol{v}_i^{\mathrm{r}}$. Thus (10.17) may be rewritten,

$$\mathrm{E}\{\hat{\boldsymbol{R}}\boldsymbol{v}_i^{\mathrm{r}}\boldsymbol{v}_j^{\mathrm{r}*}\hat{\boldsymbol{R}}\} = \bar{\lambda}_i\bar{\lambda}_j\boldsymbol{v}_i^{\mathrm{r}}\boldsymbol{v}_j^{\mathrm{r}}+\sum_{k=1}^{n_{\mathrm{t}}}\sum_{l=1}^{n_{\mathrm{t}}}(\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}})_{kl}\boldsymbol{R}_{\mathrm{r}}\boldsymbol{v}_j^{\mathrm{r}*}(\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}})_{lk}\boldsymbol{R}_{\mathrm{r}}\boldsymbol{v}_i^{\mathrm{r}}. \tag{10.18}$$

Using the above result and using the fact that $\boldsymbol{R}_{\mathrm{r}}$, $\boldsymbol{R}_{\mathrm{t}}$ and $\boldsymbol{Q}$ are Hermitian (10.16) may be simplified as,

$$\begin{aligned}\mathrm{E}\{(\lambda_i-\bar{\lambda}_i)(\lambda_j-\bar{\lambda}_j)\} &\simeq \left|\boldsymbol{v}_i^{\mathrm{r}*}\boldsymbol{R}_{\mathrm{r}}\boldsymbol{v}_j^{\mathrm{r}}\right|^2\sum_{k=1}^{n_{\mathrm{t}}}\sum_{l=1}^{n_{\mathrm{t}}}\left|(\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}})_{kl}\right|^2 \\ &= \delta_{ij}(\lambda_i^{\mathrm{r}})^2\left\|\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\right\|_{\mathrm{F}}^2.\end{aligned} \tag{10.19}$$

The δ_{ij} denotes the Kronecker delta function, $\delta_{ij} = 1$ if $i = j$ and otherwise $\delta_{ij} = 0$. Note that $\bar{\lambda}_i = \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}$, $i = 1\ldots n_{\mathrm{r}}$.

10.A.3 Limiting Eigenvalue Distribution

We now justify the limiting distribution. First we show that the elements of $\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*$ have limiting complex normal distributions which then implies the limiting eigenvalue distribution. Notice that the channel matrix may

be generated as $\boldsymbol{H} = {\boldsymbol{R}_{\mathrm{r}}}^{1/2}\boldsymbol{G}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}$ where $\boldsymbol{G}$ consists of independent identically distributed, IID, $\mathcal{CN}(0,1)$ elements. Thus we may write

$$\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^* = {\boldsymbol{R}_{\mathrm{r}}}^{1/2}\boldsymbol{G}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}\boldsymbol{Q}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}\boldsymbol{G}^*{\boldsymbol{R}_{\mathrm{r}}}^{1/2}. \tag{10.20}$$

Note, in the case of uncorrelated transmit antennas, $\boldsymbol{R}_{\mathrm{t}} = \boldsymbol{I}$, and uncorrelated equi-powered transmitters, $\boldsymbol{Q} = \boldsymbol{I}P/n_{\mathrm{t}}$, the distribution of $\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*$ is identical to that considered in [Gup65]. In this case, the asymptotic normality follows directly from the central limit theorem. In the case where $\boldsymbol{R}_{\mathrm{t}} \neq \boldsymbol{I}$ and/or $\boldsymbol{Q} \neq \boldsymbol{I}P/n_{\mathrm{t}}$ deriving the asymptotic normality may be done as follows. From an eigenvalue decomposition of ${\boldsymbol{R}_{\mathrm{t}}}^{1/2}\boldsymbol{Q}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}$ we have that $\boldsymbol{G}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}\boldsymbol{Q}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}\boldsymbol{G}^*$ is distributed as $\boldsymbol{G}'\boldsymbol{\Lambda}_{\mathrm{tq}}\boldsymbol{G}'^*$ where $\boldsymbol{G}'$ consists of IID $\mathcal{CN}(0,1)$ elements. If we first consider an element on the main diagonal of $\boldsymbol{G}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}\boldsymbol{Q}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}\boldsymbol{G}^*$, it is distributed as

$$\begin{aligned}(\boldsymbol{G}'\boldsymbol{\Lambda}_{\mathrm{tq}}\boldsymbol{G}'^*)_{ii} &= \sum_{k=1}^{n_{\mathrm{t}}} (\boldsymbol{G}')_{ik}(\boldsymbol{G}')^*_{ik}\lambda_k^{\mathrm{tq}} \\ &= \sum_{k=1}^{n_{\mathrm{t}}} (g_k^2 + g_{k+n_{\mathrm{t}}}^2)\lambda_k^{\mathrm{tq}},\end{aligned} \tag{10.21}$$

where $g_1 \ldots g_{2n_{\mathrm{t}}}$ are IID $\mathcal{N}(0,1/2)$. For the terms in the sum, the first, second and third order moments are, $m_k = \mathrm{E}\{(g_k^2 + g_{k+n_{\mathrm{t}}}^2)\lambda_k^{\mathrm{tq}}\} = \lambda_k^{\mathrm{tq}}$, $\sigma_k^2 = \mathrm{Var}\{(g_k^2 + g_{k+n_{\mathrm{t}}}^2)\lambda_k^{\mathrm{tq}}\} = c_1(\lambda_k^{\mathrm{tq}})^2$, $\varrho_k^3 = \mathrm{E}\{\left|(g_k^2 + g_{k+n_{\mathrm{t}}}^2)\lambda_k^{\mathrm{tq}} - \lambda_k^{\mathrm{tq}}\right|^3\} = c_2(\lambda_k^{\mathrm{tq}})^3$, c_1 and c_2 are finite non-zero constants. Then (10.21) is a sum of independent terms, with, under the conditions of Theorem 1, finite third order moments, satisfying

$$\begin{aligned}\lim_{n_{\mathrm{t}}\to\infty} \frac{\left(\sum_{k=1}^{n_{\mathrm{t}}} \varrho_k^3\right)^{\frac{1}{3}}}{\left(\sum_{k=1}^{n_{\mathrm{t}}} \sigma_k^2\right)^{\frac{1}{2}}} &= \frac{c_2}{c_1} \lim_{n_{\mathrm{t}}\to\infty} \frac{\left(\sum_{k=1}^{n_{\mathrm{t}}} (\lambda_k^{\mathrm{tq}})^3\right)^{\frac{1}{3}}}{\left(\sum_{k=1}^{n_{\mathrm{t}}} (\lambda_k^{\mathrm{tq}})^2\right)^{\frac{1}{2}}} \\ &= 0.\end{aligned} \tag{10.22}$$

Hence, if the generalized central limit theorem due to Liapounoff[1], [Cra46], is applied,

$$\frac{(\boldsymbol{G}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}\boldsymbol{Q}{\boldsymbol{R}_{\mathrm{t}}}^{1/2}\boldsymbol{G}^*)_{ii} - \sum_{k=1}^{n_{\mathrm{t}}} \lambda_k^{\mathrm{tq}}}{\left(\sum_{k=1}^{n_{\mathrm{t}}} (\lambda_k^{\mathrm{tq}})^2\right)^{\frac{1}{2}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \tag{10.23}$$

[1]Note that this theorem only gives sufficient conditions on the terms of the sum. If needed, the Lindeberg-Feller Theorem [Fer96] provides necessary but more complicated conditions.

as $n_{\mathrm{t}} \to \infty$. In a similar fashion it can be shown that the off-diagonal elements, $(\boldsymbol{G}\boldsymbol{R}_{\mathrm{t}}{}^{1/2}\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}{}^{1/2}\boldsymbol{G}^{*})_{ij}$, $i \neq j$, converge to zero-mean complex Gaussian distributions as n_{t} increases. Hence, all elements of $\sqrt{n_{\mathrm{t}}}(\boldsymbol{G}\boldsymbol{R}_{\mathrm{t}}{}^{1/2}\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}{}^{1/2}\boldsymbol{G}^{*} - \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\boldsymbol{I})$ have limiting complex normal distributions. From a straightforward application of Slutsky's theorem [Fer96], the elements of $\sqrt{n_{\mathrm{t}}}\boldsymbol{R}_{\mathrm{r}}{}^{1/2}(\boldsymbol{G}\boldsymbol{R}_{\mathrm{t}}{}^{1/2}\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}{}^{1/2}\boldsymbol{G}^{*} - \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\boldsymbol{I})\boldsymbol{R}_{\mathrm{r}}{}^{1/2} = \sqrt{n_{\mathrm{t}}}(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H} - \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\boldsymbol{R}_{\mathrm{r}})$ converge in distribution to complex Gaussian distributions. Now, the same arguments as in [And84] may be applied to motivate the limiting eigenvalue distribution. Notice that the equations $\boldsymbol{T} = \boldsymbol{Y}\boldsymbol{\Lambda}\boldsymbol{Y}^{*}$ and $\boldsymbol{Y}\boldsymbol{Y}^{*} = \boldsymbol{I}$ and conditions $\mathrm{Re}(\boldsymbol{Y})_{ii} \geq 0$, $\mathrm{Im}(\boldsymbol{Y})_{ii} = 0, \forall\ i$, $\lambda_1^{\mathrm{r}} > \lambda_2^{\mathrm{r}} > \cdots > \lambda_{n_{\mathrm{r}}}^{\mathrm{r}}$ define a one to one mapping between $\boldsymbol{T}$ and $\boldsymbol{Y}$ and $\boldsymbol{\Lambda}$ except for a set of measure zero. The transformation from $\boldsymbol{T}$ to $\boldsymbol{Y}$ and $\boldsymbol{\Lambda}$ is continuously differentiable and from Cramérs Theorem [Fer96] the elements of $\boldsymbol{\Lambda}$ have a limiting normal distribution.

Chapter 11

Asymptotic Mutual Information and Outage Capacity

Based on the channel eigenvalue distribution derived in the previous chapter techniques for analyzing fundamental system limitations set by information theory are derived. In particular we are interested in the mutual information and the outage capacity. Under identical asymptotic assumptions as in the previous chapter, simple expressions for these properties are derived and their accuracy as approximations in finite dimensional systems investigated.

Section 11.1 defines the properties of interest. In Section 11.2 and Section 11.3 approximations asymptotic in the number of transmit antennas are derived. Similar results are straightforward to compute under a large receiver array assumption. Finally, in Section 11.4 we investigate the accuracy of our expressions when used as approximations in finite dimensional systems.

11.1 MIMO Channel Capacity

Consider a communication system as specified above where the channel is known at the receiver, i.e. a channel with input $\boldsymbol{s}$ and output $(\boldsymbol{x}, \boldsymbol{H})$. For an information theoretic analysis of such a system we are interested

in the statistical properties of $J = J(\boldsymbol{HQH}^*)$, where we have defined the function $J(\boldsymbol{Z})$ as

$$J(\boldsymbol{Z}) = \log\det\{\boldsymbol{I} + \boldsymbol{Z}\}, \tag{11.1}$$

where $\boldsymbol{Z} = \boldsymbol{Z}^*$ and positive semi-definite. Note that since $\boldsymbol{H}$ is a random matrix $J(\boldsymbol{HQH}^*)$ is a random variable. Furthermore, if $\boldsymbol{\zeta} = [\zeta_1 \dots \zeta_n]^{\mathrm{T}}$ are used to denote the eigenvalues of $\boldsymbol{Z}$, from an eigendecomposition of $\boldsymbol{Z}$ it follows,

$$\begin{aligned} J(\boldsymbol{Z}) &= \sum_{i=1}^{n} \log(1+\zeta_i) \\ &\equiv J(\boldsymbol{\zeta}). \end{aligned} \tag{11.2}$$

Similarly the statistical properties of $J(\boldsymbol{HQH}^*)$ are determined by the eigenvalues of $\boldsymbol{HQH}^*$, $\boldsymbol{\lambda} = [\lambda_1 \dots \lambda_{n_{\mathrm{r}}}]$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n_{\mathrm{r}}}$.

Based on the characteristics of $J(\boldsymbol{HQH}^*)$, several interesting properties of the MIMO communication system may be analyzed. First, consider a fast fading scenario where for each use of the channel an independent realization of $\boldsymbol{H}$ is drawn. The average mutual information between the inputs and outputs of the channel is given by [Tel99]

$$\begin{aligned} I &= I(\boldsymbol{s}; (\boldsymbol{x}, \boldsymbol{H})) \\ &= \mathrm{E}\{J(\boldsymbol{HQH}^*)\} \\ &= \mathrm{E}\left\{\sum_{i=1}^{n_{\mathrm{r}}} \log(1+\lambda_i)\right\}. \end{aligned} \tag{11.3}$$

With the assumption of a memoryless channel the Shannon channel capacity may be found as [GJJV03],

$$C = \max_{\boldsymbol{Q}} I. \tag{11.4}$$

Second, consider a quasi-static scenario where a single $\boldsymbol{H}$ is used for the duration of one code-word, i.e. $\boldsymbol{H}$ is drawn only once. In this case the mutual information is given directly by $J(\boldsymbol{HQH}^*)$. The x percent outage capacity, $C_{x/100}$, may be defined as [Fos96, FG98],

$$\frac{x}{100} = \Pr(J(\boldsymbol{HQH}^*) < C_{x/100}). \tag{11.5}$$

To gain insights on the impact of spatial correlation on I and $C_{x/100}$, the behavior of the statistical properties of $J(\boldsymbol{HQH}^*)$ must be understood.

While difficult to compute exactly, in the next section a distribution limiting as the number transmit antennas increases is derived. This is done by using the observation that the statistical properties of $J(\boldsymbol{HQH}^*)$ are determined by the channel eigenvalues and using asymptotic eigenvalue results derived in Chapter 10.

11.2 A Limiting Distribution of $J(\boldsymbol{HQH}^*)$

If the conditions of Theorem 1 are satisfied, we know that $\sqrt{n_{\mathrm{t}}}(\boldsymbol{\lambda} - \mathrm{Tr}\{\boldsymbol{QR}_{\mathrm{t}}\}\boldsymbol{\lambda}^{\mathrm{r}})$ converges in distribution to $\mathcal{N}(\mathbf{0}, \boldsymbol{R}_{\tilde{\lambda}})$ as $n_{\mathrm{t}} \to \infty$. Also, from (11.2) we notice that $J(\boldsymbol{\lambda})$ is continuously differentiable in a neighborhood of $\mathrm{Tr}\{\boldsymbol{QR}_{\mathrm{t}}\}\boldsymbol{\lambda}^{\mathrm{r}}$. Hence, from Cramér's Theorem [Fer96], as $n_{\mathrm{t}} \to \infty$,

$$\sqrt{n_{\mathrm{t}}}(J(\boldsymbol{\lambda}) - J(\mathrm{Tr}\{\boldsymbol{QR}_{\mathrm{t}}\}\boldsymbol{\lambda}^{\mathrm{r}})) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sum_{i=1}^{n_{\mathrm{r}}} (\boldsymbol{R}_{\tilde{\lambda}})_{ii} \left(\left.\frac{\partial J(\boldsymbol{\zeta})}{\partial \zeta_i}\right|_{\zeta_i = \mathrm{Tr}\{\boldsymbol{QR}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}} \right)^2 \right), \quad (11.6)$$

where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution. In the expression of the variance in (11.6), the diagonal structure of $\boldsymbol{R}_{\tilde{\lambda}}$ has been used. Using the results of Section 10.2, the variance is straightforward to compute. This provides a limiting distribution of $J(\boldsymbol{\lambda})$ which here is summarized in the following corollary.

Corollary 2 *Assume that the conditions of Theorem 1 are satisfied, then as* $n_t \to \infty$,

$$\sqrt{n_t}(J(\boldsymbol{\lambda}) - J(\mathrm{Tr}\{\boldsymbol{QR}_t\}\boldsymbol{\lambda}^{r})) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, n_t \|\boldsymbol{QR}_t\|_F^2 \sum_{i=1}^{n_r} \left(\frac{\lambda_i^r}{1 + \mathrm{Tr}\{\boldsymbol{QR}_t\}\lambda_i^r} \right)^2 \right). \quad (11.7)$$

Note, Corollary 2 does not state how fast the distribution of $J(\boldsymbol{\lambda})$ converges to its limiting distribution, only that it does so as n_{t} approaches infinity. However, it has been observed, both from realistic channel simulations [SS00] and from channel measurements [SKO00] that this property of the MIMO channel seems to approach a Gaussian distribution also at realistic system dimensions. In [WV01] it is also shown that $J(\boldsymbol{\lambda})$ converges to a normal distribution in the case of uncorrelated columns in the

channel matrix. This is also implied by the results in [MSS03]. Combined with the result above, using a normal distribution as an approximation seems promising also in the correlated case. This idea will be further explored in the following sections.

11.3 Improving the Approximation

If Corollary 2 is used as an approximation, we have an expression that from the previous section is known to have the attractive feature of approaching the true distribution as n_{t} increases. However, while Corollary 2 might be used as an approximation directly, convergence of the mean, the mutual information of the system, is unnecessarily slow. Normally, the derivation of Cramér's theorem is based on a first order Taylor expansion of the function of interest. Here, to improve on the speed of convergence of the mutual information, a second order Taylor expansion of $J(\boldsymbol{\lambda})$ around $\mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\boldsymbol{\lambda}^{\mathrm{r}}$ is considered, see [Fer96] or Appendix C of [Por94],

$$\begin{aligned} J(\boldsymbol{\lambda}) = \sum_{i=1}^{n_{\mathrm{r}}} \Bigg(& \log(1+\mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}) \\ & + \frac{1}{1+\mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}}(\lambda_i - \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}) \\ & - \frac{1}{2}\left(\frac{1}{1+\mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}}\right)^2 (\lambda_i - \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}})^2 \\ & + \mathcal{O}(\lambda_i - \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}})^3 \Bigg). \end{aligned} \tag{11.8}$$

By applying the asymptotic first and second order eigenvalue moments derived in Section 10.2, an approximation of the average mutual information is found as,

$$\begin{aligned} I &\approx \tilde{I} \\ &= \sum_{i=1}^{n_{\mathrm{r}}} \left(\log(1+\mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}) - \frac{1}{2}\left(\frac{\|\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\|_{\mathrm{F}}\,\lambda_i^{\mathrm{r}}}{1+\mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}}\right)^2 \right). \end{aligned} \tag{11.9}$$

Notice that the approximation is still asymptotic as n_{t} increases. Maximization of the above expression with respect to $\boldsymbol{Q}$ should provide the channel capacity of the system for large n_{t} as long as the maximizing

$\boldsymbol{Q}$ satisfies the conditions of Theorem 1, see (11.4). This optimization problem will be addressed in Chapter 12.

In summary, to approximate the distribution of J, a normal distribution with a mean given by (11.9) and a variance, from (11.7),

$$\mathrm{Var}\{J\} \approx \|\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\|_{\mathrm{F}}^2 \sum_{i=1}^{n_{\mathrm{r}}} \left(\frac{\lambda_i^{\mathrm{r}}}{1 + \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}} \right)^2 \tag{11.10}$$

is proposed. From the above, this approximation should improve as n_{t} increases. This will be further investigated in Section 11.4 where the result above is compared against simulations and measurements.

Note that from the concave nature of the log det function and Jensen's inequality [CT91], the average mutual information of the MIMO channel may be upper bounded as, [Loy01],

$$\begin{aligned} I &= \mathrm{E}\left\{\log\det\{\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*\}\right\} \\ &\leq \log\det\{\boldsymbol{I} + \mathrm{E}\{\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*\}\} \\ &= \sum_{i=1}^{n_{\mathrm{r}}} \log(1 + \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}). \end{aligned} \tag{11.11}$$

This upper bound coincides with the first term of (11.9). For $n_{\mathrm{t}} >> n_{\mathrm{r}}$, and $\boldsymbol{Q} = \boldsymbol{I}P/n_{\mathrm{t}}$, the upper bound on the average mutual information derived in [SFGK00],

$$I \leq \sum_{l=1}^{n_{\mathrm{r}}} \log\left(1 + \frac{P}{n_{\mathrm{t}}}\lambda_i^{\mathrm{r}}(n_{\mathrm{t}} + n_{\mathrm{r}} - 2l + 1)\right), \tag{11.12}$$

also coincides with this term.

11.4 Verification Against Simulations and Measurements

In this section we explore how well the asymptotic approximation of the distribution of $J(\boldsymbol{\lambda})$ derived in the above sections performs for finite transmit array sizes.

In all the numerical examples studied below the transmitter covariance matrix is a scaled identity matrix, i.e. $\boldsymbol{Q} = \boldsymbol{I}P/n_{\mathrm{t}}$. This can either be seen as a constraint of the channel in which case the average mutual information coincide with the channel capacity, or that we study the average mutual information given the $\boldsymbol{Q}$ above.

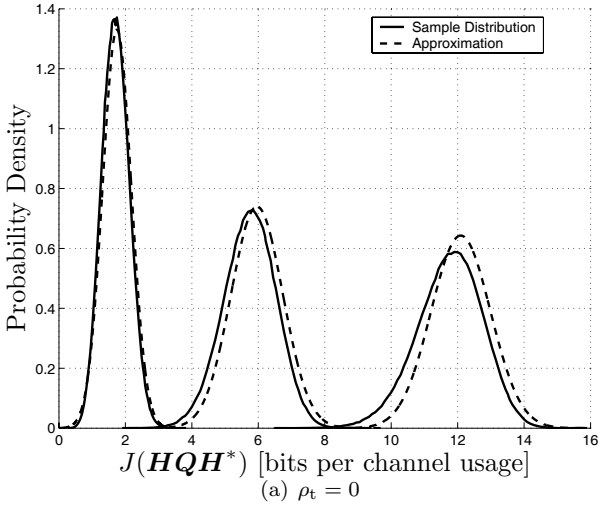
To produce transmitter and receiver covariance matrices two methods were used. Either the covariance matrices were generated from the exponential correlation model, (9.4), or, to generate more realistic channel matrices, covariance matrices estimated from MIMO channel measurements performed at the University of Bristol were used.

In these measurements, a uniform linear array with eight directional elements were used at the receiver while a uniform linear array with eight half-wavelength spaced omni-directional elements was employed at the transmitter for the indoor measurements which were performed at 5.2 GHz. The measurements are described in [MBF00]. Details regarding the estimation of the covariance matrices as well as the estimated covariance matrices themselves may be found in [YBO$^+$01, BYO01].

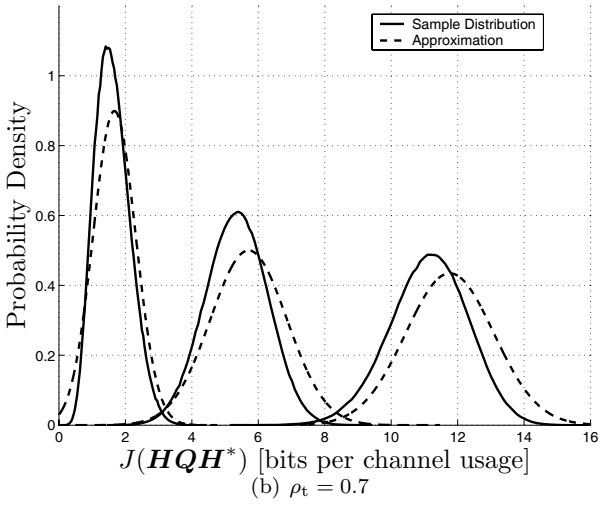
To verify the Gaussian distribution as a reasonable approximation also at realistic array sizes we compared the asymptotic distribution with sample distributions from simulations. Figure 11.1 and Figure 11.2 compare sample distributions from simulations of $J(\boldsymbol{\lambda})$ with the approximation of Section 11.3. This comparison was performed at 0, 10, 20 dB SNR with 5 transmit and 2 receive antennas. In Figure 11.1 the covariance matrices, $\boldsymbol{R}_{\mathrm{t}}$ and $\boldsymbol{R}_{\mathrm{r}}$, were generated using the exponential correlation model (9.4). In Figure 11.1(a) with $\rho_{\mathrm{t}} = 0$, i.e. without correlation at the transmit site, and in Figure 11.1(b) with $\rho_{\mathrm{r}} = 0.7$, i.e. with correlation at the transmit site. In Figure 11.2, the covariance matrices were obtained from measurements [BYO01], in this case from receiver location 3 and transmitter location 13.

From Figure 11.1(a) we observe that, without correlation at the transmit side, the Gaussian approximation appears to work well at all the simulated signal to noise ratios, even with only five transmit antennas. Note that the approximation performs better at small signal to noise ratios. When correlation is present also at the transmitter, more transmit antennas are required for the approximation to work well, see Figure 11.1(b). Still, a reasonable approximation is provided in this five transmit antenna example.

The performance of the approximation when measured covariances are used to generate the channel matrices is shown in Figure 11.2. The approximated and the simulated probability densities are also compared with sample distributions computed directly from the measured MIMO channel matrices. Since omni-directional antennas are employed at the transmitter, the transmit elements excite more scatterers than the directional elements at the receiver. The correlation between the transmit elements is therefore much lower than the correlation between the receive



(a) $\rho_{\mathrm{t}} = 0$



(b) $\rho_{\mathrm{t}} = 0.7$

Figure 11.1: Simulated and approximated probability densities of J at SNR = 0, 10, 20 dB. The system employed $n_{\mathrm{t}} = 5$ transmit antennas, $n_{\mathrm{r}} = 2$ receive antennas and the channel correlation matrices where generated using the exponential correlation model (9.4). The receive side correlation was $\rho_{\mathrm{r}} = 0.7$.

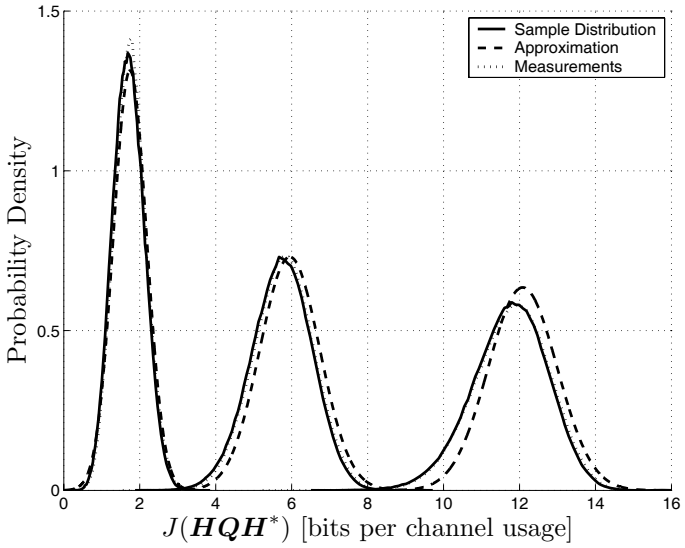


Figure 11.2: Simulated and approximated probability densities of J from measured $\boldsymbol{R}_{\mathrm{t}}$ and $\boldsymbol{R}_{\mathrm{r}}$, sample distributions of J from measured channel matrices. $n_{\mathrm{t}} = 5$, $n_{\mathrm{r}} = 2$ and SNR = 0, 10, 20 dB.

elements. As could have been expected from the previous two figures the agreement between the approximated probability densities and the sample densities from simulations is therefore good. In this case, the agreement between the measured and modeled probability densities is also excellent and the results are even hard to distinguish.

Figure 11.3 and 11.4 verify that the approximation improves as the number of transmit antennas increases. In these figures we study how two parameters of the capacity distribution evolve as the number of transmit antennas is increased. The studied parameters are the average mutual information, I, and the 5% outage capacity, $C_{0.05}$, see (11.5).

In the first of these two examples, Figure 11.3, the convergence of a $n_{\mathrm{r}} = 2$ system is studied in a low and a high SNR scenario, 0 and 20 dB. The covariance matrices were chosen according to the exponential convergence model (9.4), with $\rho_{\mathrm{t}} = \rho_{\mathrm{r}} = 0.7$. Note that the low SNR case converges faster than the high SNR case.

In the second example, Figure 11.4, the convergence of the approximation for an $n_{\mathrm{r}} = 2$ receiver is compared with the convergence of an

$n_\mathrm{r} = 4$ receiver. The systems are studied in a 10 dB scenario, this time with $\rho_\mathrm{r} = .7$ and $\rho_\mathrm{t} = 0$. As could be expected the properties of the system employing the larger receiver require more transmit antennas to converge.

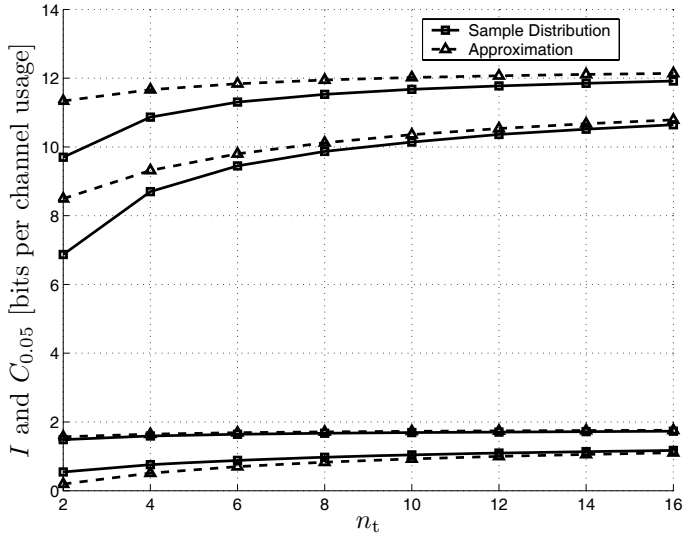


Figure 11.3: Convergence of the average mutual information and 5% outage capacity. $n_\mathrm{r} = 2$, $\rho_\mathrm{r} = 0.7$ $\rho_\mathrm{t} = 0.7$ and SNR = 0, 20 dB. The top four curves correspond to the high SNR case.

To investigate the sensitivity of the approximation to correlation, it is compared with the average mutual information and 5% outage capacities estimated from simulations as the correlation factors on both sides of the MIMO system ($\rho_\mathrm{r} = \rho_\mathrm{t}$) was varied from 0 to 0.99. Again two (high and low SNR 20, 0 dB) scenarios were studied and again five antennas were used at the transmitter and two at the receiver.

As can be expected from our previous experiments, the approximation in the low SNR scenario shows the best agreement. However, the agreement in the high SNR case is also reasonable and in both cases the qualitative behavior of the approximation matches the simulated results well.

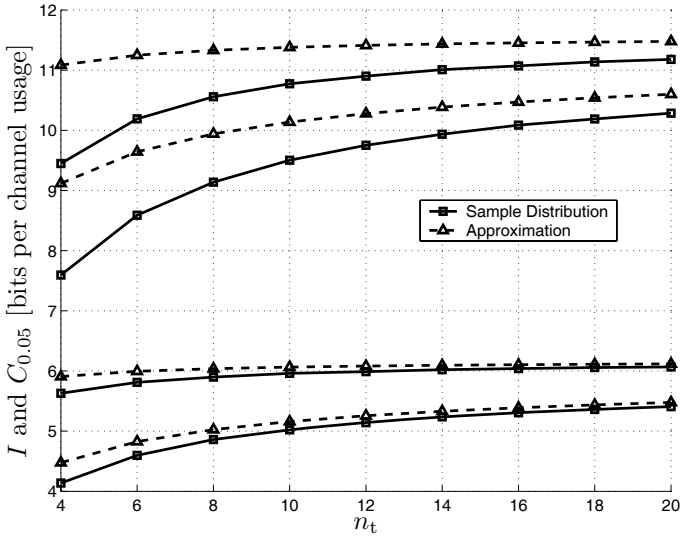


Figure 11.4: Convergence of the average mutual information and 5% outage capacity. $\rho_r = 0.7$ $\rho_t = 0$ and SNR = 10 dB. The top four curves correspond to an $n_r = 4$ case and the bottom four to an $n_r = 2$ case.

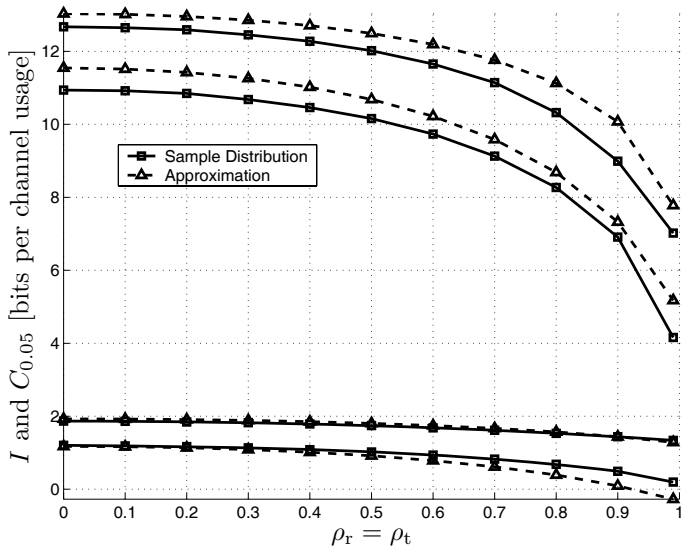


Figure 11.5: Average mutual information and 5% outage capacity. System configurations with $n_r = 2$ receive and $n_t = 5$ transmit antennas and SNR = 0, 20 dB are studied. The top four curves correspond to the high SNR case.

Chapter 12

Approximate Transmit Covariance Optimization

In this chapter we provide an example on how the approximations derived in Chapter 10 and Chapter 11 can be used in the design of MIMO systems where the transmitter has access to covariance feedback. Here an approximation of the transmit covariance that maximizes the average mutual information is considered, i.e. we attempt to solve (11.4) using the approximation of the mutual information (11.9). Similar approaches could be taken to approximately solve other optimization problems, for example to maximize some level of outage capacity.

The remainder of the chapter is organized as follows. Section 12.1 specifies the optimization problem considered in this chapter. In Section 12.2 an attempt to solve this problem is presented. The performance of our solution is illustrated with numerical examples in Section 12.3. Finally, in Section 12.4 we consider an ad-hoc improvement of our procedure that appears to produce more accurate numerical results at finite system dimensions.

12.1 Goal

The goal of the work presented herein is to approximately find the transmit covariance $\boldsymbol{Q}$ that solves (11.4) by optimizing $\tilde{I}$, the approximation of the average mutual information form presented in (11.9). That is, we

attempt to solve,

$$\begin{aligned} \boldsymbol{Q} = \arg\max_{\boldsymbol{Q}} \quad & \tilde{I} \\ \text{s.t.} \quad & \operatorname{Tr}\boldsymbol{Q} \leq P \\ & \boldsymbol{Q} \succeq \mathbf{0}, \end{aligned} \tag{12.1}$$

with the hope of finding a high performing system design with a reasonable computational burden. Note that (12.1) differs from the traditional water-filling problem [Tel99] as the transmitter only has access to the channel statistics and $\tilde{I}$ is an approximation of the true average mutual information.

This chapter is organized as follows, in Section 12.2 a technique for solving (12.1) is presented. Section 12.3 shows some numerical examples while Section 12.4 discusses what improvement can be expected if the mutual information approximation is improved.

12.2 Optimizing $\tilde{I}$

We attempt to find the $\boldsymbol{Q}$ that maximize $\tilde{I}$ using a two step procedure. First, the eigenvectors of the maximizing $\boldsymbol{Q}$ are derived and the problem is reduced to the optimization of the eigenvalues of $\boldsymbol{Q}$, q_i , $i = 1 \dots n_{\mathrm{t}}$. Secondly, we assume that the problem (12.1), when expressed in the eigenvalues of $\boldsymbol{Q}$, is convex in some volume around the maximum and that properly initialized standard convex optimization tools can be used to find the solution.

Theorem 3 *Let* $\boldsymbol{\lambda}^t = \left[\lambda_1^t \dots \lambda_{n_t}^t\right]$ *and* $\boldsymbol{q} = \left[q_1 \dots q_{n_t}\right]$ *denote the ordered eigenvalues of* $\boldsymbol{R}_t$ *and* $\boldsymbol{Q}$ *respectively. Then, the* $\boldsymbol{Q}$ *that solves (12.1) has identical eigenvectors as* $\boldsymbol{R}_t$ *and the considered problem may be reduced to only consider the eigenvalues of* $\boldsymbol{Q}$*,*

$$\begin{aligned} \boldsymbol{q} = \arg\max_{\boldsymbol{q}} \quad & \sum_i \log(1 + \lambda_i^r \textstyle\sum_j \lambda_j^t q_j) - \frac{1}{2} \frac{\lambda_i^{r2} \sum_j (\lambda_j^t q_j)^2}{(1+\lambda_i^r \sum_j \lambda_j^t q_j)^2} \\ s.t. \quad & \textstyle\sum_i q_i \leq P \\ & q_i \geq 0 \quad i = 1 \dots n_t, \end{aligned} \tag{12.2}$$

Proof: It is well known $\operatorname{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}$ and $\|\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\|_{\mathrm{F}}$ do not depend on the entire structure of $\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}$ but are only functions of the eigenvalues of $\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}$ and thus so is $\tilde{I}$. From (11.9) it is also clear that $\tilde{I}$ is a monotonic increasing function of transmit power, i.e. $\tilde{I}(\boldsymbol{Q}) < \tilde{I}(r\boldsymbol{Q})$ for any $r > 1$.

The proof presented in [SM03, Appendix B] can then be applied to prove that the $\boldsymbol{Q}$ that maximize $\tilde{I}$ has the same eigenvectors as $\boldsymbol{R}_{\mathrm{t}}$.

Rewriting (12.1) on a form that only depends on the eigenvalues of $\boldsymbol{Q}$ is straightforward using the well known results that the trace and the squared Frobenius norm of a matrix are the sum of the eigenvalues and the sum of the squared eigenvalues of that matrix as well as all eigenvalues of non-negative definite matrices are positive. ■

To solve (12.2) we assume that the problem is convex is some volume surrounding the solution so that a properly initialized standard convex optimization tool can find the optimum. While we have yet to show the convexity of (12.2) numerical experiments indicate that this approach does find the proper solution. In the numerical examples below $\boldsymbol{q}$ is initialized as a scaled version of $\boldsymbol{\lambda}^{\mathrm{t}}$.

The expression we are optimizing over, $\tilde{I}(\boldsymbol{q})$, is on a closed form, making the optimization efficient. Note that the structure of $\boldsymbol{Q}$ derived in Theorem 3 coincides with the structures derived from exact expressions of the capacity, see e.g. [JVG01, SM03, JB04]. This structure also means that $\lambda_i^{\mathrm{tq}} = \lambda_i^{\mathrm{t}} q_i$, $i = 1 \dots n_{\mathrm{t}}$.

12.3 Numerical Results

In this section some numerical results are presented to verify the performance of this approximate approach. To keep things simple, the following conventions are used in the examples.

- To model the spatial channel correlation $\boldsymbol{R}_{\mathrm{t}}$ and $\boldsymbol{R}_{\mathrm{r}}$ are exponentially correlated according to (9.4).
- The signal to noise ratio in all examples is 10 dB.
- To illustrate the potential gain of the "optimized scheme" in Section 12.2 it is compared with two reference schemes. The first reference, denoted "pure diversity" in the figures, chooses $\boldsymbol{Q} = \boldsymbol{I}P/n_{\mathrm{t}}$. This choice is suitable when the transmitter lacks channel information and can be expected to work well when the spatial correlation is small. The second reference scheme, "beamforming", maximizes the received signal to noise ratio. This is done by choosing $\boldsymbol{Q}$ so that all transmit power is concentrated on the eigenvector that corresponds to the largest eigenvalue of $\boldsymbol{R}_{\mathrm{t}}$. Such a scheme is optimal when the MIMO channel is severely correlated [JG04, JB04].

- In all figures both the approximate mutual information and the mutual information computed using Monte-Carlo methods is shown. This provides insight to the performance of the approximation and might hint how far our technique is from capacity. Results from Monte Carlo simulations are displayed with solid lines, results using approximations are displayed with dashed lines.

In the first example, Figure 12.1, the performance of the approximation and the approximative optimization scheme is studied as the number of transmit antennas increases. The studied system employs two receive antennas and correlation is present at both receiver and transmitter, $\rho_\mathrm{r} = \rho_\mathrm{t} = 0.5$. As could be expected, the results are improved as the number of transmit antennas is increased. When n_t is small, the beamforming solution is selected by our algorithm. Clearly this is not a good choice as it is outperformed by the pure diversity scheme in the simulated reality. However at larger n_t the optimized scheme is able to achieve higher average mutual information than both the other schemes.

The correlation between the antenna elements clearly affects the performance of the MIMO communication system. Earlier work indicates that the pure diversity scheme can be expected to performance well when there is little correlation among the elements of $\boldsymbol{H}$ while the beamforming solution is optimal in highly correlated scenarios [JG04]. Figure 12.2 provides an example where our algorithm combines the best of these two worlds. In this example a system consisting of eight transmitters and two receivers is considered and the average mutual information is studied at various correlation factors $\rho_\mathrm{r} = \rho_\mathrm{t}$. This example illustrates how the approximation can combine the best of both worlds, a system designed with the scheme above can be expected to provide data rates as high or higher than the pure diversity or the beamforming approach for different correlations. At some correlation factors, around $\rho_\mathrm{r} = \rho_\mathrm{t} = 0.5$ our scheme even performs considerably better. Note that in contrast to if knowledge of the channel covariance is ignored ($\boldsymbol{Q} = \boldsymbol{I}$), correlation between the antenna elements may actually improve performance of the system. This effect has also been noticed for systems where the transmitter has access to perfect channel estimates [CTKV02] and for systems with covariance feedback [IUN03].

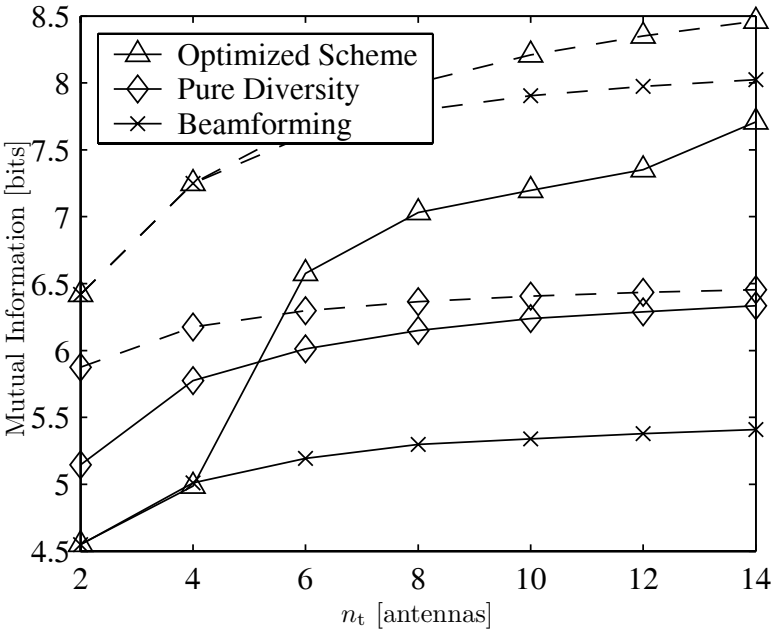


Figure 12.1: Performance of our approximatively optimized scheme. The SNR is 10 dB, $\rho_r = \rho_t = 0.5$, $n_r = 2$, and the solid and dashed curves correspond to the average mutual information computed with Monte Carlo methods and the approximation (11.9) respectively.

12.4 Ad-hoc Improvement

The above results indicate that significant gains are achievable using the method described in Section 12.2. However, the disparity between the simulated reality and the approximation is quite large, especially in cases where the transmitter only transmits on a few eigenvectors, which in Figure 12.2 is illustrated for $\rho_r = \rho_t > 0.6$ where the algorithm chooses the beamforming solution, and the performance is reduced. This is to be expected from the conditions under which (11.9) was derived under. For example these conditions are clearly violated in the case of beamforming, see Theorem 1. Due to poor performance of the approximation in such scenarios the approximative solution (12.1) is far from the solution of the

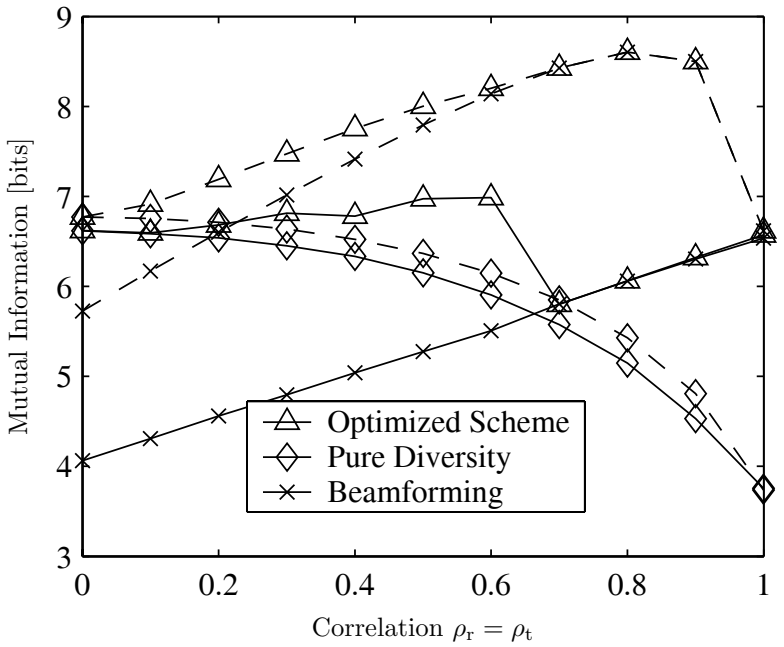


Figure 12.2: Performance of a $n_t = 8$, $n_r = 2$ system operating at a signal to noise ratio of 10 dB. The solid and dashed curves correspond to the average mutual information computed with Monte Carlo methods and the approximation (11.9) respectively.

original problem (11.4). This can be illustrated by for example comparing the optimized scheme from Figure 12.2 to a system where the transmit covariance is selected as $\boldsymbol{Q} = \boldsymbol{R}_t P/n_t$. This simple design can in the considered scenario be shown to outperform our optimized scheme in a range of correlation factors.

If the accuracy of the approximation (11.9) could be increased, the results achieved based on optimization of this approximation could be expected to see some improvements. For this particular system setup we have noticed that the accuracy of the approximation of the mutual information can be improved if the factor before the second order term of the Taylor series resulting in our approximation is increased. Reasonable results seem to be achieved if this term is amplified by a factor n_r. This

would then result in an approximation of the average mutual information (in nats),

$$\begin{aligned} I &\approx \bar{I} \\ &= \sum_{i=1}^{n_{\mathrm{r}}} \log(1 + \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}) - \frac{n_{\mathrm{r}}}{2}\left(\frac{\|\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\|_{\mathrm{F}}\,\lambda_i^{\mathrm{r}}}{1 + \mathrm{Tr}\{\boldsymbol{Q}\boldsymbol{R}_{\mathrm{t}}\}\lambda_i^{\mathrm{r}}}\right)^2. \end{aligned} \tag{12.3}$$

Since n_{r} is finite and the first order terms of the two approximations are identical, the two approximations will converge as the number of transmit antennas is large. The behavior when the number of antennas is more reasonable however is often improved from this ad-hoc procedure. Below, results achieved using the ad-hoc improvement above to optimize the MIMO antenna system are presented and discussed.

The improvement in approximation accuracy and the resulting system design are first illustrated in Figure 12.3. Here the performance of an approximatively optimized system is studied as the number of transmit antennas increases. The receiver employs two antennas and the correlation factors are chosen as $\rho_{\mathrm{r}} = \rho_{\mathrm{t}} = 0.5$, i.e. the results can be directly compared to those presented in Figure 12.1. Clearly using (12.3) instead of (11.9) in the optimization procedure produces a higher performing configuration in this scenario.

To provide further understanding to what kind of performance could be achievable using a system optimized using the proposed algorithms a few examples are presented in Figures 12.4–12.6 where the effect of spatial correlation is studied again.

In the first of these examples, presented in Figure 12.4, an equivalent scenario to that in Figure 12.2 is studied. This scenario is relatively easy, as eight transmit but only two receive antennas are employed, but by exploiting the ad-hoc improvement significant improvements both in terms of the quality of the approximation and in terms of the quality of the optimized system are still achieved. This example also illustrates that an optimal scheme may provide considerable gains in a wide range of spatial correlations as compared of resorting to one of the simpler schemes.

More difficult, and perhaps more realistic, scenarios are considered in the remaining two figures, Figure 12.5 and Figure 12.6. Here the difference between the number of antennas at the access point and at the terminal is smaller, in Figure 12.5 four transmitters and two receivers are employed, in Figure 12.6 eight transmitters and four receivers. In these scenarios, the performance of the approximation and the resulting

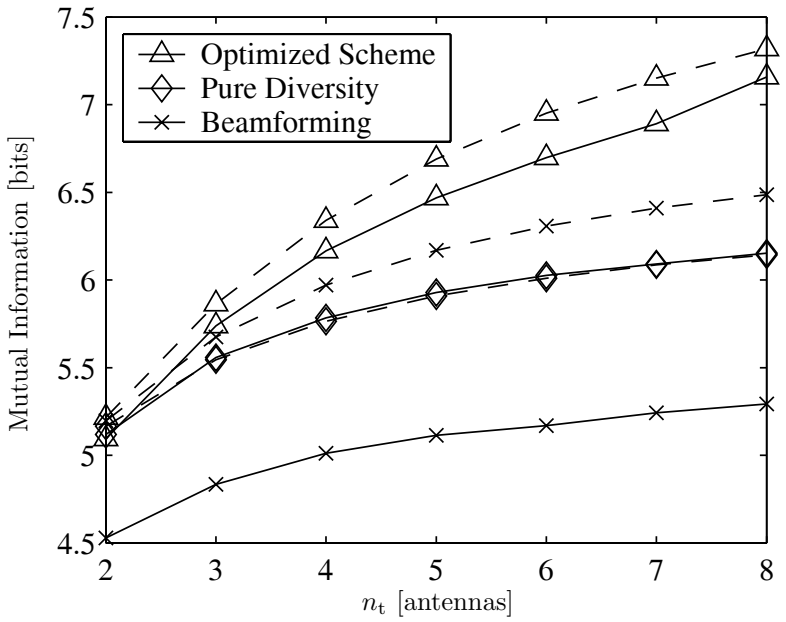


Figure 12.3: Performance of our approximatively optimized scheme using the ad-hoc method from Section 12.4 to estimate I. The SNR is 10 dB, $\rho_r = \rho_t = 0.5$, $n_r = 2$ and the solid and dashed curves correspond to the average mutual information computed with Monte Carlo methods and the approximation (12.3) respectively.

optimized system can be expected to decrease. Even so, the design resulting from the algorithm described in this section is still, in all points of simulation, able to perform at least as good, and in most cases better, than the two reference designs.

Finally, in an attempt to illustrate some of the inner workings of the algorithm, the resulting eigenvalues of $\boldsymbol{Q}$ for the scenario considered for Figure 12.4 are studied. These eigenvalues correspond to the transmit power along the eigenvectors of $\boldsymbol{R}_t$. The ordered eigenvalues of $\boldsymbol{Q}$ are presented in Table 12.1 and for convenience the corresponding eigenvalues of $\boldsymbol{R}_t$ are also included. Note that the more correlated the elements of the channel matrix become the more the larger eigenvalues of $\boldsymbol{R}_t$ increase and the more transmit power is allocated to the eigenvector corresponding

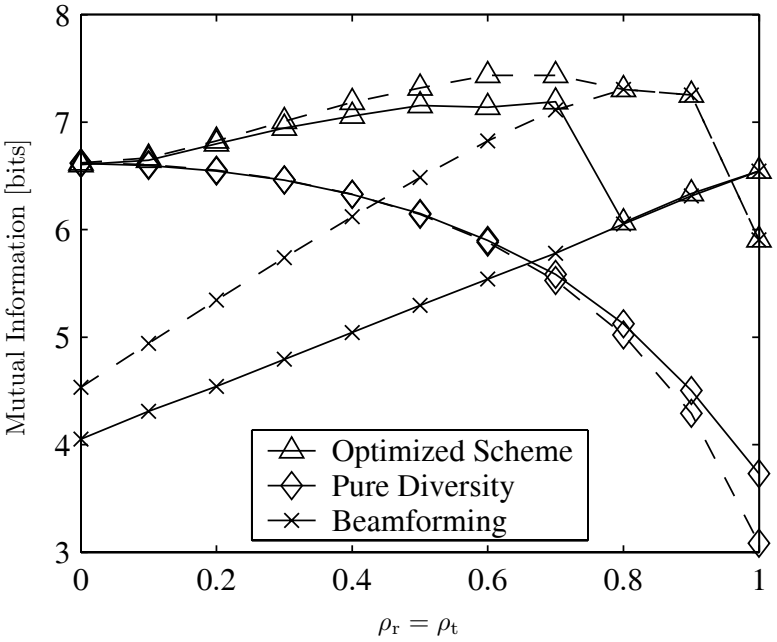


Figure 12.4: Performance of the approximately optimized system when (12.3) is used. The SNR is 10 dB, $n_r = 2$, and $n_t = 8$. Simulated and approximative results are illustrated with solid and dashed curves respectively.

the larger eigenvectors of $\boldsymbol{R}_t$. Just like in the water-filling algorithm used when the MIMO channel is perfectly known [Tel99] this results in no power allocation for certain eigenvectors.

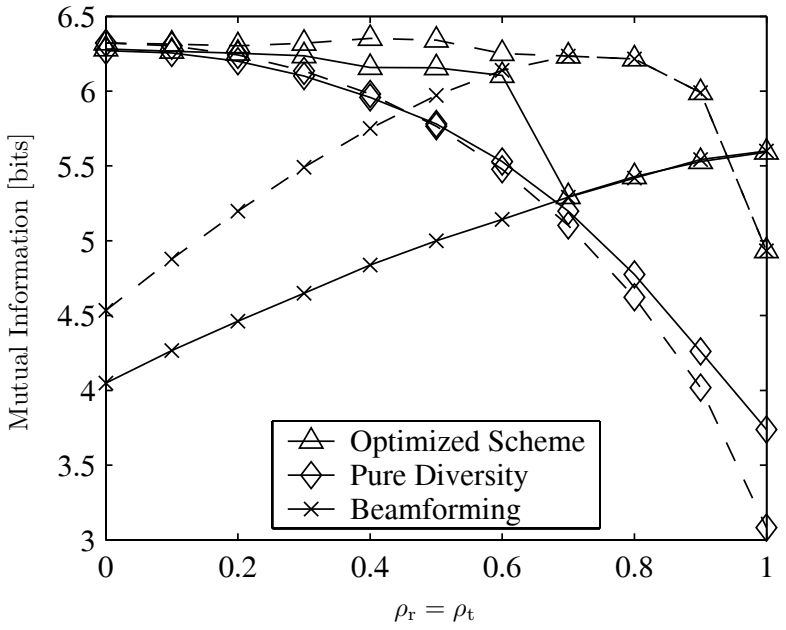


Figure 12.5: Performance of the approximately optimized system when (12.3) is used. The SNR is 10 dB, $n_r = 2$, and $n_t = 4$. Simulated and approximative results are illustrated with solid and dashed curves respectively.

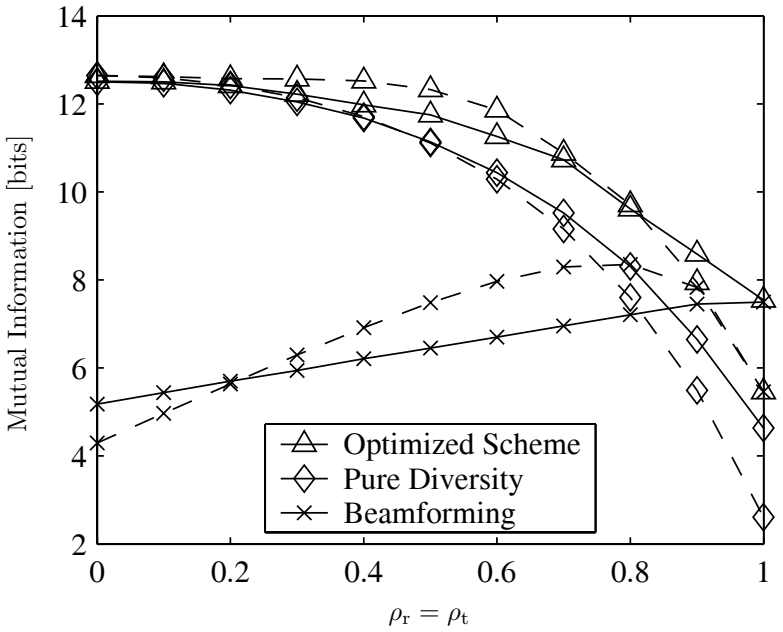


Figure 12.6: Performance of the approximately optimized system when (12.3) is used. The SNR is 10 dB, $n_r = 4$, and $n_t = 8$. Simulated and approximative results are illustrated with solid and dashed curves respectively.

(a) Eigenvalues of $\boldsymbol{Q}$

$\rho_\mathrm{r} = \rho_\mathrm{t}$	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8
0.0	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
0.2	2.81	2.70	2.35	1.60	0.54	0.00	0.00	0.00
0.4	3.87	3.77	2.36	0.00	0.00	0.00	0.00	0.00
0.6	5.12	4.88	0.00	0.00	0.00	0.00	0.00	0.00
0.8	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(b) Eigenvalues of $\boldsymbol{R}_\mathrm{t}$

$\rho_\mathrm{r} = \rho_\mathrm{t}$	λ_1^t	λ_2^t	λ_3^t	λ_4^t	λ_5^t	λ_6^t	λ_7^t	λ_8^t
0.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.2	1.45	1.32	1.16	1.01	0.88	0.78	0.72	0.68
0.4	2.11	1.64	1.20	0.88	0.68	0.55	0.48	0.44
0.6	3.15	1.83	1.04	0.64	0.45	0.34	0.29	0.26
0.8	4.88	1.55	0.62	0.33	0.21	0.16	0.13	0.12

Table 12.1: Eigenvalues of $\boldsymbol{Q}$ for the optimized scheme presented in Figure 12.4 and the corresponding eigenvalues of $\boldsymbol{R}_\mathrm{t}$.

Chapter 13

Bit and Power Loading With Mean or Covariance Feedback

In Chapter 11 and Chapter 12 transmission schemes optimal in an information theory sense were considered. Unfortunately, implementation of such a system is impractical. Therefore suboptimal schemes must be considered to exploit some of the promised gain from using multi-element antenna systems. In this chapter a transmission scheme suitable for MIMO systems such as those presented in Chapter 9 is presented and illustrated with numerical examples. While suboptimal, the scheme has the advantage of being simple and capable of exploiting some of the potential gain in using partial channel state information according to both the mean and covariance feedback models of Section 9.2.1.

The remainder of the chapter is organized in three sections. In Section 13.1 we outline a vision for an end-to-end system design and specify which parts of the system we consider in this chapter. Section 13.2 proposes a simple transmission scheme capable of exploiting some of the partial channel information provided. The performance of our scheme is illustrated by numerical examples in Section 13.3.

13.1 System Vision

The proposed transmission scheme is similar to the transmission scheme analyzed in [HtB03] and to the various versions of BLAST [Fos96, GFVW99] in that data already protected by a channel code is demultiplexed into multiple parallel spatial streams for transmission. At reception, the data symbols are detected, remultiplexed into a single data stream and decoded. This process is often called spatial multiplexing. Under certain assumptions such schemes have been shown to perform near the limits of capacity if soft detection and efficient codes are used [HtB03]. In contrast with the above schemes, we allow for different data rates and power allocation for the different spatial streams. Furthermore, linear algebra techniques are used to exploit directional properties of the MIMO channel and a maximum likelihood detector is used to take decisions on the received data symbols. This way, channel knowledge at the transmitter can be exploited to optimize the transmitted data for the channel. In the literature, spatial multiplexing schemes exploiting channel knowledge at the transmitter have been considered in e.g. [RC98, ZO03]. Unlike those schemes our transmission technique does not require perfect transmitter channel knowledge. Principal system schematics may be found in Figure 13.1.

By allowing these additional degrees of freedom in the system, schemes that exploit available channel knowledge at the transmitter can be formulated. This is the purpose of this chapter where we propose a scheme for choosing the transmission directions for the different spatial streams as well as their data rate and power allocation based on the feedback models of Chapter 9.

13.2 Spatial Multiplexing and Loading

The basic idea behind bit and power loading algorithms is to optimize the transmitted vector data so that the available transmit power is used efficiently. For the system (9.1) the optimal transmit data design when the channel is known at the transmitter is, from a capacity viewpoint [Tel99], to diagonalize $\boldsymbol{H}$, creating parallel spatial channels. This is achieved by using a singular value decomposition and to transmit Gaussian distributed data symbols with power allocated using the well known water-filling solution. This way, transmit power is allocated to the directions where it is put to best use, i.e. in directions with little attenuation more

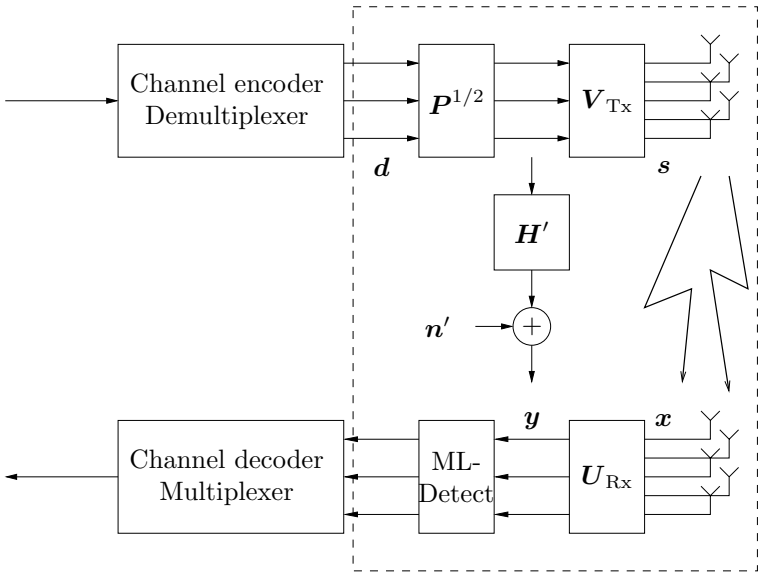


Figure 13.1: The components of a spatial multiplexing and loading MIMO system. In this chapter mainly the part within the dashed rectangle is considered.

power is allocated and higher data rates can be supported.

For practical systems, using Gaussian distributed symbols is not an option. Instead the transmitted data symbols belong to some finite alphabet, resulting in an optimization problem where different constellation sizes and transmit powers are allocated to different directions to satisfy some constraint on the quality of the received data and to maximize data rates. Here, we term this type of technique spatial loading and in the subsections below such schemes are discussed and presented for different scenarios. In all cases we attempt to optimize the transmit data rate, given some design uncoded bit error rate, BER, and transmit power constraint.

13.2.1 Spatial Bit and Power Loading

In the design of a general spatial loading system it is natural to introduce some structure to characterize the spatial properties of the channel and to

simplify the following adaptive loading. To that end, let $\boldsymbol{s} = \boldsymbol{V}_{\mathrm{Tx}}\boldsymbol{P}^{1/2}\boldsymbol{d}$ and $\boldsymbol{y} = \boldsymbol{U}_{\mathrm{Rx}}\boldsymbol{x}$, where $\boldsymbol{V}_{\mathrm{Tx}}$ and $\boldsymbol{U}_{\mathrm{Rx}}$ are unitary matrices characterizing the directivity of the spatial loading system, $\boldsymbol{P}$ is a diagonal matrix defining the power loading in the different spatial dimensions, $\boldsymbol{y}$ is the received data to be considered by the detector and $\boldsymbol{d}$ are the transmitted symbols. The introduced structure is illustrated in Figure 13.1. Based on these definitions, the effective system between the transmitter and the receiver can be modeled as,

$$\boldsymbol{y} = \boldsymbol{U}_{\mathrm{Rx}}(\boldsymbol{H}\boldsymbol{V}_{\mathrm{Tx}}\boldsymbol{P}^{1/2}\boldsymbol{d} + \boldsymbol{n}) = \boldsymbol{H}'\boldsymbol{P}^{1/2}\boldsymbol{d} + \boldsymbol{n}', \tag{13.1}$$

where $\boldsymbol{H}'$ is the effective channel matrix and $\boldsymbol{n}'$ is the effective noise of the MIMO channel between $\boldsymbol{d}$ and $\boldsymbol{y}$, see Figure 13.1. Note that since $\boldsymbol{U}_{\mathrm{Rx}}$ is unitary the effective noise is still white complex Gaussian, each element of variance one. Also, since $\boldsymbol{U}_{\mathrm{Rx}}$ and $\boldsymbol{V}_{\mathrm{Tx}}$ are invertible and the distributions of $\boldsymbol{n}$ and $\boldsymbol{n}'$ are identical the system (13.1) is equivalent to (9.1). The transmitted symbols of $\boldsymbol{d}$ are considered uncorrelated, with zero mean and normalized to variance one. The normalizations from Section 9.2 imply that $\boldsymbol{P}$ should be chosen such that $\mathrm{Tr}\,\boldsymbol{P} = P$.

The focus of this study is on the uncertainty aspect of the channel, not on the receiver algorithm or bit loading scheme being used. To simplify the interpretation of the results and the discussion below, a maximum likelihood (ML) detector is employed in all cases and the well known greedy Hughes-Hartogs algorithm [HH, Bin90] has been selected for the spatial bit and power loading. While these choices may be too computationally demanding for practical implementations, they simplify the derivation and presentation. For reference, the Hughes-Hartogs adaptive loading algorithm for parallel channels consists in principle of the following steps,

1. Try to increase the constellation size by one for all the symbols of $\boldsymbol{d}$, one at the time. Add enough power so that the bit error constraint is not violated.

2. Increase the constellation size and allocate power to the symbol requiring the least additional power in the previous step.

3. Repeat until the available power is insufficient to add more bits given the error constraint. If desirable, any remaining power can be spread over the spatial channels to improve the error rate performance of the wireless link.

Using the framework above different transmission strategies are presented below. Which one should be used depend on the amount of transmitter channel knowledge and channel characteristics. First well known transmission strategies useful when the channel is perfectly known, completely unknown or severely correlated are outlined. Second we propose schemes that maximize the data rate based on the mean and covariance feedback models of Section 9.2.1.

13.2.2 Perfect Channel Knowledge

When the transmitter has perfect channel knowledge, a well known spatial loading scheme based on the singular value decomposition of the channel matrix can be applied [RC98, ZO03]. By choosing $\boldsymbol{U}_{\mathrm{Rx}}$ and $\boldsymbol{V}_{\mathrm{Tx}}$ as the transposed conjugate of the matrix containing the left singular vectors and the matrix containing right singular vectors of the channel matrix $\boldsymbol{H}$ respectively, the channel between $\boldsymbol{d}$ and $\boldsymbol{y}$ is diagonalized and $\min(n_{\mathrm{r}}, n_{\mathrm{t}})$ non-interfering scalar channels between the transmitted and the received data are formed. Since each of these parallel channels are characterized by an SNR given by the singular values of $\boldsymbol{H}$, bit and power loading using the algorithm above is straightforward.

13.2.3 Pure Diversity

For the case when the transmitter has no knowledge of the channel it is not possible to optimize the directivity of the transmitting array. Here, we assume that a non-line of sight system is considered where the antenna elements are sufficiently separated so that the elements of $\boldsymbol{H}$ can be considered independent Rayleigh fading.

For this type of system several techniques such as BLAST [Fos96] or more sophisticated space-time coding schemes [TSC98] have been developed. While elaborate techniques are required for efficient detection of the transmitted symbols and coding is required for optimal performance we here only consider a system where the transmitter transmits uncoded symbols, $\boldsymbol{y}$, with equally distributed power, $\boldsymbol{P} = P/n_{\mathrm{t}}\boldsymbol{I}$, and the receiver uses an ML-detector to estimate the transmitted symbols.

Provided that the channel matrix elements are uncorrelated and soft detection is used, this type of transmission scheme has been shown to perform near capacity [HtB03]. However, if the elements of $\boldsymbol{H}$ are correlated, some directions in space will suffer a higher attenuation than others and transmit power will be wasted.

13.2.4 Beamforming

With access to channel statistics at the transmitter, beamforming schemes can be designed by choosing the transmitted data vectors such that the average received signal power is maximized. For example, in the case of covariance feedback according to the model in Section 9.2.1, this would mean allocating all data and power to the direction corresponding to the maximum eigenvalue of $\boldsymbol{R}_{\mathrm{t}}$, i.e. all data is transmitted over a single spatial channel. Note that this solution is optimal if the elements of $\boldsymbol{H}$ are perfectly correlated and only one spatial channel can be supported. In other cases, where $\boldsymbol{H}$ has a rank higher than one, the technique may be wasteful since available spatial dimensions are not being used.

13.2.5 Mean Feedback

When perfect channel knowledge is not available at the transmitter as in Section 13.2.2 data rate maximization becomes more complicated. For example, it is no longer possible to completely diagonalize the channel and the parallel data streams will interfere at the receiver. Here, we propose a simple sub-optimal scheme for optimizing the data rates and exploiting available transmitter channel knowledge.

Let the singular value decomposition of $\hat{\boldsymbol{H}}$ be written $\hat{\boldsymbol{H}} = \boldsymbol{U}_{\hat{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\hat{\boldsymbol{H}}}\boldsymbol{V}_{\hat{\boldsymbol{H}}}$. The transmitter and receiver directive matrices, $\boldsymbol{V}_{\mathrm{Tx}}$ and $\boldsymbol{U}_{\mathrm{Rx}}$ are selected as $\boldsymbol{V}_{\mathrm{Tx}} = \boldsymbol{V}^*_{\hat{\boldsymbol{H}}}$ and $\boldsymbol{U}_{\mathrm{Rx}} = \boldsymbol{U}^*_{\hat{\boldsymbol{H}}}$ resulting in,

$$\boldsymbol{y} = (\boldsymbol{\Lambda}_{\hat{\boldsymbol{H}}} + \tilde{\boldsymbol{H}}')\boldsymbol{P}^{1/2}\boldsymbol{d} + \boldsymbol{n}' \tag{13.2}$$

where $\tilde{\boldsymbol{H}}' = \boldsymbol{U}_{\mathrm{Rx}}\tilde{\boldsymbol{H}}\boldsymbol{V}_{\mathrm{Tx}}$ has the same distribution as $\tilde{\boldsymbol{H}}$ since $\boldsymbol{V}_{\mathrm{Tx}}$ and $\boldsymbol{U}_{\mathrm{Rx}}$ are unitary. While we do not claim that this choice of $\boldsymbol{V}_{\mathrm{Tx}}$ and $\boldsymbol{U}_{\mathrm{Rx}}$ is optimal it can be motivated in a number of ways. First, given the channel knowledge at the transmitter, the received signal power for the first symbol in $\boldsymbol{d}$, $\mathrm{E}\,|\boldsymbol{U}_{\mathrm{Rx}}\boldsymbol{H}\boldsymbol{V}_{\mathrm{Tx}}\boldsymbol{P}^{1/2}\boldsymbol{d}_1|^2$ is maximized. Here, $\boldsymbol{d}_k$ is a vector where all entries are zero except the kth entry which is identical to kth element of $\boldsymbol{d}$. Similarly, the second element is transmitted in the orthogonal direction, in $\mathbb{C}^{n_{\mathrm{t}}}$, to the first, that maximizes the received signal power and so on. Hence given this choice of directive matrices it is possible to ensure that data is transmitted in the directions where the receive conditions are likely to be the most favorable. Second, since $\boldsymbol{\Lambda}_{\hat{\boldsymbol{H}}}$ is diagonal and $\tilde{\boldsymbol{H}}'$ spatially white, the received power from the first transmitted symbol of $\boldsymbol{d}$ is maximized at the first received symbol of $\boldsymbol{y}$

while the received power at the other elements of $\boldsymbol{y}$ is minimized given the available channel information. Similarly, the received power resulting from the second symbol of $\boldsymbol{d}$ is concentrated to the second element of $\boldsymbol{y}$, and so on. Thus, while interference between the different spatial carriers is unavoidable, this choice of directive matrices in some sense minimizes it. Third, our $\boldsymbol{V}_{\mathrm{Tx}}$ coincides with the choice based on capacity arguments made in [VM01, JG04]. In these papers only the special cases of a single antenna receiver and a rank one $\hat{\boldsymbol{H}}$ are considered though. Finally, for the case $\sigma^2_{\boldsymbol{H}} \to 0$, this solution coincides with that of Section 13.2.2, and, if the elements of the channel can be considered independent Rayleigh fading, the choice is also valid when $\sigma^2_{\boldsymbol{H}} \to 1$, see Section 13.2.3.

To provide a practical spatial bit and power loading scheme it is necessary to be able compute the resulting BER for various constellation sizes and output powers efficiently. Since the communication channels are interfering this is complicated and for a computationally attractive scheme some approximations are necessary. First, we assume that the design BER is chosen so low that more than one symbol error per received vector $\boldsymbol{y}$ is rare. Thus, when computing the error rates of the ML-detector for each of the transmitted symbols in $\boldsymbol{d}$ we assume that the other symbols have been correctly detected and subtracted from the received data, i.e. error propagation is ignored. Note that this approximation can be expected to work better when there are more receive than transmit antennas, and in practice it is good to limit the number of spatial channels to $\min(n_{\mathrm{r}}, n_{\mathrm{t}})$ or less. Furthermore, it is assumed that the transmitted symbols have been Gray-encoded so that, given the low design BER, the number of bit errors can be approximated by the number of symbol errors. These approximations result in a simplified scenario where it suffices to compute the error probability of a single symbol, transmitted over a vector channel consisting of a combination of independent Rayleigh and Ricean fading elements. For this type of channels efficient techniques for computing the error probability for many types of constellations exist [AG99]. Note that since the variances of the vector channel elements are determined by $\sigma^2_{\boldsymbol{H}}$ the bit and power loading in the different directions are adapted to the uncertainty of the channel.

Using the selected $\boldsymbol{V}_{\mathrm{Tx}}$ and $\boldsymbol{U}_{\mathrm{Rx}}$ and the approximations above, performing the spatial loading can be performed in a simple and computationally efficient manner. While this algorithm is clearly suboptimal, note that for the cases when the channel information is almost perfect, or when the channel is almost completely unknown, our solution converges to well known results, see Sections 13.2.2 and 13.2.3.

13.2.6 Covariance Feedback

If covariance feedback is available, one method of optimizing the system could be to, depending on the amount of correlation, chose either the pure diversity or the beamforming, see Section 13.2.3, Section 13.2.4 and also [VM01, JG04, JB04]. Such a solution however, is not very elegant and suffers from suboptimal performance in many cases. Below we instead propose to use a spatial loading scheme to optimize the transmitted data and adapt it to the available channel information to provide a smooth transition between the cases above. When perfect channel knowledge is not available at the transmitter, optimizing the transmitted data becomes more complicated. For example, it is no longer possible to completely diagonalize the channel and the parallel data streams will interfere at the receiver.

Let $\boldsymbol{R}_{\mathrm{t}} = \boldsymbol{V}_{\mathrm{t}}\boldsymbol{\Lambda}_{\mathrm{t}}\boldsymbol{V}_{\mathrm{t}}^{*}$ and $\boldsymbol{R}_{\mathrm{r}} = \boldsymbol{V}_{\mathrm{r}}\boldsymbol{\Lambda}_{\mathrm{r}}\boldsymbol{V}_{\mathrm{r}}^{*}$ be the eigenvalue decompositions of the transmit and receive covariance matrices. The transmitter and receiver directive matrices, $\boldsymbol{V}_{\mathrm{Tx}}$ and $\boldsymbol{U}_{\mathrm{Rx}}$ are selected as $\boldsymbol{V}_{\mathrm{Tx}} = \boldsymbol{V}_{\mathrm{t}}$ and $\boldsymbol{U}_{\mathrm{Rx}} = \boldsymbol{V}_{\mathrm{r}}^{*}$ resulting in an efficient system,

$$\boldsymbol{y} = \boldsymbol{\Lambda}_{\mathrm{r}}^{1/2}\boldsymbol{G}'\boldsymbol{\Lambda}_{\mathrm{t}}^{1/2}\boldsymbol{P}^{1/2}\boldsymbol{d} + \boldsymbol{n}', \tag{13.3}$$

where $\boldsymbol{G}' = \boldsymbol{V}_{\mathrm{r}}^{*}\boldsymbol{G}\boldsymbol{V}_{\mathrm{t}}$ has the same distribution as $\boldsymbol{G}$ since $\boldsymbol{V}_{\mathrm{r}}$ and $\boldsymbol{V}_{\mathrm{t}}$ are unitary. While we do not claim that this choice of $\boldsymbol{V}_{\mathrm{Tx}}$ and $\boldsymbol{U}_{\mathrm{Rx}}$ is optimal it can be motivated in a number of ways. First, given the channel knowledge at the transmitter, the received signal power for the first symbol in $\boldsymbol{d}$, $\mathrm{E}\,|\boldsymbol{U}_{\mathrm{Rx}}\boldsymbol{H}\boldsymbol{V}_{\mathrm{Tx}}\boldsymbol{P}^{1/2}\boldsymbol{d}_1|^2$ is maximized. Here, $\boldsymbol{d}_k$ is a vector where all entries are zero except the kth entry which is identical to kth element of $\boldsymbol{d}$. Similarly the second element is transmitted in the orthogonal direction, in $\mathbb{C}^{n_{\mathrm{t}}}$, to the first, that maximizes the received signal power and so on. Hence given this choice of directive matrices it is possible to ensure that data is transmitted in the directions where the receive conditions are likely to be the most favorable. Second, this choice can be motivated by capacity arguments, see Chapter 12 or e.g. [VM01, JVG01, SM03, JB04]. Last, this choice of $\boldsymbol{U}_{\mathrm{Rx}}$ ensures that the elements of $\boldsymbol{H}'$ are uncorrelated significantly simplifying the BER computations in the following loading step of the proposed design process. Note that since $\boldsymbol{U}_{\mathrm{Rx}}$ is unitary and an ML-detector is used the choice of $\boldsymbol{U}_{\mathrm{Rx}}$ in the receiver does not affect the resulting error probability for the designed system.

To provide a practical spatial bit and power loading scheme it is necessary to be able compute the resulting BER for various constellation

sizes and output powers efficiently. We again ignore error propagation and assume that the transmitted symbols have been Gray-encoded. Then similar approximations as those made for the mean feedback case may be applied to estimate the error rates of the spatial channels, see Section 13.2.5 or [MBO04b] for more details.

Using the selected $\boldsymbol{V}_{\mathrm{Tx}}$ and $\boldsymbol{U}_{\mathrm{Rx}}$ and the approximations above, the spatial loading can be performed in a simple and computationally efficient manner. While the resulting design is clearly suboptimal, note that for completely uncorrelated channel elements, i.e. when $\boldsymbol{R}_{\mathrm{t}} = \boldsymbol{I}$ and $\boldsymbol{R}_{\mathrm{r}} = \boldsymbol{I}$, our solution converges to the well know solution of Section 13.2.3. Also, when the channel elements are perfectly correlated, our algorithm produces a beamforming solution, concentrating all transmit power on the single available spatial channel, see Section 13.2.4.

13.3 Numerical Results and Analysis

Since an analytical analysis on the performance of the method proposed in the previous section would be difficult, a numerical analysis is performed. The simulations performed in this study were simplified by limiting the bit-loading algorithm to square M-QAM constellations, where $M = 2^{2b}$ and b is integer valued. Note that results such as those in [AG99] allow a larger selection of constellations which may be desirable.

To make the simulations efficient, the ML-detector at the receiver was implemented in the form of a sphere decoder [VB99]. For reasonably sized arrays the sphere decoding algorithm is on average very efficient, significantly shortening the time required to find the ML-solution compared with a full search [JO05]. In the case of spatial loading, the sphere decoding algorithm needs to consider the different constellations transmitted in the different elements of $\boldsymbol{d}$.

13.3.1 Examples with Mean Feedback

LOS Channel with Local Scattering

Consider a line of sight scenario with significant scattering around the arrays. Provided that the antenna elements of the arrays are placed too closely to be separated based on the line of sight component of the received data, but so well separated that the fading caused by the multi-path scatterers can be considered independent, a reasonable chan-

nel model is

$$\boldsymbol{H} = \boldsymbol{H}_{\mathrm{LOS}} + \boldsymbol{H}_{NLOS}, \tag{13.4}$$

see e.g. [YBOB02]. Here, $\boldsymbol{H}_{\mathrm{LOS}}$ is a rank one matrix modeling the line of sight component, $\boldsymbol{H}_{\mathrm{NLOS}}$ is a random matrix where the elements are independently drawn from a zero-mean complex Gaussian source. In some scenarios the scattering components of the channel may be rapidly changing while the line of sight component and the statistical properties of the scattering component is changing more slowly. Hence, while keeping exact channel estimates at the transmitter may be difficult, maintaining estimates of $\boldsymbol{H}_{\mathrm{LOS}}$ and the variance of the elements of $\boldsymbol{H}_{\mathrm{NLOS}}$ could be feasible. This scenario fits well within the framework of our proposed method, $\hat{\boldsymbol{H}}$ then corresponds to $\boldsymbol{H}_{\mathrm{LOS}}$ and $\sigma^2_{\boldsymbol{H}}$ to the variance of the elements of $\boldsymbol{H}_{\mathrm{NLOS}}$.

Figure 13.2 shows the bit rate as a function of the resulting BER of a three by three antenna system where the line of sight component provides 80% of the received signal power, i.e. $\sigma^2_{\boldsymbol{H}} = 0.2$. Two methods of system design are considered, one using only the line of sight rank one channel estimate in the design of the bit and power loading, the other using our proposed method. In both cases, any remaining transmit power after the bit loading is spread evenly over the spatial carriers to ensure SNR of 15, 20 and 25 dB. By taking the unknown multi-path components into account the simple method proposed in Section 13.2.6 is capable of increasing the data rate. In the high SNR scenarios of this example the gains are significant.

To further illustrate the behavior of the algorithm, Table 13.1 shows how the transmitted bits are allocated as a function of $\sigma^2_{\boldsymbol{H}}$. As the resulting error rates vary greatly between the different channels this result gives no indication of resulting bit rates but is intended to illustrate how the bits are allocated to the different carriers. Note that when $\sigma^2_{\boldsymbol{H}} \to 0$ signal power and loaded bits are concentrated in the direction given by the line of sight channel, while when $\sigma^2_{\boldsymbol{H}} \to 1$ power and data are transmitted as evenly as possible in space given the restriction in constellation sizes. This intuitive result illustrates how this method, while simple, is capable of adapting the transmitted data according to the transmitter channel knowledge available.

Rayleigh Fading with Imperfect Channel Estimates

In this example the elements of both $\hat{\boldsymbol{H}}$ and $\tilde{\boldsymbol{H}}$ are drawn independently from Gaussian distributions making the channel $\boldsymbol{H}$ Rayleigh fading. This

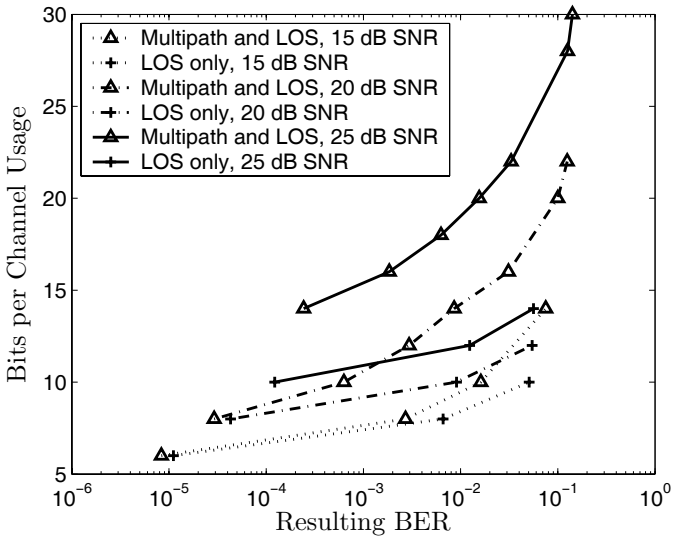


Figure 13.2: Mean feedback in a line of sight (LOS) scenario with local scattering, i.e $\hat{\boldsymbol{H}}$ is rank one. Three $n_{\mathrm{r}} = n_{\mathrm{t}} = 3$ antenna systems with $\sigma^2_{\boldsymbol{H}} = 0.2$ were considered with SNR 15, 20 and 25 dB respectively. The beamforming scheme of Section 13.2.4 is compared with the optimized scheme of Section 13.2.5. In contrast with the beamforming scheme which only uses the spatial channel associated with the LOS component, the optimized scheme also exploits channels created by the multi-path propagation resulting in improved data rate performance.

	$\sigma^2_{\boldsymbol{H}}$										
Channel	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1:	8	8	8	8	6	6	6	6	6	6	4
2:	0	2	2	2	4	4	4	4	4	4	4
3:	0	0	2	2	4	4	4	4	4	4	4
4:	0	0	2	2	2	2	4	4	4	4	4

Table 13.1: Bit allocation, design SNR 20 dB, design bit error rate 0.003. Note that for the case of an uncorrelated channel and a rank one channel our method converges to well known solutions.

could model an indoor non line of sight channel with well separated antenna elements at both arrays. A four transmit, six receive antenna system designed for an SNR of 20 dB is considered.

Figure 13.3 shows bit rate versus resulting BER at various degree of uncertainty of the channel estimate, $\sigma^2_{\boldsymbol{H}}$. Any remaining power after the bit loading is spread evenly over the carriers to make the comparison fair. The figure clearly illustrates the capability of the proposed method to connect well known results for the design of system with perfect channel knowledge at the transmitter, $\sigma^2_{\boldsymbol{H}} = 0$, to designs used when no instantaneous channel knowledge, $\sigma^2_{\boldsymbol{H}} = 1$, is available.

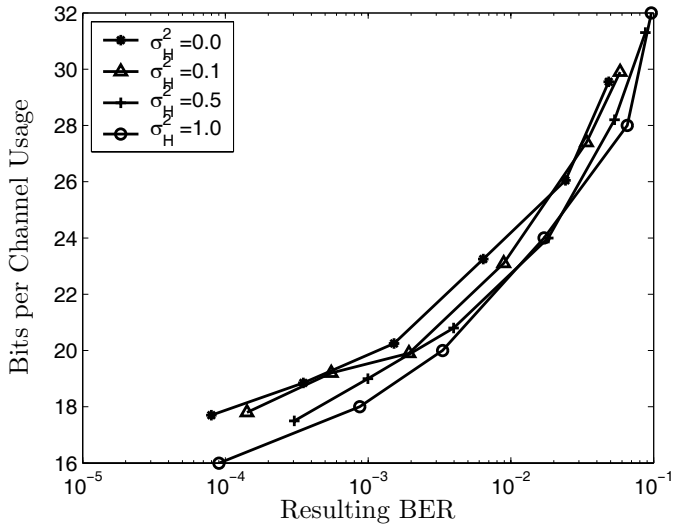


Figure 13.3: Mean feedback with an independently Rayleigh fading channel, $n_{\mathrm{r}} = 6$, $n_{\mathrm{t}} = 4$, SNR 20 dB. The data rate performance improves as $\sigma^2_{\boldsymbol{H}}$ approaches 0 and the transmitter channel estimate becomes more accurate.

Finally, the performance of the approximations leading up to the bit loading algorithm are evaluated. Figure 13.4 shows the resulting BER as a function of the design BER. To be able to evaluate the performance of the approximations, the remaining power after the bit allocation is not allocated to any spatial carrier. As the approximations used in deriving the proposed method are underestimating the probability of error, the design BER is lower than the resulting BER. While this implies that the

design BER must be chosen lower than that required by the system, the proposed method achieves balancing in the loading of the different spatial channels. Furthermore, notice that as $\sigma^2_{\boldsymbol{H}}$ approaches zero, the approximations improve. This is expected as small $\sigma^2_{\boldsymbol{H}}$ means less interference and error propagation between the channels making the approximation to ignore these effects more justified.

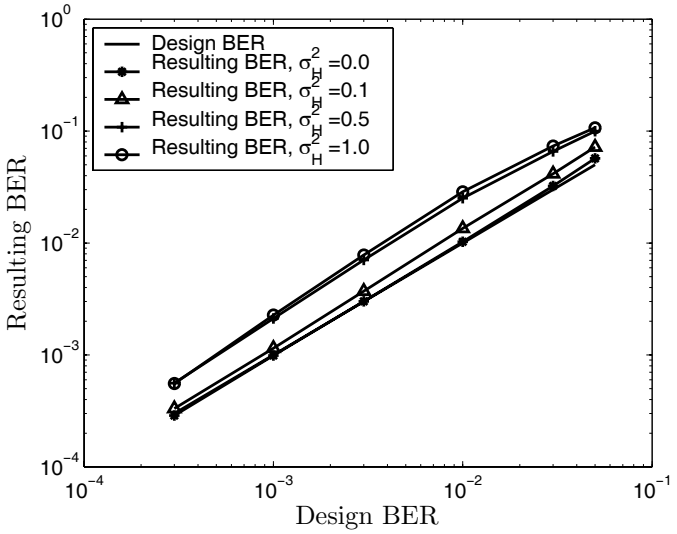


Figure 13.4: Mean feedback with independently Rayleigh fading channel, $n_{\mathrm{r}} = 6$, $n_{\mathrm{t}} = 4$, design SNR 20 dB. Performance of the BER approximations at various degrees of channel uncertainty. Perfect channel knowledge ($\sigma^2_{\boldsymbol{H}} = 0$) removes the inter-channel interference improving the error rate approximation.

13.3.2 Examples with Covariance Feedback

While the algorithm presented herein is applicable to general $\boldsymbol{R}_{\mathrm{r}}$ and $\boldsymbol{R}_{\mathrm{t}}$, to keep things simple, exponential correlation is assumed in all examples below, i.e. the correlation matrices, $\boldsymbol{R}_{\mathrm{t}}$ and $\boldsymbol{R}_{\mathrm{r}}$ are selected according to (9.4).

Data Rate Performance

Figure 13.5 illustrates how the proposed method can exploit some of the available information to increase the data rate of the system. The data rate performance of the algorithm is compared with the two simpler schemes of Sections 13.2.3 and 13.2.4. In this downlink scenario an $n_t = 6$, $n_r = 3$ antenna system with $\rho_r = \rho_t = 0.8$ is considered operating at SNR of 15 and 25 dB. During the loading procedure described in Section 13.2.6, the number of spatial channels was limited to three and to make the comparison fair in terms of transmit power, any residual power left after the loading procedure has been added evenly over those channels.

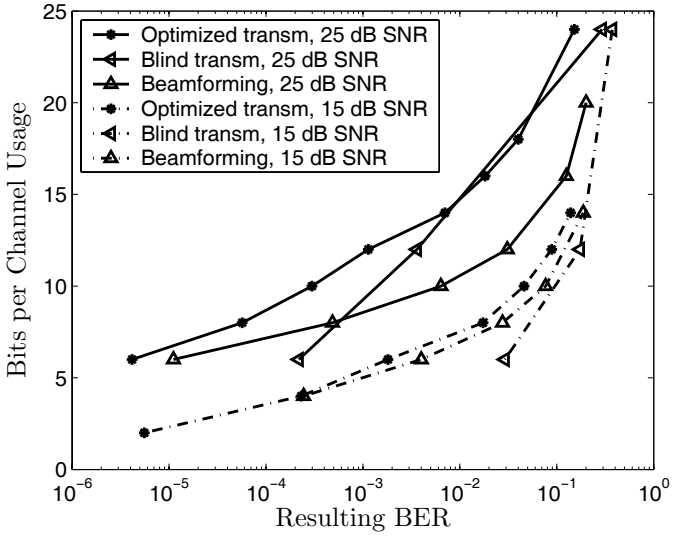


Figure 13.5: Covariance feedback with spatially correlated channels, $\rho_r = \rho_t = 0.8$, $n_t = 6$, $n_r = 3$. Bit error rate performance at 15 and 25 dB SNR. Note that our technique, denoted "optimized transm" see Section 13.2.6, outperform both the "blind" transmission scheme, see Section 13.2.3, and the "beamforming" scheme, see Section 13.2.4.

From the figure, notice how the algorithm presented herein is capable of outperforming the pure diversity and beamforming methods. This illustrates that this method, while simple, has the capability to exploit some of the information in the available channel statistics to increase the

data rates of the system.

Bit Loading Strategy

To further illustrate the behavior of the proposed algorithm, Table 13.2 shows how the transmitted bits are allocated as a function of $\rho_r = \rho_t$ for this $n_r = n_t = 4$ scenario. While the design BER in the example is 0.003 for the entire table, as the resulting error rates vary between the different channels, this table is not intended to indicate the resulting bit rates of a practical system but to show how the bits are distributed over the different spatial channels. Note that when $\rho_r = \rho_t \to 0$ signal power and transmitted data are transmitted as evenly as possible in space given the restriction in constellation sizes. On the other hand, when $\rho_r = \rho_t \to 1$ power and loaded bits are concentrated in the direction with the lowest attenuation. This intuitive result illustrates how the proposed method is capable of adapting the transmitted data to the available channel knowledge and providing a transition between the extreme cases of an uncorrelated channel and a perfectly correlated channel.

	$\rho_r = \rho_t$										
Channel	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1:	6	6	6	6	6	8	8	8	8	6	8
2:	6	6	6	6	6	6	6	6	4	4	0
3:	6	6	6	6	6	4	4	4	2	2	0
4:	6	6	6	6	4	4	4	2	2	0	0

Table 13.2: Bit allocation, covariance feedback with spatially correlated channels, design SNR 25 dB, design BER 0.003, $n_t = n_r = 4$. Note that for the cases of no correlation and perfect correlation our technique converges to well known solutions.

Performance of the Approximation

Finally, the performance of the approximations leading up to the bit loading algorithm are evaluated. Figure 13.6 shows the resulting BER as a function of the design BER for an $n_t = 6$, $n_r = 3$ system operating at an SNR of 20 dB. The number of spatial channels was limited to three. To be able to evaluate the performance of the approximations, the remaining power after the bit allocation is not allocated to any spatial channel. As the approximations used in deriving the proposed method are underestimating the probability of error, the design BER is lower than the

resulting BER. While this means that the design BER must be chosen lower than that required by the system, the design achieves balancing in the loading between the different spatial channels. Furthermore, notice that as $\rho_r = \rho_t$ approaches one, the approximations improve. This is expected as large $\rho_r = \rho_t$ means that almost all power and bits are allocated to a single channel and thus there is less interference and error propagation between the spatial channels and the approximation of ignoring these effects improves.

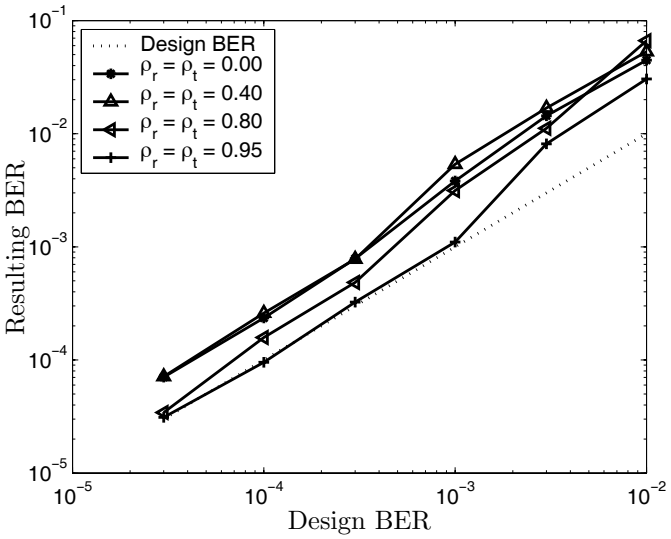


Figure 13.6: Covariance feedback with spatially correlated channels, $n_t = 6$, $n_r = 3$, 20 dB SNR. Performance of the bit error rate, BER, approximation. More correlation means that fewer spatial channels are used resulting in reduced error propagation.

Chapter 14

Concluding Remarks and Future Work

In this part of the thesis analysis and design techniques for MIMO communication systems were considered. Both practical and more theoretical transmission schemes were studied using statistical channel models taking into account spatial correlation and channel estimate uncertainties. Approximations of various system parameters were analytically derived, allowing for insights into the impact of channel impairments and suitable countermeasures. In this chapter we summarize and discuss our main results and some opportunities for further work are also outlined.

14.1 Conclusions

14.1.1 Channel Eigenvalues

In Chapter 10, we derived limiting eigenvalue distributions for MIMO channels suffering from spatially correlated fading under the assumption that the number of transmit antennas is large. Subject to certain conditions on the correlation between the channel elements, it is shown that the limiting distribution is Gaussian and closed form expressions for the mean and variance were derived. Simulations indicate that the derived asymptotic eigenvalue distributions can be used to approximate the true distribution for a wide range of realistic scenarios.

While the proposed techniques apply to systems where the number

of transmit antennas is large, analogous results for systems with large receiver arrays are straightforward to obtain.

14.1.2 Mutual Information and Outage Capacity

As an application of the asymptotic channel eigenvalue distributions, an approximation of the mutual information and outage capacity of the MIMO channel was derived. It was shown that as the number of transmit antennas grows, the outage capacity can be obtained from a cumulative distribution function of a limiting normal distribution. Closed form expressions for the mean and variance were derived and the asymptotic results were compared against sample results from simulations. At reasonable transmit correlations the large array results were found to provide reasonable approximations of the true properties even for realistic array sizes.

14.1.3 Transmit Covariance Optimization

In Chapter 12 a method for approximate optimization of the mutual information of a MIMO communication system was presented based on the asymptotic results of Chapter 11. As indicated by numerical experiments this may be a promising approach as the design resulting from the proposed algorithm appears to perform well in a wide range of realistic system scenarios.

The approximations of the mutual information and outage capacity illustrate the usefulness of the derived asymptotic eigenvalue results. While the approximation appears to match simulated results well, at the same time it is simple enough to provide insight to, for example, the MIMO channel capacity. Since several other properties of MIMO communication systems depend on the channel eigenvalues we believe that the asymptotic eigenvalue distribution presented in this paper may prove useful in, for example, MIMO design problems.

14.1.4 Spatial Multiplexing with Bit and Power Loading

Chapter 13 proposes and evaluates a computationally efficient method for designing a spatially multiplexed MIMO communication system with partial channel state information at the transmitter. Results from simulations indicate that the algorithm provides gains compared to pure diver-

sity techniques and methods that simply maximize received SNR. When correlation feedback is considered, for cases where the elements of the channel matrix are perfectly correlated or uncorrelated, the scheme converges to well known solutions, providing a seamless transition between beamforming techniques and transmission schemes used over unknown independent Rayleigh fading channels. Similarly, when channel feedback with uncertainty is considered, the technique provides the expected solutions in the extreme cases of perfectly known or completely unknown channel responses.

14.2 Discussion and Future Work

This section discusses some aspects of the analysis and design techniques that could be improved. We also outline some opportunities for further work.

14.2.1 The Impact of Spatial Correlation

The analysis and design techniques proposed and analyzed in Chapter 10 to Chapter 12 provide insight to the impact of spatial correlation and suitable countermeasures. However, some issues still require more attention. For example, it remains to be shown whether the global optimum of the problem (12.1) really is found. Also, as is illustrated by the improved performance due to the ad-hoc method in Section 12.4, better approximations of the mutual information may be achievable which could improve performance further. Optimization of the transmit covariance often results in a solution that severely violates the conditions the approximation of underlying eigenvalue distribution is derived under. It is therefore rather surprising that the mutual information approximation works as well as it does over a large spectrum of correlation factors and system configurations. Note that if systems with a large number of receive antennas were considered instead this problem would be alleviated.

In some scenarios it may be more attractive to optimize the system with respect to the outage capacity [Fos96]. This could be achieved using similar techniques as those applied in Chapter 12 using the approximation of the outage capacity derived in Chapter 11.

14.2.2 Practical System Design

For a practical implementation of techniques similar to those proposed in Chapter 14.1.4 significant amounts of theoretical work remain. For example, the computational complexity of the technique on both transmit and receive side of the system must be addressed. On the transmitter side, if a frequency selective channel with OFDM is considered, there will be parallel channels not only in space but also in frequency. Optimally bit and power loading is applied in both space and frequency [ZO03]. As the number of parallel channels then would be very large, the use of less demanding algorithms than Hughes-Harthogs needs to be considered. On the receiver side, using maximum likelihood detection may be computationally prohibitive. Though recent methods such as sphere decoding [VB99] on average drastically reduce the complexity of such a detector [JO05], if the number of spatial channels is large, suboptimal detectors need to be considered.

Also from a more theoretical viewpoint there are a number of issues to address. First, in a real performance analysis the performance of the entire system including the channel code needs to be considered. Second, for the best performance the channel decoder should have access not to decisions taken by a symbol detector but to the likelihoods of the transmitted data symbols. These two aspects were taken into account in [HtB03] where it was shown that a system similar to that considered in Chapter 13 with equal bit and power allocation can perform near capacity on a spatially independent Rayleigh fading channel. In that work a list sphere decoder was used to estimate the likelihoods and a turbo-code used to protect against noise and deep fades. A similar scheme, taking partial channel state information into account could be devised using the methods derived herein. Such a transmission scheme would have the potential of high spectral efficiency while robust against spatial correlation or uncertain channel estimates.

Another central aspect of the spatial loading algorithm that could be improved is the error estimation. The technique employed in Chapter 14.1.4 is sensitive to error propagation between the different spatial channels and it is restricted to two special cases of the more general channel model (9.2). Also, the technique does allow linear space time block coding schemes where the transmission of a symbol is spread over several spatial channels and channel usages. These types of aspects have been taken into account in [BMO05, BMO04] providing a natural continuation of the work presented in this thesis.

Part III

Epilogue

Chapter 15

Thesis Conclusions

This thesis presents analysis and design techniques for wireless communications with diversity. Using statistical models the impact of various channel impairments can be studied efficiently and suitable countermeasures designed. In particular we considered fundamental limitations and practical transmission schemes for mobile satellite broadcast and terrestrial multiple-input multiple-output systems.

Mobile satellite communication is characterized by severe temporal correlation with deep long-term fades. To overcome channel outages we considered transmission schemes allowing coded temporal and spatial diversity. Based on a realistic statistical model an efficient technique to estimate the service availability was derived. This allowed several elaborate design schemes to be developed adding insights to which service requirements can be satisfied at what cost. Combined with an analysis of different techniques to multiplex the spatial diversity, insights to the achievable spectral efficiency are provided.

Maintaining reliable channel estimates on the transmit side of a multiple-input, multiple-output communication system may be challenging. Therefore, designs relying on properties that change on a slower time scale may be attractive. Using an asymptotic assumption we analyzed fundamental properties and limitations of a channel modeled according to a statistical non-line of sight indoor model. A more practical scheme to adapt transmitted data to available transmitter channel knowledge was also presented. While simple, numerical examples demonstrated its ability to exploit some of the available information.

Correlation in the channel fading is an impairment that generally

leads to reduced system performance. This is the case with the temporal correlation in mobile satellite communications as well as the spatial correlation in MIMO systems. In this work we have shown that if the fading statistics are taken into account in the design, the negative impact of correlation can be significantly reduced and in some scenarios even prove beneficial. Thus, these aspects are critical in an optimization process leading to a reliable and efficient system.

To provide insights to the impact of correlation and other impairments, our goal has been to derive analysis and design techniques that are simple and efficient. This also makes our results suitable for extension to multi-user design and other full-scale system considerations. To achieve this goal, our results are based on statistical models of manageable complexity. Therefore these models only include the aspects that we believe have the largest impact on the parameters of interest. Thus, the presented analytical methods should be seen as a complement and starting point for more detailed link simulations and measurement campaigns.

Bibliography

[AdCS98] Jaouhar Ayadi, Elisabeth de Carvalho, and Dirk T.M. Slock. Blind and semi-blind maximum likelihood methods for FIR multichannel identification. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3185–3188, May 1998.

[AG99] Mohamed-Slim Alouini and Andrea J. Goldsmith. A unified approach for calculating error rates of linearly modulated signals over generalized fading channels. *IEEE J. Select. Areas Commun.*, 47(9):1324–1334, September 1999.

[Ala98] Siavash M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE J. Select. Areas Commun.*, 16(8):1451–1458, October 1998.

[And63] Theodore Wilbur Anderson. Asymptotic theory for principal component analysis. *Ann. Math. Statist.*, 34(1):122–148, 1963.

[And84] Theodore Wilbur Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons Inc., 2 edition, 1984.

[Arr] Arraycomm. San Jose, California USA. [Online] Available: www.arraycomm.com.

[BG96] Claude Berrou and Alain Glavieux. Near optimum error correcting coding and decoding: Turbo-codes. *IEEE Trans. Commun.*, 44(10):1261–1271, October 1996.

[Bin90] John A.C. Bingham. Multicarrier modulation for data transmission: An idea whose time has come. *IEEE Commun. Mag.*, pages 5–14, May 1990.

[BMO+02] Mats Bengtsson, Cristoff Martin, Björn Ottersten, Ben Slimane, and Per Zetterberg. Recent advances on MIMO processing in the SATURN project. In *Proc. IST Mobile Communications Summit*, June 2002.

[BMO04] Svante Bergman, Cristoff Martin, and Björn Ottersten. Bit and power loading for spatial multiplexing using partial channel information. In *ITG Workshop on Smart Antennas*, April 2004.

[BMO05] Svante Bergman, Cristoff Martin, and Björn Ottersten. Performance analysis for maximum likelihood detection of high rate space-time codes. *IEEE Trans. Signal Processing*, 2005. Submitted to.

[BPS98] Ezio Biglieri, John Proakis, and Schlomo Shamai (Shitz). Fading channels: Information-theoretic and communications aspects. *IEEE Trans. Inform. Theory*, 44(6):2619–2692, October 1998.

[Bri95] Robert D. Briskman. Satellite DAB. *Int. J. Satellite Commun.*, 13:259–266, July 1995.

[BSB+89] M.A. Beach, A.C. Swales, A. Bateman, D.J. Edwards, and J.P. McGeehan. A diversity combining antenna array for land mobile satellite communications. In *Vehicular Technology Conference*, volume 2, pages 749–756, June 1989.

[BT02] Lars Erling Bråten and Terje Tjelta. Semi-Markov multistate modeling of the land mobile propation channel for geostationary satellites. *IEEE Trans. Antennas Propagat.*, 50(12):1795–1802, December 2002.

[BYO01] Mats Bengtson, Kai Yu, and Björn Ottersten. Single and dual multi-sensor channel characterisation – analysis and models. Technical report, IST-SATURN, November 2001. IST-1999-10322 SATURN D523, Part 1, ver 2. [Online] Available: http://www.ist-saturn.org.

[CCF+04] N. Chuberre, G. E. Corazza, MG Francon, C. Nussli, C. Selier, A. Vanelli-Coralli, and P. Vincent. Satellite digital multimedia broadcasting for 3G and beyond 3G systems. Technical report, IST MAESTRO, June 2004. [Online] Available: http://ist-maestro.dyndns.org.

[Cim85] Leonard J. Cimini. Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Trans. Commun.*, COM-33(7):665–675, July 1985.

[CP96] Keith M. Chugg and Andreas Polydoros. MLSE for an unknown channel — Part I: optimality considerations. *IEEE Trans. Commun.*, 44(7):836–846, July 1996.

[Cra46] Harald Cramér. *Mathematical Methods of Statistics*. Princeton University Press, 1946.

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[CTKV02] Chen-Nee Chuah, David Tse, Josph M. Kahn, and Reinaldo A. Valenzuela. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Trans. Inform. Theory*, 48:637–650, March 2002.

[Cup69] Vladimir Cuperman. An upper bound for the error probability on the Gilbert channel. *IEEE Trans. Commun. Technol.*, 17(5):532–535, October 1969.

[Dav02] Faramaz Davarian. Sirius satellite radio: Radio entertainment in the sky. In *IEEE Aerospace Conference Proceedings*, pages 3–1031 – 3–1035, March 2002.

[dCS97] Elisabeth de Carvalho and Dirk T.M. Slock. Cramér-Rao bounds for semi-blind, blind and training sequence based channel estimation. In *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 129–132, 1997.

[DGMD04] Axel Dumeur, Philippe Godlewski, Philippe Martins, and Phillipe Debreux. On the performance of DVB-T networks for indoor and mobile reception. In *Vehicular Technology Conference*, September 2004.

[Ede89] Alan Edelman. *Eigenvalues and Condition Numbers of Random Matrices*. PhD thesis, Massachusetts Institute of Technology, 1989.

[EHS+04] Harald Ernst, Thomas Hack, Sandro Scalise, Christophe Loeillet, and Rolv Mithassel. Ku-mobil: Satellite multimedia services for cars in Ku-band. In *European Workshop on Mobile/Personal Satcoms and Advanced Satellite Mobile Systems Conference*, 2004.

[Ell63] E. O. Elliot. Estimates of error rates for codes on burst-noise channels. *Bell Syst. Tech J.*, 42:1977–1997, 1963.

[Eri00] Ericsson. Ericsson and Mannesman rollout the world's first GSM adaptive antenna base stations. [Online] Available: http://www.ericsson.com/press/archive/-2000Q2/20000417-0026.html, 2000. Ericsson Press Release.

[ESH04] Harald Ernst, Sandro Scalise, and Guy Harles. Measurement and modeling of the land mobile satellite channel at Ku-band. *IEEE Trans. Veh. Technol.*, 2004. Preprint.

[ESS04] Harald Ernst, Luca Sartorello, and Sandro Scalise. Transport layer coding for the land mobile satellite channel. In *Vehicular Technology Conference*, May 2004.

[ETS01] ETSI, 650 Route des Lucioles, F-06921 Sophia Antipolis Cedex, France. *ETSI EN 300 401; Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers*, 1.3.3 edition, 2001.

[ETS04a] ETSI, 650 Route des Lucioles, F-06921 Sophia Antipolis Cedex, France. *ETSI EN 300 744 V1.5.1 (2004-11); Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television*, 2004. [Online] Available: http://www.etsi.org.

[ETS04b] ETSI, 650 Route des Lucioles, F-06921 Sophia Antipolis Cedex, France. *ETSI EN 302 304 V1.1.1 (2004-11); Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H)*, 2004. [Online] Available: http://www.etsi.org.

[ETS04c] ETSI, 650 Route des Lucioles, F-06921 Sophia Antipolis Cedex, France. *ETSI ES 201980 V2.1.1 (2004-06); Digital Radio Mondiale (DRM); System Specification*, 2004. [Online] Available: http://www.etsi.org.

[Fer96] Thomas S. Ferguson. *A Course in Large Sample Theory.* Chapman & Hall, 1996.

[FG98] Gerard J. Foschini and Michael J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6(3):311–335, March 1998.

[FGF+97] Fernando Pérez Fontán, Jorge Pereda González, Maria José Sedes Ferreiro, M. Ángeles Vázquez-Castro, Sergio Buanomo, and Jose Pedro V. Poires Baptista. Complex envelope three-state Markov model based simulator for the narrow-band LMS channel. *Int. J. Satellite Commun.*, 15(1):1–15, January 1997.

[FJK+02] Christof Faller, Biing-Hwang Juang, Peter Kroon, Hui-Ling Lou, Sean A. Ramprashad, and Carl-Erik W. Sundberg. Technical advances in digital audio radio broadcasting. *Proc. IEEE*, 90(8):1303–1333, August 2002.

[FKKK01] S. Fischer, S. Kudras, V. Kühn, and K.D. Kammeyer. Analysis of diversity effects for satellite communication systems. In *Proceedings IEEE Global Telecommunications Conference*, pages 2759–2763, 2001.

[Fos96] Gerard J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Technical Journal*, 1996.

[FVCC+01] Fernando Pérez Fontán, Maryan Vázquez-Castro, Cristina Enjamio Cabado, Jorge Pita García, and Erwin Kubista. Statistical modelling of the LMS channel. *IEEE Trans. Veh. Technol.*, 50(6):1549–1567, November 2001.

[Gal62] R. G. Gallager. Low-density parity-check codes. *IRE Trans. Inform. Theory*, pages 21–28, January 1962.

[GFVW99] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky. Detection algorithm and initial laboratory results using V-BLAST space-time communications architecture. *Electron. Lett.*, 35(1):14–16, January 1999.

[GG94] Riccardo De Gaudenzi and Filippo Gianetti. Analysis of an advanced satellite digital audio broadcasting system and complementary terrestrial gap-filler single frequency network. *IEEE Trans. Veh. Technol.*, 43(2):194–210, May 1994.

[GJJV03] Andrea Goldsmith, Syed Ali Jafar, Nihar Jindal, and Sriram Vishwanath. Capacity limits of MIMO channels. *IEEE J. Select. Areas Commun.*, 21(5):684–701, June 2003.

[Gra81] Alexander Graham. *Kronecker Products and Matrix Calculus: with Applications*. Ellis Horwood Limited, Chichester, England, Chichester, England, 1981.

[Gup65] R. P. Gupta. Asymptotic theory for principal component analysis in the complex case. *J. Ind. Stat. Assoc.*, 3(283):97–106, 1965.

[HH] Dirk Hughes-Hartogs. Ensemble modem structure for imperfect transmission media. United States Patents, 4,679,227 (July 1987), 4,731,816 (March 1998), 4,833,706 (May 1998).

[HM00] Martin Haardt and Werner Mohr. The complete solution for third-generation wireless communications: Two modes on air, one winning strategy. *IEEE Personal Commun. Mag.*, 7(6):18–24, December 2000.

[HtB03] Bertrand M. Hochwald and Stephan ten Brink. Achieving near-capacity on a multiple-antenna channel. *IEEE Trans. Commun.*, 51(3):389–399, March 2003.

[HVG91] Yoshihiro Hase, Wolfhard J. Vogel, and Julius Goldhirsh. Fade durations derived from land-mobile-satellite measurements in Australia. *IEEE Trans. Commun.*, 39(5):664–668, May 1991.

[ITU01] ITU. *Recommendation ITU-R BO.1130-4 (04/01), Systems for digital satellite broadcasting to vehicular, portable and fixed receivers in the bands allocated to BSS (sound) in the frequency range 1400-2700 MHz*, April 2001.

[ITU03] ITU. *Recommendation ITU-R P.681-6, Propagation data required for the design of earth-space land mobile telecommunication systems*, 2003.

[IUN03] Michel T. Ivrlač, Wolfgang Utschick, and Josef A. Nossek. Fading correlations in wireless MIMO communication systems. *IEEE J. Select. Areas Commun.*, 21(5):819–828, June 2003.

[JB04] Eduard A. Jorswieck and Holger Boche. Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback. *IEEE Trans. Wireless Commun.*, 3(5):1543–1553, September 2004.

[JG04] Syed Ali Jafar and Andrea Goldsmith. Transmitter optimization and optimality of beamforming for multiple antenna systems. *IEEE Trans. Wireless Commun.*, 3(4):1165–1175, July 2004.

[JMO03] Joakim Jaldén, Cristoff Martin, and Björn Ottersten. Semidefinite programming for detection in linear systems – optimality conditions and space-time decoding. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2003.

[JO05] Joakim Jaldén and Björn Ottersten. On the complexity of sphere decoding in digital communications. *IEEE Trans. Signal Processing*, 53(4):1474–1484, April 2005.

[JS88] Peter H. M. Janssen and Petre Stoica. On the expectation of the product of four matrix-valued Gaussian random variables. *IEEE Trans. Automat. Contr.*, 33(9):867–870, September 1988.

[JSO02] George Jöngren, Mikael Skoglund, and Björn Ottersten. Combining beamforming and orthogonal space-time block coding. *IEEE Trans. Inform. Theory*, 48:611–627, March 2002.

[JVG01] Syed Ali Jafar, Sriram Vishwanath, and Andrea Goldsmith. Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback. In *ICC*, volume 7, pages 2266–2270, June 2001.

[KKM97] Yoshio Karasawa, Kazuhiro Kimura, and Kenichi Minamisono. Analysis of availability improvement in LMSS

by means of satellite diversity based on three-state propagation channel model. *IEEE Trans. Veh. Technol.*, 46(4):1047–1056, November 1997.

[KuM04] Mobile Ku-band receiver demonstrator. ESA [online] Available: http://telecom.esa.int, 2004.

[Lay01] David H. Layer. Digital radio takes to the road. *IEEE Spectr.*, July 2001.

[LB98] Chun Loo and John S. Butterworth. Land mobile satellite channel, measurements and modeling. *Proc. IEEE*, 86(7):1442, 1998.

[LCD+91] Erich Lutz, Daniel Cygan, Michael Dippold, Frank Dolainsky, and Wolfgang Papke. The land mobile satellite communication channel – recording, statistics, and channel model. *IEEE Trans. Veh. Technol.*, (2):375–386, May 1991.

[Lee04] Seorim Lee. Satellite DMB in Korea. Invited Speaker, ASMS/EMPS 2004, [Online] Available: http://telecom.esa.int, September 2004.

[LGW00] Hui-Ling Lou, M. Julia Fernández-Getino García, and Vijitha Weerackody. FEC scheme for a TDM-OFDM based satellite radio broadcasting system. *IEEE Trans. Broadcast.*, 46(1):60–67, March 2000.

[LK02] Sergey Loyka and Ammar Kouki. New compound upper bound on MIMO channel capacity. *IEEE Commun. Lett.*, 6(3):96–98, March 2002.

[Loy01] Sergey L. Loyka. Channel capacity of MIMO architecture using the exponential correlation matrix. *IEEE Commun. Lett.*, 5(9):369–371, September 2001.

[Lut96] Erich Lutz. A Markov model for correlated land mobile satellite channels. *Int. J. Satellite Commun.*, 14(4):333–339, July 1996.

[Mac04] David MacKay. "David Mackay: Homepage". [Online]. Available: http://www.inference.phy.cam.ac.uk/mackay/, 2004.

[Mar89] Bev Marks. RDS – the way ahead for radio. *IEE Review*, pages 127–129, April 1989.

[Mar02] Cristoff Martin. MIMO channel characterization and joint channel estimation and detection in digital communications. Technical report, Licentiate Thesis, KTH, September 2002.

[Mar03] Bev Marks. TPEG – what is it all about? *EBU Diffusion*, November 2003.

[MBB+00] Aris L. Moustakas, Harold U. Baranger, Leon Balents, Anirvan M. Sengupta, and Steven H. Simon. Communication through a diffuse medium: Coherence and capacity. *Science*, 287:287–290, January 2000.

[MBC05] Mobile Broadcasting Corporation, Tokyo, Japan. [Online] Available: www.mbco.co.jp, 2005.

[MBD89] Mordechai Mushkin and Israel Bar-David. Capacity and coding for the Gilbert-Elliot channels. *IEEE Trans. Inform. Theory*, 35(6):1277–1290, November 1989.

[MBF00] Darren P. McNamara, Mark A. Beach, and P.N. Fletcher. Initial characterisation of multiple-input multiple-output (MIMO) channels for space-time communication. In *Vehicular Technology Conference*, volume 3, pages 1193–1197, Fall 2000.

[MBO04a] Cristoff Martin, Svante Bergman, and Björn Ottersten. Simple spatial multiplexing based on imperfect channel estimates. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2004.

[MBO04b] Cristoff Martin, Svante Bergman, and Björn Ottersten. Spatial loading based on channel covariance feedback and channel estimates. In *Proceedings European Signal Processing Conference*, September 2004.

[MGO04a] Cristoff Martin, Alexander Geurtz, and Björn Ottersten. File based mobile satellite broadcast systems: Error rate computation and QoS based design. In *Vehicular Technology Conference*, September 2004.

[MGO04b] Cristoff Martin, Alexander Geurtz, and Björn Ottersten. Packet coded mobile satellite broadcast systems: Error rate computations and quality of service based design. In *European Workshop on Mobile/Personal Satcoms and Advanced Satellite Mobile Systems Conference*, pages 37–44, September 2004.

[MGO04c] Cristoff Martin, Alexander Geurtz, and Björn Ottersten. Spectrally efficient mobile satellite real-time broadcast with transmit diversity. In *Vehicular Technology Conference*, September 2004.

[MGO05] Cristoff Martin, Alexander Geurtz, and Björn Ottersten. Statistical analysis and optimal design of mobile satellite broadcast with diversity. *IEEE Trans. Veh. Technol.*, 2005. Submitted to.

[Mic02] Richard A Michalski. Method for calculating link margin in a satellite system employing signal diversity. In *International Communications Satellite Systems Conference*. AIAA, May 2002.

[MN97] David J. C. MacKay and R. M. Neal. Near Shannon limit performance of low density parity check codes. *Electron. Lett.*, 33(6):457–458, March 1997.

[MN02] Richard A. Michalski and Duy Nguyen. A method for jointly optimizing two antennas in a diversity satellite system. In *International Communications Satellite Systems Conference*. AIAA, May 2002.

[MO00a] Cristoff Martin and Björn Ottersten. Joint channel estimation and detection for interference cancellation in multichannel systems. In *Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing*, 2000.

[MO00b] Cristoff Martin and Björn Ottersten. On robustness against burst unsynchronized co-channel interference in semi-blind detection. In *Asilomar Conference on Signals, Systems and Computers*, 2000.

[MO02] Cristoff Martin and Björn Ottersten. Analytic approximations of eigenvalue moments and mean channel capacity for

MIMO channels. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages III–2389 – III–2392, May 2002.

[MO03] Cristoff Martin and Björn Ottersten. Approximate transmit covariance optimization of MIMO systems with covariance feedback. In *Asilomar Conference on Signals, Systems and Computers*, 2003.

[MO04] Cristoff Martin and Björn Ottersten. Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading. *IEEE Trans. Wireless Commun.*, 3(4):1350–1359, July 2004.

[MOB05] MOBAHO!, MBCO content web site. [Online] Available: www.mobaho.com, 2005.

[MSS99] Jun Mitsugi, Mutsumu Serizawa, and Nobuyasa Sato. S-band digital mobile satellite broadcast system. In *Vehicular Technology Conference*, pages 2755–2759, 1999.

[MSS03] Aris L. Moustakas, Steven H. Simon, and Anirvan M. Sengupta. MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis. *IEEE Trans. Inform. Theory*, 49(10):2545–2561, October 2003.

[MTHK99] Ryu Miura, Toyohisa Tanaka, Akio Horie, and Yoshio Karasawa. A DBF self-beam steering array antenna for mobile satellite applications using beam-space maximal-ratio combining. *IEEE Trans. Veh. Technol.*, 48(3):665–675, May 1999.

[Mül02a] Ralf R. Müller. On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels. *IEEE Trans. Inform. Theory*, pages 2086–2091, July 2002.

[Mül02b] Ralf R. Müller. A random matrix model of communication via antenna arrays. *IEEE Trans. Inform. Theory*, pages 2495–2506, September 2002.

[Nar03] Ravi Narasimhan. Spatial multiplexing with transmit antenna and constellation selection for correlated MIMO fading channels. *IEEE Trans. Signal Processing*, 51(11):2829–2838, November 2003.

[Nat88] Jon E. Natvig. Evaluation of six medium bit-rate coders for the pan-european digital mobile radio system. *IEEE J. Select. Areas Commun.*, 6(2):324–331, February 1988.

[PB99] Cecilio Pimitel and Ian F. Blake. Enumeration of Markov chains and burst error statistics for finite state channel models. *IEEE Trans. Veh. Technol.*, 48:415–428, March 1999.

[PGNB04] Arogyaswami J. Paulraj, Dhananjay A. Gore, Rohit U. Nabar, and Helmut Bölcskei. An overview of MIMO communications — A key to gigabit wireless. *Proc. IEEE*, 92(2):198–218, February 2004.

[PK94] Arogyaswami J. Paulraj and Thomas Kailath. Increasing capacity in wireless broadcast systems using distributed transmission/directional reception. United States Patent, 5,345,599, September 1994.

[PN98] Arogyaswami J. Paulraj and Boon Chong Ng. Space-time modems for wireless personal communications. *IEEE Personal Commun. Mag.*, pages 36–48, February 1998.

[Por94] Boaz Porat. *Digital Processing of Random Signals*. Prentice-Hall Inc., 1994.

[PP97] Arogyaswami J. Paulraj and Constantinos B. Papadias. Space-time processing for wireless communications. *IEEE Signal Processing Mag.*, November 1997.

[Pro00] Theodor A. Prosch. Digital audio broadcasting (DAB) over non-stationary satellite systems. *Int. J. Satellite Commun.*, 18(6):433–456, November 2000.

[Pro01] John G. Proakis. *Digital Communications*. McGraw-Hill, 2001.

[Rap95] Juha Rapeli. UMTS: Targets, system concept and standardization in a global framework. *IEEE Personal Commun. Mag.*, 2(1):20–28, February 1995.

[RC98] Gregory G. Raleigh and John M. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Trans. Inform. Theory*, 46(3):357–366, March 1998.

[REKH02] I. Rulands, H. Ernst, J. Kunisch, and G. Harles. Feasibility study of a mobile Ku-band terminal: Final report. Technical report, ESA, 2002. [Online]. Available: http://www.telecom.esa.int.

[RZ93] Richard Rebhan and Jens Zander. On the outage probability in single frequency networks for digital broadcasting. *IEEE Trans. Broadcast.*, 39(4):395–401, December 1993.

[SFGK00] Da-Shan Shiu, Gerard J. Foschini, Michael J. Gans, and Joseph M. Kahn. Fading correlation and its effect on the capacity of multielement antenna systems. *IEEE Trans. Commun.*, 48(3):502–512, March 2000.

[SGP02] Mikael Skoglund, Jochen Giese, and Stefan Parkvall. Code design for combined channel estimation and error protection. *IEEE Trans. Inform. Theory*, 48(5):1162–1171, May 2002.

[Sha48] Claude Elwood Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423, 623–656, July, October 1948.

[Sir05a] Sirius Satellite Radio Inc. New York, USA. [Online] Available: www.siriusradio.com, 2005.

[Sir05b] Sirius Satellite Radio. Sirius satellite radio ends year with more than 1.1 million subscribers. [Online] Available: http://www.sirius.com, January 2005. Press Release.

[SKE+02] Sandro Scalise, Jürgen Kunisch, Harald Ernst, Jasper Siemons, Guy Harles, and Jürgen Hörle. Measurement campaign for the land mobile satellite channel in Ku-band. In *European Workshop on Mobile/Personal Satcoms*, September 2002.

[SKO00] Rickard Stridh, Peter Karlsson, and Björn Ottersten. MIMO channel capacity on a measured indoor radio channel at 5.8 GHz. In *Asilomar Conference on Signals, Systems and Computers*, pages 733–737, October 2000.

[SM03] Steven H. Simon and Aris L. Moustakas. Optimizing MIMO antenna systems with channel covariance feedback. *IEEE J. Select. Areas Commun.*, 21(3):406–417, April 2003.

[SS97] Petre Stoica and Torsten Söderström. Eigenelement statistics of sample covariance matrix in the correlated data case. *Digital Signal Processing*, 7(2):136–143, April 1997.

[SS00] Quentin H. Spencer and A. Lee Swindlehurst. Some results on channel capacity when using multiple antennas. In *Vehicular Technology Conference*, pages 681–688, 2000.

[SS01] Igal Sason and Shlomo Shamai (Shitz). On improved bounds on the decoding error probability of block codes over interleaved fading channels, with applications to Turbo-like codes. *IEEE Trans. Inform. Theory*, 47(6):2275–2299, September 2001.

[SS03] O. Simeone and U. Spagnolini. Combined linear pre-equalization and BLAST equalization with channel correlation feedback. *IEEE Commun. Lett.*, 7(10):487–489, October 2003.

[STE01] S.R. Saunders, C. Tzaras, and B.G. Evans. Physical-statistical methods for determining state transition probabilities in mobile-satellite channel models. *Int. J. Satellite Commun.*, 19(6):207–222, November 2001.

[SV04] M. W. Shelley and J. Vazquez. Low profile scanning antennas for satcom "on-the-move". In *European Workshop on Mobile/Personal Satcoms and Advanced Satellite Mobile Systems Conference*. ESA, 2004.

[Tel99] Emre Telatar. Capacity of multi-antenna Gaussian channels. *Europ. Trans. on Telecom.*, 10(6):585–595, November 1999.

[TSC98] Vahid Tarokh, Nambi Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communication: Performance criterion and code construction. *IEEE Trans. Inform. Theory*, 44:744–765, March 1998.

[TUM05] TU Media Corporation, South Korea [Online] Available: www.tu4u.com, 2005.

[TVP96] Shilpa Talwar, Mats Viberg, and Arogyaswami Paulraj. Blind separation of synchronous co-channel digital signals using an antenna array – part I: Algorithms. *IEEE Trans. Signal Processing*, 44(5):1184–1197, May 1996.

[VB99] E. Viterbo and J. Boutros. A universal lattice decoder for fading channels. *IEEE Trans. Inform. Theory*, 45(5):1639–1642, July 1999.

[VCFS02] Maryan Vázquez-Castro, Fernando Pérez Fontán, and S.R. Saunders. Shadowing correlation assessment and modeling for satellite diversity in urban environments. *Int. J. Satellite Commun.*, 20(2):151–166, March 2002.

[VD92] Branka Vucetic and Jun Du. Channel modeling and simulation in satellite mobile communication systems. *IEEE J. Select. Areas Commun.*, 10(8):1209–1218, October 1992.

[vdVTP97] Alle-Jan van der Veen, Shilpa Talwar, and Arogyaswami Paulraj. A subspace approach to blind space-time signal processing for wireless communication systems. *IEEE Trans. Signal Processing*, 45(1):173–190, January 1997.

[VM01] Eugene Visotsky and Upamanyu Madhow. Space-time transmit precoding with imperfect feedback. *IEEE Trans. Inform. Theory*, 47(6):2632–2639, September 2001.

[VP04] Jani Väre and Matti Puputti. Soft handover in terrestrial broadcast networks. In *IEEE International Conference on Mobile Data Management (MDM'04)*, pages 236–242, 2004.

[WE71] S. B. Weinstein and Paul M. Ebert. Data transmission by frequency-division multiplexing using the discrete Fourier transform. *IEEE Trans. Commun. Technol.*, COM-19(5):628–634, October 1971.

[Win87] Jack H. Winters. On the capacity of radio communication systems with diversity in a Rayleigh fading environment. *IEEE J. Select. Areas Commun.*, SAC-5(5):871–878, June 1987.

[WLS99] Vijitha Weerackody, Hui-Ling Lou, and Zulfiquar Sayeed. A code division multiplexing scheme for satellite digital audio

broadcasting. *IEEE J. Select. Areas Commun.*, 17(11):1985–1999, November 1999.

[WM99] Leif Wilhelmsson and Laurence B. Milstein. On the effect of imperfect interleaving for the Gilbert–Elliot channel. *IEEE Trans. Commun.*, 47(5), May 1999.

[WV01] C. S. Withers and R. Vaughan. The distribution and percentiles of channel capacity for multiple arrays. In *Proceedings of the 11th Virginia Tech/MPRG Symposium on Wireless personal Communications*, pages 141–152, 2001.

[WW04] James Wang and Jack H. Winters. An embedded antenna for mobile DBS. In *Vehicular Technology Conference*, September 2004.

[XM 05] XM Satellite Radio. XM satellite radio surpasses 3.2 million subscribers. [Online] Available: http://www.xm-radio.com, January 2005. Press Release.

[XMR05] XM Satellite Radio Inc. Washington, DC, USA. [Online] Available: www.xmradio.com, 2005.

[XZG04] Pengfei Xia, Shengli Zhou, and Georgios B. Giannakis. Adaptive MIMO-OFDM based on partial channel state information. *IEEE Trans. Signal Processing*, 52(1):202–213, January 2004.

[YBO+01] Kai Yu, Mats Bengtsson, Björn Ottersten, Peter Karlsson, and Mark Beach. Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements. In *Proceedings IEEE Global Telecommunications Conference*, pages 156–160, 2001.

[YBOB02] Kai Yu, Mats Bengtsson, Björn Ottersten, and Mark Beach. Narrowband MIMO channel modeling for LOS indoor scenarois. In *XXVIIth Triennial General Assembly of International Union of Radio Science*. URSI, August 2002.

[YW95] James R. Yee and Edward J. Weldon. Evaluation of the performance of error-correcting codes on a Gilbert channel. *IEEE Trans. Commun.*, 43(8):2316–2323, August 1995.

[ZO03] Xi Zhang and Björn Ottersten. Power allocation and bit loading for spatial multiplexing in MIMO systems. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2003.