Universidade de Vigo
Signal Theory and Communications Department

UNIVERSIDADE
DE VIGO

DOCTORAL DISSERTATION
INTERNATIONAL MENTION

Cognitive and Signal Processing Techniques for Improved Spectrum Exploitation in Wireless Communications

Author:
Alberto Rico Alvariño

Directed by:
Carlos Mosquera Nartallo

2014

DOCTORAL DISSERTATION
INTERNATIONAL MENTION

Cognitive and Signal Processing Techniques for Improved Spectrum Exploitation in Wireless Communications

Author:
Alberto Rico Alvarino

Directed by:
Carlos Mosquera Nartallo

EXAMINATION COMMITTEE

Board members: Ángel Lozano (President)

Roberto López-Valcarce (Secretary)

Laura Cottatellucci

Thesis defense: Vigo, June 30, 2014

This edition: May 4, 2014

Agradecementos

O desenvolvemento desta tese de doutoramento non tería sido posible sen a axuda directa e indirecta dun gran número de persoas da miña contorna. O obxectivo desta breve sección é darlles o recoñecemento que merecen, mais a miña memoria de peixe e o sprint final na redacción desta tese poden causar que me esqueza dalgunha delas. Tentarei que non sexa así.

En primeiro lugar, o meu agradecemento ao meu director de tese, Carlos Mosquera, que me leva apoiando e guiando dende o meu primeiro día no grupo de investigación, e especialmente nos momentos nos que as cousas non saían coma un querería (léase rexeitamento masivo de artigos, especialmente ó comezo da tese). O meu seguinte agradecemento vai dirixido a Nuria González, que foi a persoa que me convenceu de unirme ao grupo de investigación e embarcarme na tarefa de realizar unha tese de doutoramento. Agradezo tamén as discusións con outros profesores do grupo de procesado de sinal para comunicacións durante os seminarios quasi-semanais que tivemos durante estes anos: Roberto López-Valcarce, Pedro Comesaña, e especialmente Fernando Pérez-González, que contribuíu de xeito máis activo á investigación desenvolta nesta tese.

Agradezo á Fundación Pedro Barrié de la Maza o soporte económico, loxístico e moral que me permitiu realizar a miña estancia na University of Texas at Austin, durante a cal desenvolvín unha parte substancial desta tese de doutoramento, e que me permitiu traballar nun dos mellores grupos de investigación en comunicacións sen fíos.

I am extremely grateful to Prof. Robert W. Heath Jr. for giving me the opportunity to join his group, and for guiding my research during my stay at UT Austin. It was a real pleasure to work with an amazing person like Robert, who definitely helped me develop new ideas present in this thesis.

Non podo esquecerme de agradecerlle ós meus compañeiros de laboratorio o bo ambiente creado durante todos estes anos: Alberto, David, Massimo, Jorge, Dani, Bamrung, Roi, e especialmente Suso, co que colaborei máis estreitamente na miña investigación.

O meu seguinte agradecemento (¿ou debería dicir desculpa?) é para a miña familia, que sempre me apoiou nos meus (máis ou menos frecuentes) momentos de desesperación e pesimismo

ante diferentes dificultades. Unha mención especial merece a miña moza Marta, que demostrou ter unha paciencia case infinita ó aturarme a diario durante a nosa estadía en Austin e San Diego, e neste último ano en Vigo.

Abstract

The increasing demand of wireless data makes the electromagnetic spectrum a scarce and valuable resource. Wireless networks should exploit its available spectrum in the most efficient way to maximize the amount of data they can support. Alternatives to increase the spectral efficiency of wireless networks include the design of new spectrum access paradigms and the operation of existing wireless links in the most efficient way. In this thesis we present two different approaches for increasing the spectral efficiency of wireless networks.

The first approach is based on a cognitive-inspired spectrum access paradigm in which a secondary user is allowed to access the frequency bands owned by a primary user. This cognitive access has to be properly designed to avoid any possible degradation on the primary service. We advocate the use of the so-called overlay cognitive radio paradigm, in which primary and secondary users are allowed to cooperate. More precisely, we focus on the cases where the secondary transmitter knows the primary message, and spectrum access is allowed as long as the primary user's service is not affected. We analyze overlay access over both point to point and broadcast primary networks, and incorporate some practical impairments into the problem.

The second part of the thesis is focused on the design of adaptive transmission strategies under changing channel conditions, in a process that is known as link adaptation. The time varying nature of fading channels causes changes in the maximum rate they can support and, therefore, a transmitter that wants to efficiently exploit the wireless medium should be able to adapt its transmission rate depending on the channel conditions. We first design adaptation techniques for mobile satellite communications, where the main challenge is the inaccuracy of channel state information at the transmitter due to the large propagation delay. We propose the use of adaptation techniques based on statistical knowledge of the channel, and show how to exploit some degree of channel reciprocity even in frequency division duplexing systems. We move afterwards to a different setting, and analyze the link adaptation problem in multiuser multiple input multiple output orthogonal frequency division multiplexing systems with limited feedback of the channel state. Adaptation is performed by combining machine learning techniques, greedy algorithms and interference estimation due to imperfect channel state information.

Publications

The following is a list of journal and conference publications that have been produced as a result of the work on this thesis.

Journal publications

1. **Alberto Rico-Alvariño** and Robert W. Heath Jr. *Learning-Based Adaptive Transmission for Limited Feedback Multiuser MIMO-OFDM.* IEEE Transactions on Wireless Communications, 2014 (to appear).

2. **Alberto Rico-Alvariño**, Carlos Mosquera, and Fernando Perez-Gonzalez. *Overlay Cognitive Transmission in a Multicarrier Broadcast Network with Dominant Line of Sight Reception.* IEEE Transactions on Wireless Communications, November 2012.

Conference publications

1. **Alberto Rico-Alvariño**, Jesus Arnau and Carlos Mosquera *Balancing Closed and Open Loop CSI in Mobile Satellite Link Adaptation.* Proc. ASMS/SPSC 2014 - Livorno, Italy, Sep 2014 *Submitted.*

2. **Alberto Rico-Alvariño**, Robert W. Heath Jr. and Carlos Mosquera *FER Prediction with Variable Codeword Length.* Proc. ICASSP, Florence, Italy, May 2014.

3. **Alberto Rico-Alvariño**, Jesus Arnau and Carlos Mosquera *Statistical Cross Layer Adaptation in Fast Fading Mobile Satellite Channels.* Proc. GLOBECOM - Symposium on Selected Areas in Communications, Atlanta, GA, Dec 2013.

4. **Alberto Rico-Alvariño** and Robert W. Heath Jr. *Link Adaptation in MIMO-OFDM with Practical Impairments.* Proc. Asilomar Conference on Signals, Systems, and Com-

puters, Pacific Grove, CA, USA, Nov 2013.

5. **Alberto Rico-Alvariño** and Robert W. Heath Jr. *Learning Based Link Adaptation in Multiuser MIMO-OFDM.* Proc. EUSIPCO 2013, Marrakech, Morocco, Sep 2013.

6. **Alberto Rico-Alvariño** and Carlos Mosquera. *Analytical Characterization of the Single Frequency Network Gain Using Effective SNR Metrics.* Proc. ISWCS 2013, Ilmenau, Germany, Aug 2013.

7. **Alberto Rico-Alvariño** and Carlos Mosquera. *Optimum Training for CSI Acquisition in Cognitive Radio Channels.* Proc. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, Nov 2012.

8. **Alberto Rico-Alvariño** and Carlos Mosquera. *Overlay Spectrum Reuse in a Broadcast Network: Covering the Whole Grayscale of Spaces.* Proc. IEEE DySpan, Bellevue, Washington, USA, Oct 2012.

9. Jesus Arnau, **Alberto Rico-Alvariño**, and Carlos Mosquera. *Adaptive Transmission Techniques for Mobile Satellite Links.* Proc. ICSSC 2012, Ottawa, Canada, Sep 2012. **Best Professional Paper Award**

10. **Alberto Rico-Alvariño** and Carlos Mosquera. *On the Effect of Echoes in Hybrid Terrestrial-Satellite Single Frequency Networks: Analysis and Countermeasures.* Proc. ASMS/SPSC 2012, Baiona, Spain, Sep 2012.

11. **Alberto Rico-Alvariño**, Carlos Mosquera, and Fernando Pérez-González. *Overlay Spectrum Reuse in a Multicarrier Broadcast Network: Single Receiver Analysis.* Proc. IEEE SPAWC, Cesme, Turkey, Jun 2012.

12. **Alberto Rico-Alvariño**, Carlos Mosquera, and Fernando Pérez-González. *Overlay Spectrum Reuse in a Multicarrier Broadcast Network: Coverage Analysis.* Proc. IEEE SPAWC, Cesme, Turkey, June 2012.

13. **Alberto Rico-Alvariño** and Carlos Mosquera. *Overlay Cognitive Transmission in OFDM Point to Point Systems Exploiting Primary Feedback.* Proc. CIP, Baiona, Spain, May 2012.

14. **Alberto Rico-Alvariño**, Carlos Mosquera, and Fernando Pérez-González., *On the Coexistence of Primary and Secondary Transmitters in a Broadcast Network.* Proc. CogART, Barcelona, Spain, Oct 2011.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Contents

1.1 Motivation

In the last years there has been an increasing demand of mobile data, and it is forecast to keep growing in the next years. According to [1], there has been an increase of around 81% in traffic during 2013. To accommodate this growth, modern wireless standards have been deployed to increase the spectral efficiency of wireless networks. During the development of this thesis, there have been transitions from UMTS to LTE in cellular, from IEEE 802.11n to IEEE 802.11ac in wireless local area networks (WLAN), from DVB-T (and even analog television) to DVB-T2 in terrestrial television broadcasting, and from DVB-S to DVB-S2/DVB-RCS in satellite broadcasting. On top of the advances in modulation and coding that allowed an increase of

the spectral efficiency, these standards are also characterized by the addition of novel network structures, such as single frequency network (SFN) operation in LTE and DVB-T/DVB-T2, small cells in LTE, multiuser multiple input multiple output (MIMO) in IEEE 802.11ac and support of mobile receivers in DVB-S2.

These improvements in spectrum exploitation, however, do not seem enough to be able to provide the ten-fold increase in traffic that is forecast between 2014 and 2017 [1], thus envisioning the need for research in the direction of more efficient physical layer techniques and network architectures. Also, some of the recently developed technologies have still room for improvement in non-standard procedures, such as scheduling or adaptive modulation and coding. Research efforts should be made in both directions to provide additional improvements in spectral efficiency.

In this thesis we explore both directions and propose different techniques to improve spectrum exploitation. In the first part we propose spectrum access techniques based on the cognitive radio paradigm to allow secondary access to parts of spectrum allocated to a primary user. This access has to be carefully planned to be able to use the wireless media without the disruption of the primary service. In the second part we design link adaptation algorithms for two scenarios where the use of traditional techniques may result in suboptimal spectrum exploitation. The two scenarios are the mobile satellite channel, where the large round-trip time delay causes channel state information (CSI) to be inaccurate, and multiuser MIMO orthogonal frequency division multiplexing (OFDM) with limited feedback, where the inter-user interference has to be carefully taken into account in the adaptation method.

1.2 Cognitive Radio

1.2.1 Introduction

A cognitive radio is an intelligent radio device that is capable of adapting its behavior depending on its environment. Since the introduction of this idea in [2], there has been an increased interest in exploring the potential benefits of this kind of devices. Possibly the best known example of cognitive radio is the dynamic spectrum access to underutilized frequency bands (known as *white spaces*), mostly used for terrestrial television broadcasting. The television broadcaster, known as *primary user* in cognitive radio jargon, must not suffer any service degradation due to the operation of the cognitive or *secondary user*. Thus, the secondary user must identify the unused frequency bands with high reliability in order to prevent interruptions in the primary service. This approach to dynamic spectrum access has been a major topic of research during

the last decade, and some countries have already decided to allow this kind of access. For example, the Federal Communications Commission (FCC) in the United States of America has already approved dynamic spectrum access by means of a combination of spectrum sensing and a geolocation database that contains the unused spectrum bands [3]. In Europe, the Office of Communications (OFCOM) in the United Kingdom is launching a pilot for white space spectrum access [4].

The concept of cognitive radio was recently extended to more sophisticated spectrum access techniques where primary and secondary users can interact in a variety of manners. These techniques can be classified in three groups, according to [5]:

- **Interweave cognitive radio** This dynamic spectrum access paradigm covers the classic white spaces approach. The cognitive user has to obtain information about the occupancy of the different channels by means of a geolocation database [6] or by spectrum sensing [7]. Under this setting, the secondary user does not transmit in the same channels as the primary user, but detects the unused frequency bands to use them.

- **Underlay cognitive radio** The underlay cognitive radio paradigm allows secondary transmissions in the same time-frequency resources as the primary user, thus increasing the interference temperature suffered by the primary receivers. The transmit power of the secondary user is usually constrained by the maximum tolerable interference at the primary receivers [8,9]. Other approaches to underlay transmission include the concentration of the interference in a small part of the spectrum, so it can be filtered out by the primary receiver [10], or the transmission in some spatial dimensions such that a multiple antenna primary receiver is able to remove the interference [11].

- **Overlay cognitive radio** Systems based on overlay cognitive radios allow some kind of cooperation between the primary and secondary transmitters. In this setting, it is usually assumed that the secondary transmitter acts as a relay of the primary signal, or that the secondary transmitter knows the primary signal in a non-causal way. The primary and secondary transmitters use the same time and frequency resources, so there is interference between them. The knowledge of the primary signal can be used to alleviate the interference in two different forms: first, the primary signal can be transmitted by the secondary user to increase the signal to interference and noise ratio (SINR) of the primary link; second, interference cancellation techniques can be performed at the secondary transmitter to increase the quality of the secondary link.

In this thesis we focus on the overlay cognitive radio paradigm. Based on information-theoretical results on the so-called cognitive interference channel, we propose and analyze the

application of the overlay cognitive radio paradigm to practical scenarios. We consider overlay transmission over broadcast[1] and point to point primary users, and we take into account some effects that appear in practical scenarios. We focus on the issues created by imperfect (or none) CSI at the secondary transmitter. This lack of CSI will affect both the transmission of the primary signal as well as the possibility of performing interference cancellation.

1.2.2 Prior work

The information theoretical foundation of overlay cognitive radio is on the so-called *cognitive channel*. The cognitive channel is a two user interference channel where the secondary transmitter knows (causally or non-causally) the message to be transmitted by the primary one. A diagram depicting this kind of channel is shown in Figure 1.1.

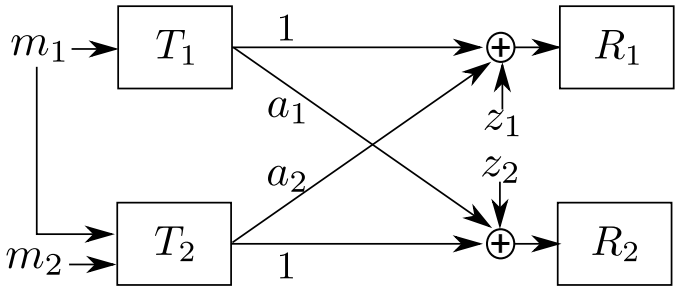


Figure 1.1: Cognitive radio channel: primary and secondary transmitters T_1 and T_2 communicate messages m_1 and m_2 to receivers R_1 and R_2, respectively. The received signals are the output of an interference channel with noise z_1 and z_2 and interfering gains a_1 and a_2. The secondary transmitter knows the primary message.

The capacity region [12,13] of this channel is achieved by transmitting the primary message to increase the capacity of the primary link, as well as by using Costa precoding [14] (also known as dirty paper coding) to remove the interference at the secondary receiver. Some practical conditions for the cognitive channel were introduced in [15], where the operation of the secondary transmitter is constrained so the primary user is not affected by its presence. These conditions are, in short,

1. The rate of the primary system cannot be affected by the operation of the secondary user.
2. The primary receiver uses a single-user decoder, as in the case where the secondary transmitter is not present. For example, decoding the secondary message and subtracting it from the received signal is not allowed.

[1]Throughout the thesis, the term *broadcast* system will denote a communication scenario where multiple receivers are decoding the same message, unlike the information-theoretical *broadcast channel*.

We follow these two principles in the designs described in this thesis.

One of the main criticisms to overlay cognitive radio is the feasibility of knowing the primary message before its transmission. Although it might seam unfeasible in practice, there are several scenarios where this knowledge can be possible:

- In some cases, the secondary transmitter does not know the primary message in a non-causal way, but acts as a relay of it instead. The presence of a relay can boost the achievable rate of the primary system, and use the capacity gain to transmit its own message [16, 17].
- In a broadcast network organized as an SFN, multiple base stations transmit the same data in a synchronized way. In this setting, we could think of a secondary transmitter as an additional SFN transmitter conveying its own information on top of the primary one [18].
- In point to point systems with automatic repeat request (ARQ), a message is retransmitted if the first transmission is incorrectly decoded. In this case, a potential secondary user might have decoded the message in the first transmission. When a retransmission occurs, the secondary user knows the primary message [5].
- In a more general setting, a system can be designed so that the primary transmitter shares its message with the secondary transmitter through a backbone connection. Although in a different setting, this idea can be seen to be similar to coordinated multipoint (COMP) [19].

Imperfect CSI in the cognitive radio channel introduces serious drawbacks in the system design. We could think of modifying the primary receiver to estimate the direct and interfering channels, and feed them back to the secondary transmitter. This approach, however, would require modifying the primary receiver, which is undesirable. Let us focus on the diagram of Figure 1.1. Imperfect CSI has the following effect on cognitive radio channels:

- If the secondary transmitter does not transmit the primary message, knowledge on the magnitude of a_2 is needed to determine the amount of interference the primary receiver is receiving. For example, small values of a_2 would allow a higher transmit power at the secondary transmitter.
- If the secondary transmitter transmits the primary message, knowledge on the phase of a_2 is needed to align the primary and secondary contributions of the primary signal so that the interference is constructive. For example, if $a_2 = -1$ and the secondary transmitter transmits the primary message with the same power as the primary transmitter, then

the interference between the primary and secondary replicas will be destructive and the received signal to noise ratio (SNR) will be zero.

- Knowledge of a_1 is needed if the secondary transmitter uses dirty paper coding to remove the interference caused by the primary transmitter at the secondary receiver. Although this value could be fed back by the secondary receiver, in some cases this is not possible (e.g., if the secondary system is of broadcast nature).

This thesis is focused on analyzing and proposing solutions for the CSI acquisition problem. In some cases, where some reduced amount of feedback from the primary transmitter is allowed, we analyze and optimize this information interchange to maximize the efficiency of the secondary link. In the cases where the primary receiver is not capable of feeding back any CSI, we design transmission schemes that are able to cope with unknown channels.

1.2.3 Contributions

In this thesis we consider two different scenarios: in the first one, the primary system is a broadcast network, where users inside the coverage area receive the same message. In this setting, there is no possible feedback from the primary receivers; in the second one, the primary system is point to point, and a reduced amount of feedback is allowed. In both cases, we analyze the problem of imperfect CSI at the secondary transmitter. The contributions on this first part of the thesis can be summarized as follows:

- **C1.1) Performance analysis and design of overlay cognitive radio systems over an SFN with line of sight propagation** The overlay cognitive radio paradigm can be applied to an SFN, where the secondary transmitter can gain access to the primary message and synchronization signals. Unlike [18], we take into account the effect of transmitting the primary signal when the phase of the interfering channel is not known. We consider wideband transmission using OFDM, and model the channels as a pure line of sight component (one-tap channel). The joint transmission of the primary signal from both transmitters creates an artificial multipath, or frequency selectivity. We show that this artificial multipath cannot be neglected in most SNR regimes, and derive the optimum power allocation at the secondary transmitter to circumvent this problem. We assume that the secondary receiver is inside the primary coverage area, so it can decode the primary message and, in consequence, remove its interference. We consider the case of a single primary receiver, as well as the case of infinite primary receivers in the coverage area.

- **C1.2) Extension to non line of sight channels** We extend the previous results and analyze the case of channels with different degrees of multipath, and in different SNR regimes. We resort to the use of packet error rate prediction metrics, or effective SNR metrics, to perform this analysis. We conclude that the negative effect of artificial multipath appears in scenarios with high SNR and strong line of sight propagation. We also consider the case of transmitters using space time coding, so that the artificial multipath does not appear.

- **C1.3) Extension to secondary receivers outside the primary coverage area** A second extension is considered for secondary receivers located outside the primary coverage area. If the secondary receiver is near the coverage edge, it is going to receive a substantial amount of interference from the primary transmitter without being able to decode it. This is usually called a *gray space*, placed between the *black space*, or coverage area, and *white space*, where the effect of the primary transmitter is negligible. In gray spaces, the secondary transmitter can extend the primary coverage area by transmitting the primary message, so that the secondary receiver can decode the interference. We analyze optimum transmission strategies for black, gray and white spaces under the overlay cognitive radio paradigm, for unicast and broadcast secondary users.

- **C1.4) Optimum power allocation for OFDM-based primary user with SNR feedback from the primary receiver** If the primary system is unicast, then a purely cooperative secondary transmitter can adjust the phase of its transmission so that the received signals at the primary receiver are coherently added, thus increasing the capacity of the primary link. If the primary system is capable of performing link adaptation, then this increment in capacity will be translated into an increment in the transmission rate and, therefore, less transmit resources will be used to convey the same amount of data. The unused resources can be used by the secondary transmitter to convey its own information. We analyze a scenario where the primary system uses OFDM, so the secondary transmitter can perform power allocation among the subcarriers to maximize the rate of the primary link. We assume that the primary receiver performs periodic SNR measurements and feeds back this information to the primary transmitter. This information can be also received by the secondary transmitter, and be exploited to acquire CSI.

- **C1.5) Optimum transmit signals for CSI acquisition** If the primary receiver feeds back to the primary transmitter the magnitude and phase observed by the received signal, then this information can be exploited to obtain CSI at the secondary transmitter. A simple procedure was described in [15] to obtain CSI from the feedback information. We analyze this procedure under an estimation theoretic approach, and derive optimum

transmit signals that minimize the CSI estimation error for single input single output (SISO), MIMO and time-varying channels. We show that the original procedure of [15] is suboptimal in the mean squared error sense.

1.3 Link Adaptation

1.3.1 Introduction

Link adaptation is the process of dynamically changing the transmission parameters according to the channel state. Depending on the particular scenario, different adaptations might result of interest. For example, power control can be applied in fading channels to keep a constant bit rate, in what is known as *channel inversion* [20], user scheduling can be performed depending on the channel quality, etc.

One problem that is particularly interesting, and that is usually performed on modern wireless communication systems, is rate adaptation. Wireless channels are of fading nature and, therefore, the rate they can support changes over time. A communication system that wants to maximize its spectral efficiency should somehow track the channel changes and adapt its transmission rate accordingly. This rate selection is usually performed by the definition of a discrete set of available coding rates and modulations, or modulation and coding schemes (MCS). Rate adaptation is usually referred to as adaptive modulation and coding (AMC). Most modern communication systems, including cellular technologies (3GPP LTE [21], IEEE 802.16 [22]), wireless local area networks (IEEE 802.11 [23]) and satellite communication standards (DVB-S2 [24], DVB-RCS [25]) support different MCS and, therefore, enable the use of AMC.

Performing AMC requires the knowledge of some sort of channel state information (CSI) at the transmitter (CSIT). The transmitter can gain access to this information by means of feedback from the receiver, in what is known as *closed loop* CSI, or by exploiting the pilots present in the incoming signal if there is channel reciprocity, and estimating the channel in an *open loop* way. The former CSI acquisition technique is widely supported by modern communication standards, while the latter can only be exploited if the duplexing is performed in the time domain.

Once the transmitter has gained access to CSI, it is still not trivial how to select the most appropriate MCS. For example, a problem arises when CSIT is inaccurate as a result of estimation errors, or simply due to the delay induced by the wireless channel. In these cases, there is usually a tradeoff between throughput maximization and outage minimization [26]. Most modern communication standards include some sort of automatic repeat request (ARQ) mechanism that transforms the outage probability into an average number of retransmissions

or, equivalently, into an average delay. MCS selection is particularly difficult in mobile satellite communications, where the large round trip time increases the CSI delay and decreases the maximum allowable number of retransmissions.

MCS selection is also difficult in some scenarios where CSI is timely. For example, systems using MIMO-OFDM suffer from a large dimensionality of the CSI. Unlike single carrier systems, which are usually characterized by a single SNR value, multicarrier systems have one SNR value for each subcarrier and spatial stream. Reducing the dimensionality by averaging the SNR values is usually not a good approach, since average SNR is not a good quality metric in these systems. The difficulty of performing MCS selection is increased when CSI is imperfect due to limited feedback.

In this thesis we design MCS selection strategies for mobile satellite communications and MIMO-OFDM systems. Particularly, we focus on the forward and return links of mobile satellite systems with statistical and delayed CSI, and the downlink of an IEEE 802.11ac-like system with limited feedback CSI.

1.3.2 Prior work

Classic work on link adaptation focused on narrowband fading. For example, the modulation and power were dynamically adjusted with constraints on the uncoded bit error rate (BER) and average transmit power to maximize the spectral efficiency [27, 28]. The extension of these approaches to coded transmission resulted in more complicated analytical expressions, and usually required the use of BER approximations [29, 30]. The MCS selection problem in [27–30] is essentially a unidimensional problem that consists of assigning an SNR interval to each MCS. A different approach to link adaptation can be seen in works like [31, 32], where the transmit rate is modified without taking into account the current channel state, but only with ACK/NAK information.

Mobile satellite communications

Prior work on satellite communications focused on selecting the MCS based on thresholds on the estimated SNR value. In [33–35] thresholds for MCS selection were designed taking into account the CSI error caused by imperfect SNR estimation, but the effect of delay in a mobile environment was not analyzed. The obtained adaptation strategy is based on adaptation with hysteresis, i.e., the MCS to be used in the next transmission is selected depending on the estimated SNR value and the current MCS. In [36] it is proposed to use a backoff margin to account for the inaccuracy of the SNR estimate, including both delay and estimation error. This

backoff margin can be obtained online, from the ACK/NAK interchange, or can be obtained by simulations. In [37] it is proposed to use only a fraction of the available MCS to perform link adaptation with the objective of increasing the robustness of the adaptation.

In this thesis we follow different approaches to adaptation in forward and return links. In the forward link we exploit statistical information about the channel to maximize the throughput subject to a packet error rate constraint. This maximization is performed by suggesting different transmission strategies, namely multilayer coding and use of different MCS in each retransmission. In the return link we propose the use of open loop CSI to obtain timely information about the channel.

MIMO-OFDM

Link adaptation in systems with multiple channels is challenging due to the higher dimensionality of the CSI. The channel state cannot be characterized with a single SNR value in systems using OFDM or MIMO with spatial multiplexing. The reason is that different symbols in the same codeword experience different SNR values. Because of the complicated mapping between codeword error and symbol error, the average SNR may not contain enough information to permit effective adaptation [38]. An alternative is to map the set of SNR values (one for each carrier and spatial stream) to one effective SNR [39–43]. The effective SNR is defined as the SNR for an additive white Gaussian noise (AWGN) channel to experience the same frame error rate (FER) or packet error rate (PER) as the fading channel under study. Effective SNR metrics are defined as a Kolmogorov mean [44] of the SNR values with some parameters that are fitted according to empirical results. The WiMAX forum, for example, recommends the exponential effective SNR metric as the default method for FER prediction [45] and link adaptation in IEEE 802.16e. The effective SNR metrics lead to adaptation algorithms in the form of look-up tables, where each effective SNR value is associated to an MCS. Some works like [46] make use of effective SNR metrics to develop link adaptation algorithms in single user scenarios. Having a fixed mapping between effective SNR and MCS is not ideal due to the impact of practical impairments like non-linearities, non-Gaussian noise or implementation dependent parameters, like Viterbi truncation depth or number of rounds in a turbo decoder.

Data-driven approaches provide a solution to the problem of mapping an appropriate FER to the set of SNR values. Based on empirical observations of the SNR values and their associated FER, learning algorithms have been designed to select the proper MCS for each channel realization [47–52]. This classification task is performed by machine learning algorithms like K-nearest neighbors or support vector machines (SVM) [53]. These algorithms are usually described as non-parametric, since they do not assume any model that maps SNR values to FER, but try

to learn it from empirical data. In these non-parametric approaches there exists a tradeoff between adjusting the model to the training samples and producing a smooth function describing the model, which is known as the bias-variance tradeoff [54]. Learning can be performed online [50–52] or offline [47, 49], and is usually based on a low dimensional feature set containing a subset of the ordered SNR values [55]. This data-driven formulation was shown to outperform the effective SNR in [47], and is resiliant to practical impairments like non-Gaussian noise or amplifier non-linearities [51].

Prior work on learning-based link adaptation focused on single user scenarios [47–52]. The extension of these approaches to the more general case of having multiple users served at the same time by the use of space division multiplexing (SDM) is not trivial due to the interaction between user selection, mode selection, precoding, and MCS selection. For example, the link adaptation technique in [47] requires running the classification algorithm for all possible number of spatial streams (NSS) and selecting the NSS leading to a higher throughput. Applying a similar strategy in the multiuser case would require an additional exhaustive search over the choice of users and the number of spatial streams per user.

Prior work on learning-based link adaptation also did not consider the impact of limited feedback precoding [47–51]. In multiuser MIMO (MU-MIMO) systems, limited feedback creates quantization error that results in residual interuser interference [56]. The resulting interference makes performance a function of the feedback quality. Therefore, a smart multiuser link adaptation algorithm should predict the interuser interference and move from aggressive multiuser transmission to more conservative modes depending on the feedback quality [57].

Link adaptation may be performed taking as input more information than SNR values. For example, codeword length or noise distribution affect the performance of a given MCS. Previous work using FER prediction dealt with the variable codeword length in various ways. For example, [31, 46, 58, 59] assumed perfect knowledge of the coded bit error rate (CBER), and from that value they calculated the corresponding FER. Other previous work using FER predictors assumed constant frame length [30, 47, 48, 52, 60], which is not realistic under modern communication standards.

In this thesis we develop a learning-based link adaptation technique for multiuser MIMO-OFDM. Multiuser transmission is achieved by means of linear precoding from limited feedback information. We also consider the effect of codeword length and noise distribution in the adaptation algorithm.

1.3.3 Contributions

In the second part of the thesis we present link adaptation algorithms for mobile satellite communications and multiuser MIMO-OFDM. The contributions of this part of the thesis are summarized as follows:

- **C2.1) Link adaptation techniques for the forward link of mobile satellite channels using ARQ** CSI acquisition in the forward link of mobile satellite channels is rather inaccurate due to the large round trip delay. For a geostationary satellite, for example, this delay is around half a second. In many scenarios, and even for low speeds, the channel coherence time is lower than this value. On top of this problem, the channel variation is larger than the usual in terrestrial systems due to the sensitivity to blockages of the line of sight propagation path. Thus, it is possible to suffer very large SNR variations in a small period of time.

  In this thesis we propose to perform link adaptation based on statistical channel information, and try to exploit the diversity created by the large SNR variations. We consider systems that allow the use of retransmissions to incorporate time diversity. We present two different approaches that allow exploiting the good channel instants while incorporating an outage constraint on the system. The first one is based on the use of two-layer coding, and the second one on the use of different MCS for different retransmission index. The design of the two-layer coding approach is based on the knowledge of statistical information about the channel, which allows designing the power weighting of the two layers and the MCS selection. The second method is based on the use of different MCS in different retransmissions. This technique is also designed based on statistical information on the channel, but we present a method to estimate the necessary parameters from the exchange of ACK/NAK.

- **C2.2) Link adaptation for the return link of mobile satellite channels exploiting open loop and closed loop CSI** The return link of mobile satellite channels suffers from the same problem of CSI inaccuracy as the forward link if operating in closed loop mode. If adaptation is performed in open loop, large channel variations can be detected by observing the forward channel, as the objects that cause blockage are always placed next to the terminal. Fading due to multipath, however, is not detectable by observing the incoming singal if duplexing is performed on the frequency domain. Thus, in some cases the open loop CSI might outperform the closed loop one, and vice versa.

  In this thesis we propose a method to automatically combine both CSI values depending on the observed ACK/NAK values. The link adaptation algorithm is derived as the stochastic

gradient descent solution of an unconstrained optimization problem which sets the observed FER to a desired target value.

- **C2.3) Link adaptation for multiuser MIMO-OFDM with limited feedback** Performing link adaptation in a multiantenna and multiuser system is challenging because of the coupling between precoding, user selection, spatial mode selection and use of limited feedback about the channel. The problem is exacerbated by the difficulty of selecting the proper modulation and coding scheme when using OFDM.

  In this thesis we present a data-driven approach to link adaptation for multiuser MIMO-OFDM systems. A machine learning classifier is used to select the modulation and coding scheme, taking as input the SNR values in the different subcarriers and spatial streams. A new approximation is developed to estimate the unknown interuser interference due to the use of limited feedback. This approximation allows to obtain SNR information at the transmitter with a minimum communication overhead. A greedy algorithm is used to perform spatial mode and user selection with affordable complexity, without resorting to an exhaustive search. The proposed adaptation is studied in the context of the IEEE 802.11ac standard, and is shown to schedule users and adjust the transmission parameters to the channel conditions as well as to the rate of the feedback channel.

- **C2.4) Link adaptation with practical impairments** Link adaptation is usually performed by exploiting SNR information only. In a realistic environment, however, there might be other factors, which we denominate practical impairments, that affect the performance of a certain MCS. For example, higher layers can deliver to the physical layer packets of different size, which changes the codeword size (and, in consequence, the FER) for the same MCS. Also, although it is usually assumed that the noise is Gaussian, in many realistic environments this might not be the case, and noise is better modeled by the use of a generalized Gaussian distribution.

  In this thesis we present two methods to incorporate practical impairments into link adaptation, particularly into the FER prediction problem. The first method is inspired on machine learning techniques, and includes the practical impairments as additional features to the FER prediction problem. We observe that a combination of effective SNR and machine learning offers a good performance. The second method is explicitly designed to incorporate different codeword length into the FER prediction problem, and is based on classic estimation theory and an analytical expression that relates the FER of two different codeword lengths.

Part	Chapter	Contributions
I	Chapter 2	C1.1
I	Chapter 3	C1.2
I	Chapter 4	C1.3
I	Chapter 5	C1.4, C1.5
II	Chapter 6	C2.1, C2.2
II	Chapter 7	C2.3
II	Chapter 8	C2.4

Table 1.1: Summary of contributions and chapters

1.4 Structure of the Thesis

The main content of this thesis is structured in 7 chapters, divided into two parts. Part I, which includes Chapters 2 to 5, contains the contributions on cognitive radio. Part II, which includes Chapters 6 to 8, contains the contributions on link adaptation. Chapter 9 contains the conclusion of the thesis and some comments on future lines of work. The distribution of contributions in the different chapters is summarized in Table 1.1.

1.5 Notation

Any non-standard notation used in this thesis is defined for the particular chapter at the point where the symbols first occur. For reader's reference, we also include a comprehensive list of the notation in Table 1.2.

Symbol	Description
$\circledast$	Circular convolution operator
$\mathcal{U}(a,b)$	Uniform random distribution with support $[a,b]$
$\mathcal{CN}(\boldsymbol{\mu},\mathbf{R})$, $\mathcal{N}(\boldsymbol{\mu},\mathbf{R})$	(Complex circular) Gaussian random distribution of mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$
$\lvert\alpha\rvert, \angle\alpha$	Absolute value and phase of α
χ^2_n	χ-squared distribution with n degrees of freedom
$\nabla_{\mathbf{x}} f(\mathbf{x}_0)$	Gradient of the function f evaluated in $\mathbf{x}_0$
$\nabla^2_{\mathbf{x}} f(\mathbf{x}_0)$	Hessian matrix of the function f evaluated in $\mathbf{x}_0$
$\mathbb{E}[\cdot]$	Expectation operator
$\mathbb{V}\mathrm{ar}[\cdot]$	Variance operator
$\mathbb{P}[A]$	Probability of event A
$\mathrm{adj}(\mathbf{A})$	Adjugate matrix of $\mathbf{A}$
$\mathrm{vec}(\mathbf{A})$	Column-wise vectorization of $\mathbf{A}$
$\mathrm{diag}(\mathbf{a})$	Diagonal matrix with diagonal equal to $\mathbf{a}$
$\det(\mathbf{A})$, $\mathrm{tr}(\mathbf{A})$	Determinant and trace of $\mathbf{A}$
$\odot$	Hadamard product
$\otimes$	Kronecker product
$\mathbb{X}^{a\times b}$	Set of all matrices with a rows and b columns with entries in $\mathbb{X}$
$(\cdot)^T$, $(\cdot)^*$	Transpose and conjugate transpose
$\lVert\cdot\rVert_\ell$ (resp. $\lVert\cdot\rVert$)	norm ℓ (resp. norm 2)
$\mathbf{0}_L$	Zero $L\times 1$ vector or $L\times L$ matrix
$\mathbf{1}_L$	$L\times 1$ all-ones vector
$\mathbf{I}_L$	Identity matrix of size $L\times L$

Table 1.2: Notation used in this Thesis.

Part I

Overlay Cognitive Radio

Chapter 2

Overlay Cognitive Radio over Broadcast Networks with Line of Sight Reception

Contents

2.1 Introduction

Recently, there has been an increased interest for learning the potential of those cognitive radio (CR) systems where the secondary transmitter has knowledge of the primary message, in what is known as the *overlay paradigm* [5]. This prior knowledge of the primary transmission can be exploited by the secondary users to convey their own information when accessing primary user spectrum in an efficient way, while preserving the primary user's quality of service (QoS). Hence, the usefulness of the knowledge of the primary message is twofold: on the one hand, the degradation of the primary user link due to the insertion of a secondary signal can be compensated by the secondary transmitter by using a fraction of its available power to transmit the primary message, keeping the signal to noise ratio (SNR) at the primary receivers above a given threshold; on the other hand, since the secondary transmitter knows the primary message, some kind of interference cancellation scheme can be applied, like dirty paper coding (DPC) [14].

However, the knowledge of the primary signal by the secondary transmitter is hard to justify, and, therefore, limited to a small quantity of practical cases [5]. In this chapter, similarly to [18], we introduce another practical scenario where the knowledge of the primary signal is possible: in broadcasting systems working as a single frequency network (SFN), e.g., the European digital video broadcasting - terrestrial (DVB-T) based service, deployed in many countries worldwide, the primary signal is sent via satellite (or other kind of distribution network) to some major transmitters, which need to apply the corresponding delay to keep the synchronization required by the SFN mode. Thus, a potential secondary transmitter might also gain access to the primary signal, keeping time and frequency synchronization with the primary transmitters and, therefore, join the primary network. The ultimate goal is to overlay the secondary information on the primary signal which can be decoded by secondary receivers, while preserving and possibly reinforcing the quality of service of the primary network (see Figure 4.1) without any modification on the primary receivers. Thus, the present work is focused on the cognitive spectrum reuse of the frequency bands used by any broadcast system working as an SFN, as they are specially interesting due to the high amount of bandwidth that these services are allocated, the possibility of accessing the primary message, and also due to the good propagation conditions of these frequency bands.

Although following a different principle, this idea has been developed in [67], where by resorting to game theory principles, the primary transmitter adapts its power to the overlaid secondary use of the spectrum to keep its QoS. The case of a secondary user that is aware of the primary message is studied in [16] following a similar game-theoretic formulation, where it is proposed to transmit the primary message over a fraction of the primary transmission resources

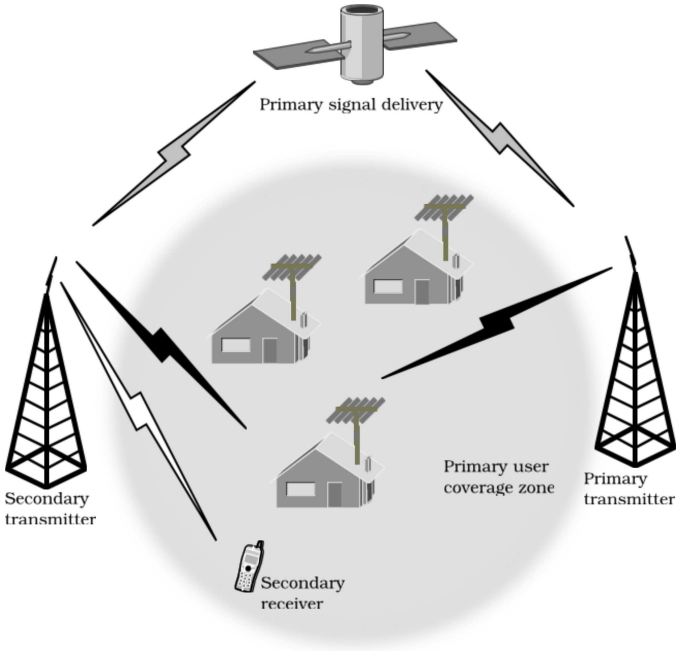


Figure 2.1: The secondary transmitter conveys the primary signal (black rays), which is delivered via a distribution network (gray rays). The secondary transmitter overlays the secondary message (white ray) on top of the primary one.

(in this case, frequency bands) with the objective of incrementing its capacity and, therefore, release the remaining spectrum for the secondary transmission while keeping the original primary transmitter rate. Here we will consider that the transmit power of the primary user is fixed, as usual in current broadcast networks, and we will introduce the possibility of using a fraction of the secondary power to transmit the primary message. The objective of this transmission is not to free some transmission resources of the primary user, but to control the interference due to the transmission of the secondary message.

Although the case of secondary transmitters with knowledge of the primary signal has been addressed from an information theoretic point of view, see e.g., [12,13,15] among others, there is still an important gap between *capacity-achieving* models and practical implementations where successful spectrum reuse is expected to be achieved. In short, some of the main issues to address are:

Metric for primary QoS. A usual metric for the QoS of the primary user is the capacity of a transmitter-receiver pair: if this capacity is greater than the transmission rate of the primary user, then the primary communication is not compromised [15]. If the channel is of time-varying nature, the QoS is measured in terms of probability of outage for a given transmission rate [9].

However, in broadcasting scenarios, coverage areas become the relevant metric as a result of the achieved bit error rates.

Primary user reinforcement. Even in the absence of a secondary information signal, the simple transmission of the primary signal from a secondary transmitter will not necessarily improve the primary service quality, since echoes can degrade performance as it is well-known in current SFN deployments [68], unless proper countermeasures can be taken. This effect is especially noticeable in those systems with a dominant line of sight (LOS) component, and almost negligible in high scattering environments. In practical cases, the degradation coming from the secondary echo could be higher than the power gain due to the extra contribution of the secondary transmitter. This type of problems is expected to be mitigated in the future with new standards such as DVB-T2 [69], which include some precoding schemes such as Alamouti space-time coding or constellation rotation. On the other side, some specific channels, such as Rayleigh fading channels, benefit from the diversity created by SFN deployments, as illustrated in [68]. In this chapter, we will model both the primary and secondary channels as a pure LOS component, which is indeed the case for which a higher degradation is expected, according to [68]. In Chapter 3 we extend this scenario to accommodate non-LOS channels as well.

Interference cancellation techniques. In many cases practical interference mitigation techniques at the transmitter exploiting side information cannot be directly applied, as they require knowledge of the channel state. In [70] it was shown that the uncertainty in the channel phase suffices to decrease the achievable capacity of the secondary link dramatically. Interference cancellation can be also performed at the secondary receiver, provided the interfering power is strong enough [71], as proposed in [72].

Given the widespread current use of DVB-T, we will focus on this multicarrier technology as support for the primary signal, and show how an appropriate secondary transmission of the primary signal can reinforce the original QoS. This is a first step towards a cognitive secondary transmitter which additionally includes a secondary information signal while preserving the primary user coverage area.

The remaining of the chapter is organized as follows: in Section 2.2 we introduce the notation and the analytical expressions to be used afterwards. In the next sections, the problem is treated in an incremental way, using the aforementioned analytical expressions as quality metrics for the primary system: in Section 2.3, a pure cooperative secondary user that tries to maximize a primary receiver QoS is studied, and practical transmission strategies are derived; in Section 2.4 the case of a secondary user maximizing its own transmission rate in presence of a single secondary receiver is presented; Section 2.5 completes the study, introducing the restriction of preserving the original coverage zone of the primary user. In Section 2.6 the

analytical expressions are verified by means of software simulations and hardware measurements. Finally, the conclusions are presented in Section 2.7.

2.2 Problem statement

Throughout the chapter we will assume that the links from both primary and secondary transmitters to a given primary receiver can be modeled[1] as additive white Gaussian noise (AWGN) channels, so the equivalent baseband received signal after the cyclic prefix (CP) removal can be written as

$$y_n = \left(\delta_n + \gamma e^{-j\theta} f_{n-n_0}\right) \circledast x_n + \rho e^{-j\theta} s_{n-n_0} + w_n \tag{2.1}$$

where the equivalent channel was normalized to set the channel from the primary transmitter to δ_n, while γ, θ and n_0 are the relative amplitude, phase and delay of the primary signal contribution sent from the secondary transmitter, $\circledast$ denotes the circular convolution operator, x_n denotes the n-th sample of the primary signal (normalized to have unit power), ρ denotes the relative amplitude of the secondary signal $s_n \sim \mathcal{CN}(0,1)$, assumed to be white Gaussian[2], sent from the secondary transmitter, and $w_n \sim \mathcal{CN}\left(0,\sigma^2\right)$ is a sample of white Gaussian noise. As an additional degree of freedom, the secondary transmitter is allowed to (circularly) filter the primary signal with a transmit filter f_n. The convenience of this filtering will be illustrated in the remaining of the chapter. In the discrete Fourier transform (DFT) domain, the previous relation reads for a given carrier k as

$$\begin{aligned} Y_k &= \left(1 + \gamma e^{-j(2\pi k n_0/N+\theta)} F_k\right) X_k + \\ &\quad + \rho e^{-j(2\pi k n_0/N+\theta)} S_k + W_k \;\; k = 1, ..., N \end{aligned} \tag{2.2}$$

where X_k, S_k, F_k and W_k denote the N-DFT of x_n, s_n, f_n and w_n, respectively, with N the number of carriers. Figure 5.1 summarizes the system model.

For the sake of simplicity, we will assume perfect channel estimation[3] and frequency synchronization in the analytical derivations, and an overall channel length shorter than the CP.

[1] The simple AWGN channel can be a good approximation, specially for those receivers with rooftop antennas (very common in terrestrial television broadcasting), which allow the existence of a strong LOS propagation path.

[2] This can be a good approximation, for example, in the case of a secondary transmitter using an OFDM waveform, where the time-domain signal is generated by combining a relatively large number of independent random variables (the symbols on the different carriers). The gaussianity is maintained in the DFT domain provided both primary and secondary waveforms are not identical (for example, by using different FFT sizes or CP lengths).

[3] We are assuming that the primary waveform carries some pilot symbols (which are also transmitted by the secondary transmitter) to perform the channel estimation, so the *equivalent channel* $\left(1 + \gamma e^{-j(2\pi k n_0/N+\theta)} F_k\right)$ can be accurately estimated at the primary receivers.

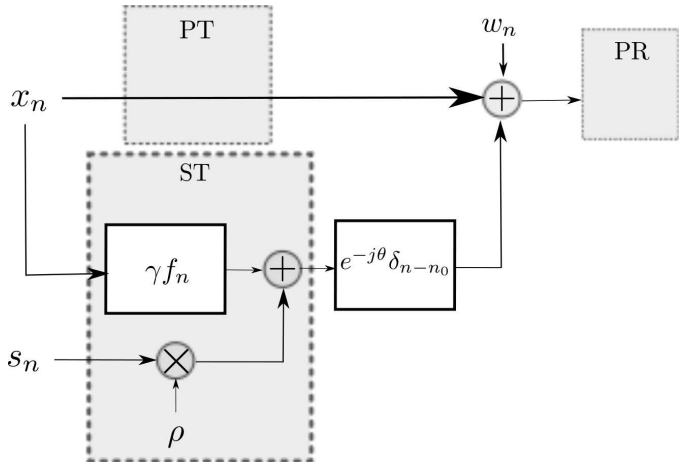


Figure 2.2: System model: the Secondary Transmitter (ST) knows the message x_n of the Primary Transmitter (PT). The ST filters the primary signal with the filter γf_n, and scales the secondary message s_n with ρ. The signal received by the Primary Receiver (PR) is described by equations (2.1) and (2.2).

Moreover, we will consider a quadrature phase shift keying (QPSK) constellation in the primary system, as the derived analytical bounds are easier to deal with. However, these results will be extended to higher order constellations and practical synchronization schemes by means of hardware measurements.

Unlike previous approaches to similar problems that use a capacity-based quality metric for the primary system [9, 18], we propose to analyze the performance of the primary system by means of the Chernoff bound (CB) for the uncoded bit error rate (BER) or, equivalently, by the exponential effective signal to noise ratio metric (EESM) [39], one of the metrics used orthogonal frequency division multiplexing (OFDM) systems with adaptive coding and modulation (ACM) [41], and one of the physical layer abstraction methods proposed in IEEE 802.16 [73]. The expression for the effective Signal to Noise Ratio (SNR) using the EESM metric is[4] $\Upsilon_{eff} = -2\log(\eta)$, where

$$\eta = \frac{1}{N}\sum_{k=1}^{N} e^{-\Upsilon|H_k|^2/2} = \frac{1}{N}\sum_{k=1}^{N} e^{-\beta|H_k|^2} \tag{2.3}$$

is the expression for the CB. From (2.2) we have that $H_k = 1 + \gamma e^{-j(2\pi k n_0/N+\theta)} F_k$ is the

[4]The general expression for the EESM is $\Upsilon_{eff} = -\lambda \log\left(\frac{1}{N}\sum_{k=1}^{N} e^{-\Upsilon|H_k|^2/\lambda}\right)$, being λ a degree of freedom that depends on the particular modulation and coding scheme [39]. In this chapter we will set $\lambda = 2$, as it is the value for the CB of the BER of a QPSK, although results can be easily extended to other values of λ to accommodate other constellations.

equivalent channel seen by the k-th carrier at a given receiver, so

$$|H_k|^2 = 1 + \gamma_k^2 + 2\gamma_k \cos\left(\theta + 2\pi k n_0/N\right) \tag{2.4}$$

where $\gamma_k \triangleq \gamma F_k$ is assumed to be real, without loss of generality, and $\Upsilon \triangleq \frac{1}{\sigma^2+\rho^2}$ denotes the SNR of the system in absence of the secondary transmitter conveying the primary message, which is constant along all the carriers due to the AWGN assumption. Moreover, we have defined $\beta = \Upsilon/2$ for the sake of simplicity.

In the following, we will assume that the value of the relative amplitude γ is deterministic, as it can be obtained by means of a propagation model or by measurements, and model θ as a uniform random variable (RV) $\theta \sim U\left(0, 2\pi\right]$ as it is not possible to determine the exact phase difference between echoes θ. Note that the metric η as defined in (2.3) is a RV, so a deterministic figure of merit for a primary receiver is obtained after substituting (2.4) in (2.3) and averaging[5] over θ:

$$\begin{aligned} \eta\left(\boldsymbol{\gamma}, \rho\right) &= \frac{1}{N}\sum_{k=1}^{N} \mathbb{E}_\theta\left[e^{-\beta\left(1+\gamma_k^2+2\gamma_k\cos(\theta+2\pi k n_0/N)\right)}\right] \\ &= \frac{e^{-\beta}}{N}\sum_{k=1}^{N} e^{-\beta\gamma_k^2}\frac{1}{2\pi}\int_0^{2\pi} e^{-2\beta\gamma_k cos(\theta)}\mathrm{d}\theta \\ &= \frac{e^{-\beta}}{N}\sum_{k=1}^{N} e^{-\beta\gamma_k^2} I_0(2\beta\gamma_k) \end{aligned} \tag{2.5}$$

where $I_0(\cdot)$ is the zero-th order modified Bessel function of the first kind, $\mathbb{E}_X[\cdot]$ denotes the expectation operator over the RV X, and $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_N]^T$. As the obtained expression does not depend on the time difference n_0, there is no need to make any assumption about this value. This CB-based metric η will be recurrent throughout the chapter, and will appear as the optimization objective in Section 2.3, and as a design constraint in Sections 2.4 and 2.5.

In order to obtain a relationship between the CB and the definition of the coverage zone, which is determined by the coded BER, we introduce the following analytical bound for the

[5]If we assume a static channel model, the value θ will not change for different OFDM blocks in given receiver, but only change among different receivers. Thus, in order to make the quality metric process ergodic in every receiver, a different random phase component could be applied to every OFDM block at the secondary transmitter (similarly to [74]), so the long-term average η seen by a single receiver is the expected value of η, even for a static channel scenario.

BER after Viterbi for DVB-T, taken from [75]:

$$BER \leq \frac{1}{4} \sum_{d=d_{min}}^{\infty} c_d \eta\left(\boldsymbol{\gamma}, \rho\right)^d \tag{2.6}$$

with d_{min} the minimum Hamming distance of the convolutional code, and c_d the total input weight due to an error event at distance d from the all-zero path.

Unlike previous work on BER minimization via power allocation for coded OFDM [76, 77], we will use the CB to get analytical results of potential interest for practical designs. Obviously, it cannot be said that the obtained results are *optimum* in the sense that we are dealing with BER bounds, not the BER itself. However, due to the impossibility of obtaining a closed form expression for the BER, the most reasonable approach is to derive optimum transmission strategies with respect to the BER bound, and afterwards analyze them in a real environment, by means of software simulation and hardware measurements.

2.3 Optimum power allocation for a purely cooperative secondary user

In this section, we will obtain the optimum carrier power allocation (with respect to the metric η in (2.5)) for a secondary transmitter that cooperates by minimizing the BER of a single primary receiver as a first approach, without inserting a secondary message[6]. For a primary receiver location where the ratio between the powers coming from the secondary and primary transmitters is γ^2, the minimization of the CB (2.5) reads as

$$\text{minimize} \quad \sum_{k=1}^{N} e^{-\beta\gamma_k^2} I_0(2\beta\gamma_k) \quad \text{subject to} \quad \frac{1}{N}\sum_{k=1}^{N} \gamma_k^2 \leq \gamma^2. \tag{2.7}$$

This is a non-convex problem over N variables, which makes numerical methods difficult to apply. However, as shown in Appendix 2.A, those points of the form $\boldsymbol{\gamma} = [\mathbf{0}^T_{N(1-\phi)}\ K\mathbf{1}^T_{N\phi}]^T$, (where $\mathbf{1}_p$ and $\mathbf{0}_q$ denote the all-ones row vector of p elements and the all-zeros column vector of q elements, respectively) with K such that the power constraint is met with equality, and with a fraction of active carriers ϕ such that $N\phi$ is an integer, are critical points of the Lagrangian of the proposed optimization problem. For this type of solutions, the optimization problem (2.7)

[6]This case is of special interest, as it provides the solution to the optimum power weighting $\boldsymbol{\gamma}$ given a total power γ^2 allocated to the primary waveform at the secondary transmitter.

can be recast as[7]

$$\begin{array}{ll} \text{minimize} & (1-\phi)+\phi e^{-\beta\gamma^2/\phi} I_0\left(\frac{2\beta\gamma}{\sqrt{\phi}}\right) \\ \text{subject to} & 0 \leq \phi \leq 1. \end{array} \tag{2.8}$$

As shown in Appendix 2.B, the asymptotic solution of (2.8) for large SNR values is $\phi = \frac{\gamma^2}{4}$, which forces to allocate $\gamma_k^2 = 4$ to the corresponding fraction of carriers. Following (2.4), in this case we have that $|H_k| \geq 1$, so no carrier suffers from an SNR loss with respect to the scenario without a secondary transmitter. Note that the optimum solution is only dependent on the fraction of active carriers, and not on their specific locations, due to the symmetry of the problem. In any case, the unidimensional problem (2.8) is computationally tractable as opposed to (2.7).

We have evaluated the analytical bound for the BER in (2.6) for those solutions found in (2.8): Figure 2.3 shows that the proposed method always decreases the BER bound, even when the *unfiltered* approach leads to a huge degradation, thus showing the importance of the proposed filtering, which intends to reduce the degradation due to the presence of SFN echoes. Interestingly, the solution $\phi = \frac{\gamma^2}{4}$ is quite a good approximation to the optimum value of the fraction of active carriers, especially for the higher SNR case.

2.4 Optimum power distribution for a single primary receiver

In this section, we will focus on the strategy the secondary transmitter must follow in order to maximize its own capacity subject to a controlled degradation of the primary service at a given receiver. We will assume that the secondary users are able to use some kind of interference mitigation techniques so the capacity of the secondary link is equivalent to that in absence of the primary transmitter. As we explained previously, the use of DPC techniques [14] would require channel knowledge at the transmitter [70] and, therefore, a feedback channel to convey that information, whereas the use of successive interference cancellation (SIC) at the receivers is more likely to be performed. A similar idea was developed in [9, 72], where the use of opportunistic interference cancellation (OIC) was shown to dramatically increase the secondary user rate. In our case, we will assume that interference cancellation can be always performed, as the secondary user is expected to be in the primary user coverage area. Therefore, our channel model will be an interference Z channel [5], where the secondary message is treated as noise by the primary receivers, and the primary interference can be completely canceled out by the

[7] Here, we are assuming that the number of carriers is large enough to approximate the fraction of active carriers by any real number in the interval $[0, 1]$. If the resulting optimum value of ϕ is such that $N\phi$ is not an integer, the loss of performance taking $\lfloor N\phi \rfloor$ as the number of active carriers will be negligible.

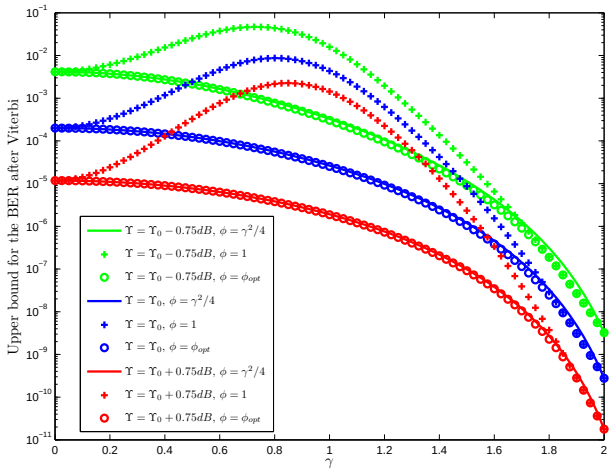


Figure 2.3: Analytical bound for the BER for QPSK, convolutional rate 2/3, with different secondary transmission approaches: no filtering ($\phi = 1$), filtered with $\phi = \gamma^2/4$ and with $\phi = \phi_{opt}$ as found by `fminbnd`. Υ_0 denotes the SNR that the bound predicts for the quasi error free (QEF - BER after Viterbi decoding of $2\,10^{-4}$) threshold for the system under analysis, which is $\Upsilon_0 \approx 5.6dB$.

secondary receivers.

In the design of practical multicarrier receivers it is sometimes assumed that the noise power is constant for all the carriers. If this is the case, the fact of transmitting with high power in a few carriers will be a source of narrowband interference, which is very harmful to OFDM transmission [78]. In consequence, we will restrict the design of secondary signals to those with constant power along the carriers, although the proposed methodology can be extended to the general case.

Let us denote by P the secondary received power (normalized by the primary one) at a given location, that has to be split between the primary $\left(\frac{1}{N}\sum_{k=1}^{N}\gamma_k^2 = \gamma^2\right)$ and secondary (ρ^2) signals. Note that the flat spectrum constraint for the secondary message turns the maximization of the capacity equivalent to the maximization of the power allocated to the secondary message ρ^2, so introducing a power constraint and a constraint on the primary user CB η in (2.5), we

can formulate the optimization problem as

$$\begin{aligned} &\text{minimize} && -\rho \\ &\text{subject to} && \frac{1}{N}\sum_{k=1}^{N} e^{-\frac{1+\gamma_k^2}{\psi+2\rho^2}} I_0\left(\frac{2\gamma_k}{\psi+2\rho^2}\right) \leq \eta_0 \\ & && \rho^2 + \frac{1}{N}\sum_{k=1}^{N}\gamma_k^2 \leq P \end{aligned} \tag{2.9}$$

where η_0 is the constraint on the CB η, and $\psi \triangleq 2\sigma^2$, leading to $\beta = \frac{1}{2(\sigma^2+\rho^2)} = \frac{1}{\psi+2\rho^2}$.

In Section 2.3 it was shown that, for a given allocated average power of a purely cooperative secondary user to the primary user message $\gamma^2 = \frac{1}{N}\sum_{k=1}^{N}\gamma_k^2$, the optimum power distribution consisted on concentrating the power in a fraction ϕ of carriers, leaving the remaining fraction $1-\phi$ set to zero. Again, for a sufficiently large number of carriers, we can approximate the fraction ϕ by a real number in the interval $[0,1]$, so problem (2.9) can be rephrased as

$$\begin{aligned} &\text{minimize} && -\rho \\ &\text{subject to} && e^{-\frac{1}{\psi+2\rho^2}}\left((1-\phi)+\phi e^{-\frac{\gamma^2/\phi}{\psi+2\rho^2}} I_0\left(\frac{2\gamma}{\sqrt{\phi}(\psi+2\rho^2)}\right)\right) \leq \eta_0 \\ & && \rho^2+\gamma^2 \leq P \\ & && 0 \leq \phi \leq 1. \end{aligned} \tag{2.10}$$

With this simplification we have reduced the number of variables from $N+1$ (the N variables γ_k to perform the power weighting, and ρ) to three. Furthermore, we can reduce the number of variables to two by approximating ϕ by its asymptotic optimum (and heuristic) value $\left(\phi = \frac{\gamma^2}{4}\right)$ for the sake of analytical tractability. With this last simplification, the CB constraint in (2.10) can be rewritten as

$$f(\gamma,\rho) = e^{-\frac{1}{\psi+2\rho^2}}\left(\left(1-\frac{\gamma^2}{4}\right)+\frac{\gamma^2}{4}e^{-\frac{4}{\psi+2\rho^2}} I_0\left(\frac{4}{(\psi+2\rho^2)}\right)\right) - \eta_0 \leq 0 \tag{2.11}$$

or, equivalently,

$$\gamma^2 \geq 4\frac{1-\eta_0 e^{\frac{1}{\psi+2\rho^2}}}{1-e^{-\frac{4}{\psi+2\rho^2}} I_0\left(\frac{4}{\psi+2\rho^2}\right)}. \tag{2.12}$$

The solution to this problem presents a different behavior depending on the values of the SNR in absence of the secondary transmitter, $\Upsilon_{NS} \triangleq 2/\psi$, and the received power from the secondary transmitter P, as detailed next.

2.4.1 Moderate values of P

For non-extreme values of P, if $\gamma^2 \leq 4$, the approximate optimum value of γ is the one that maximizes ρ while meeting constraint (2.12) and, therefore, is the value obtained from (2.12) with equality, so the BER restriction is active and the remaining power is used to transmit the secondary information. By substituting $\gamma^2 = P - \rho^2$ in (2.12) we can obtain the value of ρ as the root of the following equation:

$$\rho^2 = P - 4\frac{1 - \eta_0 e^{\frac{1}{\psi+2\rho^2}}}{1 - e^{-\frac{4}{\psi+2\rho^2}} I_0\left(\frac{4}{\psi+2\rho^2}\right)}. \tag{2.13}$$

If $\gamma^2 = P - \rho^2 > 4$, then the obtained solution is not valid, as $\phi > 1$. In such a case the solution would be obtained by forcing $\phi = 1$ and $\rho^2 + \gamma^2 = P$ in problem (2.10), so the desired value of ρ^2 is the root of

$$e^{-\frac{1+P-\rho^2}{\psi+2\rho^2}} I_0\left(\frac{2\sqrt{P-\rho^2}}{\psi+2\rho^2}\right) - \eta_0 = 0. \tag{2.14}$$

2.4.2 $P \to 0$

For small values of P the solution will be strongly dependent on the SNR in absence of the secondary transmitter Υ_{NS}. Let us define Υ_0 as the value of SNR such that the BER constraint is met with equality in absence of the secondary transmitter, i.e., $e^{-\frac{\Upsilon_0}{2}} = \eta_0$. Equivalently, we define $\psi_0 \triangleq \frac{2}{\Upsilon_0} = \frac{-1}{\log(\eta_0)}$. We will restrict our analysis to those receivers in the original coverage region, i.e., $\Upsilon_{NS} \geq \Upsilon_0$.

$\Upsilon_{NS} > \Upsilon_0$

In this case, as $\psi < \psi_0$, we have that $e^{-\frac{1}{\psi+2P}} \leq \eta_0$ for sufficiently small values of P. Therefore, the secondary transmitter can allocate all the available power to the secondary message without violating the BER constraint, i.e, its optimum allocated power to the secondary message is $\rho^2 = P$. This could be the case of a primary receiver operating at a very high SNR, or a low-power secondary user.

$\Upsilon_{NS} = \Upsilon_0$

In this case, as the BER constraint is met with equality, we have that $e^{-\frac{1}{\psi_0+2P}} > \eta_0$, so the CB constraint is not fulfilled if all the power P is allocated to the secondary message. Following expression (2.11) and from the definition of ψ_0, we have $f(0,0) = 0$. For $\rho \approx 0$, $\gamma \approx 0$ and as $\nabla_{\gamma,\rho} f(0,0) = \mathbf{0}$, we can approximate the CB constraint (2.11) by its second order Taylor polynomial:

$$f(\gamma,\rho) = \frac{1}{2}[\gamma\ \rho]\nabla^2_{\gamma,\rho} f(0,0)\,[\gamma\ \rho]^T \tag{2.15}$$

where $\nabla_{\gamma,\rho} f(\gamma_0,\rho_0)$ denotes the gradient of the function f evaluated in (γ_0,ρ_0), and $\nabla^2_{\gamma,\rho} f(\gamma_0,\rho_0)$ denotes the Hessian matrix evaluated in the same point. In this case, the Hessian evaluated in $(0,0)$ is a diagonal matrix with entries

$$\frac{\partial^2 f}{\partial \gamma^2}(0,0) = \frac{1}{2} e^{-5/\psi_0}\left(I_0\left(\frac{4}{\psi_0}\right) - e^{4/\psi_0}\right) \tag{2.16}$$

$$\frac{\partial^2 f}{\partial \rho^2}(0,0) = \frac{4e^{-1/\psi_0}}{\psi_0^2}. \tag{2.17}$$

The maximum value of ρ will be obtained when both the CB constraint and the power constraint are met with equality. Therefore, the solution is obtained by equating (2.15) to zero and substituting $\gamma^2 = P - \rho^2$, so the following equality arises:

$$\frac{\rho^2}{P} = \frac{\frac{\partial^2 f}{\partial \gamma^2}(0,0)}{\frac{\partial^2 f}{\partial \gamma^2}(0,0) - \frac{\partial^2 f}{\partial \rho^2}(0,0)} = \frac{\psi_0^2\left(e^{4/\psi_0} - I_0\left(\frac{4}{\psi_0}\right)\right)}{e^{4/\psi_0}\left(\psi_0^2+8\right) - \psi_0^2 I_0\left(\frac{4}{\psi_0}\right)}. \tag{2.18}$$

2.4.3 $P \to \infty$

For high values of P, the high power coming from the secondary transmitter makes the primary contribution negligible. In this case it can be easily seen that the optimum filtering of the primary signal leads to $\phi = 1$, so we can write the CB constraint as

$$\eta(\gamma,\rho) = e^{-\frac{\gamma^2}{2\rho^2}} \leq \eta_0, \tag{2.19}$$

so the optimum value of ρ^2 will be obtained when (2.19) and the power constraint are met with equality, so we arrive to

$$\frac{\rho^2}{P} = \frac{1}{1 - 2\log(\eta_0)} = \frac{1}{1+\Upsilon_0}. \tag{2.20}$$

Table 2.1: Values for the design parameters ϕ, γ^2 and ρ^2 for the different cases under study.

Case	ϕ	γ^2	ρ^2
P Moderate, $\gamma^2 < 4$	$\gamma^2/4$	$P-\rho^2$	Root of (2.13)
P Moderate, $\gamma^2 \geq 4$	1	$P-\rho^2$	Root of (2.14)
$P \to 0$, $\Upsilon_{NS} > \Upsilon_0$	N/A	0	P
$P \to 0$, $\Upsilon_{NS} = \Upsilon_0$	$\gamma^2/4$	$P-\rho^2$	(2.18)
$P \to \infty$	1	$\frac{P\Upsilon_0}{1+\Upsilon_0}$	$\frac{P}{1+\Upsilon_0}$

Note that this is the case when both noise and primary user power are negligible, so the constraint for the secondary user is to keep the ratio between primary and secondary messages over the limit SNR value, $\frac{\gamma^2}{\rho^2} = \Upsilon_0$.

The analytical power allocation results are summarized in Table 2.1 for the different cases.

2.4.4 Results

We will show the values of the secondary message power ρ^2 for receivers with different margins with respect to the necessary SNR for QEF reception, obtained with the analytical approximation $\phi = \gamma^2/4$. These results will be compared with those obtained with the optimum value $\phi = \phi_{opt}$ in order to check the accuracy of the approximation, and with those forcing $\phi = 1$, thus showing the importance of the unequal power weighting. These two latter approximation are obtained by `MATLAB fmincon` applied to the problem (2.10), with ϕ a degree of freedom and $\phi = 1$, respectively. The approximation $\phi = \gamma^2/4$ is obtained[8] following the P *moderate* entries in Table 2.1.

In the simulations the selected convolutional code rate is 2/3 again, for which the bound (2.6) predicts a value of $\Upsilon_0 \approx 5.6dB$ for a BER of $2 \cdot 10^{-4}$, being $\eta_0 \approx 0.16$.

In Figure 2.4 it is shown the evolution of ρ with the total available power for moderate values of P and three different SNR values, with $\Upsilon_{NS,dB} = 10\log_{10}(\Upsilon_{NS})$. Obviously, as all the three cases have the same CB restriction, the one with the higher Υ_{NS} will require a lighter support from the secondary transmitter and, therefore, ρ^2 will be higher. It is also noticeable that the evolution of ρ^2 (in both the $\phi = \gamma^2/4$ and ϕ_{opt} cases) has two differentiated regimes: the *low power* regime, where all the secondary power can be allocated to the secondary message without breaking the BER constraint and, therefore, in this region $\rho^2 = P$; and the *moderate*

[8] An additional check has to be performed: If the obtained value meets $\rho^2 > P$, then all the available power can be allocated to the secondary message $\rho^2 = P$, and the CB constraint will be met with strict inequality.

power regime. Note also that the case of $\Upsilon_{NS} = \Upsilon_0$ does only admit the *moderate power* regime, as the BER constraint is met with equality even in absence of the secondary transmitter. The solution for $\phi = 1$ has a slightly different behavior:

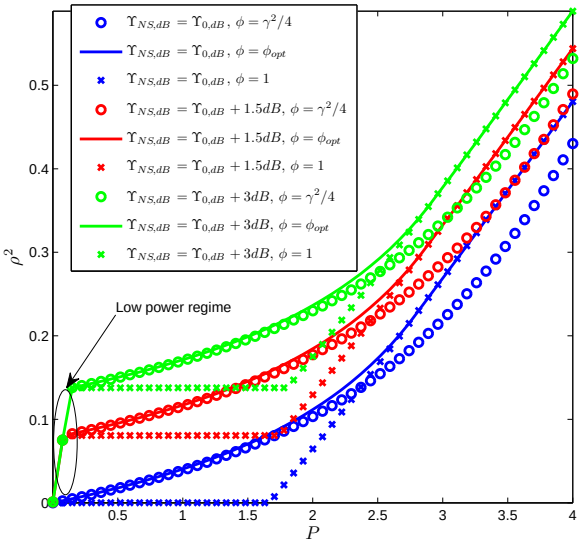


Figure 2.4: Power (seen at reception) allocated to the secondary signal at the secondary transmitter as a function of the received secondary power with respect to the primary one..

- For the cases where all the power can be allocated to the secondary message without breaking the BER constraint, the solution is the same as in the other approximations. If this region does not exist (for $\Upsilon_{NS} = \Upsilon_0$) the optimum value of ρ is zero for a large range of values of P.
- For moderate values of P, an increment on the value of P is not reflected in the value of ρ, as allocating some power to the primary message would increase the BER bound.
- For high values of P, the value of ρ increases with P. In this region, the value of ρ is obtained as the root of (2.14), and approximates the optimum solution as P increases.

It is also noticeable that the solution with $\phi = \gamma^2/4$ offers very little degradation with respect to the optimum value of ϕ for small values of P, while the solution for $\phi = 1$ offers a good performance for larger values. Therefore, a near-optimum solution could be obtained just by solving the $\phi = 1$ and $\phi = \gamma^2/4$ problems, and choosing the one whose performance is better, which is substantially less computationally expensive than solving the more general problem. The degradation due to the presence of echoes is transcendent for a large range of values of P,

specially the lower ones. In this region, the importance of the proposed filtering is clear, as it allows the secondary transmitter to achieve a non-zero rate.

In Figure 2.5 the accuracy of the $P \to \infty$ and $P \to 0$ expressions for ρ^2/P is shown. For moderate values of P, it is also shown that if the target receivers have $\Upsilon_{NS} > \Upsilon_0$, then the fraction of available power used for the secondary transmission can be quite high for low values of P and then it has to decrease. In fact, in the low power regime, all the available power can be allocated to the secondary message without breaking the BER constraint, as previously stated. It can be also seen that the family of curves for $\Upsilon_{NS} > \Upsilon_0$ tend to approach the $\Upsilon_{NS} = \Upsilon_0$ curve as Υ_{NS} approaches Υ_0.

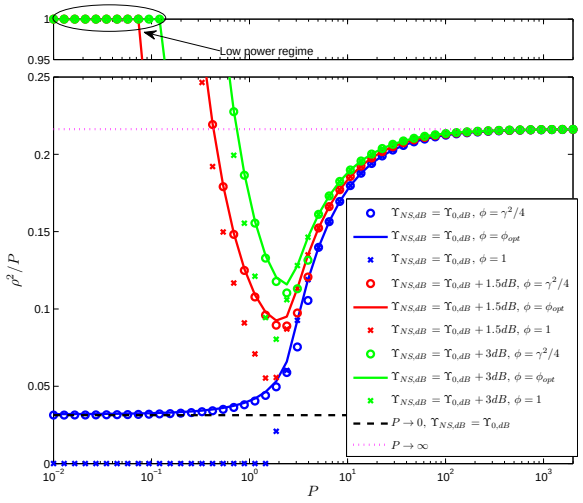


Figure 2.5: Fraction of power used for the transmission of the secondary message as a function of the total received power from the secondary transmitter with respect to the primary one.

2.5 Coverage analysis

In this section we will extend the previously obtained results to the case of having several primary receivers in different reception states (i.e., different values of P and Υ_{NS}), as expected in a realistic broadcast scenario. As we will see next, obtaining a solution is more involved than just considering a *worst case* primary receiver.

Let us define the *transmit mask* $\tilde{\boldsymbol{\gamma}} = [\tilde{\gamma}_1, ..., \tilde{\gamma}_N]$, and the *secondary ratio* $\tilde{\rho}$ as the transmit parameters such that $\frac{1}{N}||\tilde{\boldsymbol{\gamma}}||_2^2 + \tilde{\rho}^2 \leq 1$, so we can write $\boldsymbol{\gamma}(\mathbf{x}) = \sqrt{P(\mathbf{x})}\tilde{\boldsymbol{\gamma}}$ and $\rho(\mathbf{x}) = \sqrt{P(\mathbf{x})}\tilde{\rho}$, where $P(\mathbf{x})$, $\boldsymbol{\gamma}(\mathbf{x})$ and $\rho(\mathbf{x})$ denote the same quantities as in previous sections with the insertion of a parameter that indicates the position $\mathbf{x}$ (in polar coordinates $\mathbf{x} = (r, \theta)$, for convenience) of

a receiver located at $\mathbf{x}$. Similarly, we introduce the modified metric

$$\eta(\mathbf{x}, \tilde{\boldsymbol{\gamma}}, \tilde{\rho}) = \frac{1}{N} \sum_{k=1}^{N} e^{-\frac{1+\gamma_k^2(\mathbf{x})}{\psi(\mathbf{x})+2\rho^2(\mathbf{x})}} I_0 \left(\frac{2\gamma_k(\mathbf{x})}{\psi(\mathbf{x}) + 2\rho^2(\mathbf{x})} \right) \tag{2.21}$$

that extends (2.5) by adding the location parameter $\mathbf{x}$. With this extension, $\eta(\mathbf{x}, \mathbf{0}, 0)$ denotes the same metric in the absence of a secondary transmitter.

We will constrain the secondary user to keep (at least) the original coverage area of the primary system so the licensed service is not compromised. For the sake of simplicity, we will only consider those points within the coverage zone that are aligned with the primary and secondary transmitters, and have the two transmitters at the same side. This is equivalent to assuming receivers with perfectly pointed antennas with a gain of $-\infty$ dB for all angular directions (except 0°). Thus, the points that are affected by the secondary user and, therefore, the points we must take into account in the coverage constraint can be written in polar coordinates as $\mathcal{C}_0 = \{(r, \theta) \,|\, r \in [r_s, r_0],\, \theta = \theta_0\}$, where r_0 is the radius of the coverage zone, assumed to be a circle centered on the primary transmitter, and (r_s, θ_0) denotes the secondary transmitter location. This scenario is depicted in Figure 2.6.

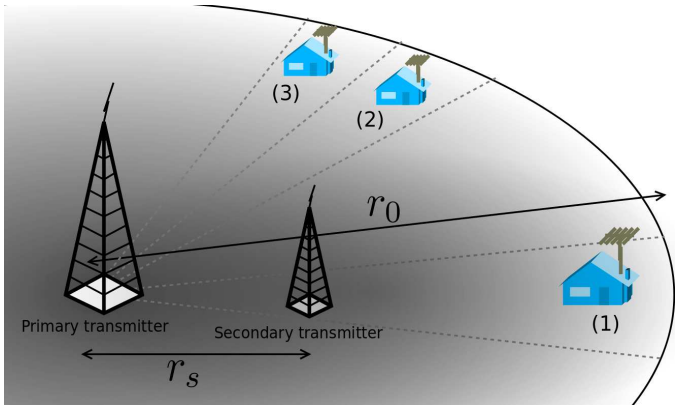


Figure 2.6: Coverage diagram. Due to the assumption on the perfect directivity of the antennas, receiver (1) is affected by the secondary transmitter, but receivers (2) and (3) are not.

Unfortunately, the problem of maximizing the secondary rate subject to a constraint on the primary coverage area is analytically intractable. However, and in order to show the effects of having several receivers under very different reception characteristics, we will study first a simplified two-user scenario, where one of the primary receivers is located near the secondary transmitter, and the other one far from it. As we will see afterwards, this two-user scenario is quite a good approximation to the solution to the complete coverage scenario, which has to be obtained numerically.

2.5.1 Two different receivers

In the proposed scenario we are likely to find two receivers that are in extremely different reception situations. For instance, if the secondary transmitter is located far from the coverage edge and its transmit power is much smaller than that of the primary transmitter, those receivers in the limit of the coverage zone will have an active CB constraint ($\eta(\mathbf{x}, \mathbf{0}, 0) = \eta_0$), and a value of $P(\mathbf{x}) \to 0$, whereas the receivers near the secondary transmitter will have a value of $P(\mathbf{x}) \to \infty$. We will study this case as a simplification of the general case covering receivers under many different values of P.

Let us denote as $\mathbf{x}_n$ the position of the receiver that is near the secondary transmitter ($P(\mathbf{x}_n) \to \infty$), and as $\mathbf{x}_f$ the position of the receiver that is far from the secondary transmitter ($P(\mathbf{x}_f) \to 0$). Even for this simple case, the optimum fraction of active carriers for the nearby receiver is $\phi = 1$, and for the far-off receiver is $\phi = \gamma^2(\mathbf{x}_f)/4 \approx 0$. For the sake of analytical tractability, we will restrict our analysis to two-level solutions for the primary power weighting, i.e., solutions of the form $\tilde{\boldsymbol{\gamma}} = [\tilde{\gamma}_1 \mathbf{1}_{N\phi}^T \ \tilde{\gamma}_2 \mathbf{1}_{N(1-\phi)}^T]^T$, with neither $\tilde{\gamma}_1$ nor $\tilde{\gamma}_2$ necessarily zero. In Appendix 2.C it is shown that a fraction of power

$$\tilde{\rho}^2 = \frac{\psi_0^2 \left(e^{4/\psi_0} - I_0\left(\frac{4}{\psi_0}\right)\right)}{e^{4/\psi_0}\left(\Upsilon_0 \left(\psi_0 - 2\right)^2 + \psi^2 + 8\right) - \left(\Upsilon_0 + 1\right)\psi_0^2 I_0\left(\frac{4}{\psi_0}\right)} \tag{2.22}$$

can be allocated to the secondary signal in this scenario, with $\tilde{\gamma}_2 = \Upsilon_0 \tilde{\rho}$, $\tilde{\gamma}_1^2 \to \infty$ and $\phi \to 0$. As we will see in the following section, this simplified scenario is a good approximation to the general one, where all the receivers in the coverage zone are taken into account.

2.5.2 Conditions for $\tilde{\rho} > 0$

Under the diverse reception characteristics present in this problem, it is not trivial to prove if the optimization problem has a solution with a non-zero power for the secondary user. Here, we propose two different sufficient conditions and one necessary condition for the attainment of a value of $\tilde{\rho} > 0$. We will focus on the case where the receiver with a minimum SNR is $\Upsilon_{min} = \Upsilon_0$, as $\Upsilon_{min} > \Upsilon_0$ implies that every point in the coverage zone taken into account has an SNR margin, and, therefore, the insertion of a secondary signal is always possible.

Proposition 2.1 (Sufficient condition 1)**.** *If for every point $\mathbf{x}_0 \in \mathcal{C}_0$ such that the BER constraint is active, i.e, $\eta(\mathbf{x}_0, \mathbf{0}, 0) = \eta_0$ we have that $\psi(\mathbf{x}_0) > 1$, then there exists $\tilde{\boldsymbol{\gamma}}, \tilde{\rho} > 0$ such that $\eta(\mathbf{x}, \tilde{\boldsymbol{\gamma}}, \tilde{\rho}) \leq \eta_0, \forall \mathbf{x} \in \mathcal{C}_0$.*

Proof. We will check the local optimality of the point $(\tilde{\boldsymbol{\gamma}}, \tilde{\rho}) = \mathbf{0}$ for all points in the coverage zone $\mathcal{C}_0$. For those points such that $\eta(\mathbf{x}, \tilde{\gamma}, \tilde{\rho}) < \eta_0$, i.e, the BER constraint is not active, the problem can be (locally) seen as an unconstrained optimization problem[9] and, therefore, a feasible direction with $\tilde{\rho} > 0$ exists.

For the points where the BER constraint is active, the solution is more involved.

In order to check the local optimality of the point $\mathbf{0}$, the whole optimization problem has to be recast because of the lack of regularity of the problem in the point $\mathbf{0}$. It can be easily seen that $\nabla_{\tilde{\gamma}}\eta(\mathbf{x}_0, \mathbf{0}, 0) = \mathbf{0}$, and $\frac{\partial}{\partial\tilde{\rho}}\eta(\mathbf{x}_0, \mathbf{0}, 0) = 0$, so the gradient of the restriction with respect to the design variables in the point $\mathbf{0}$ is null and, therefore, no constraint qualification can be applied such that strong duality holds and the KKT conditions are no longer valid.

The cause of this lack of regularity is the symmetry of the problem around 0; we are allowing negative values of ρ and γ_i, but as ρ always appear in the form of ρ^2, and γ_i squared or multiplying the argument of $I_0(\cdot)$ (which is an even function), the result is not affected. We can recast the problem with the insertion of the variables $\hat{\gamma}_i = \gamma_i^2$, $\hat{\rho} = \rho^2$, with $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_0 \dots \hat{\gamma}_{N-1}]$. The problem now reads as

$$\begin{array}{ll} \text{minimize} & -\hat{\rho} \\ \text{subject to} & \frac{1}{N}\sum_{k=0}^{N-1} e^{-\frac{1+\hat{\gamma}}{\psi+2\hat{\rho}}} I_0\left(\frac{2\sqrt{\hat{\gamma}}}{\psi+2\hat{\rho}}\right) \leq \eta_0 \\ & -\hat{\rho} \leq 0 \\ & -\hat{\boldsymbol{\gamma}} \preceq \mathbf{0} \\ & \hat{\rho} + \frac{1}{N}\sum_{k=0}^{N-1}\hat{\gamma}_k \leq P. \end{array}$$

If we are in the point $\hat{\boldsymbol{\gamma}} = \mathbf{0}$, $\hat{\rho} = 0$, then all constraints are active except the power constraint. The KKT conditions are

$$\begin{bmatrix} 0 \\ . \\ . \\ . \\ 0 \\ -1 \end{bmatrix} + \frac{\lambda_1}{N}\begin{bmatrix} -\frac{e^{-\frac{1}{\psi}}(\psi-1)}{\psi} \\ . \\ . \\ . \\ -\frac{e^{-\frac{1}{\psi}}(\psi-1)}{\psi} \\ \frac{2e^{-\frac{1}{\psi}}}{\psi^2} \end{bmatrix} - \begin{bmatrix} \lambda_{2,1} \\ . \\ . \\ . \\ \lambda_{2,N} \\ \lambda_{2,N+1} \end{bmatrix} = \mathbf{0}. \tag{2.23}$$

where λ_1 is the Lagrange multiplier of the BER restriction, and $\lambda_{2,i}$ are the Lagrange multipliers of the restrictions on the sign of $\hat{\gamma}_i$ and $\hat{\rho}$.

[9] Except in the degenerate case of $P = 0$, where the power constraint will be active.

The last equation can be easily solved just by setting

$$\lambda_1 = N\frac{\psi^2\left(\lambda_{2,N+1}+1\right)}{2e^{-\frac{1}{\psi}}}, \tag{2.24}$$

which is a positive number.

The other N equations must meet the condition

$$-N\frac{e^{-\frac{1}{\psi}}\left(\psi-1\right)}{\psi} = \frac{\lambda_{2,i}}{\lambda_1}, \tag{2.25}$$

so, as $\boldsymbol{\lambda} \succeq \mathbf{0}$, if $\psi \geq 1$ the left part of the equality is positive and, therefore, the point $\mathbf{0}$ does not meet the necessary conditions for optimality, which implies that a value of $\tilde{\rho} > 0$ is feasible. □

This result is similar to the one shown in Appendix 2.A, where it is proved that, when trying to minimize the primary BER, the point $\boldsymbol{\gamma} = 0$ was a local minimum if $\beta = \psi^{-1} < 1$, and a local maximum if $\beta = \psi^{-1} > 1$. In fact, if $\psi < 1$, the point $\mathbf{0}$ is the only feasible point in its neighborhood, so there is no feasible direction where the objective function decreases (in fact, there is no feasible direction). If $\psi > 1$, then any direction with $\hat{\gamma}_i > 0$ will make the BER to decrease and, therefore, open a *gap* for the insertion of the secondary signal.

This proposition states that very robust signaling waveforms (the ones that are able to provide the target BER with an SNR of less than 3dB) are easily enforced and, therefore, the insertion of a secondary message is possible. Unfortunately, the studied system (DVB-T with QPSK and convolutional rate 2/3) requires a SNR of 5.7dB, which is a value of $\psi \approx 0.55$.

Even when there is no feasible point in the neighborhood of $(\rho, \boldsymbol{\gamma}) = 0$ with $\rho > 0$, we will prove that, under some conditions, a value of $\tilde{\rho} > 0$ can be attained.

Proposition 2.2 (Sufficient condition 2)**.** *If for every point $\mathbf{x} \in \mathcal{C}_0$ the relationship between the received powers from the secondary and primary transmitters is such that $P(\mathbf{x}) > \frac{4}{N}$, then there exists $\tilde{\boldsymbol{\gamma}}, \tilde{\rho} > 0$ such that $\eta\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}, \tilde{\rho}\right) \leq \eta_0, \forall \mathbf{x} \in \mathcal{C}_0$*

Proof. We will follow the reasoning of proposition 2.1: we will find a point $(\tilde{\boldsymbol{\gamma}}, \tilde{\rho})$ such that $\eta\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}, \tilde{\rho}\right) < \eta_0 \forall \mathbf{x} \in \mathcal{C}_0$ and $\frac{1}{N}||\tilde{\boldsymbol{\gamma}}||_2^2 + \tilde{\rho}^2 < 1$. If this point exists, then the optimization problem is (locally) equivalent to an unconstrained optimization problem, so a value of $\tilde{\rho} > 0$ is attainable.

Let $\mathbf{x}_0 = \arg\min_{\mathbf{x}\in\mathcal{C}}\left\{P(\mathbf{x})\right\}$. For this point, we will analyze the effect of one carrier in the global BER. If we want to reduce the BER contribution of the carrier k, the value of the sum

term with γ_k must be strictly less than setting $\gamma_k = 0$, this is

$$e^{-\frac{1}{\psi(\mathbf{x}_0)}} > e^{-\frac{1+\gamma_k^2(\mathbf{x}_0)}{\psi(\mathbf{x}_0)}} I_0\left(\frac{2\gamma_k(\mathbf{x}_0)}{\psi(\mathbf{x}_0)}\right), \tag{2.26}$$

or, equivalently

$$e^{\frac{\gamma_k^2(\mathbf{x}_0)}{\psi(\mathbf{x}_0)}} > I_0\left(\frac{2\gamma_k(\mathbf{x}_0)}{\psi(\mathbf{x}_0)}\right). \tag{2.27}$$

This equation cannot be solved analytically (for values of $\gamma_k(\mathbf{x}_0) \neq 0$), so we will use the fact that

$$e^{x_1} \geq I_0(x_2) \, \forall x_1 \geq x_2 \geq 0 \tag{2.28}$$

with equality only for $x_1 = x_2 = 0$ to give a value of $\gamma_k(\mathbf{x}_0)$ that meets the inequality. If we set

$$\frac{\gamma_k^2(\mathbf{x}_0)}{\psi(\mathbf{x}_0)} \geq \frac{2\gamma_k(\mathbf{x}_0)}{\psi(\mathbf{x}_0)} \implies \gamma_k(\mathbf{x}_0) \geq 2, \tag{2.29}$$

then the inequality (2.27) will be met. Note that this result is independent of $\psi(\mathbf{x}_0)$.

For the other points $\mathbf{x} \neq \mathbf{x}_0$ in the coverage area, as $P(x) \geq P(x_0) \, \forall x \in \mathcal{C}_0$, $\gamma_k(\mathbf{x}) = \frac{P(x)}{P(x_0)} \gamma_k(\mathbf{x}_0) \geq \gamma_k(\mathbf{x}_0)$, they will also meet inequality (2.27). Therefore, if $P(\mathbf{x}_0) > 4/N$, both the power and the BER constraint will not be active, so a value of $\tilde{\rho} > 0$ is attainable.

□

Note that this result is highly dependent on the number of carriers: if this number is larger, the quantity of power we can save by nullying all the carriers but one will be larger and, therefore, the necessary value of P will be smaller. This means that if the number of carriers is large enough, a value of $\tilde{\rho} > 0$ can be always achieved without an increment of the available power.

Lemma 2.1. *Let $\beta > 0$. Then,*

$$e^{\gamma_0^2 \beta} > I_0(2\beta\gamma_0) \implies e^{\gamma^2 \beta} > I_0(2\beta\gamma), \, 0 < \gamma_0 \leq \gamma. \tag{2.30}$$

Proof. See Appendix 2.D. □

Proposition 2.3 (Necessary condition)**.** *Let $\mathcal{X} = \{\mathbf{x} \in \mathcal{C}_0 \,|\, \eta(\mathbf{x}, \mathbf{0}, 0) = \eta_0\}$. If there exists $\tilde{\boldsymbol{\gamma}}, \tilde{\rho} > 0$ feasible, then $\forall \mathbf{x}_0 \in \mathcal{X}$, $e^{\frac{NP(\mathbf{x}_0)}{\psi(\mathbf{x}_0)}} > I_0\left(2\sqrt{NP(\mathbf{x}_0)}/\psi(\mathbf{x}_0)\right)$.*

Proof. Let us assume that $\exists \mathbf{x}_0 \in \mathcal{C}_0$ such that $e^{\frac{NP(\mathbf{x}_0)}{\psi(\mathbf{x}_0)}} \leq I_0\left(2\sqrt{P(\mathbf{x}_0)}/\psi(\mathbf{x}_0)\right)$ and $\tilde{\boldsymbol{\gamma}}, \tilde{\rho} > 0$.

Then, as η is a strictly monotonic increasing function of $\tilde{\rho}$,

$$\eta\left(\mathbf{x}_0, \tilde{\boldsymbol{\gamma}}, 0\right) < \eta_0 = \eta\left(\mathbf{x}_0, 0, 0\right), \tag{2.31}$$

or, equivalently, and dropping the spatial indexing $\mathbf{x}_0$

$$\sum_{k=0}^{N-1} e^{-\frac{1+\gamma_k^2}{\psi}} I_0\left(\frac{2\gamma_k}{\psi}\right) < \sum_{k=0}^{N-1} e^{\frac{1}{\psi}}, \tag{2.32}$$

so at least one of the summands of the left part of the equation has to meet inequality (2.27)

$$e^{\frac{\gamma_k^2}{\psi}} > I_0\left(\frac{2\gamma_k}{\psi}\right). \tag{2.33}$$

However, using Lemma 2.1,

$$e^{\frac{\gamma_1^2}{\psi}} > I_0\left(\frac{2\gamma_1}{\psi}\right) \implies e^{\frac{\gamma_2^2}{\psi}} > I_0\left(\frac{2\gamma_2}{\psi}\right) \forall \gamma_2 \geq \gamma_1, \tag{2.34}$$

and, using the fact that $\tilde{\rho} > 0$ implies that $\gamma_k^2 < NP \forall k$,

$$e^{\frac{NP}{\psi}} > I_0\left(\frac{2\sqrt{NP}}{\psi}\right), \tag{2.35}$$

which contradicts our assumption.

□

These two last results show the fact that the insertion of a secondary signal is possible if and only if the secondary transmitter is able to enforce all the primary receivers with an active BER constraint by using a fraction of its available power, but not all of it. Therefore, we conclude that for values of $P > 4/N$ the insertion of the secondary message is always possible and for values of $P < P_0/N$ is impossible, being P_0 a nonzero value such that $e^{P_0/\psi_0} = I_0\left(2\sqrt{P_0}/\psi_0\right)$. For the value of ψ_0 predicted for the studied system, $P_0 \approx 1.67$.

2.5.3 Numerical approach and results

The extension of the previous optimization problems to the complete coverage zone implies the insertion of an infinite number of CB constraints (one for each of the infinite points in the coverage zone), so the problem can be seen to be a semi-infinite program (SIP), i.e., an optimization problem with a finite number of design variables, but an infinite number of constraints. This

problem is intractable due to the high dimensionality of the problem.[10]

We can reduce this dimensionality by grouping the N amplitude values $\tilde{\gamma}_1$, ..., $\tilde{\gamma}_N$ in M groups $\mathcal{G}_1$, ..., $\mathcal{G}_M$, such that the power allocation will be constant in each group, this is, $\tilde{\gamma}_j = \tilde{\gamma}_k \, \forall \, \tilde{\gamma}_j, \, \tilde{\gamma}_k \in \mathcal{G}_i$. Similarly to (2.9), we can rewrite the problem as

$$\begin{aligned} \text{minimize} \quad & -\tilde{\rho} \\ \text{subject to} \quad & \tilde{\eta}(\mathbf{x}, \tilde{\boldsymbol{\gamma}}, \tilde{\rho}, \boldsymbol{\phi}) \leq \eta_0 \, \forall \, \mathbf{x} \in \mathcal{C}_0, \\ & \tilde{\rho}^2 + \sum_{i=1}^{M} \tilde{\gamma}_i^2 \phi_i \leq 1 \\ & \sum_{i=1}^{M} \phi_i = 1 \\ & \phi_i \geq 0 \end{aligned} \tag{2.36}$$

where ϕ_i is the fraction of carriers in the i-th group, $\phi_i = \frac{|\mathcal{G}_i|}{N} \leq 1$, $|\mathcal{X}|$ denotes the cardinality of set $\mathcal{X}$, $\boldsymbol{\phi} = [\phi_1, ..., \phi_M]$ and

$$\tilde{\eta}(\mathbf{x}, \tilde{\boldsymbol{\gamma}}, \tilde{\rho}, \boldsymbol{\phi}) = \sum_{i=1}^{M} \phi_i e^{-\frac{1+\gamma_k^2(\mathbf{x})}{\psi(\mathbf{x})+2\rho^2(\mathbf{x})}} I_0 \left(\frac{2\gamma_k(\mathbf{x})}{\psi(\mathbf{x}) + 2\rho^2(\mathbf{x})} \right). \tag{2.37}$$

In this problem, the number of variables is $2M$: the secondary ratio $\tilde{\rho}$, the amplitudes for the different groups $\tilde{\gamma}_1$, ..., $\tilde{\gamma}_M$ and the corresponding fractions of carriers ϕ_1, ..., ϕ_{M-1}. The remaining fraction can be computed as $\phi_{M-1} = 1 - \sum_{i=1}^{M-1} \phi_i$. We will also assume that there is a large enough number of carriers in every group, so $0 \leq \phi_i \leq 1$, with $\phi_i \in \mathbb{R}$.

`MATLAB` function `fseminf` was used to obtain the solution of the optimization problem. This algorithm, of the discretization type [79], is based on a quasi-Newton sequential quadratic programming (SQP) algorithm applied to a finite number of constraints, as a result of the discretization of the semi infinite constraint. This optimization method will return a local minimum, but as the problem is not convex we cannot guarantee global optimality. In order to overcome this problem, the optimization algorithm was run 2,000 times for each pair of problem complexity and secondary position (M, r_s) with different initial random points, selecting afterwards the solution that provided the lowest value on the objective function.

Other parameters that describe the scenario (height of transmitters and receivers, transmit power...) are shown in Table 2.2, with the Okumura-Hata propagation model equations taken from [80]. In Figure 2.7 the obtained results are compared with those corresponding to a single

[10] For instance, this problem for a DVB-T system operating in the 8K-Mode will have 8193 variables, although this number can be slightly lower if we take into account the guard bands, for example.

Table 2.2: Parameters of the proposed scenario.

Parameter	**Value**
Height of primary transmitter	324m
EIRP of primary transmitter	70dBm
Position of primary transmitter	r=0Km, $\theta = 0$
Height of secondary transmitter	40m
EIRP of secondary transmitter	36dBm
Position of secondary transmitter	$r = r_s$ (Variable), $\theta = 0$
Height of receivers	30m
Thermal Noise Power	-105dBm
Propagation model	Modified Okumura-Hata, Urban Model
Discretization step for the SIP solver	200m

primary receiver on the border of the coverage zone (which might be thought to be a *worst case*, but as previously seen, the solution is more involved), and the two user scenario previously described by the numerical evaluation of (2.22).

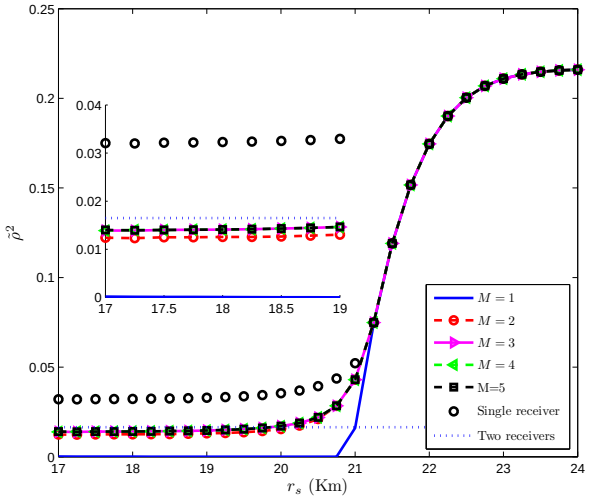


Figure 2.7: Fraction of transmit power of the secondary transmitter allocated to the secondary message as a function of the secondary transmitter position.

It can be seen that the lower values of r_s suffer from quite a large degradation with respect to the single user case, while for higher values this difference does not exist. The cause of this difference resides in the variability of P among the different receivers: while for low r_s values those receivers near the secondary transmitter have $P \to \infty$ and those near the coverage limit have $P \to 0$, in the high r_s case all the receivers that are affected by the secondary transmitter

experience relatively high values of P. Not surprisingly, the degradation in the low r_s zone is much more reduced if we compare the actual result with the simplified two-user scenario, as it is closer to the studied case. These results imply that the insertion of two receivers in very different situations reduces the power allocated to the secondary message, while incrementing this number of receivers (even to infinity) does not change the result too much.

For higher values of r_s, all the affected receivers have large values of P, so the optimum fraction of active carriers is one for all of them, and the obtained solution is equivalent to the *worst case* single-receiver solution. In this region the proposed approximation with two users is not realistic, as even the users on the border receive a much higher power contribution from the secondary transmitter, as previously pointed out.

With respect to the complexity of the problem (the number M of groups), for the lower values of r_s, similarly to the single receiver case, $M = 1$ results in a null power allocated to the secondary message, whereas for values of $M > 3$ no additional gain is attained. Note that the $M = 1$ case is equivalent to the transmission without the proposed power weighting in the frequency domain, which use is shown once again to be mandatory in order to achieve a nonzero rate for the secondary user. Moreover, the solution $M = 2$ (which was shown to be optimum for the single user case) suffers only a slight degradation with respect to $M = 3$. For higher values of r_s, the solution is to perform a uniform power allocation for the primary message, so the optimum number of groups is $M = 1$ and, therefore, further gain is not achieved by incrementing the order of the problem.

2.6 Bound verification: software and hardware simulations

In the previous sections, the transmit parameters of the secondary system have been designed according to the BER bound (2.6), due to the impossibility of finding a closed form expression for the actual BER. The objective of this section is to verify the aforementioned bound, thus providing an empirical proof of the previous theoretical results.

Computer simulations and hardware measurements were conducted to validate the proposed power allocation for the secondary transmitter. Hardware tests were performed in order to check the potential negative effects that the proposed transmission technique could have on the synchronization and estimation stages of a real receiver. The measurement set-up is described in Figure 2.8[11]. In Figure 2.9 it can be seen that, although the bound is not remarkably tight, its

[11] Due to hardware constraints, the power weighting was performed in the time domain, by means of a 32-ray equivalent baseband channel, where one ray was used to emulate the primary contribution, and the remaining 31 to perform the frequency power weighting. The results were averaged in both cases for 50 different pairs (n_0, θ)

use as a performance metric for the design of the proposed filtering provides a clear improvement in the primary link quality with respect to the simple transmission of the primary message, and, therefore, the achievable rate of the secondary system is going to be larger[12].

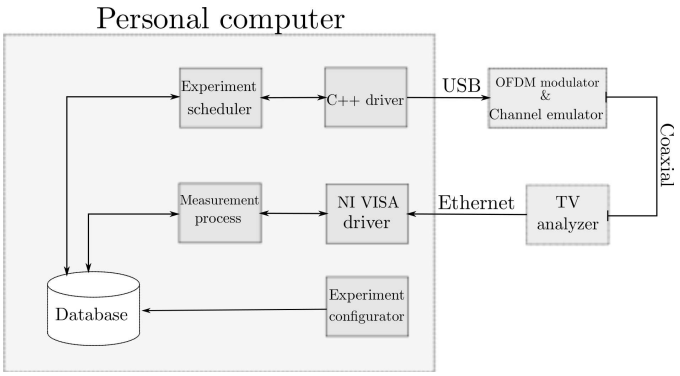


Figure 2.8: Hardware measurements set-up. The OFDM signal was generated with the DekTec DTU-215 USB-2 VHF/UHF modulator [63], which allows to simulate a 32 rays baseband equivalent channel (by defining the delay, amplitude and phase of each ray), and the addition of Gaussian noise. The BER was measured with Rohde & Schwarz ETL TV Analyzer [64], and captured with `MATLAB` via the National Instruments (NI) VISA driver. The experiments (CNR, channel model, number of measurements...) are configured, inserted into a relational database, and finally executed by the experiment scheduler.

In order to show the usefulness of the proposed filtering when dealing with higher order constellations, hardware tests were run also for a 64-QAM constellation, with the corresponding results shown in Figure 2.10. It can be seen that the filtered approach outperforms the non-filtered transmission in all the scenarios except for the $\gamma = 0.5$ one, where some artifacts were found. These effects are expected to disappear when using higher order transmission filters or weighting directly in the DFT domain. A similar behavior was obtained for a 16-QAM constellation, although the results are omitted due to space constraints.

2.7 Conclusions

In this chapter we considered the application of the overlay cognitive radio paradigm to a broadcast Single Frequency Network. Given the fact that the primary user Quality of Service is not simply a function of the Signal to Noise Ratio, our approach has taken into account the possible degradation of the primary service in strong line of sight environments due to

of delay and phase differences between primary and secondary contributions

[12] As the carrier to noise ratio (CNR) required for a given BER performance is going to be lower for the filtered transmission, the secondary user is allowed to allocate more power to the secondary message and, therefore, achieve a larger rate.

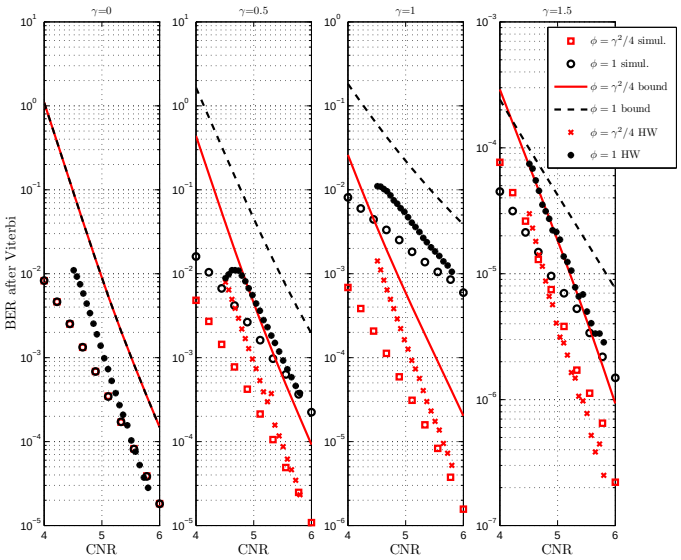


Figure 2.9: Analytical bounds, simulation and hardware (HW) results for multiple CNR and γ values. DVB-T waveform with Constellation: QPSK, Code Rate: 2/3. The CNR is calculated prior to the transmission of the secondary user, i.e. $\beta = \frac{1}{2}\Upsilon = \frac{1}{2}10^{(CNR+0.33)/10}$ [65]

the impossibility a of coherent combination of the primary waveforms. Optimum transmission strategies with respect to analytical BER bounds have been derived and analyzed via software simulations. The proposed approach was further verified by means of BER measurements in an actual hardware receiver. These modified transmission schemes were applied in order to maximize the transmission rate of a secondary user operating at the same frequency and location as the primary user. The primary QoS is assured by means of a coverage analysis whereby the BER is restricted to be above a given threshold. Spectrum reuse is successfully achieved without requiring any modification on the primary users, and with no cooperation with the primary transmitters.

The content in this chapter is an extended version of a paper published in IEEE Transactions on Wireless Communications [81] together with Prof. Carlos Mosquera and Prof. Fernando Perez-Gonzalez. Preliminary versions of the paper were also presented in SPAWC 2012 [82, 83] and CogArt 2011 [84].

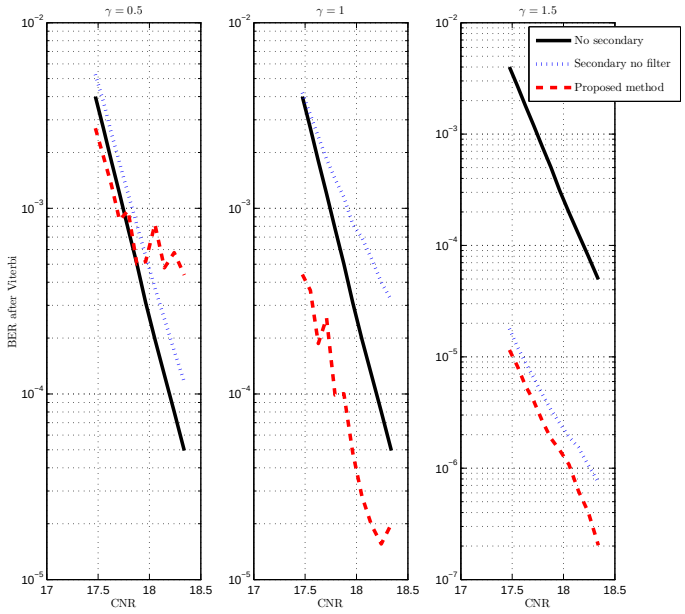


Figure 2.10: Hardware tests for a 64-QAM 2/3 DVB-T waveform. The proposed method (filtering with $\phi = \gamma^2/4$) is compared with the unfiltered approach ($\phi = 1$) and with the scenario without the secondary transmitter ($\gamma = 0$).

Appendix 2.A Optimality conditions for the optimization problem

The associated Karush-Kuhn-Tucker (KKT) conditions to problem (2.7) are

$$2e^{-\beta\gamma_k^2}\left(-\beta\gamma_k I_0(2\beta\gamma_k) + \beta I_1(2\beta\gamma_k)\right) + 2\lambda\gamma_k = 0 \ \forall k, \tag{2.38}$$

$$\lambda\left(\sum_{k=1}^{N}\gamma_k^2 - N\gamma^2\right) = 0, \quad \lambda \geq 0. \tag{2.39}$$

We will distinguish two cases: *1)* when $\sum_{k=1}^{N}\gamma_k^2 - N\gamma^2 < 0$, so λ is forced to be zero in order to meet condition (2.39), and *2)* when $\sum_{k=1}^{N}\gamma_k^2 - N\gamma^2 = 0$, so λ is not forced to be zero (we will refer to the constraint as *active* in that case).

Non-active constraint

In this case, we have $\lambda = 0$, so the resulting condition is

$$2e^{-\beta\gamma_k^2}\left(-\beta\gamma_k I_0(2\beta\gamma_k) + \beta I_1(2\beta\gamma_k)\right) = 0 \ \forall k \implies \\ \implies \gamma_k I_0(2\beta\gamma_k) = I_1(2\beta\gamma_k) \ \forall k. \tag{2.40}$$

Proposition 2.4. *The nontrivial solutions for (2.40) are in the interval* $\sqrt{\frac{\beta-1}{\beta}} \leq \gamma_k \leq 1$ *for* $\beta > 1$. *For* $\beta \leq 1$, *the only solution is* $\gamma_k = 0$.

Proof. We will assume $\gamma_k \neq 0$. Using (2.45), we can write $I_0(2\beta\gamma_k) \geq \frac{1}{\beta\gamma_k} I_1(2\beta\gamma_k)$. Combining this inequality with (2.40) we obtain $I_0(2\beta\gamma_k) \geq \frac{1}{\beta} I_0(2\beta\gamma_k)$, so $\beta \geq 1$.

Starting with (2.46), we have that $I_1^2(2\beta\gamma_k) > I_0(2\beta\gamma_k) I_2(2\beta\gamma_k)$, which together with (2.45) leads to

$$\gamma_k^2 I_0^2(2\beta\gamma_k) > I_0(2\beta\gamma_k) I_2(2\beta\gamma_k). \tag{2.41}$$

Finally, combining equations (2.45) and (2.40), we have that $I_2\left(2\beta\gamma_k\right) = \left(1 - \frac{1}{\beta}\right) I_0\left(2\beta\gamma_k\right)$, so (2.41) reads as $\gamma_k^2 I_0^2(2\beta\gamma_k) > \left(1 - \frac{1}{\beta}\right) I_0^2\left(2\beta\gamma_k\right)$, or, equivalently $\gamma_k > \sqrt{\frac{\beta-1}{\beta}}$.

□

Proposition 2.5. *The nontrivial solutions for (2.40) are not local minima of the optimization problem.*

Proof. In order to be a local minimum, the Hessian matrix of the objective function has to be positive definite. The Hessian is a diagonal matrix with elements

$$L(\gamma_k) = \left(L(\boldsymbol{\gamma})\right)_{k,k} = 2e^{-\beta\gamma_k^2} \times \\ \times \left(-4\beta^2\gamma_k I_1(2\beta\gamma_k) + \ \beta^2 I_2(2\beta\gamma_k) + \left(2\beta\gamma_k^2 + \beta - 1\right) I_0(2\beta\gamma_k)\right). \tag{2.42}$$

Moreover, we have that

$$\begin{aligned} -4\beta\gamma_k I_1(2\beta\gamma_k) + \beta I_2(2\beta\gamma_k) + & \\ \beta\left(2\beta\gamma_k^2 + \beta - 1\right) I_0(2\beta\gamma_k) & \overset{(i)}{=} \\ (-2\beta\gamma_k^2 - 1 + b) I_0(2\beta\gamma_k) + \beta I_0(2\beta\gamma_k) & \overset{(ii)}{=} \\ (-2\beta\gamma^2 + 2\beta - 2) I_0(2\beta\gamma_k) & \end{aligned} \tag{2.43}$$

where (i) derives from (2.40) and (ii) from (2.45) and (2.40).

As all the elements must be positive if $\boldsymbol{\gamma}$ is a local minimum, and since I_0 is strictly positive, the condition for the minimum is $\gamma_k < \sqrt{\frac{\beta-1}{\beta}}$, which contradicts proposition 2.4. □

Therefore, those points with some $\gamma_k \neq 0$ and inactive power constraint are not local minimum of the optimization problem.

Active constraint

In this case, we have the following necessary conditions for the point $\boldsymbol{\gamma}$ to be optimal

$$-2\beta\gamma_k e^{-\beta\gamma_k^2} I_0(2\beta\gamma_k) + 2\beta I_1(2\beta\gamma_k) e^{-\beta\gamma_k^2} + 2\lambda\gamma_k = 0 \qquad \forall i = 1, ..., N, \lambda \geq 0. \tag{2.44}$$

The condition is met if $\gamma_k^2 = 0$, as $I_1(0) = 0$. If $\gamma_k^2 \neq 0$, we can rewrite (2.44) as $\lambda = \beta e^{-\beta\gamma_k^2} \left(I_0(2\beta\gamma_k) - \frac{1}{\gamma_k} I_1(2\beta\gamma_k) \right) \forall\ k$, so it can be seen that those points of the form $\boldsymbol{\gamma}_M = [\mathbf{0}_{N-M}\ k\mathbf{1}_M]$ (or their corresponding permutations) where the power constraint is active are critical points of the Lagrangian.

As the function $\lambda(\gamma_k) = \beta e^{-\beta\gamma_k^2} \left(I_0(2\beta\gamma_k) - \frac{1}{\gamma_k} I_1(2\beta\gamma_k) \right)$ is non-injective, there are some points $\gamma_1 \neq \gamma_2$ such that $\lambda(\gamma_1) = \lambda(\gamma_2)$. However, these points were found to be local maxima of the objective function by checking the second order necessary conditions for optimality.

Regarding the second order conditions, some of the points under study can be local maxima, whereas others are local minima. As we are optimizing over the whole set of points, it is expected that the solution will lead to a global optimum.

2.A.1 Properties of the Bessel functions

$$I_v(t) = I_{v-2}(t) - \frac{2(v-1)}{t} I_{v-1}(t), \tag{2.45}$$

$$I_1^2(t) > I_0^2(t) I_2^2(t). \tag{2.46}$$

Appendix 2.B Asymptotic optimum value for ϕ

We start with $f(\phi) = (1-\phi)+\phi e^{-\beta\gamma^2/\phi} I_0(2\beta\gamma/\sqrt{\phi})$. In order to find a minimum of this function, we take its derivative

$$\begin{aligned} &\tfrac{d}{d\phi}\left(f(\phi)\right) = e^{-\beta\gamma^2/\phi}\times \\ &\times\left(I_0(2\beta\gamma/\sqrt{\phi})\left(1+\tfrac{\beta\gamma^2}{\phi}\right) - I_1(2\beta\gamma/\sqrt{\phi})\tfrac{\beta\gamma}{\sqrt{\phi}}\right) - 1. \end{aligned} \tag{2.47}$$

For high SNR, if we use the asymptotic approximation for the Bessel function $I_{0,1}(x) \approx \frac{e^x}{\sqrt{2\pi x}}$, and make the variable change $\alpha = \frac{\gamma}{\sqrt{\phi}}$, after equating (2.47) to zero we have

$$e^{2\beta\alpha}\left(\frac{1-\beta\alpha+\beta\alpha^2}{\sqrt{4\pi\beta\alpha}}\right) = e^{\beta\alpha^2}. \tag{2.48}$$

For asymptotically large β, the expression between parenthesis can be ignored, so the remaining expression is $e^{2\beta\alpha} = e^{\beta\alpha^2}$. Therefore, we have $\alpha = 2$, which leads to a value of $\phi = \frac{\gamma^2}{4}$. Note that this expression is only valid for values of $\gamma < 2$. In fact, if $\gamma > 2$, the solution of the problem is to transmit over all carriers with equal power, i.e., $\phi = 1$.

Appendix 2.C Two different receivers

If we constrain the frequency power weighting to have only two different levels, we can write the Chernoff bound as

$$\begin{aligned} \eta\left(\phi,\tilde{\gamma}_1,\tilde{\gamma}_2,\tilde{\rho},\mathbf{x}\right) = \phi e^{-\frac{1+\gamma_1^2(\mathbf{x})}{\psi(\mathbf{x})+2\rho^2(\mathbf{x})}} I_0\left(\frac{2\gamma_1(\mathbf{x})}{\psi(\mathbf{x})+2\rho^2(\mathbf{x})}\right) + \\ (1-\phi)e^{-\frac{1+\gamma_2^2(\mathbf{x})}{\psi(\mathbf{x})+2\rho^2(\mathbf{x})}} I_0\left(\frac{2\gamma_2(\mathbf{x})}{\psi(\mathbf{x})+2\rho^2(\mathbf{x})}\right), \end{aligned} \tag{2.49}$$

where $\phi \in [0,1] \subset \mathbb{R}$ since we are assuming a large enough number of carriers, and $\gamma_i(\mathbf{x}) = \sqrt{P(\mathbf{x})}\tilde{\gamma}_i$, $\rho(\mathbf{x}) = \sqrt{P(\mathbf{x})}\tilde{\rho}$. We will try to find a solution $(\phi,\tilde{\gamma}_1,\tilde{\gamma}_2,\tilde{\rho})$ that fulfills the BER constraint at both receivers even with the insertion of a secondary signal, i.e., $\eta\left(\phi,\tilde{\gamma}_1,\tilde{\gamma}_2,\tilde{\rho},\mathbf{x}\right) \leq \eta_0$, $\forall\mathbf{x} \in \{\mathbf{x}_n, \mathbf{x}_f\}$. For the nearby receiver, the signal coming from the primary transmitter will be negligible with respect to the secondary transmission, so we have that

$$\eta\left(\phi,\tilde{\gamma}_1,\tilde{\gamma}_2,\tilde{\rho},\mathbf{x}_n\right) = \phi e^{-\frac{\tilde{\gamma}_1^2}{2\tilde{\rho}^2}} + (1-\phi)e^{-\frac{\tilde{\gamma}_2^2}{2\tilde{\rho}^2}} \tag{2.50}$$

just by taking the limit $P(\mathbf{x}) \to \infty$ in (2.49).

Let us define $\tilde{\gamma}_m = \min\{\tilde{\gamma}_1, \tilde{\gamma}_2\}$. Then $\eta(\phi, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\rho}, \mathbf{x}_n) < e^{-\frac{\tilde{\gamma}_m^2}{2\tilde{\rho}^2}}$, so if we set $e^{-\frac{\tilde{\gamma}_m^2}{2\tilde{\rho}^2}} = \eta_0$, then $\tilde{\gamma}_m^2/\tilde{\rho}^2 = \Upsilon_0$, the CB constraint will be met.

With this restriction, we can write a simplified CB constraint for the distant receiver using[13] $\phi = \gamma^2/4$ and $\gamma_2 = \sqrt{\Upsilon_0}\rho$, with $\gamma^2 = \phi\gamma_1^2$ the total power spent in the carriers with amplitude $\gamma_1 = 2$, as

$$\eta(\gamma, \rho, \mathbf{x}_f) = \frac{\gamma^2}{4} e^{-\frac{5}{\psi_0 + 2\rho^2}} I_0\left(\frac{4}{\psi_0 + 2\rho^2}\right) + \left(1 - \frac{\gamma^2}{4}\right) e^{-\frac{1+\Upsilon_0\rho^2}{\psi_0+2\rho^2}} I_0\left(\frac{2\rho\sqrt{\Upsilon_0}}{\psi_0 + 2\rho^2}\right). \tag{2.51}$$

Let us define $f(\gamma, \rho) = \eta(\gamma, \rho, \mathbf{x}_f) - \eta_0$. As $P \to 0$, we have that $\gamma \to 0$ and $\rho \to 0$, so we can write $f(\gamma, \rho) = \frac{1}{2}[\gamma\ \rho]\nabla^2_{\gamma,\rho} f(0,0)[\gamma\ \rho]^T$, being $\nabla^2_{\gamma,\rho} f(0,0)$ a diagonal matrix with entries

$$\frac{\partial^2 f}{\partial \gamma^2}(0,0) = \frac{1}{2} e^{-5/\psi_0}\left(I_0\left(\frac{4}{\psi_0}\right) - e^{4/\psi_0}\right) \tag{2.52}$$

$$\frac{\partial^2 f}{\partial \rho^2}(0,0) = \frac{(4 - 2\Upsilon_0(\psi_0 - 1))e^{-1/\psi_0}}{\psi_0^2}. \tag{2.53}$$

The maximum value of ρ will be obtained when the power constraint is met with equality. In this case $\gamma^2 + \rho^2 + \left(1 - \frac{\gamma^2}{4}\right)\Upsilon_0\rho^2 = P$, so $\gamma^2 = \frac{4\left((\Upsilon_0+1)\rho^2 - P\right)}{\Upsilon_0\rho^2 - 4} \approx P - (\Upsilon_0 + 1)\rho^2$, where the last approximation holds provided $\Upsilon_0\rho^2$ is small enough with respect to 4. With these expressions, we get to the desired equation

$$\tilde{\rho}^2 = \frac{\rho^2}{P} = \frac{\frac{\partial^2 f}{\partial \gamma^2}(0,0)}{(\Upsilon_0 + 1)\frac{\partial^2 f}{\partial \gamma^2}(0,0) - \frac{\partial^2 f}{\partial \rho^2}(0,0)} = \frac{\psi_0^2\left(e^{4/\psi_0} - I_0\left(\frac{4}{\psi_0}\right)\right)}{e^{4/\psi_0}\left(\Upsilon_0(\psi_0 - 2)^2 + \psi_0^2 + 8\right) - (\Upsilon_0 + 1)\psi_0^2 I_0\left(\frac{4}{\psi_0}\right)}. \tag{2.54}$$

Appendix 2.D Proof of Lemma 2.1

Firstly, we will prove a proposition that will be used later.

Proposition 2.6. *Let $\beta > 0, \beta \in \mathbb{R}$, $K \in \mathbb{Z}$. If $\frac{\beta^K}{K!} < 1$, then $\frac{\beta^{K+\Delta}}{(K+\Delta)!} < 1$, $\forall \Delta \in \mathbb{N}$*

[13] With this simplification, $\gamma_1 = 2$, and as $\gamma_2 \approx 0, \gamma_m = \gamma_2$. In the following, we will omit the position indexing ($\mathbf{x}$), as we are only taking into account the far-off receiver.

Proof. The proof is quite straightforward. If we prove that $\frac{\beta^K}{K!} < 1 \implies \frac{\beta^{K+1}}{(K+1)!} < 1$ the proposition is proved by induction.

First, we characterize a property of K

$$\beta^K < K! < K^K \implies \beta < K, \tag{2.55}$$

so

$$\frac{\beta^{K+1}}{(K+1)!} = \frac{\beta}{K+1}\frac{\beta^K}{K!} < \frac{\beta}{K+1} < 1 \tag{2.56}$$

□

Proof of Lemma. We aim to prove that

$$e^{\gamma_0^2\beta} > I_0\left(2\beta\gamma_0\right) \implies e^{\gamma^2\beta} > I_0\left(2\beta\gamma\right), 0 < \gamma_0 \leq \gamma \tag{2.57}$$

Let us define

$$F\left(\gamma\right) = e^{\beta\gamma^2} - I_0\left(2\beta\gamma\right). \tag{2.58}$$

We will prove that $F\left(\gamma\right) > 0 \implies F'\left(\gamma\right) > 0$, which is a sufficient condition for (2.57). We will use the series expansion of both the exponential and Bessel function [85]

$$e^x = \sum_{K=0}^{\infty} \frac{x^K}{K!} \tag{2.59}$$

$$I_0\left(x\right) = \sum_{K=0}^{\infty} \frac{\left(\frac{1}{4}x^2\right)^K}{(K!)^2}, \tag{2.60}$$

so we can write (2.58) as

$$F\left(\gamma\right) = \sum_{K=0}^{\infty} \gamma^{2K}\left(\frac{\beta^K}{K!}\left(1 - \frac{\beta^K}{K!}\right)\right) = \sum_{K=0}^{\infty} \gamma^{2K} b_K, \tag{2.61}$$

with $b_K = \frac{\beta^K}{K!}\left(1 - \frac{\beta^K}{K!}\right)$, and its derivative as

$$F'\left(\gamma\right) = \sum_{K=0}^{\infty} \frac{2K}{\gamma}\gamma^{2K}\left(\frac{\beta^K}{K!}\left(1 - \frac{\beta^K}{K!}\right)\right). \tag{2.62}$$

Using proposition 2.6 we can write

$$F(\gamma) = \sum_{K=0}^{K_0} \gamma^{2K} b_K + \sum_{K=K_0+1}^{\infty} \gamma^{2K} b_K, \tag{2.63}$$

where $b_K \leq 0 \, \forall K \leq K_0$ and $b_K \geq 0 \, \forall K > K_0$. Therefore, if $F(\gamma) > 0$,

$$\begin{aligned} F'(\gamma) &= \sum_{K=0}^{K_0} \frac{2K}{\gamma} \gamma^{2K} b_K + \sum_{K=K_0+1}^{\infty} \frac{2K}{\gamma} \gamma^{2K} b_K \\ &> \sum_{K=0}^{K_0} \frac{2K_0}{\gamma} \gamma^{2K} b_K + \sum_{K=K_0+1}^{\infty} \frac{2(K_0+1)}{\gamma} \gamma^{2K} b_K \\ &> \frac{2K_0}{\gamma} \left(\sum_{K=0}^{K_0} \gamma^{2K} b_K + \sum_{K=K_0+1}^{\infty} \gamma^{2K} b_K \right) = \frac{2K_0}{\gamma} F(\gamma) > 0 \end{aligned} \tag{2.64}$$

□

Chapter 3

Broadcast Networks with Multiple Transmitters: Extension of Performance Results to NLOS Scenarios

Contents

3.1 Introduction

In the previous chapter we investigated the insertion of a secondary transmitter in a multicarrier broadcast network modeling the propagation channel from each both primary and secondary transmitters as frequency flat. One of the main results was the fact that the insertion of a cooperative secondary transmitter (i.e., a transmitter that is not transmitting the secondary message) might not increase the performance of primary receivers. In this chapter we extend this result to include general Rician channels in different signal to noise ratio (SNR) regimes. The additive white Gaussian noise (AWGN) (treated in Chapter 2) and Rayleigh channels are obtained as special cases of the Rice fading channel.

In this chapter we focus on the effect of inserting a secondary transmitter conveying the same information as the primary one. The results on this chapter can be extended to include a secondary message, just like in Chapter 2, but also used on its own to calculate performance and coverage of single frequency network (SFN) deployments. For example, it is usually assumed that SFN offers a performance gain due to the reception of signal from two or more transmitters [18, 86]. In this chapter, we show that this power gain does not always translate directly into a performance gain due to the presence of SFN echoes. We quantify the effect of the artificial multipath by the use of effective SNR metrics, as well as two different methods to overcome this degradation, namely the use of space-time codes, and the use of a transmit filter like the one in Chapter 2.

The remaining of the chapter is structured as follows: Section 3.2 describes the system model; Section 3.3, 3.4 and 3.5 analyze the AWGN, Rayleigh and Rice channels, respectively; we analyze AWGN and Rayleigh channels separately from Rice, although the two former can be obtained from the latter; Section 3.6 introduces a general result for fading channels in high and low SNR regimes; Section 3.7 and 3.8 present two techniques to improve the SFN performance: the use of Alamouti space time codes, and the use of a filter at the secondary transmitter, similarly to Chapter 2; Section 3.9 presents the numerical results; Section 3.10 concludes the chapter.

3.2 System model

Throughout the chapter, we will compare a scenario with a single transmitter (the *only-primary* scenario) with an SFN operation comprising the presence of another transmitter (the *secondary* transmitter). The system under study uses orthogonal frequency division multiplexing (OFDM) for data transmission. We will assume perfect synchronization, perfect channel estimation, and

an overall channel length shorter than the cyclic prefix (CP) length. Therefore, the received baseband signal after CP removal can be written as

$$y_n = (h_n + g_n) \circledast x_n + w_n \tag{3.1}$$

where h_n denotes the channel from the original transmitter, g_n the channel from the secondary transmitter, x_n is the (normalized) time-domain signal, $w_n \sim \mathcal{CN}\left(0, \sigma^2\right)$, and $\circledast$ denotes the circular convolution operator. Therefore, in the Discrete Fourier Transform (DFT) domain, (3.1) reads as

$$Y_k = (H_k + G_k) X_k + W_k \tag{3.2}$$

with H_k, G_k, X_k and W_k the N points discrete Fourier transform (DFT) of h_n, g_n, x_n and w_n.

The average SNR metric (ASM) of the complete system (3.2) is

$$\bar{\gamma}_S \triangleq \frac{1}{N} \sum_{k=1}^{N} \frac{|H_k + G_k|^2}{\sigma^2} \approx \frac{1}{N} \sum_{k=1}^{N} \frac{|H_k|^2 + |G_k|^2}{\sigma^2} \tag{3.3}$$

where the approximation holds if both H_k and G_k are independently drawn from a zero mean probability distribution[1], so $E\{H_k G_k\} = 0$. In the same way, the ASM in absence of the secondary transmitter is

$$\bar{\gamma}_0 \triangleq \frac{1}{N} \sum_{k=1}^{N} \frac{|H_k|^2}{\sigma^2}, \tag{3.4}$$

so the SFN gain (we will call it average SFN gain (ASG)) can be defined as

$$\Delta\bar{\gamma} \triangleq \frac{\bar{\gamma}_S}{\bar{\gamma}_0} \approx 1 + \frac{\sum_{k=1}^{N} |H_k|^2}{\sum_{k=1}^{N} |G_k|^2}, \tag{3.5}$$

so we have that $\Delta\bar{\gamma} \geq 1$, leading to a positive ASG (in dB). This is the usual approximation when calculating the SFN gain [86].

The performance of multicarrier systems, however, is not just a direct function of the ASM, but also of the distribution of the SNR on the different carriers. Effective SNR metrics (ESM) have been developed [39, 40] with the purpose of predicting the performance (in terms of packet error rate - PER, or frame error rate - FER) of a multicarrier system in the presence of a frequency selective channel. The effective SNR $\hat{\gamma}$ can be written as a function of the SNR of the

[1]This is a good approximation even in the case of systems working with a strong line of sight, as a uniform phase term in the signal received from one of the transmitters makes the resulting ASM to follow (3.3).

N carriers (γ_i, $i = 1, ..., N$) as

$$\hat{\gamma} \triangleq \Phi^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}\Phi\left(\gamma_k\right)\right) \tag{3.6}$$

with

$$\gamma_k = \frac{|H_k + G_k|^2}{\sigma^2} \tag{3.7}$$

where the function Φ is chosen as a concave increasing function, or convex decreasing function.

In particular, the mutual information ESM (MIESM) has been found to be of special interest because of its accuracy in predicting the PER [41]. The function Φ associated to the MIESM[2], taken from [73], is

$$\Phi\left(\gamma\right) = \frac{1}{M\log_2 M}\sum_{m=1}^{M}\mathbb{E}_U\left[\log_2\left(\sum_{k=1}^{M}e^{-\frac{|X_m - X_k + U| - |U|^2}{1/\gamma}}\right)\right] \tag{3.8}$$

where $U \sim \mathcal{CN}\left(0, 1/\gamma\right)$, and X_m, $m = 1, ..., M$ are the complex constellation points. (3.8) can be approximated as [41]

$$\Phi\left(\gamma\right) = \sum_{l=1}^{L}\phi_l e^{-\beta_l\gamma} \tag{3.9}$$

where $\sum_{l=1}^{L}\phi_l = 1$, and $\phi_l \geq 0$ and $\beta_l \geq 0$ are parameters that have to be properly chosen in order to fit the actual value of (3.8).

Following this ESM approach, we define the effective SFN gain (ESG) as

$$\Delta\hat{\gamma} \triangleq \frac{\hat{\gamma}_S}{\hat{\gamma}_0}, \tag{3.10}$$

with

$$\hat{\gamma}_S \triangleq \Phi^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}\Phi\left(\frac{|H_k + G_k|^2}{\sigma^2}\right)\right) \tag{3.11}$$

and

$$\hat{\gamma}_0 \triangleq \Phi^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}\Phi\left(\frac{|H_k|^2}{\sigma^2}\right)\right). \tag{3.12}$$

As ESM are good PER predictors, those scenarios with $\Delta\hat{\gamma} > 1$ will benefit from the insertion of a secondary transmitter, whereas those with $\Delta\hat{\gamma} < 1$ offer a worse performance than the single transmitter case. In the remaining of the chapter, we will calculate the ESG for different channel

[2]For the sake of simplicity, (3.8) is different (in a constant term) from the original expression in [73], but the overall ESM is the same.

models.

3.3 AWGN channel

In the AWGN case, we can set $H_k = 1$, $k = 1, ..., N$, and

$$G_k = \alpha e^{-j\left(\theta + 2\pi n_0 \frac{k}{N}\right)} \tag{3.13}$$

where α accounts for the different amplitude of G_k, n_0 accounts for the delay between the two contributions, and θ is the difference between phases.

Since $H_k = 1$, we have

$$\hat{\gamma}_0 = \bar{\gamma}_0 = \frac{1}{\sigma^2}. \tag{3.14}$$

The calculation of $\hat{\gamma}_S$ is more involved. First, note that

$$|H_k + G_k|^2 = 1 + \alpha^2 + 2\alpha \cos\left(\theta + 2\pi n_0 \frac{k}{N}\right), \tag{3.15}$$

so in the degenerate case of $n_0 = 0, \theta = 0$ we have $|H_k + G_k|^2 = (1+\alpha)^2 \,\forall\, k$, and if $n_0 = 0, \theta = \pi$, $|H_k + G_k|^2 = (1-\alpha)^2 \,\forall\, k$. However, for usual values of n_0, the N different arguments of the cosine

$$\theta + 2\pi n_0 \frac{k}{N},\ k = 1 \ldots,\ N, \tag{3.16}$$

will conform an approximately uniform sampling of the interval $(0, 2\pi]$, so we can write for a sufficiently large number of carriers

$$\Phi\left(\hat{\gamma}_S\right) = \frac{1}{N} \sum_{k=1}^{N} \Phi\left(\frac{|H_k + G_k|^2}{\sigma^2}\right) \approx \mathbb{E}_a\left[\Phi\left(\frac{1 + \alpha^2 + 2\alpha\cos\left(a\right)}{\sigma^2}\right)\right] \tag{3.17}$$

with $a \sim \mathcal{U}\left(0, 2\pi\right]$. Substituting (3.9) in (3.17) we arrive to

$$\Phi\left(\hat{\gamma}_S\right) = \frac{1}{2\pi} \int_0^{2\pi} \sum_{l=1}^{L} \phi_l e^{-\beta_l\left(1+\alpha^2+2\alpha\cos(a)\right)/\sigma^2} da = \sum_{l=1}^{L} \phi_l e^{-\beta_l\left(1+\alpha^2\right)/\sigma^2} I_0\left(\frac{2\beta_l\alpha}{\sigma^2}\right) \tag{3.18}$$

with $I_0\left(\cdot\right)$ the zeroth order modified Bessel function of the first kind.

In order to gain insight on the implications of (3.18), we will focus on the case[3] with $L = 1$ (or, equivalently, the exponential ESM - EESM [40]), so

$$\Phi^{-1}(x) = -\frac{1}{\beta}\log(x) \tag{3.19}$$

where we denote $\beta \triangleq \beta_1$ for the sake of simplicity. Note that $\phi_1 = 1$, so a simple form for $\hat{\gamma}_S$ is obtained by applying (3.19) to (3.18)

$$\hat{\gamma}_S = -\frac{1}{\beta}\log\left(e^{-\beta(1+\alpha^2)/\sigma^2} I_0\left(\frac{2\beta\alpha}{\sigma^2}\right)\right) = \frac{1+\alpha^2}{\sigma^2} - \frac{1}{\beta}\log\left(I_0\left(\frac{2\beta\alpha}{\sigma^2}\right)\right). \tag{3.20}$$

The ESG for the AWGN channel is readily obtained by dividing (3.20) by (3.14):

$$\Delta\hat{\gamma} = 1 + \alpha^2 - \frac{\sigma^2}{\beta}\log\left(I_0\left(\frac{2\beta\alpha}{\sigma^2}\right)\right). \tag{3.21}$$

This expression has two clearly differentiated components:

- The term $1 + \alpha^2 = \Delta\bar{\gamma}$ represents the average power gain due to the two different components.
- The term $\frac{\sigma^2}{\beta}\log\left(I_0\left(\frac{2\beta\alpha}{\sigma^2}\right)\right)$ represents the degradation caused by the transformation of a flat fading channel into a multipath one.

Therefore, the ESG will be positive (in dB) if

$$e^{C\alpha^2} > I_0(2C\alpha) \tag{3.22}$$

with $C = \frac{\beta}{\sigma^2}$. Although the actual value of the ESG has to be computed numerically, we will analyze the asymptotic ESG in the low and high SNR regimes.

3.3.1 High SNR

In the high SNR regime, we can approximate [85]

$$I_0(x) = \frac{1}{\sqrt{2\pi x}} e^x, \tag{3.23}$$

[3]Unfortunately, we are not able to provide closed form expressions for $L > 1$ due to the impossibility of obtaining a closed form inverse function for (3.9).

so the limit ESM reads as

$$\lim_{\sigma^2\to 0}\Delta\hat{\gamma}=\lim_{\sigma^2\to 0}1+\alpha^2-2\alpha+\frac{\sigma^2}{2\beta}\log\left(2\pi\frac{2\beta\alpha}{\sigma^2}\right)=(1-\alpha)^2. \tag{3.24}$$

In this case, values of $\alpha<2$ will lead to a lower ESM value, with two transmitters performing worse than just a single one. Note that, in this high SNR case, the performance is always decreased with respect to having only the transmitter from which more power is received (a scenario with $\alpha=2$, for example, can be transformed into a scenario with $\alpha=1/2$ just by changing the roles of primary and secondary transmitters).

3.3.2 Low SNR

In this case, we can approximate for $x\approx 0$ [85]

$$I_0(x)=1+\frac{1}{4}x^2 \tag{3.25}$$

and

$$\log(1+x)=x, \tag{3.26}$$

leading to

$$\lim_{\sigma^2\to\infty}\Delta\hat{\gamma}=\lim_{\sigma^2\to\infty}1+\alpha^2-\frac{\sigma^2}{4\beta}\left(\frac{2\beta\alpha}{\sigma^2}\right)^2=1+\alpha^2=\Delta\bar{\gamma}, \tag{3.27}$$

so in the low SNR regime the power gain is much more important than the channel degradation, and the ESG is always positive.

This difference can be contrasted by the propositions in Section 3.6, where it is shown that for low SNR, the EESM tends to the average SNR, and for high SNR, the EESM tends to the minimum SNR.

3.4 Rayleigh channel

In the case of a Rayleigh channel, we have $H_k \sim \mathcal{CN}\left(0, \sigma_h^2\right)$, $G_k \sim \mathcal{CN}\left(0, \sigma_g^2\right)$. We obtain now the ESM for a generic Rayleigh channel $F_k \sim \mathcal{CN}\left(0, \sigma_f^2\right)$. We have that

$$\Phi\left(\hat{\gamma}\right) = \frac{1}{N}\sum_{k=1}^{N}\Phi\left(\frac{|F_k|^2}{\sigma^2}\right) \approx \mathbb{E}_{F_k}\left[\Phi\left(\frac{|F_k|^2}{\sigma^2}\right)\right] = \mathbb{E}_{F_k}\left[\sum_{l=1}^{L}\phi_l e^{-\beta_l\frac{|F_k|^2}{\sigma^2}}\right] \tag{3.28}$$

where the approximation holds for a sufficiently large number of carriers. We can write $|F_k|^2 = \frac{\sigma_f^2}{2}X$, with $X \sim \chi_2^2$, so

$$\begin{aligned}\Phi\left(\hat{\gamma}\right) &= \sum_{l=1}^{L}\phi_l\int_0^{+\infty}\frac{1}{2}e^{-\frac{x}{2}}e^{-\frac{\beta_l\sigma_f^2 x}{2\sigma^2}}dx \\ &= \sum_{l=1}^{L}\phi_l\frac{1}{1+\beta_l\sigma_f^2/\sigma^2} = \sum_{l=1}^{L}\phi_l\frac{1}{1+\beta_l\bar{\gamma}}\end{aligned} \tag{3.29}$$

as the average SNR reads as

$$\bar{\gamma} = \frac{1}{N}\sum_{k=1}^{N}\frac{|F_k|^2}{\sigma^2} \approx \frac{\sigma_f^2}{\sigma^2} \tag{3.30}$$

for large N. Therefore, as $H_k + G_k \sim \mathcal{CN}\left(0, \sigma_h^2 + \sigma_g^2\right)$, we have for the EESM metric

$$\Delta\hat{\gamma} = \frac{\log\left(1+\beta\frac{\sigma_g^2+\sigma_h^2}{\sigma^2}\right)}{\log\left(1+\beta\frac{\sigma_h^2}{\sigma^2}\right)} = \frac{\log\left(1+\beta\bar{\gamma}_S\right)}{\log\left(1+\beta\bar{\gamma}_0\right)}. \tag{3.31}$$

In this case, as the Rayleigh channel can be thought to be the sum of infinite multipath components, the insertion of the secondary transmitter does not cause the channel degradation as in the AWGN case, so the power gain always provides $\Delta\hat{\gamma} > 1$.

3.5 Rician channel

In the Rice case, both channels have a line of sight (LOS) and a non line of sight (NLOS) component.

First, we will study the single transmitter case, with $H_k \sim \mathcal{CN}\left(\mu_h, \sigma_h^2\right)$. Under the suffi-

ciently large number of carriers assumption, we have that

$$\Phi\left(\hat{\gamma}_0\right) = \frac{1}{N}\sum_{k=1}^{N}\Phi\left(\frac{|H_k|^2}{\sigma^2}\right) \approx \mathbb{E}_x\left[\Phi\left(\frac{x^2}{\sigma^2}\right)\right] \tag{3.32}$$

where x is Rician distributed with parameters $\nu = |\mu_h|$ and $\sigma_x^2 = \sigma_h^2/2$. Therefore, we can write

$$\begin{aligned}\Phi\left(\hat{\gamma}_0\right) &= \int_0^\infty \sum_{l=1}^{L}\phi_l e^{-\frac{\beta_l x^2}{\sigma^2}} f_x\left(x\right)dx \\ &= \sum_{l=1}^{L}\phi_l\int_0^\infty e^{-\frac{\beta_l x^2}{\sigma^2}}\frac{x}{\sigma_x^2}e^{-\frac{x^2+\nu^2}{2\sigma_x^2}}I_0\left(\frac{x\nu}{\sigma_x^2}\right)dx \\ &= \sum_{l=1}^{L}\phi_l e^{-\frac{\nu^2}{2\sigma_x^2}}\frac{1}{\sigma_x^2}\int_0^\infty x e^{-\frac{x^2}{2}\left(\frac{2\beta_l}{\sigma^2}+\frac{1}{\sigma_x^2}\right)}I_0\left(\frac{x\nu}{\sigma_x^2}\right)dx.\end{aligned} \tag{3.33}$$

Now, if we denote $p_l = \sqrt{\frac{2\beta_l}{\sigma^2}+\frac{1}{\sigma_x^2}}$, $a = \frac{\nu}{\sigma_x^2}$, we can follow [87] and write

$$\int_0^\infty x e^{-p_l^2x^2/2}I_0\left(ax\right)dx = \frac{1}{p_l^2}e^{\frac{a^2}{2p_l^2}}Q\left(a/p_l,0\right) = \frac{1}{p_l^2}e^{\frac{a^2}{2p_l^2}} \tag{3.34}$$

with $Q(x,0) = 0\,\forall\, x$ the Marcum Q-function, so (3.33) can be written as

$$\Phi\left(\hat{\gamma}_0\right) = \sum_{l=1}^{L}\phi_l\frac{1}{\sigma_x^2p_l^2}\exp\left(-\frac{\nu^2}{2\sigma_x^2}\left(1-\frac{1}{\sigma_x^2p_l^2}\right)\right), \tag{3.35}$$

or, as

$$\sigma_x^2p_l^2 = \frac{2\beta_l\sigma_x^2}{\sigma^2}+1 \tag{3.36}$$

we have that

$$\Phi\left(\hat{\gamma}_0\right) = \sum_{l=1}^{L}\phi_l\frac{1}{\frac{2\beta_l\sigma_x^2}{\sigma^2}+1}\exp\left(-\frac{\nu^2}{2\sigma_x^2}\left(1-\frac{1}{\frac{2\beta_l\sigma_x^2}{\sigma^2}+1}\right)\right), \tag{3.37}$$

or, equivalently

$$\Phi\left(\hat{\gamma}_0\right) = \sum_{l=1}^{L}\phi_l\frac{\sigma^2}{\beta_l\sigma_h^2+\sigma^2}\exp\left(-\frac{\nu^2\beta_l}{\beta_l\sigma_h^2+\sigma^2}\right). \tag{3.38}$$

Note that if we force $\nu = 0$ in (3.38), we arrive to the expression for the Rayleigh channel, 3.29), and if we force $\sigma_h^2 = 0$, we arrive to the expression for the AWGN channel (3.14). Like in the previous cases, we will obtain a closed form expression for the EESM, in order to gain

insight on the problem. In this case

$$\begin{aligned}\hat{\gamma}_0 &= -\frac{1}{\beta}\log\left(\frac{\sigma^2}{\beta\sigma_h^2+\sigma^2}\exp\left(-\frac{\nu^2\beta}{\beta\sigma_h^2+\sigma^2}\right)\right) \\ &= \frac{\nu^2}{\beta\sigma_h^2+\sigma^2}+\frac{1}{\beta}\log\left(1+\beta\frac{\sigma_h^2}{\sigma^2}\right).\end{aligned} \tag{3.39}$$

Now, if we define $\bar{\gamma}_N \triangleq \frac{\sigma_h^2}{\sigma^2}$ the average SNR due to the multipath component, and $\bar{\gamma}_L \triangleq \frac{\nu^2}{\sigma^2}$ the average SNR caused by the direct component, we can write

$$\hat{\gamma}_0 = \frac{\bar{\gamma}_L}{\beta\bar{\gamma}_N+1}+\frac{1}{\beta}\log\left(1+\beta\bar{\gamma}_N\right), \tag{3.40}$$

so we can find two different contributions to the ESM

- The LOS component $\frac{\bar{\gamma}_L}{\beta\bar{\gamma}_N+1}$ is similar to the one in AWGN, but in this case the NLOS contribution acts as an additional noise source (it could be thought as a *self-interference* term).
- The NLOS component $\frac{1}{\beta}\log\left(1+\beta\bar{\gamma}_N\right)$ is the same as in the Rayleigh case.

At this point, it is not clear what the effect of $\bar{\gamma}_N$ is in the performance of the system. The LOS contribution to (3.40) decreases with $\bar{\gamma}_N$, but the logarithmic term increases with $\bar{\gamma}_N$. The question is equivalent to determining whether a system with direct and multipath components performs better than a system with only a direct component. In the following, we analyze this effect in the high and low SNR regimes.

3.5.1 NLOS effect in low SNR

We will write $\bar{\gamma}_N = s\tilde{\gamma}_N$, $\bar{\gamma}_L = s\tilde{\gamma}_L$, and calculate the ESM gain with respect to an AWGN channel with only the LOS component. In this case

$$\begin{aligned}\Delta\hat{\gamma} &= \lim_{s\to 0}\frac{\frac{s\tilde{\gamma}_L}{\beta s\tilde{\gamma}_N+1}+\frac{1}{\beta}\log\left(1+\beta s\tilde{\gamma}_N\right)}{s\tilde{\gamma}_L} \\ &= \lim_{s\to 0}\frac{1}{\beta s\tilde{\gamma}_N+1}+\frac{1}{\beta s\tilde{\gamma}_L}\log\left(1+\beta\bar{s}\tilde{\gamma}_N\right) \\ &= 1+\frac{\tilde{\gamma}_N}{\tilde{\gamma}_L} = 1+\frac{\bar{\gamma}_N}{\bar{\gamma}_L} = \Delta\bar{\gamma}\end{aligned} \tag{3.41}$$

so in the low SNR regime, once again, we verify that the power gain is much more important than the channel degradation.

3.5.2 NLOS effect in high SNR

Following the same approach as in the low SNR case,

$$\Delta\hat{\gamma} = \lim_{s\to\infty} \frac{\frac{s\tilde{\gamma}_L}{\beta s\tilde{\gamma}_N+1} + \frac{1}{\beta}\log\left(1+\beta s\tilde{\gamma}_N\right)}{s\tilde{\gamma}_L} = 0. \tag{3.42}$$

As in the high SNR regime the ESM $\hat{\gamma}$ increases logarithmically while the ASM $\bar{\gamma}$ increases linearly.

Now, we proceed to calculate the ESM in an SFN scenario with two transmitters. In this case, we have that $H_k = \mathcal{CN}\left(1,\sigma_h^2\right)$ and $G_k = \mathcal{CN}\left(\alpha e^{-j\left(\theta+2\pi n_0\frac{k}{N}\right)},\sigma_g^2\right)$. Like in the AWGN case, we will assume that the phase term in G_k conforms an approximately uniform sampling of the interval $[0,2\pi)$, so we have $a\sim\mathcal{U}\left[0,2\pi\right)$, and, therefore, the distribution of $Z_k = G_k + H_k$ conditioned on a is $Z_k = \mathcal{CN}\left(1+\alpha e^{-ja},\sigma_g^2+\sigma_h^2\right)$. With this, we can approximate

$$\Phi\left(\hat{\gamma}_S\right) = \frac{1}{N}\sum_{k=1}^{N}\Phi\left(\frac{|Z_k|^2}{\sigma^2}\right) \approx \mathbb{E}_Z\left[\frac{|Z^2|}{\sigma^2}\right]. \tag{3.43}$$

We will solve the expectation by conditioning on a

$$\Phi\left(\hat{\gamma}_S\right) = \int_0^{2\pi}\mathbb{E}_{Z|a}\left[\frac{|Z^2|}{\sigma^2}\middle| a\right] f_a(a)da. \tag{3.44}$$

Now, note that $\mathbb{E}_{Z|a}\left\{\frac{|Z^2|}{\sigma^2}\middle| a\right\}$ is a particular case of (3.32), so a closed form expression follows (3.38) with $\nu = |1+\alpha e^{-ja}|$ and the variance of the NLOS component is $\sigma_z^2 = \sigma_g^2+\sigma_h^2$ instead of σ_h^2. Therefore $\Phi\left(\hat{\gamma}_S\right)$ can be obtained by just averaging over a as

$$\begin{aligned}
\Phi\left(\hat{\gamma}_S\right) &= \frac{1}{2\pi}\int_0^{2\pi}\sum_{l=1}^{L}\phi_l\frac{\sigma^2}{\beta_l\sigma_z^2+\sigma^2}\exp\left(-\frac{\nu^2\beta_l}{\beta_l\sigma_z^2+\sigma^2}\right)da \qquad (3.45)\\
&= \sum_{l=1}^{L}\phi_l\frac{\sigma^2}{\beta_l\sigma_z^2+\sigma^2}\exp\left(-\frac{\beta_l\left(1+\alpha^2\right)}{\beta_l\sigma_z^2+\sigma^2}\right)\frac{1}{2\pi}\int_0^{2\pi}\exp\left(-\frac{2\alpha\beta_l\cos\left(a\right)}{\beta_l\sigma_z^2+\sigma^2}\right)da\\
&= \sum_{l=1}^{L}\phi_l\frac{\sigma^2}{\beta_l\sigma_z^2+\sigma^2}\exp\left(-\frac{\beta_l(1+\alpha^2)}{\beta_l\sigma_z^2+\sigma^2}\right)I_0\left(\frac{2\beta_l\alpha}{\beta_l\sigma_z^2+\sigma^2}\right).
\end{aligned}$$

Finally, $\hat{\gamma}_S$ for $L=1$ can be written as

$$\hat{\gamma}_S = \frac{1}{\beta}\log\left(1+\beta\frac{\sigma_z^2}{\sigma^2}\right) + \frac{1+\alpha^2}{\beta\sigma_z^2+\sigma^2} - \frac{1}{\beta}\log\left(I_0\left(\frac{2\beta\alpha}{\beta\sigma_z^2+\sigma^2}\right)\right) \tag{3.47}$$

which has three different terms:

- $\frac{1}{\beta}\log\left(1+\beta\frac{\sigma_z^2}{\sigma^2}\right)$ reflects the ESM gain due to the NLOS component.
- $\frac{1+\alpha^2}{\beta\sigma_z^2+\sigma^2}$ includes the power gain $(1+\alpha^2)$ due to the insertion of LOS component coming from the secondary transmitter, and includes the channel degradation due to the presence of an NLOS component of power σ_z^2.
- $\frac{1}{\beta}\log\left(I_0\left(\frac{2\beta\alpha}{\beta\sigma_z^2+\sigma^2}\right)\right)$ reflects the channel degradation due to the LOS component of the secondary transmitter.

Finally, if we assume the same power ratio between the LOS and NLOS components in both transmitters (the K Rician factor), so $H_k \sim \mathcal{CN}\left(1, K^{-1}\right)$ and $G_k \sim \mathcal{CN}\left(\alpha e^{-j\left(\theta+2\pi n_0\frac{k}{N}\right)}, \alpha^2 K^{-1}\right)$, then the ESG can be written as

$$\Delta\hat{\gamma} = \frac{\frac{1}{\beta}\log\left(1+\beta\frac{1+\alpha^2}{K\sigma^2}\right)+\frac{1+\alpha^2}{\beta\frac{1+\alpha^2}{K}+\sigma^2}-\frac{1}{\beta}\log\left(I_0\left(\frac{2\beta\alpha}{\beta\frac{1+\alpha^2}{K}+\sigma^2}\right)\right)}{\frac{1}{\beta\bar{K}^{-1}+\sigma^2}+\frac{1}{\beta}\log\left(1+\beta\frac{K^{-1}}{\sigma^2}\right)}. \tag{3.48}$$

3.6 Effect of SNR regime for general frequency selective channels

From the previous sections we can extract that in the low SNR regime the SFN gain is always greater than zero, while in the high SNR regime the results depend on the underlying channel structure. In this section, we present a general result that explains the behavior of a family of ESMs (including the EESM) for general frequency selective channels in high and low SNR: for low SNR values, the EESM tends to the average SNR, while for high SNR values it tends to the minimum SNR.

Proposition 3.1. *Let*

$$\hat{\gamma} = \Phi^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}\Phi\left(\gamma_k\right)\right) \tag{3.49}$$

be an ESM with the following properties:

$$\left.\frac{\partial\Phi\left(\gamma\right)}{\partial\gamma}\right|_{\gamma=0} \neq 0 \tag{3.50}$$

$$\lim_{t\to\infty} \frac{\Phi\left(t\left(x+\epsilon\right)\right)}{\Phi\left(tx\right)} = 0 \ \ \forall \epsilon > 0. \tag{3.51}$$

$$\lim_{t\to 0} \frac{\Phi^{-1}\left(at\right)}{\Phi^{-1}\left(t\right)} = 1 \ \ \forall a > 0. \tag{3.52}$$

Then,

- *if* $\gamma_k \to 0 \ \ \forall \gamma_k$ *then* $\frac{\hat{\gamma}}{\bar{\gamma}} = 1$, *with* $\bar{\gamma} \triangleq \frac{1}{N}\sum_{k=1}^{N} \gamma_k$
- *if* $\gamma_k \to +\infty \ \ \forall \gamma_k$ *then* $\frac{\hat{\gamma}}{\gamma_{min}} = 1$, *with* $\gamma_{min} \triangleq \min_k \{\gamma_k\}$.

Proof. The proof for the low SNR case is straightforward, and follows after applying a Taylor expansion of the ESM (3.49) around 0 and using property (3.50). We can study the high SNR case by writing $\gamma_k \triangleq t\tilde{\gamma}_k$, $\gamma_{\min} \triangleq t\tilde{\gamma}_{\min}$ and making t tend to infinity, so

$$\begin{aligned}
\lim_{t\to\infty} \frac{\hat{\gamma}}{\gamma_{\min}} &= \lim_{t\to\infty} \frac{\Phi^{-1}\left(\frac{1}{N}\sum_{k=1}^{N} \Phi\left(t\tilde{\gamma}_k\right)\right)}{t\tilde{\gamma}_{\min}} \\
&= \lim_{t\to\infty} \frac{\Phi^{-1}\left(\Phi\left(t\tilde{\gamma}_{\min}\right)\frac{1}{N}\sum_{k=1}^{N} \frac{\Phi\left(t\tilde{\gamma}_k\right)}{\Phi\left(t\tilde{\gamma}_{\min}\right)}\right)}{t\tilde{\gamma}_{\min}} \\
&\overset{(i)}{=} \lim_{t\to\infty} \frac{\Phi^{-1}\left(\frac{1}{N}\Phi\left(t\tilde{\gamma}_{\min}\right)\right)}{\Phi^{-1}\left(\Phi\left(t\tilde{\gamma}_{\min}\right)\right)} \\
&\overset{(ii)}{=} 1
\end{aligned} \tag{3.53}$$

where (i) is due to (3.51) and (ii) is due to (3.52). □

3.7 Alamouti preprocessing

The use of Alamouti space time codes (STC) in multiple input single output (MISO) processing [88] can be used to overcome the channel degradation problem. With this kind of precoding, present in state of the art standards like DVB-T2 [69], the secondary transmitter does not convey the same message as the primary one, but a slightly modified constellation point.

After the processing performed at the receiver, which requires to estimate the channel from both transmitters separately, the resulting SNR at the k-th carrier is

$$\gamma_k = \frac{|H_k|^2 + |G_k|^2}{\sigma^2}. \tag{3.54}$$

Note that the two channel contributions are added after the modulus squared operation, as opposed to (3.7). This preprocessing ensures that the SNR at each carrier is greater than or equal to that in absence of the secondary transmitter, so a positive ESG is always attained.

Now, we derive the expression for the ESM (3.6) in the Rician scenario with primary and secondary transmiters (for the primary-only scenario, the ESM follows (3.45)). By combining (3.6) and (3.54) we have that

$$\hat{\gamma}_S = \Phi^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}\Phi\left(\frac{|H_k|^2+|G_k|^2}{\sigma^2}\right)\right) \tag{3.55}$$

so, if we assume a sufficiently large number of carriers, we can write

$$\Phi\left(\hat{\gamma}_S\right) = \mathbb{E}_{H,G}\left[\Phi\left(\frac{|H|^2+|G|^2}{\sigma^2}\right)\right] \tag{3.56}$$

with $H \sim \mathcal{CN}\left(1,\sigma_h^2\right)$ and $G \sim \mathcal{CN}\left(\alpha,\sigma_g^2\right)$. Substituting Φ by its approximation (3.9) we arrive to

$$\begin{aligned}\Phi\left(\hat{\gamma}_S\right) &= \sum_{l=1}^{L}\phi_l\mathbb{E}_{H,G}\left[e^{-\beta_l|H_k|^2}e^{-\beta_l|G_k|^2}\right] \\ &= \sum_{l=1}^{L}\phi_l\mathbb{E}_{H}\left[e^{-\beta_l|H_k|^2}\right]E_G\left[e^{-\beta_l|G_k|^2}\right]\end{aligned} \tag{3.57}$$

since H and G are assumed to be independent. The two expected values in (3.57) can be solved following the same procedure as in (3.32) to obtain

$$\begin{aligned}\Phi\left(\hat{\gamma}_S\right) = \sum_{l=1}^{L}\phi_l\frac{\sigma^2}{\beta_l\sigma_h^2+\sigma^2}\exp\left(-\frac{\beta_l}{\beta_l\sigma_h^2+\sigma^2}\right)\times \\ \times\frac{\sigma^2}{\beta_l\sigma_g^2+\sigma^2}\exp\left(-\frac{\beta_l\alpha^2}{\beta_l\sigma_g^2+\sigma^2}\right).\end{aligned} \tag{3.58}$$

If we set $L=1$ a closed form expression for γ_H can be found as

$$\begin{aligned}\hat{\gamma}_S &= \frac{1}{\beta}\log\left(1+\beta\frac{\sigma_h^2}{\sigma^2}\right)+\frac{1}{\beta\sigma_h^2+\sigma^2}+ \\ &\quad+\frac{\alpha^2}{\beta\sigma_g^2+\sigma^2}+\frac{1}{\beta}\log\left(1+\beta\frac{\sigma_g^2}{\sigma^2}\right) \\ &= \hat{\gamma}_0+\hat{\gamma}_T\end{aligned} \tag{3.59}$$

with

$$\hat{\gamma}_T \triangleq \Phi^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}\Phi\left(\frac{|G_k|^2}{\sigma^2}\right)\right) \tag{3.60}$$

the ESM of an only-secondary scenario. It is clear that. since

$$\Delta\hat{\gamma} = 1 + \frac{\hat{\gamma}_T}{\hat{\gamma}_0} \geq 1. \tag{3.61}$$

the system performance is always improved.

Although it can seem that this distributed MISO processing clearly solves the channel degradation problem, it presents some serious drawbacks:

- **Standard dependency** Receivers have to be designed to be able to perform the necessary processing in order to properly obtain the transmitted symbols, thus requiring that the standard they are based on includes Alamouti STC as an option. If this is not the case, the standard should be updated to include it, which is usually undesirable. For example, DVB-SH does not support this kind of MISO processing.
- **Increased overhead** As the receiver has to estimate the channels with both the primary and secondary transmitters, the pilot density has to be doubled with respect to the SISO operation, which significantly increases the signaling overhead.
- **Increased complexity** The receiver has to include additional hardware, mainly the presence of additional multipliers for the Alamouti processing, and the duplication of the channel estimation stage.
- **Extension to more transmitters** The extension of the Alamouti STC (which is a full-diversity rate-one STC) to more than two transmitters (or transmit antennas) is not possible without a rate loss [89]. Moreover, the pilot density should be increased proportionally to the number of transmitters, which is not scalable in practice.

In the next section we introduce an alternative method to avoid the channel degradation problem that, despite not achieving the same gain as the Alamouti preprocessing, does not present the previous drawbacks.

3.8 Pre-filtering

In this section we will explain how an appropriate filtering at the secondary transmitter can improve the performance of the system and avoid the channel degradation problem. This is

the approach presented in Chapter 2 to introduce a secondary signal. We will illustrate the usefulness of this filtering for the AWGN channel, although simulations will be performed for general Rician channels.

If we assume an uniform phase distribution, we can write

$$T_k \triangleq |H_k + G_k|^2 = 1 + \alpha^2 + 2\alpha \cos(a_k) \tag{3.62}$$

with $a_k \sim \mathcal{U}(0, 2\pi]$ independent and identically distributed. Note that the channel degradation problem is caused by some of the T_k suffering a *destructive interference* so $T_k < 1 = |H_k|^2$. In fact, the values of T_k are contained in the interval $\left[(1-\alpha)^2, (1+\alpha)^2\right]$, so $T_k > 1 = |H_k|^2 \,\forall\, K$, i.e., an SNR gain in every carrier can be assured if $\alpha \geq 2$. Unfortunately, the parameter α cannot be modified by the secondary transmitter, as it would require to increase the transmit power or change the secondary location, which is usually not possible.

However, if $\alpha < 2$, the secondary transmitter could perform a filtering in the time domain (or, alternatively, a power weighting in the DFT domain) so the power is concentrated in a fraction of carriers. If we denote by F_k the DFT response of the filter, the channel model (3.2) reads for the k-th carrier as

$$Y_k = (H_k + G_k F_k) X_k + W_k = \left(1 + e^{ja_k} \alpha F_k\right) X_k + W_k. \tag{3.63}$$

We will consider the following structure for the filter F_k:

- A fraction $1 - \frac{\alpha^2}{4}$ of the carriers will be weighted with $F_k = 0$. As the ESM does not depend on the particular position of this carriers, we will assume, without loss of generality, that the first $N\left(1 - \frac{4}{\alpha^2}\right)$ carriers are nulled by the transmit filter F_k. Obviously, the power spent on these carriers is zero, and $T_k = |H_k|^2 = 1$.
- The remaining carriers (this is, a fraction $\varphi \triangleq \frac{\alpha^2}{4}$ of them) are weighted with $F_k = \frac{2}{\alpha}$. The average power consumption of this group of carriers is $\frac{4}{\alpha^2}$, and we have that

$$T_k = |1 + 2e^{ja_k}|^2 \geq 1 = |H_k|^2. \tag{3.64}$$

Note that with this approach we assure that no carrier suffers a power loss ($T_k \geq 1 \,\forall\, k$), and the transmit power is not increased, as

$$\frac{1}{N} \sum_{k=1}^{N} |F_k|^2 = \varphi \frac{4}{\alpha^2} = 1, \tag{3.65}$$

so the power consumption is the same as in the absence of F_k (or, equivalently, $F_k = 1 \,\forall\, k$).

This filter was shown in Chapter 2 to be optimal in the high SNR regime for the AWGN channel, while for lower SNR values the two-level filter was shown to be optimal, but the optimum fraction of active carriers is no longer $\frac{\alpha^2}{4}$ and has to be computed numerically.

Now, we proceed to calculate the ESM of a system where the secondary transmitter uses this kind of filtering. Similarly to (3.32), we have

$$\begin{aligned} \hat{\gamma}_S &= \Phi^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}\Phi\left(\frac{\left|H_k + F_k G_k\right|^2}{\sigma^2}\right)\right) \\ &= \Phi^{-1}\left(\frac{1}{N}\left(\sum_{k=1}^{N\varphi}\Phi\left(\frac{\left|1+2e^{ja_k}\right|^2}{\sigma^2}\right) + \sum_{N\varphi+1}^{N}\Phi\left(\frac{1}{\sigma^2}\right)\right)\right). \end{aligned} \tag{3.66}$$

Once again, if we assume a sufficiently large number of carriers[4], we can approximate (3.66) by

$$\begin{aligned} \Phi\left(\hat{\gamma}_S\right) &\approx (1-\varphi)\sum_{l=1}^{L}\phi_l e^{-\frac{\beta_l}{\sigma^2}} + \varphi\sum_{l=1}^{L}\phi_l E_a\left\{e^{-\frac{\beta_l\left(1+2e^{ja}\right)}{\sigma^2}}\right\} \\ &= \sum_{l=1}^{L}\phi_l \exp\left(-\frac{\beta_l}{\sigma^2}\right)\left((1-\varphi) + \varphi\exp\left(-\frac{4\beta_l}{\sigma^2}\right) I_0\left(\frac{4\beta_l}{\sigma^2}\right)\right). \end{aligned} \tag{3.67}$$

Note that in the only-primary scenario we have

$$\Phi\left(\hat{\gamma}_0\right) = \sum_{l=1}^{L}\phi_l \exp\left(-\frac{\beta_l}{\sigma^2}\right), \tag{3.68}$$

and since $e^x \geq I_0(x)$ [85],

$$(1-\varphi) + \varphi\exp\left(-\frac{4\beta_l}{\sigma^2}\right) I_0\left(\frac{4\beta_l}{\sigma^2}\right) \leq 1, \tag{3.69}$$

we have that $\hat{\gamma}_S \geq \hat{\gamma}_0$, as Φ is a monotonic decreasing function.

Like in the previous cases, we will provide a closed-form expression for $L = 1$. It can be easily seen that the EESM reads as

$$\hat{\gamma}_S = \frac{1}{\sigma^2} - \frac{1}{\beta}\log\left(1 - \varphi + \varphi\exp\left(-\frac{4\beta_l}{\sigma^2}\right) I_0\left(\frac{4\beta_l}{\sigma^2}\right)\right) \tag{3.70}$$

[4]In this case, the *sufficiently large* approximation must hold not only for the overall system, but also for each of the two groups of carriers.

and the ESG is

$$\Delta\hat{\gamma} = 1 - \frac{\sigma^2}{\beta}\log\left(1-\varphi+\varphi\exp\left(-\frac{4\beta_l}{\sigma^2}\right)I_0\left(\frac{4\beta_l}{\sigma^2}\right)\right) \tag{3.71}$$

which is always greater than or equal to one, since (3.69) holds.

This result for the AWGN channel can be extended to a general Rician channel following similar steps. The ESM of the two transmitter system follows

$$\hat{\gamma}_S = -\frac{1}{\beta}\log\left((1-\varphi)\frac{\sigma^2}{\beta\sigma_h^2+\sigma^2}\exp\left(-\frac{\beta}{\beta\sigma_h^2+\sigma^2}\right)+\varphi\frac{\sigma^2}{\beta\sigma_z^2+\sigma^2}\exp\left(-\beta\frac{1+\frac{\alpha^2}{\varphi}}{\beta\sigma_z^2+\sigma^2}\right)I_0\left(\frac{2\beta\left(\alpha/\sqrt{\varphi}\right)}{\beta\sigma_z^2+\sigma^2}\right)\right) \tag{3.72}$$

while the ESG assuming the same Rician K factor from both transmitter is

$$\Delta\hat{\gamma} = \frac{-\frac{1}{\beta}\log\left((1-\varphi)\frac{\sigma^2}{\beta K^{-1}+\sigma^2}\exp\left(-\frac{\beta}{\beta K^{-1}+\sigma^2}\right)+\varphi\frac{\sigma^2}{\beta K^{-1}(1+\alpha^2/\varphi)+\sigma^2}\exp\left(-\beta\frac{1+\frac{\alpha^2}{\varphi}}{\beta K^{-1}(1+\frac{\alpha^2}{\varphi})+\sigma^2}\right)I_0\left(\frac{2\beta(\alpha/\sqrt{\varphi})}{\beta K^{-1}(1+\frac{\alpha^2}{\varphi})+\sigma^2}\right)\right)}{\frac{1}{\beta K^{-1}+\sigma^2}+\frac{1}{\beta}\log\left(1+\beta\frac{K^{-1}}{\sigma^2}\right)}. \tag{3.73}$$

These two formulas can be obtained following the same steps that led to (3.45).

This preprocessing offers some clear advantages with respect to the Alamouti preprocessing, but it also presents some drawbacks:

- The power gain is smaller than the one attained with Alamouti STC.
- The fraction of active carriers φ depends on the value of α (the relative amplitude between the primary and secondary components) and, therefore, on the position of the receiver. The design of the filter is more involved if several receivers with a huge range of values of α are present.

3.9 Results

The obtained closed form expressions for the EESM were verified in OFDM channels with a finite number of carriers, correlated carriers in the frequency domain (as a result of a channel with a limited length in the time domain) and various values of n_0 and θ. Although simulations with several values of α (the relative amplitude between contributions) and β (the EESM parameter) were conducted, the provided results are shown for values of $\alpha = 1$ (same power coming from both primary and secondary transmitters) and $\alpha = \sqrt{1/5}$, and $\beta = \beta_{QPSK} \approx 0.64$, which is the value obtained via mean squared error (MSE) fitting of the MIESM function (3.8) for a QPSK constellation by the parametrized function (3.9) with $L = 1$. All the channels were generated in the time domain with an impulse response of $\frac{N}{4}$ samples, that would be the case of an OFDM system working with a 1/4 CP.

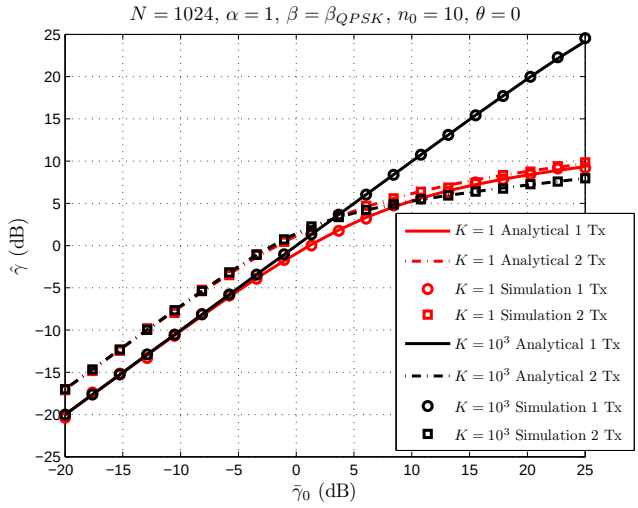
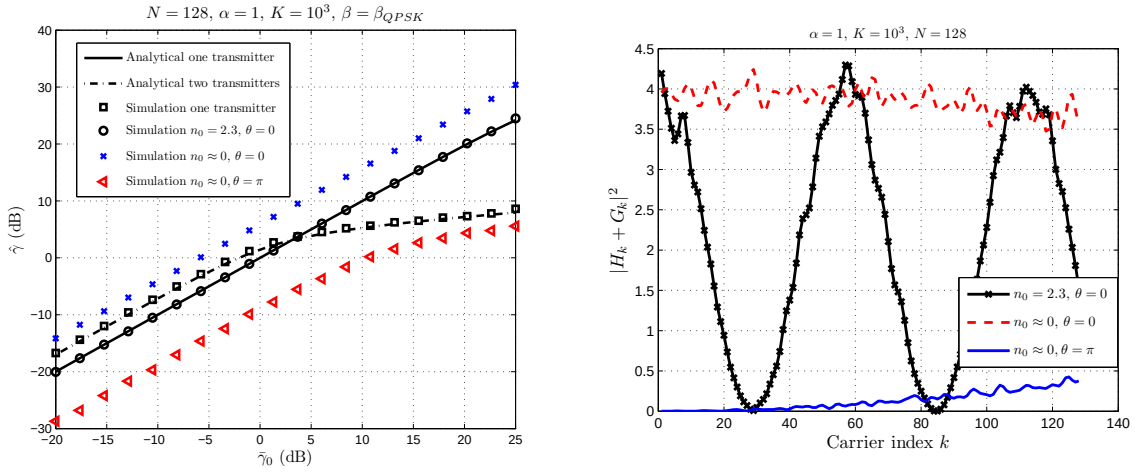


Figure 3.1: Analytical and simulated effective SNR as a function of the initial average SNR in the presence of one and two transmitters, and different K factors.

In Figure 3.1 the analytical expressions (3.45) and (3.33) are compared with the actual EESM obtained by simulation[5] for different K factors, and in the single transmitter (1 Tx) and two transmitters (2 Tx) case. In this case, the number of carriers was $N = 1024$, the selected value for n_0 was of 10 samples, and the phase difference $\theta = 0$, although other simulations were conducted for non-extreme values of n_0 with similar results. It can be seen that the analytical expressions fit almost perfectly the simulation results. Note that the scenario with a high value of K and two transmitters, that could seem the most favorable *a priori*, is indeed the worst case scenario in the high SNR regime due to the strong impact of channel degradation. Moreover, it can be seen that the analytical approximation is accurate even in the presence of correlation among carriers (introduced by the limited length of the channel in the time domain).

In Figure 3.2a a similar simulation is shown, but now with $K = 10^3$, $N = 128$ carriers and different values of n_0 and θ. Recall that the approximation for the computation of the ESM contribution of the two line of sight components consisted on assuming a uniform phase distribution, which is not a good approximation in the cases of $n_0 \approx 0$, as briefly depicted in Section 3.3. In this case, it can be clearly seen that the analytical approximation is no longer valid for the case $n_0 \approx 0$, as it produces a constant positive or negative interference, depending on the value of θ and, therefore, the relative phase in the different carriers cannot be approximated by a uniform distribution, as seen in Figure 3.2b. It is also remarkable that the

[5]Note that the channel is random but the obtained analytical approximations are deterministic, so each point in the simulations correspond to one different realization of the channels H and G, i.e., no averaging is performed among different realizations.



(a) Effective SNR.

(b) Channels.

Figure 3.2: Analytical and simulated effective SNR as a function of the initial average SNR in the presence of one and two transmitters for different delay n_0 and phase θ values.

analytical approximation is still quite accurate even for $N = 128$ carriers (that may contradict the *sufficiently large number of carriers* assumption) and a small delay as $n_0 = 2.3$ samples.

In Figure 3.3 we show the evolution of the SFN gain with the average SNR for different K values. As predicted, in the low SNR regime the gain is positive and equal for all channels, while for higher SNR values the behavior is highly dependent on the Rician factor: while for low K values the gain is always positive and rapidly tends to zero, for stronger LOS environments there is an increasing loss at moderate SNR values. Note that the gain in all Rician channels tends to zero as the SNR increases due to the fact that, as shown in Section 3.5.2, in the high SNR regime the ESM increases logarithmically and, therefore, the behavior is similar to that of a Rayleigh channel.

The contour plot in Figure 3.4 shows the ESG in different multipath (values of K) and average SNR scenarios, and characterizes the region where a positive gain is attained by the insertion of the secondary transmitter, i.e., those points where the *power gain* is more important than the *channel degradation*. It can be seen that Rayleigh or low-SNR channels are easily enforced, but those receivers with large values of both K and $\bar{\gamma}_0$ are degraded after the insertion of the secondary transmitter. These results provide an analytical justification of the measurements in [90], and predict that the degradation in strong line of sight environments does not happen in the low SNR regime[6]. It is also observed that the case with a low power secondary transmitter

[6] Although in a somehow different scenario, a similar result was obtained in [15] for the cognitive radio channel, where it was shown that channel state information (CSI) is not that important in the low SNR regime, as long

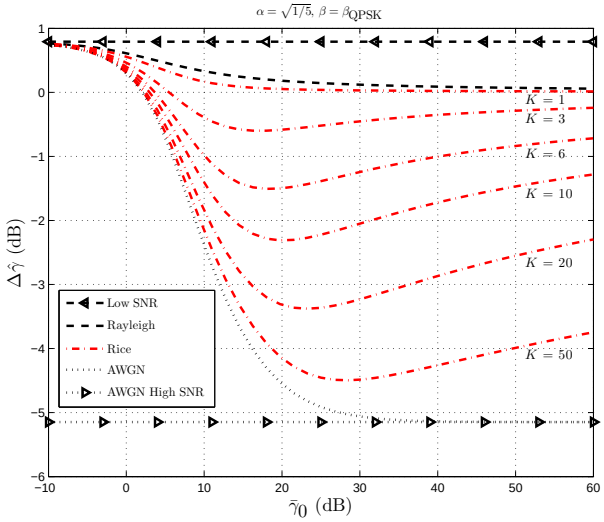
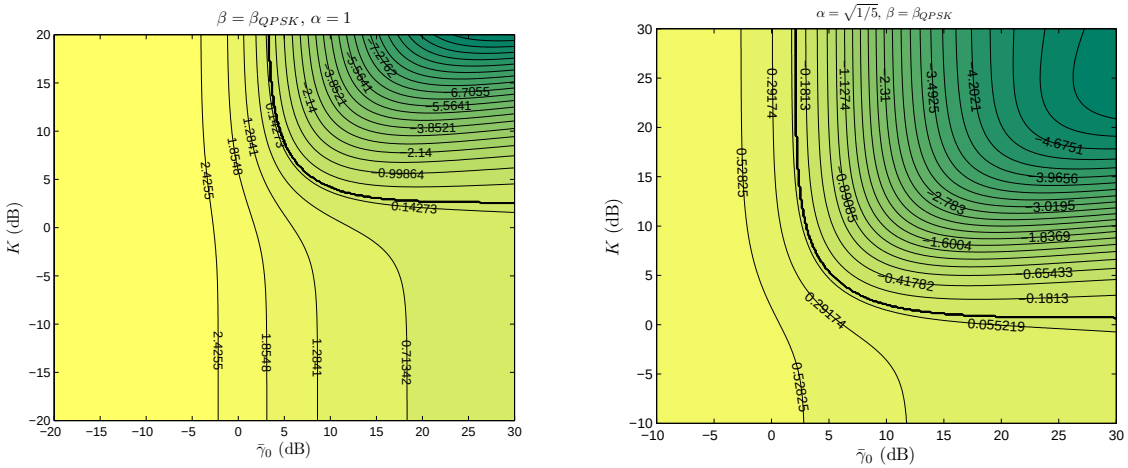


Figure 3.3: Evolution of the SFN gain $\Delta\hat{\gamma}$ with the average SNR, for different K factors.

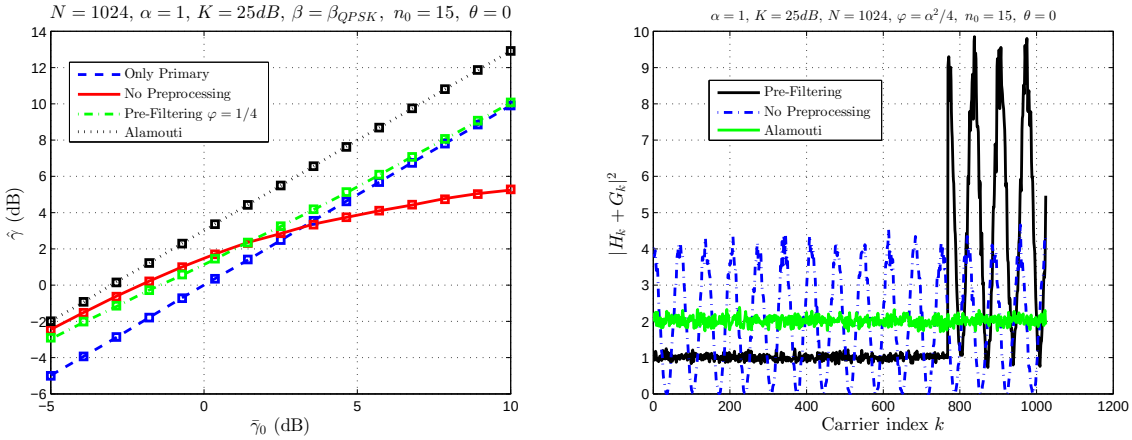


(a) $\alpha = 1$.

(b) $\alpha = \sqrt{1/5}$.

Figure 3.4: Contour plot of the ESM gain $\Delta\hat{\gamma}$ (in dB) for different values of average SNR and K factor. The thick line delimites the region with a negative ESG.

has a larger region with negative ESG, but the degradation is smaller in the high SNR - high K region (observe the numbers in the contour plot in the upper-right part).



(a) Analytical (squares) and simulated (lines) ESM. (b) Channel response

Figure 3.5: Effective SNR as a function of the initial average SNR in the presence of one and two transmitters for different scenarios.

We focus now on the countermeasures to prevent the channel degradation. We illustrate first the effect of the different approaches in the resulting channel seen by a given receiver. In Figure 3.5b we show a plot of the channels obtained with the different approaches in a scenario with $\alpha = 1$, i.e., same power coming from the primary and the secondary transmitters, and $K = 25dB$. As we are assuming an almost pure LOS scenario, the squared modulus of the primary and secondary channels is similar for all carriers, and approximately equal to 1 (these channels are not shown for the sake of clarity). The effects on the channel of the different preprocessing strategies are:

- With no preprocessing, the channels are directly added *in the air* with different phases, thus resulting in a sequence of carriers suffering negative and positive interference. As both primary and secondary channels are approximately equal to 1, the squared modulus of the sum channel is concentrated between $\max_k \left\{|H_k + G_k|^2\right\} \approx 4$ and $\min_k \left\{|H_k + G_k|^2\right\} \approx 0$. It is clear that those carriers with $|H_k + G_k|^2 < 1$ will achieve a lower SNR than in the only-primary scenario, thus potentially harming the performance of the system.

- With the proposed pre-filtering, part of the carriers are nulled at the secondary, so the resulting channel is equal to that in the only-primary scenario. All the available power is concentrated in a group of carriers (from $k = 769$ to $k = 1024$), being the filter coefficient

as the channels from both transmitters can be considered to be ergodic, which can be seen to be similar to our *uniform phase* assumption.

in these carriers set to $F_k = 2$. Similarly to the non-preprocessing case, the channels are added prior to the squared-modulus operation, so the resulting channel is concentrated between $\max_k \left\{|H_k + F_k G_k|^2\right\} \approx 9$ and $\min_k \left\{|H_k + F_k G_k|^2\right\} \approx 1$. Note that in this case the minimum value is approximately equal to 1, which was the channel value in the only-primary scenario, so no carrier suffers an SNR loss (up to the random but weak multipath component).

- With Alamouti preprocessing, the squared modulus of both channels are added following (3.54). Therefore, the resulting channel is almost flat (except for the weak multipath component), and approximately equal to 2, as the result of $|H_k|^2 + |G_k|^2 \approx 2$.

In Figure 3.5a we show the analytical and simulated ESM as a function of the average SNR $\bar{\gamma}_0$ for the different approaches. It can be seen that the Alamouti preprocessing clearly outperforms the other strategies, specially for high SNR, and that the insertion of the secondary transmitter with no preprocessing leads to a lower ESM than the only-primary scenario (this implies that the channel degradation is larger than the power gain). For low SNR values the channel degradation is not that important, and the insertion of a secondary transmitter clearly increases the performance of the system.

The pre-filtering approach provides a smaller gain than the Alamouti preprocessing, but can be seen to clearly avoid the channel degradation problem, as the pre-Filtering curve is always above the *No Preprocessing* one.

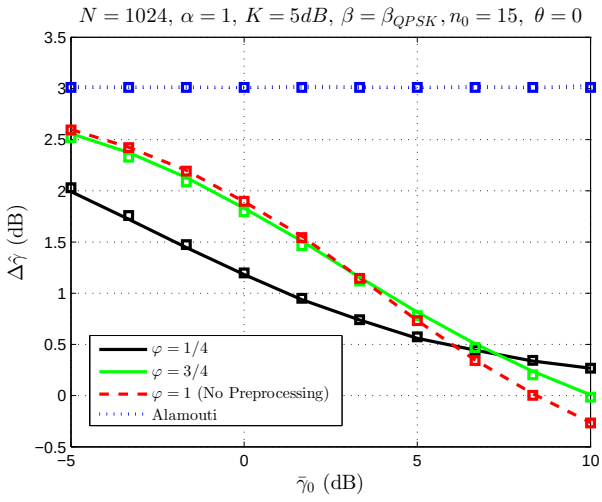


Figure 3.6: Analytical (squares) and simulated (lines) effective SNR as a function of the initial average SNR for different scenarios. The pre-filtering is performed for different values of φ.

In Figure 3.6 we present similar results for a scenario with a stronger multipath component

($K = 5dB$). In this case we are representing the ESG $\Delta\hat{\gamma}$ instead of the ESM $\hat{\gamma}$ for different values of fraction of active carriers φ as a function of the average SNR in an only-primary scenario. It can be seen that the selection of φ is quite involved, as it depends on the SNR working point: for low SNR $\varphi = 1$ provides a higher gain, for medium SNR (around $5dB$) the curve $\varphi = 3/4$ is above the others, while for high SNR the best option is $\varphi = 1/4$. It is also noticeable that the losses due to the channel degradation are much smaller than the ones shown in Figure 3.5a: note that for $\bar{\gamma}_S = 10$dB there is a loss of approximately $0.25dB$ for the No Preprocessing curve, while Figure 3.5a shows a loss of approximately 5dB for the same SNR value.

3.10 Conclusions

In this chapter we derived simple closed form expressions for the SFN gain based on effective SNR metrics. These expressions are applicable to scenarios with one and two transmitters in Rician channels, and provides an analytical justification of previous empirical results that showed a performance degradation due to the presence of SFN echoes. We conclude that this degradation is specially noticeable in those receivers with both a high SNR and a strong line of sight reception. We also introduce two possible techniques that can be used to alleviate the degradation caused by the insertion of a secondary transmitter. These analytical results could be useful for SFN planning, design of cooperative communications services or cognitive radio architectures, among others.

This chapter contains the results of two conference papers published in ASMS 2012 [91] and ISWCS 2013 [92] with Prof. Carlos Mosquera.

Chapter 4

Broadcast Networks: Covering Gray Spaces

Contents

4.1 Introduction

In Chapter 2 we analyzed the insertion of a secondary transmitter in a broadcast network with dominant line of sight reception, and a secondary receiver that is able to decode the primary information and, therefore, subtract it from the received signal. In Chapter 3 we extended the analysis to the case of Rician channels. In this chapter, we extend the analysis of Chapter 2 to the case of secondary receivers that are placed outside the primary user coverage area, so decoding of the primary message is not possible.

In this chapter we assume that the secondary transmitter is not able to use dirty paper coding [14], as it would require almost perfect channel state information at the transmitter [70]. However, interference cancellation can be performed at the secondary receiver, provided it is able to decode the primary message. This approach has been followed in [18] to study a secondary cellular system located inside the coverage area of a primary broadcast network, so interference cancellation is always possible. A similar analysis was performed in Chapter 2. In this chapter we study the case of secondary access in different positions with respect to the coverage zone of a broadcast network, so interference cancellation is not always possible unless the secondary transmitter expanded the coverage zone.

We will treat the previous idea by resorting to the concept of different *spaces*, similarly to [7]:

- **White Spaces** The secondary system is far from the primary coverage zone, so its performance is noise or power limited.
- **Black Spaces** The secondary system is inside the primary coverage zone, so it is able to decode the primary message and, therefore, cancel the interference.
- **Gray Spaces** The secondary system is outside the primary coverage zone, but it receives a considerably large amount of interference (that is not able to decode in principle) from the primary transmitter. Moreover, the interference caused to the primary system has to be properly controlled due to the proximity to the coverage zone.

The remaining of the chapter is structured as follows: Section 4.2 introduces the system model; Section 4.3 presents the coding and decoding strategies for interference cancellation; Section 4.4 presents the power allocation problem for a single secondary receiver; Section 4.5 extends the previous result to multiple secondary receivers; Section 4.6 presents some comments on practical impairments; finally, Section 4.7 concludes the chapter.

4.2 System model

Consider a single primary transmitter (PT) transmitting at a rate R_p using P_p units of power, which is communicating with N_r primary receivers (PRs).

A secondary transmitter (ST) with a total available power P_s is inserted in the network. We will assume that the ST has prior knowledge of the primary message, so the total power P_s has to be shared between the powers γ and ρ, allocated to the primary and secondary messages, respectively, so $\gamma + \rho \leq P_s$. The knowledge of the primary message can be obtained, for example, if the primary system operates as a Single Frequency Network (SFN), so the contents are delivered by a distribution network (DN), and the ST can obtain the primary waveform just by connecting to this DN [18]. The objective of the ST is to communicate with one or more secondary receivers (SRs) at the highest possible rate while meeting a set of interference constraints regarding the correct reception of the primary message at the PRs.

If we denote as $h_{p,i}$ the (complex) channel gain from the PT to the i-th PR, then the capacity of the associated link, in absence of the ST, can be written as

$$C_{P,i}^0 = C\left(\frac{|h_{p,i}|^2 P_p}{\sigma^2}\right) \tag{4.1}$$

with σ^2 the noise variance, assumed to be constant at all the receivers, and $C(x) \triangleq \log_2(1+x)$. After the insertion of the ST, we will assume that the capacity can be expressed as

$$C_{P,i} = C\left(\frac{|h_{p,i}|^2 P_p + |h_{s,i}|^2 \gamma}{\sigma^2 + |h_{s,i}|^2 \rho}\right) \tag{4.2}$$

with $h_{s,i}$ the channel gain from the ST to the i-th PR. Note that we are assuming that the primary signal contributions coming from both primary and secondary transmitters are added, that would be the case of a cooperative transmission using an Alamouti space time code (STC) [88], for example. Note that if the PT and ST transmit the same waveform (i.e., without STC) the channel gain will be of the form $|h_{p,i}P_p + \gamma h_{s,i}|^2$, so the interference will be constructive or destructive depending on the relative phase between channels [15]; in the case of an orthogonal frequency division multiplexing (OFDM) system, which has to be used in SFNs, this interference is going to be constructive on some carriers and destructive in others, so an alternative analysis like the one in Chapters 2-3 should be performed. In any case, we will restrict our analysis to the use of STC so the powers coming from both primary and secondary transmitter can be assumed to be *added*, like in (4.2).

The PT is conveying data at a rate R_p, so the i-th PR is able to decode the primary message

provided that $C_{P,i} \geq R_p$ or, equivalently

$$\gamma \geq \frac{1}{|h_{s,i}|^2}\left(\Upsilon_0\sigma^2 - |h_{p,i}|^2 P_p\right) + \rho\Upsilon_0 = \rho\Upsilon_0 - \mathcal{M}_i,\ i = 1, ..., N_r \tag{4.3}$$

with $\Upsilon_0 \triangleq 2^{R_p} - 1$ the required signal to interference plus noise ratio (SINR) for a correct reception of the primary message, and N_r the number of secondary receivers. We denoted

$$\mathcal{M}_i \triangleq \frac{1}{|h_{s,i}|^2}\left(|h_{p,i}|^2 P_p - \Upsilon_0\sigma^2\right),\ i = 1, ..., N_r \tag{4.4}$$

for the sake of clarity. Note that $|h_{s,i}|^2\mathcal{M}_i$ can be thought to be the *power margin* of the i-th PR before the insertion of the ST: $\mathcal{M}_i > 0$ means that the PR lies inside the primary coverage area, $\mathcal{M}_i < 0$ means that the PR lies outside the coverage area, and the receivers with $\mathcal{M}_i = 0$ are located at the coverage area edge. We consider that the primary service is not compromised if the N_r linear constraints in (4.3) are met or, equivalently, if the interference constraint is met in the worst PR case:

$$\gamma \geq \max_{i=1, ..., N_r} \rho\Upsilon_0 - \mathcal{M}_i, \tag{4.5}$$

so we must only take into account the receiver with a lower $\mathcal{M}_i$, $\mathcal{M} \triangleq \min_{i=1, ..., N_r} \mathcal{M}_i$ or, equivalently, the one with a lower *power margin*. This idea is similar to that of a *critical TV receiver* in [18].

In Figure 4.1 there is a plot of the proposed system model.

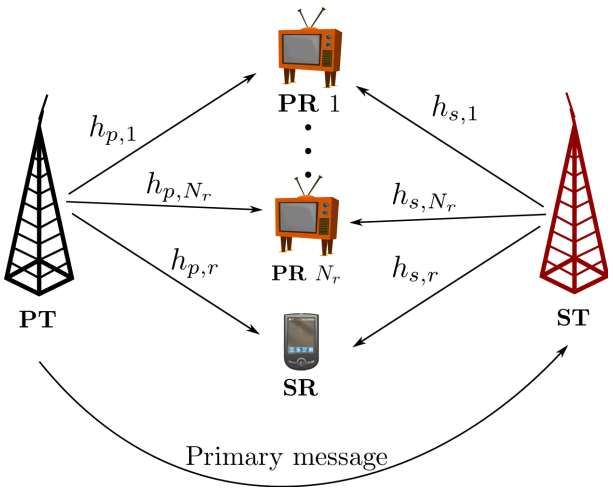


Figure 4.1: System model: the primary and secondary systems interfere with each other, so the ST has to exploit the available side information (which can be obtained by connecting the ST to a Distribution Network [18], for example) to control the interference caused to the primary system.

4.3 Coding/decoding strategies

Now, we proceed to characterize the capacity of the secondary link. It is usually assumed in the literature that the ST uses dirty paper coding (DPC) [14] to cancel the (a priori known) interference caused by the PT. However, the use of DPC requires complete channel state information (CSI) at the transmitter, both in magnitude and phase [70]. Therefore, we will assume that the available CSI at the ST is similar to the channel quality indicator (CQI) present in LTE [93], or obtained by means of field level measurements, so no phase information is available.

The channel seen from the SR is a multiple access channel (MAC), so the capacity of a secondary link C_s depends on the decodability of the primary message. In this section, we will fix the power allocation values (γ, ρ) and characterize the different decoding strategies a SR can use to recover the secondary message. Note that the power constraint $\gamma + \rho \leq P_s$ has to be met in all the cases and, as we will see next, sometimes this constraint is met with strict inequality at the optimum power allocation point. In the following, as depicted in Figure 4.1, we denote as $h_{s,r}$ and $h_{p,r}$ the channels from the ST and PT to the SR under study, respectively.

4.3.1 Treat as noise

With this approach, the secondary receiver treats the primary interference as an additional noise source. Although this decoding strategy can be used in any interference regime, it is the only possible decoding procedure if the interference is weak even in absence of the ST transmitting the secondary message, i.e., the signal to interference plus noise ration (SINR) of the primary signal at the SR is below the required SINR for decoding:

$$\frac{|h_{p,r}|^2 P_p + |h_{s,r}|^2 \gamma}{\sigma^2} < \Upsilon_0. \tag{4.6}$$

If this is the case, the primary rate R_p lies outside the capacity region of the MAC, so the SR has to treat the primary signal as noise. This strategy leads to the secondary link capacity

$$C_s = C\left(\frac{\rho |h_{s,r}|^2}{\sigma^2 + |h_{p,r}|^2 P_p + |h_{s,r}|^2 \gamma}\right). \tag{4.7}$$

4.3.2 Strong interference

If the primary interference is strong enough, even in presence of the secondary message

$$\frac{|h_{p,r}|^2 P_p + |h_{s,r}|^2 \gamma}{\sigma^2 + \rho |h_{s,r}|^2} \geq \Upsilon_0 \tag{4.8}$$

the SINR of the primary signal at the SR is above the threshold SINR, and, therefore, the SR is be able to decode the primary message and subtract it from the received signal, so the secondary link capacity is

$$C_s = C\left(\frac{\rho |h_{s,r}|^2}{\sigma^2}\right). \tag{4.9}$$

4.3.3 Medium interference

In this regime, the primary message is decodable prior to the transmission of the secondary message, but after its insertion is not:

$$\frac{|h_{p,r}|^2 P_p + |h_{s,r}|^2 \gamma}{\sigma^2 + \rho |h_{s,r}|^2} < \Upsilon_0 \leq \frac{|h_{p,r}|^2 P_p + |h_{s,r}|^2 \gamma}{\sigma^2}, \tag{4.10}$$

i.e., the SINR of the primary signal is above the threshold SINR if we do not take into account the interference caused by the insertion of the secondary message, and below the required SINR if we include this additional noise term.

This region corresponds to the classical *time-sharing* segment of the MAC. As *time-sharing* would imply the existence of synchronization between PT and ST, which is not desirable, the ST must resort to the use of superposition coding (SC) like in [72], which achieves the same capacity region, with the corresponding secondary link capacity given by

$$C_s = C\left(\frac{(\rho + \gamma)\,|h_{s,r}|^2 + |h_{p,r}|^2 P_p}{\sigma^2}\right) - R_p. \tag{4.11}$$

Note that this strategy requires the ST to split its secondary power ρ into ρ_1 and ρ_2 units of power, allocated to two different sources $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. In such a case, ρ_1 is chosen to meet

$$C\left(\frac{|h_{s,r}|^2 \gamma + |h_{p,r}|^2 P_p}{\sigma^2 + |h_{s,r}|^2 \rho_1}\right) \geq R_p \tag{4.12}$$

while $\rho_2 = \rho - \rho_1$. The encoding process is performed as follows: encode the message from

source $\mathcal{S}_1$ into the codeword x_1 at a rate

$$R_1 = C\left(\frac{|h_{s,r}|^2\rho_1}{\sigma^2}\right) \tag{4.13}$$

with ρ_1 the value obtained from (4.12) with equality; encode the message from source $\mathcal{S}_2$ into the codeword x_2 at a rate

$$R_2 = C\left(\frac{|h_{s,r}|^2\rho_2}{\sigma^2 + |h_{s,r}|^2\rho_1 + |h_{p,r}|^2 P_p}\right). \tag{4.14}$$

Decoding is performed as follows: first, decode x_2 treating x_1 and the primary message as noise, subtracting it from the received signal; then, decode the primary message treating x_1 as noise, subtracting it from the received signal; finally, decode x_1. It can be seen that the three rates (4.12), (4.13) and (4.14) lie inside the three dimensional MAC region, so the proposed decoding sequence is possible. The total rate is given by

$$R_1 + R_2 = C\left(\frac{(\rho+\gamma)\,|h_{s,r}|^2 + |h_{p,r}|^2 P_p}{\sigma^2}\right) - R_p. \tag{4.15}$$

4.4 Optimum power allocation for a single secondary receiver

As the capacity function is defined in a piecewise way and, therefore, is not differentiable, we propose to solve three different optimization problems (one for each of the three decoding strategies) and afterwards select the one that leads to a higher capacity. The two last decoding strategies conform problems with a different associated *region constraint* (4.8) and (4.10), and the three of them share both a *power constraint* $\gamma + \rho \leq P_s$ and N_r *interference constraints* (or a worst case constraint (4.5)). In order to define in a compact way the *region constraints* we denote

$$\mathcal{M}^s \triangleq \frac{1}{|h_{s,r}|^2}\left(|h_{p,r}|^2 P_p - \Upsilon_0\sigma^2\right) \tag{4.16}$$

as the *power margin* for the primary signal at the SR under study. Note that, unlike $\mathcal{M}$, which is related to the interference constraints at the primary receivers, we will use this parameter to constrain the operation of the different decoding strategies at the SR to the corresponding interference regimes.

Now, we proceed to enunciate the three different problems:

4.4.1 Treat as noise

With this decoding strategy, we can enunciate the optimization problem as

$$\begin{array}{rlcl} \max. & f(\rho,\gamma) & \triangleq & \dfrac{\rho|h_{s,r}|^2}{\sigma^2+\gamma|h_{s,r}|^2+|h_{p,r}|^2P_p} \\ \text{s.t.} & \gamma+\rho & \leq & P_s \\ & \gamma & \geq & \rho\Upsilon_0-\mathcal{M} \\ & \gamma & \geq & 0 \\ & \rho & \geq & 0. \end{array} \tag{4.17}$$

This is a linear-fractional program, which can be easily converted to a linear program (LP) by the Charnes-Cooper transformation [94]. It is clear that if $\mathcal{M}^s > 0$, i.e., if the primary signal is decodable at the SR in absence of the ST, or if $\mathcal{M} \leq \mathcal{M}^s$, i.e., if the reinforcement of one of the primary receivers forces the SR to be able to decode the primary signal, then the SR could operate in the medium or strong interference regimes, thus leading to a higher capacity. Also, if some $\mathcal{M}_i < 0$ (so $\mathcal{M} < 0$) we are introducing a constraint on some PRs that are not able to decode the primary signal prior to the insertion of the ST - these receivers could be thought to be lying outside the *primary coverage area*, and the ST would be constrained to extend that coverage area.

Note that the objective function f is increasing with ρ and decreasing with γ, so the allocation of power to the primary message decreases the capacity of the secondary user. Therefore, it would be desirable to allocate power only to the secondary message, but the presence of an active interference constraint forces to allocate $\Delta\gamma = \Upsilon_0\Delta\rho$ units of power to the primary message for every $\Delta\rho$ units of power allocated to the secondary message or, equivalently, to move in the direction $\mathbf{x}$ given by

$$\mathbf{x} = \begin{bmatrix} \Delta\rho \\ \Delta\gamma \end{bmatrix} = k \begin{bmatrix} 1 \\ \Upsilon_0 \end{bmatrix}. \tag{4.18}$$

We can check if the objective function is increasing in the direction dictated by the constraint just by calculating the gradient ∇f of the objective function

$$\begin{bmatrix} \frac{\partial f}{\partial \rho} \\ \frac{\partial f}{\partial \gamma} \end{bmatrix} = \nabla f = \frac{|h_{s,r}|^2}{\sigma^2+\gamma|h_{s,r}|^2+|h_{p,r}|^2P_p} \times \begin{bmatrix} 1 \\ -\dfrac{|h_{s,r}|^2}{\sigma^2+\gamma|h_{s,r}|^2+|h_{p,r}|^2P_p} \end{bmatrix} \tag{4.19}$$

so if $\nabla f^T \mathbf{x} > 0$ the function is increasing in the constraint direction or, equivalently, if

$$\Upsilon(\gamma, \rho) < \frac{1}{\Upsilon_0} \tag{4.20}$$

with

$$\Upsilon(\gamma, \rho) = \frac{|h_{s,r}|^2 \rho}{\sigma^2 + \gamma |h_{s,r}|^2 + |h_{p,r}|^2 P_p} \tag{4.21}$$

the current SINR of the secondary link. Therefore, we conclude that allocating extra power $\mathbf{x}$ to both primary and secondary signal components in the direction dictated by the interference constraint is beneficial if the current SINR of the secondary link meets $\Upsilon(\gamma, \rho) < \frac{1}{\Upsilon_0}$. If this is not the case, then the optimum power allocation leaves some power without being used. A plot explaining the geometrical interpretation of this constraint is shown in Figure 4.2.

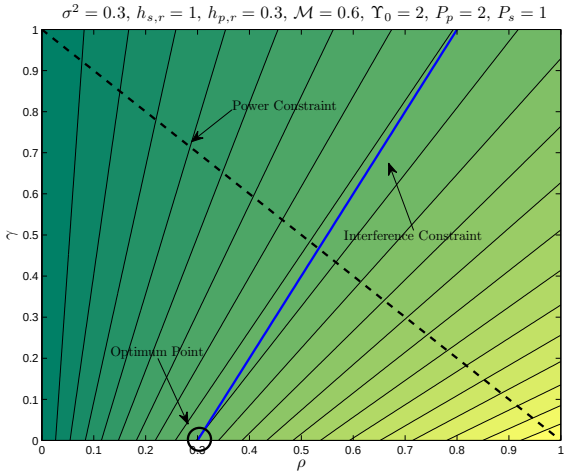


Figure 4.2: Level curves for the objective function f (those zones with lighter colors achieve a larger capacity) and constraints of the optimization problem. If we are using an *active set* method [66], for example, starting at point $(\rho, \gamma) = (0, 0)$, on the first iteration we will get to the point $(0.3, 0)$. At this point, the interference constraint forces us to move along the direction in (4.18), which leads to lower levels of the objective function. Therefore, the optimum point is reached at $(\rho, \gamma) = (0.3, 0)$.

4.4.2 Strong interference

In this case, the optimization problem is stated as

$$\begin{aligned} \text{max.} \quad & \frac{\rho|h_{s,r}|^2}{\sigma^2} && (4.22)\\ \text{s.t.} \quad & \gamma+\rho \leq P_s && (4.23)\\ & \gamma \geq \rho\Upsilon_0 - \mathcal{M} && (4.24)\\ & \gamma \geq \rho\Upsilon_0 - \mathcal{M}^s && (4.25)\\ & \gamma \geq 0 \\ & \rho \geq 0 \end{aligned}$$

which is a LP. Note that if $\mathcal{M}^s < \mathcal{M}$ then this region is forced, as the *worst case* PR has a smaller *power margin* than the SR; this could be the case of having a SR inside the primary coverage zone, for example. It is clear that in this region the ST is going to use all its available power, so by forcing an active power constraint (4.23) we can jointly write the interference (4.24) and region (4.25) constraints as

$$\rho \leq \frac{P_s + \min\{\mathcal{M}, \mathcal{M}^s\}}{1+\Upsilon_0}, \tag{4.26}$$

so the optimum value of ρ is

$$\rho = \min\left\{P_s, \frac{P_s + \min\{\mathcal{M}, \mathcal{M}^s\}}{1+\Upsilon_0}\right\}. \tag{4.27}$$

Note that if $P_s < -\min\{\mathcal{M}, \mathcal{M}^s\}$ (4.27) is negative, then the problem is infeasible, as the available secondary power is not enough to allow the SR (or the worst case PR) to decode the primary message. If this is not the case, a rate

$$C_s = \min\left\{C\left(\frac{P_s|h_{s,r}|^2}{\sigma^2}\right), C\left(\frac{|h_{s,r}|^2}{\sigma^2}\frac{P_s + \max\{\mathcal{M}, \mathcal{M}^s\}}{1+\Upsilon_0}\right)\right\} \tag{4.28}$$

is achievable.

4.4.3 Medium interference

Finally, the third decoding strategy can be solved by the following optimization problem

$$\begin{aligned}
\text{max.} \quad & \frac{(\rho+\gamma)\,|h_{s,r}|^2}{\sigma^2} && (4.29)\\
\text{s.t.} \quad \gamma+\rho \;&\leq P_s && (4.30)\\
\gamma \;&\geq \rho\Upsilon_0 - \mathcal{M} && (4.31)\\
\gamma \;&\leq \rho\Upsilon_0 - \mathcal{M}^s && (4.32)\\
\gamma \;&\geq -\mathcal{M}^s && (4.33)\\
\gamma \;&\geq 0\\
\rho \;&\geq 0.
\end{aligned}$$

Remark that the feasibility of this problem requires $\mathcal{M}^s \leq \mathcal{M}$, $-\mathcal{M}^s \leq P_s$ and $P_s\Upsilon_0 \geq \mathcal{M}^s$. If we assume an active power constraint (4.30), the region (4.32) and interference (4.31) constraints can be written as

$$\frac{\Upsilon_0 P_s - \mathcal{M}^s}{1+\Upsilon_0} \geq \gamma \geq \frac{\Upsilon_0 P_s - \mathcal{M}}{1+\Upsilon_0} \tag{4.34}$$

so after taking into account the other region constraint (4.33), if we set $\gamma = \max\left\{-\mathcal{M}^s, \frac{\Upsilon_0 P_s - \mathcal{M}}{1+\Upsilon_0}\right\}$ and the problem is feasible then the achievable capacity is

$$C_s = C\left(\frac{|h_{s,r}|^2 P_s + |h_{p,r}|^2 P_p}{\sigma^2}\right) - R_p. \tag{4.35}$$

Note that the medium and strong interference regimes lead to different power allocation policies, since in the former the primary message is not directly decodable, whereas in the latter it is. However, it can be seen that the optimum power allocation in both regimes leads to the same capacity for the secondary system. In order to prove it, we will assume that $P_s \geq -\mathcal{M}^s \geq -\mathcal{M}$ and $P_s\Upsilon_0 \geq \mathcal{M}^s$, which is the condition for joint feasibility of (4.29) and (4.22). Therefore, by substituting (4.16) in (4.28), the capacity in strong interference can be written as

$$\begin{aligned}
C_{S,Strong} &= C\left(\frac{|h_{s,r}|^2}{\sigma^2}\frac{P_s+\mathcal{M}^s}{1+\Upsilon_0}\right) && (4.36)\\
&= \log_2\left(1+\frac{|h_{s,r}|^2 P_s - \Upsilon_0\sigma^2 + |h_{p,r}|^2 P_p}{\sigma^2\,(1+\Upsilon_0)}\right)\\
&= \log_2\left(\frac{|h_{s,r}|^2 P_s + \sigma^2 + |h_{p,r}|^2 P_p}{\sigma^2\,(1+\Upsilon_0)}\right).
\end{aligned}$$

Now, since $R_p = \log_2(1+\Upsilon_0)$, we can write (4.35) as

$$\begin{aligned} C_{S,Medium} &= \log_2\left(1+\frac{|h_{s,r}|^2P_s+|h_{p,r}|^2P_p}{\sigma^2}\right) \\ &\quad -\log_2(1+\Upsilon_0) \\ &= \log_2\left(\frac{\sigma^2+|h_{s,r}|^2P_s+|h_{p,r}|^2P_p}{\sigma^2(1+\Upsilon_0)}\right) \\ &= C_{S,Strong}. \end{aligned} \tag{4.37}$$

Note that the implications of this equality are quite significant, as a secondary transmission link working with this decoding strategy can be transformed into an equivalent system (i.e., with the same capacity) working with a strong interference and, therefore, the coverage area of the primary system could be extended.

4.4.4 Numerical results

The power allocation results were evaluated in different scenarios to obtain the resulting secondary user capacity. In order to get insight on the implications of the secondary system position, the results are shown as a function of this location. We assume that the PT is located at $\mathbf{x}_p = (0,0)$, the ST at $\mathbf{x}_s = (r,0)$ and the SR at $\mathbf{x}_r = (r+d,0)$. We also assume a free space propagation loss model, so the attenuation between two different points separated x units of distance can be written as x^2/k, with $k = \left(\frac{\lambda}{4\pi}\right)^2$. In the numerical results, we have set $d = 2$ and $k = 2$.

In order to properly describe the interference constraint, we define

$$\Delta P_i \triangleq |h_{p,i}|^2P_p - \Upsilon_0\sigma^2, \tag{4.38}$$

so $\mathcal{M}_i = \frac{1}{|h_{s,i}|^2}\Delta P_i$, and we define $\Delta P \triangleq \min\{\Delta P_i\}$.

If we restrict our secondary system to keep the same original coverage area, the worst case receiver would be located at the edge of the coverage zone, so $|h_{p,i}|^2P_p/\sigma^2 = \Upsilon_0$ and, therefore, $\Delta P = 0$, i.e., the worst case receiver has a zero power margin. If this is the case, the existence of a white space is impossible, strictly speaking, as even an infinitesimal extra interference coming from the ST would cause this extreme *worst case* receiver to fail.

The capacity results are shown in Figure 4.3 for different positions of the ST. The coverage limit ≈ 4.3 is the position of the ST for which the SR crosses the coverage edge ≈ 6.3. It can be seen that the optimum power allocation forces to be in the medium/high interference regimes, while the *treat as noise* scenario tends to the limit capacity of $\log_2\left(1+\Upsilon_0^{-1}\right)$, as depicted in

Section 4.4. Moreover, the system is quite far away from its performance in a true *white space*, with a capacity of $C_{noPU} = \log_2\left(1 + \frac{|h_{s,r}|^2 P_s}{\sigma^2}\right)$.

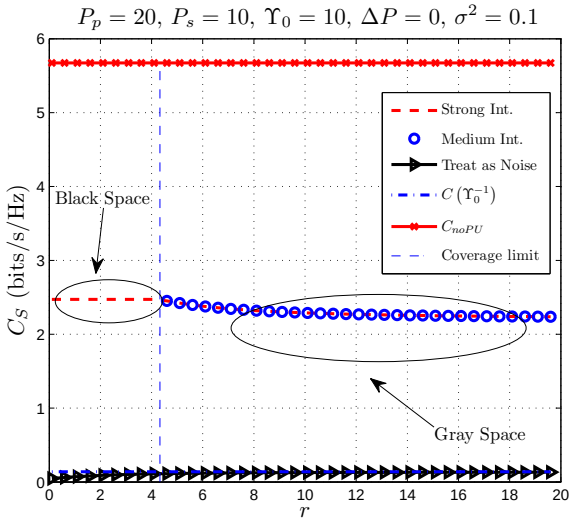


Figure 4.3: Capacity for the different decoding strategies with no coverage margin ($\Delta P = 0$).

Now, we propose to include a *power margin* that allows the existence of a smooth transition from the *black space* to a true *white space*. This could be the case, for example, of knowing the absence of PRs near the coverage edge. In this scenario, if the ST is inside the coverage zone we set $\mathcal{M} = 0$, as a PR could be located near the ST and, therefore, have a large $|h_{s,i}|^2$ value. When the ST is outside the coverage zone, the worst case receiver leads to a value of $\mathcal{M} = \frac{1}{|h_{s,i}|^2}\Delta P$, with $|h_{s,i}|^2$ calculated following the free space loss formula. Note that as the ST moves away from the coverage area $|h_{s,i}|^2 \to 0$, so the interference constraint (4.5) tends to $\gamma \geq -\infty$ for a positive ΔP, which is indeed a true *white space*, as there is no limit for the power allocated to the secondary message. In Figure 4.4 the achievable capacity results are shown for the three different decoding strategies. It can be seen that the behavior of the functions in the gray space is quite curious, as two clearly different zones exist: in the first one, near the coverage limit, the optimum power allocation forces to *expand* the coverage area and work in the medium-strong decoding region, thus *blackening* this gray space. However, as gray tends to white, the optimum decoding strategy consists on treating the primary interference as noise. Finally, when the ST is located very far from the coverage area, the secondary system capacity tends to the white space capacity.

In Figures 4.5a and 4.5b we show the fraction of power allocated to the primary and secondary messages at the ST, respectively. It is remarkable the big difference between the medium and strong interference optimum power allocation, as they lead to quite different values,

but to the same achievable rate for the ST. Another interesting result is the evolution of the power allocated to the primary and secondary message when using the *treat as noise* decoding: inside the coverage area this power is constant, and it can be seen that no power remains unused. When the ST leaves the coverage area no power is allocated to the primary message ($\gamma = 0$), and ρ starts to grow rapidly. In this zone (gray space), the capacity increases substantially with the distance as a result of two different effects: as both the ST and SR are moving away from the coverage area, the interference coming from the PT is clearly reduced, and also the value of $\mathcal{M}$ increases, so more power can be allocated to the secondary message without breaking the interference constraint. When the interference constraint allows to allocate all the power to the secondary message (white space), it can be seen that the capacity continues to increase, but in this case much slower, as we have only one of the two effects in the gray space: the decrease of the interference coming from the PT.

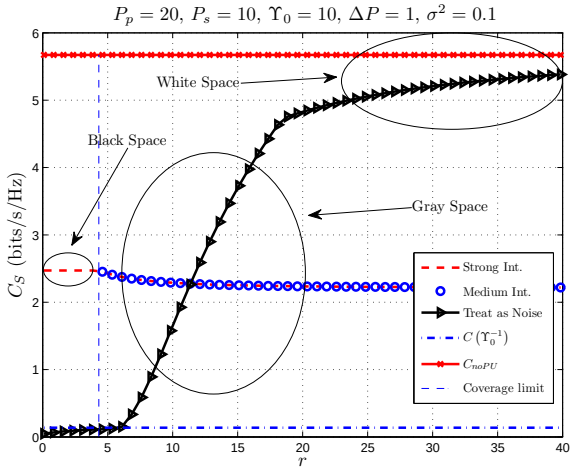
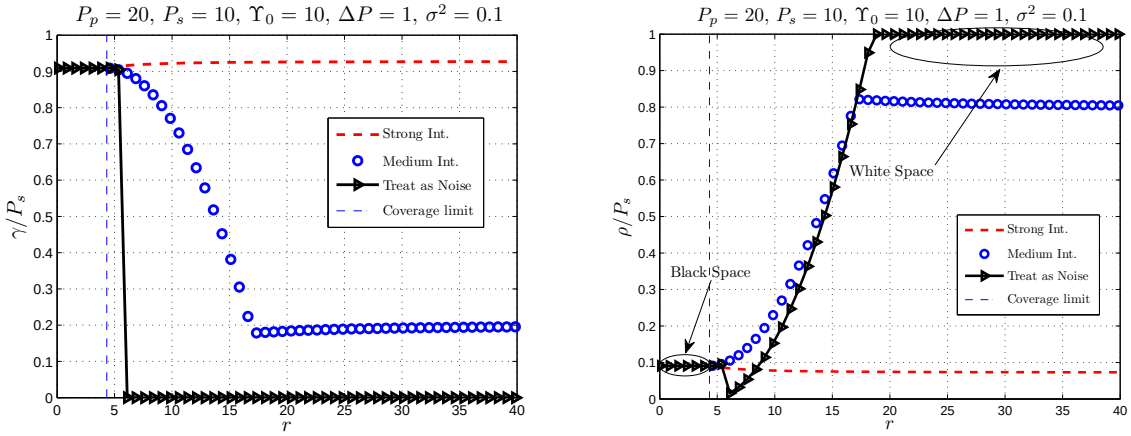


Figure 4.4: Capacity for the different power decoding strategies with a coverage margin ($\Delta P = 1$).

4.5 Optimum power allocation for multiple receivers

In this section we will study the effect of having multiple SRs on the capacity of the system. Depending on the nature of the secondary system, we can clearly distinguish two different scenarios:

- **Unicast** If this is the case, each SR is interested on a different message. An example of this kind of system could be a cellular network.



(a) Primary message.

(b) Secondary message.

Figure 4.5: Power allocation as a function of the position of the ST.

- **Multicast/Broadcast** Multiple SRs are interested in the same message. The main issue with this kind of systems is to adapt the transmission rate of the ST to deal with SRs in very different situations. An example of this type of system could be the insertion of a local TV transmitter *overlaid* on a global one, or the use of cellular-based multicast service like 3GPP multimedia broadcast multicast service (MBMS) in LTE [95].

Note that the unicast case can be solved just by considering one SR at a time (i.e., applying time division multiplexing - TDM), and afterwards adopting the optimum power allocation and decoding strategy for that SR. The case of having multiple multicast groups can be reduced to the broadcast case just by applying TDM again. However, the broadcast case is more complicated.

Consider a broadcast scenario with N_s SRs. We will extend the notation from the previous section to denote as h_{s,r_k} and h_{p,r_k} the channels from the ST and PT to the k-th SR, for $k = 1 ..., N_s$. In this broadcast analysis, we will only take into account the *treat as noise* and *strong interference* decoding strategies, as the *medium interference* one achieves the same capacity as the latter, and it requires an optimum power splitting (into ρ_1 and ρ_2) which heavily depends on the channels seen by a given receiver, thus being unpractical for multiple receivers. We will denote as

$$\mathcal{D}_k \in \{\mathcal{N}, \mathcal{S}\} \tag{4.39}$$

the decoding strategy of the k-th SR, where $\mathcal{D}_k = \mathcal{N}$ denotes that the k-th SR is decoding the secondary message treating the primary one as noise, and $\mathcal{D}_k = \mathcal{S}$ denotes a strong interference

operation. We denote the capacity under both operation modes as

$$C_k^{\mathcal{N}}(\rho,\gamma) = C\left(\frac{|h_{s,r_k}|^2\rho}{\sigma^2 + |h_{p,r_k}|^2 P_p + |h_{s,r_k}|^2\gamma}\right) \tag{4.40}$$

for the $\mathcal{N}$ operation and

$$C_k^{\mathcal{S}}(\rho,\gamma) = C\left(\frac{|h_{s,r_k}|^2\rho}{\sigma^2}\right) \tag{4.41}$$

for the $\mathcal{S}$ operation. The latter requires the constraint

$$f_k^{\mathcal{S}}(\rho,\gamma) \triangleq \Upsilon_0\rho - \gamma - \mathcal{M}_{r_k}^s \leq 0 \tag{4.42}$$

to be met, where we have defined

$$\mathcal{M}_{r_k}^s \triangleq \frac{1}{|h_{s,r_k}|^2}\left(|h_{p,r_k}|^2 P_p - \Upsilon_0\sigma^2\right). \tag{4.43}$$

Our objective is to obtain, given N_s receivers, the maximum rate R_s such that no PR is compromised and the N_s SRs are able to decode the secondary message. In the single receiver approach we had to solve a different optimization problem for each of the three decoding strategies. In this case, after dropping the *medium interference* strategy, we have that every receiver can operate with two different decoding procedures. Thus, our design variables are, in this case, the power allocation weights (ρ,γ) as well as the vector of decoding strategies $\boldsymbol{\mathcal{D}} = (\mathcal{D}_1, ..., \mathcal{D}_{N_s}) \in \{\mathcal{N},\mathcal{S}\}^{N_s}$:

$$\begin{array}{lll} \text{max.} & & \min_{k=1,...,N_s}\left\{C_k^{\mathcal{D}_k}(\rho,\gamma)\right\} \\ \text{s.t.} & \gamma+\rho & \leq P_s \\ & \gamma & \geq \rho\Upsilon_0 - \mathcal{M} \\ & f_k^{\mathcal{D}_k}(\rho,\gamma) & \leq 0,\ k=1,...,N_s \\ & \gamma & \geq 0 \\ & \rho & \geq 0 \end{array} \tag{4.44}$$

where we have introduced a *dummy constraint* $f_k^{\mathcal{N}}(\rho,\gamma) \triangleq 0$ for clarity in the notation. Note that for a given $\boldsymbol{\mathcal{D}}$, the optimization problem (4.44) is a generalized linear fractional program [94], which is quasiconvex and, therefore, can be efficiently solved.

Thus, the main problem is to choose the optimum value of $\boldsymbol{\mathcal{D}}$ among the 2^{N_s} possible vectors, which could be computationally infeasible for a large number of receivers. However, note that:

- For a given power allocation (ρ, γ), if a receiver is able to decode the primary message, then treating it as noise is suboptimal.
- If the j-th user chooses to decode the primary message, i.e., $\mathcal{D}_j = \mathcal{S}$, then the constraint $f_j^{\mathcal{S}} \leq 0$ has to be met, and can be rewritten as $\gamma \geq \Upsilon_0 \rho - \mathcal{M}_{r_j}^s$, so if $f_j^{\mathcal{S}} \leq 0$ then $f_k^{\mathcal{S}} \leq 0$ for k such that $\mathcal{M}_{r_k}^s \geq \mathcal{M}_{r_j}^s$. This means that if a given user is able to decode the primary message, then all the receivers with a larger *power margin* will also be able to decode it.

These two facts have a very strong implication, as it suffices to try at most[1] $N_s + 1$ different decoding strategies, so the complexity of the problem is substantially reduced. If we assume that the receivers are ordered in decreasing order of $\mathcal{M}_{r_k}^s$, i.e., $\mathcal{M}_{r_k}^s \geq \mathcal{M}_{r_j}^s \, \forall j \geq k$, then only the following strategy vectors can be optimum

$$\begin{aligned} \boldsymbol{\mathcal{D}}^{(0} &= (\mathcal{N}, \mathcal{N}, ..., \mathcal{N}) \\ \boldsymbol{\mathcal{D}}^{(1} &= (\mathcal{S}, \mathcal{N}, ..., \mathcal{N}) \\ \boldsymbol{\mathcal{D}}^{(2} &= (\mathcal{S}, \mathcal{S}, ..., \mathcal{N}) \\ &... \\ \boldsymbol{\mathcal{D}}^{(N_s} &= (\mathcal{S}, \mathcal{S}, ..., \mathcal{S}) \end{aligned} \tag{4.45}$$

so it suffices to solve $N_s + 1$ quasiconvex optimization problems to obtain the optimum power allocation and strategy vector.

4.5.1 Numerical results

We have numerically solved the optimization problem in (4.44) in different scenarios. In all the simulations we have set $P_p = 20$, $\sigma^2 = 0.1$, $\Upsilon_0 = 10$, and a free space propagation model was assumed, with $k = 2$. The positions of the PRs were generated following an uniform distribution inside the coverage area, leaving a small portion as a protection area that allows the existence of a white space, similarly to the single receiver study. The positions of the SRs were also generated following an uniform distribution in circles of different dimensions, depending on the scenario.

Figure 4.6 shows a secondary system positioned near the primary coverage edge, with the SRs concentrated near the ST. In this case, the optimum decoding strategy is to decode the primary message for all the SRs, so the primary coverage area is clearly extended. Moreover, due to the proximity of the SRs to the ST, a relatively high rate is achieved.

[1] This number can be further reduced if the interference constraint forces some of the SRs to be able to decode the primary message.

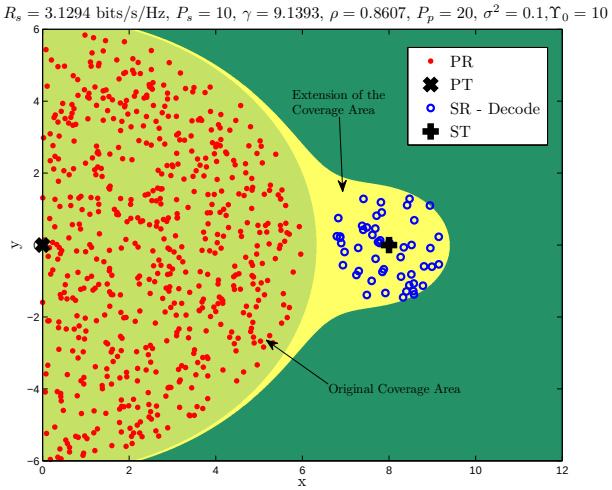


Figure 4.6: Initial and final coverage area for a scenario with a ST with SRs concentrated near the transmitter. In this case, the optimum decoding strategy is to decode the primary transmitter in all the receivers, so the coverage area is expanded.

The picture changes dramatically if the SRs can be situated far away from the ST. In Figure 4.7 the optimum decoding strategy is $\mathcal{N}$ for the SRs situated far away from the primary coverage area, and $\mathcal{S}$ for those clearly inside it. In this case, there are some SRs that are situated far from the ST and near the coverage edge, so the signal coming from the ST is very attenuated and they receive a large amount of interference (which are not able to decode) from the PT. The achievable rate in this case is much smaller and the primary coverage area is slightly diminished.

If we move the secondary system far from the primary coverage edge, we end up in a white space scenario, which is shown in Figure 4.8. In this case all the power is allocated to the secondary message and the optimum decoding strategy is to treat as noise the very weak interference coming from the PT. The primary coverage area is also reduced in this case.

A case of special interest can be the one where the insertion of the ST is clearly beneficial for the primary system: in many real-world cases a broadcaster (e.g. TV operators) needs to extend its coverage area to give its service to some new receivers. In this case, the primary operator could incur in a relatively high expense to cover a small number of new receivers. Following the overlay paradigm, the primary operator could allow the existence of a secondary system with the condition of extending the primary coverage zone to a set of new PRs. This scenario is shown in Figure 4.9, which is identical to that in Figure 4.6 except for the presence of PRs outside the original coverage area. The resulting coverage area can be seen to be substantially larger

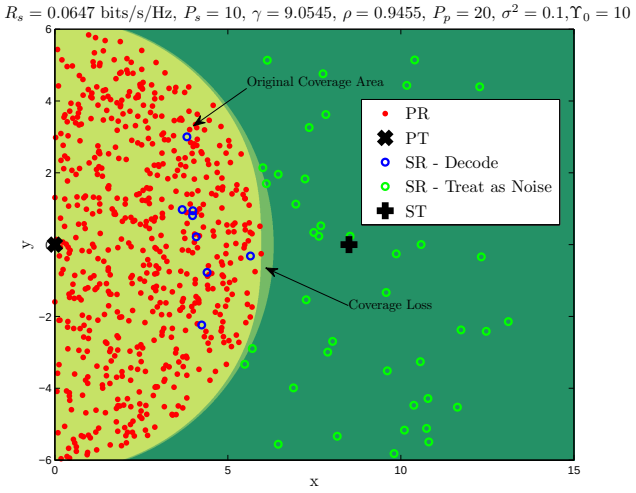


Figure 4.7: Initial and final coverage area for a scenario with a ST with some SRs situated far away from the transmitter. In this case, the optimum decoding strategy is to decode the primary transmitter for the receivers inside the final coverage zone, and to treat it as noise for those outside it. The final coverage area is slightly smaller than the initial one.

than the one in Figure 4.6 so the new PRs lie inside it. This larger coverage area forces the ST to allocate more power to the primary message, so a smaller rate is achieved for the secondary system.

4.6 Realistic impairments: directive antennas, propagation models and receiver overloading

4.6.1 Introduction

In this section we present some practical impairments that might arise in a realistic scenario. Throughout this chapter, we assumed that the receivers used omnidirectional antennas and a deterministic free space loss model was used in the simulations. Another practical problem appears when the transmitters and receivers are located nearby so that the received power is too large. In this case, the power amplifiers of the receivers suffer from overloading, and reception quality is severely degraded due to non linearities. This is a problem that has been studied in the scenario of coexistence between television and cellular systems [96]. In this section we focus on the effect of receiver overloading, that arises in a black space scenario. The analysis in this section is focused on transmit and received signal power, so the notation is slightly different

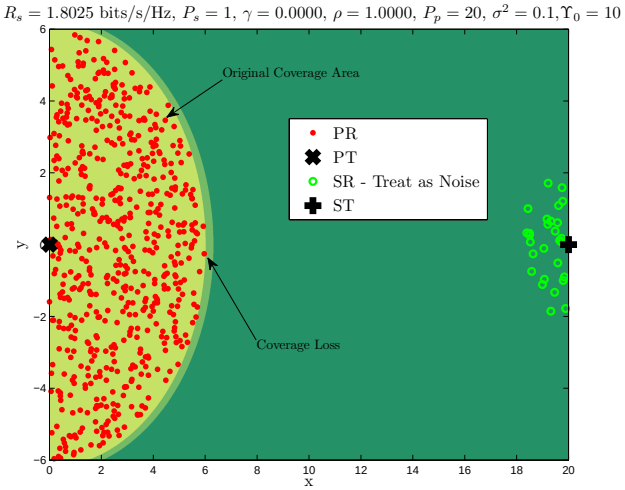


Figure 4.8: Initial and final coverage area for a ST operating in a white space. In this case, all the power can be allocated to the secondary message without breaking the interference constraint. The transmit power of the secondary transmitter has been diminished to $P_s = 1$ to force the existence of the *white space* near the primary coverage zone.

from the previous sections.

4.6.2 Problem statement

Consider a scenario with a single primary transmitter, located at $\mathbf{x}_p = \mathbf{0} \in \mathbb{R}^2$ that transmits with an equivalent isotropic radiated power (EIRP) of P_p milliwatts and a secondary transmitter located at $\mathbf{x}_s \in \mathbb{R}^2$ with an EIRP of P_s. In this section, we denote with capital P the transmit power in milliwatts, and with lowercase p the power in dBm[2]. For example, $P_s \triangleq 10^{p_s/10}$ and $P_p \triangleq 10^{p_p/10}$. We use capital R for the received power in milliwatts, and lowercase r for the received power in dBm. In general, depending on the scenario, the secondary transmitter can use all its available power to transmit the secondary message (in an underlay scenario) or split the power between ρ and γ, such that $P_s = \rho + \gamma$.

For a given primary receiver located at $\mathbf{x}_r$, the primary message is received with power

$$R_p(\mathbf{x}_r) = R_{p,p}(\mathbf{x}_r) + R_{s,p}(\mathbf{x}_r) \tag{4.46}$$

where $R_{p,p}$ and $R_{s,p}$ denote the received power of the primary message, received from the primary

[2]Although the usual notation is the opposite (capital P for dBm), we prefer this one for consistency with the previous sections.

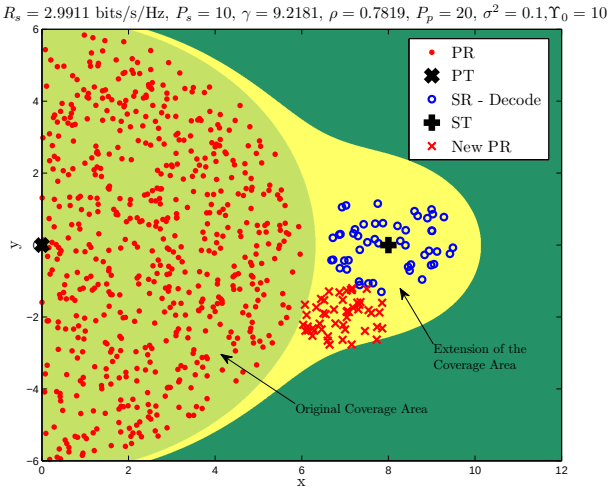


Figure 4.9: Initial and final coverage zone for a ST forced to enlarge the primary coverage area by the insertion of *New PRs*.

and secondary transmitters, respectively. These quantities can be readily obtained in dBm as

$$r_{p,p}(\mathbf{x}_r) \triangleq p_p + g(\theta_p) - \ell(\mathbf{x}_p, \mathbf{x}_r) \tag{4.47}$$

and

$$r_{s,p}(\mathbf{x}_r) \triangleq 10\log_{10}(\gamma) + g(\theta_s) - \ell(\mathbf{x}_s, \mathbf{x}_r) \tag{4.48}$$

with $\ell(\mathbf{x}_1, \mathbf{x}_2)$ the (random) propagation losses in dB between a transmitter located at $\mathbf{x}_1$ and a receiver located at $\mathbf{x}_2$, and $g(\theta)$ the receive antenna gain in dB for an incoming angle θ. Note that in general the incoming angles for the signal coming from the primary and secondary transmitters, θ_p and θ_s respectively, are not equal. In practice, the primary receive antennas are aimed at the primary transmitter to maximize the amount of received power, so we assume $\theta_p = 0$. The incoming angle for the secondary transmitter can be calculated as the angle between $\mathbf{x}_s$ and $\mathbf{x}_p$:

$$\theta_s = \cos^{-1}\left(\frac{\mathbf{x}_s^T \mathbf{x}_r}{\|\mathbf{x}_s\|_2 \|\mathbf{x}_p\|_2}\right). \tag{4.49}$$

We define in the same way the received power for the secondary message:

$$r_s(\mathbf{x}_r) = 10\log_{10}(\rho) + g(\theta_s) - \ell(\mathbf{x}_s, \mathbf{x}_r). \tag{4.50}$$

The signal to interference plus noise ratio (SINR) at a primary receiver can be written as

$$\Upsilon\left(\mathbf{x}_r\right) = \frac{R_p\left(\mathbf{x}_r\right)}{R_s\left(\mathbf{x}_r\right) + \Omega\left(\mathbf{x}_r\right)} \tag{4.51}$$

with Ω the receiver noise power in milliwatts. The total received power is

$$R\left(\mathbf{x}_r\right) = R_p\left(\mathbf{x}_r\right) + R_s\left(\mathbf{x}_r\right) \tag{4.52}$$

where the noise power was assumed to be negligible with respect to the total received power from the primary and secondary transmitters.

In general, the propagation loss ℓ is modeled as a random variable, so following the usual approach in broadcasting scenarios [96] we can define the overloading probability as

$$\mathcal{O}\left(\mathbf{x}_r\right) \triangleq \mathbb{P}\left[r\left(\mathbf{x}_r\right) > r_0\right]. \tag{4.53}$$

To assess the effect of overloading on coverage probability, we assume that an overloaded receiver is not able to successfully decode the primary message. This degradation is caused by the non-linearities introduced in the signal by the excessive amount of received power. Thus, the coverage probability for the primary user is defined as

$$\mathcal{C}\left(\mathbf{x}_r\right) \triangleq \mathbb{P}\left[\left(r\left(\mathbf{x}_r\right) \leq r_0\right) \cap \left(\Upsilon\left(\mathbf{x}_r\right) \geq \Upsilon_0\right)\right] \tag{4.54}$$

with r_0 the overloading threshold, and Υ_0 the necessary SINR for decoding the primary message.

4.6.3 Power allocation

The coverage probability of a primary receiver is highly dependent on the power allocation at the secondary transmitter between the primary and secondary messages γ and ρ. In general, the optimum power allocation depends on the position of the closest primary receiver and the strategy used by the secondary receiver to decode the secondary message, as seen in the previous sections.

If the secondary transmitter is placed inside the primary coverage area, then it is likely to find a primary receiver close to it. This is the scenario in which overloading is more likely to occur. In such a case, the power received from the primary transmitter and the noise can be considered negligible with respect to the one received from the secondary transmitter, so the

SINR can be approximated as

$$\Upsilon\left(\mathbf{x}_r\right) = \frac{R_p\left(\mathbf{x}_r\right)}{R_s\left(\mathbf{x}_r\right) + \Omega\left(\mathbf{x}_r\right)} \approx \frac{\gamma}{\rho}, \tag{4.55}$$

i.e., the SINR equals the ratio between the power allocated to the primary and secondary messages. Therefore, the power ratio between primary and secondary messages must meet $\frac{\gamma}{\rho} \geq \Upsilon_0$. The optimum power allocation for a transmitter operating in a *black space* is $\frac{\gamma}{\rho} = \Upsilon_0$, as shown in the previous sections.

We do not discuss in more depth the power weighting between primary and secondary messages, as the aim of this section is to analyze the effect of overloading in cognitive radio networks. In the simulations, we set the ratio $\frac{\gamma}{\rho} > \Upsilon_0$, so the coverage holes created by the secondary transmitters are exclusively due to receiver overloading.

4.6.4 Propagation and antenna model

The SINR and overloading probability and, consequently, the coverage probability, are highly dependent on the geometry of the receive antenna. In this section we describe the models for the propagation losses and the antenna gain.

The propagation model we consider is a modified Okumura-Hata model, taken from [80]. Under this model, the pathloss is modeled as a Gaussian random variable in dB (a lognormal random variable in natural units), and its median and typical deviation depend on parameters like the distance between transmitter and receiver, the height of both antennas, etc.

Let us denote by $d \triangleq \|\mathbf{x}_r - \mathbf{x}\|$ the distance between a receiver located at $\mathbf{x}_r$ and a transmitter located at $\mathbf{x}$. The transmit and receive antennas have a height of H_T and H_r, respectively. We define the minimum and maximum heights as $H_m \triangleq \min\{H_T, H_r\}$ and $H_b = \max\{H_T, H_r\}$. The median pathloss $\mu \triangleq \mathbb{E}[L]$ is given by (4.56), where a, b and α are defined in (4.57-4.59). The standard deviation of the propagation losses $\sigma \triangleq \sqrt{\mathbb{E}\left[(\mu - L)^2\right]}$ is given by (4.60) as a function of the distance d.

The antenna gain for the primary receivers is shown in Figure 4.10, and corresponds to a standard Yagi antenna, frequently used for the reception of terrestrial television.

4.6.5 Results

We simulated a simple scenario with a primary transmitter and a secondary transmitter, placed at $\mathbf{x}_s = (0, 4)$km. The EIRP of the primary transmitter was set to $p_p = 60$dBm, and that of the

$$\mathbb{E}\left[\ell\left(\mathbf{x}_r,\mathbf{x}\right)\right]=\begin{cases}32.4+20\log_{10}(f)+10\log_{10}(d^2+(H_b-H_m)^2/10^6) & \text{if } d\leq 0.04\text{km}\\ 69.6+26.2\log_{10}(f)-13.82\log_{10}\max\left\{30,H_b\right\}+ & \\ \left[44.9-6.55\log_{10}(\max\left\{30,H_b\right\})\right]\log_{10}(d)^{\alpha}-a\left(H_m\right)-b\left(H_b\right) & \text{d } >0.04\text{km}\end{cases} \tag{4.56}$$

$$a(H_m)=(1.1\log_{10}(f)-0.7)\min\{10,H_m\}-(1.56\log_{10}(f)-0.8)+\max\{0,20\log_{10}(H_m/10)\} \tag{4.57}$$

$$b(H_b)=\min\left\{0,20\log_{10}\left(H_b/30\right)\right\} \tag{4.58}$$

$$\alpha=\begin{cases}1 & \text{if } d\leq 20\text{km}\\ 1+\left(0.14+1.87\cdot 10^{-4}f+1.07\cdot 10^{-3}H_b\right)\left(\log_{10}\frac{d}{20}\right)^{0.8} & \text{if } d>20\text{km}\end{cases} \tag{4.59}$$

$$\sigma=\begin{cases}3.5 & \text{if } d\leq 0.04\text{km}\\ 3.5+\frac{12-3.5}{0.1-0.04}\left(d-0.04\right) & \text{if } 0.04\text{km}<d\leq 0.1\text{km}\\ 12 & \text{if } 0.1\text{km}<d\leq 0.2\text{km}\\ 12+\frac{9-12}{0.6-0.2}\left(d-0.2\right) & \text{if } 0.2\text{km}<d\leq 0.6\text{km}\\ 9 & \text{if } 0.2\text{km}<d\leq 0.6\text{km}\end{cases} \tag{4.60}$$

---

secondary transmitter to $10\log_{10}(\gamma)$ =50dBm, $10\log_{10}(\rho)$ = 35dBm. The limit SINR for decoding was set to Υ_0 = 10dB, so the *transmit SINR* of the secondary transmitter $10\log_{10}\left(\frac{\gamma}{\rho}\right)$ = 15dB is enough for decoding the primary message in the neighborhood of the secondary transmitter with a 5dB margin. The frequency of operation is 830MHz.

The height of the primary transmitter was set to 100m, while the height of the secondary transmitter and all primary receivers was set to 20m. Propagation over the roofs in an urban scenario was always assumed.

The noise power was set to Ω =-98dBm, and the overloading limit to r_0 =-15dBm (the same limit used in [96]).

With these simulation parameters the average SINR $\mathbb{E}\left[\Upsilon\right]$ map shown in Figure 4.11 was obtained. It can be seen that in the neighborhood of the secondary transmitter the SINR is reduced but it is always above the correct reception threshold. Note that the irregular shapes of the curves are due to the directivity of the receive antenna, as the receivers with an incoming angle θ_s corresponding to a secondary lobe will be more affected than others with θ_s situated in a null gain angle. This effect is not observed in Figures 4.6-4.9 due to the assumption of omnidirectional antennas. Also, the receivers that are aiming at the secondary transmitter (those with $\mathbf{x}_r=(x_0,0)$ with $x_0>4$) are even more affected by the presence of the secondary

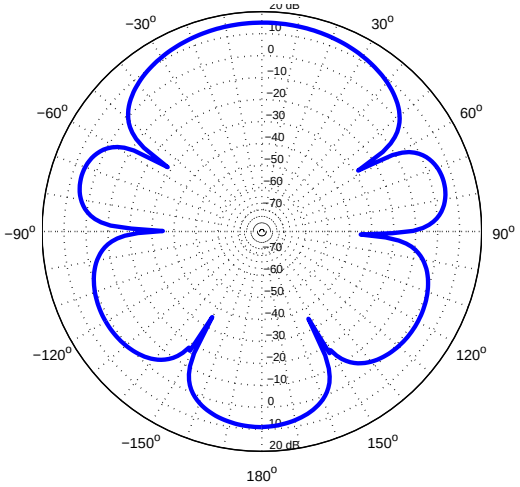


Figure 4.10: Antenna gain for the primary receivers, corresponding to a Yagi antenna.

transmitter.

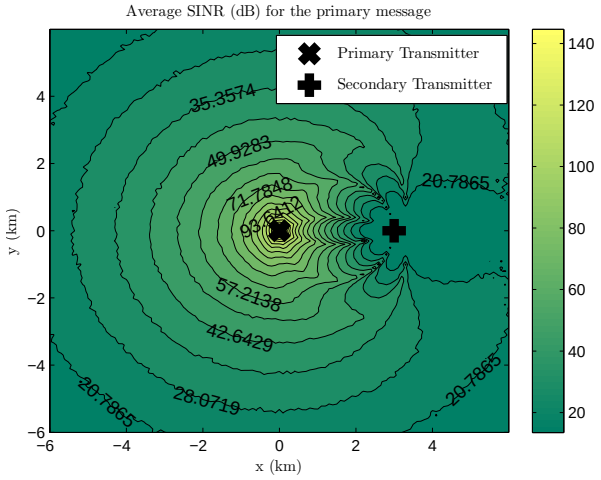


Figure 4.11: Average SINR in decibels for the primary message. The SINR next to the secondary transmitter is reduced, but is always kept above the desired threshold.

In Figure 4.12 we show a map with the evolution of the overloading probability. It can be seen that, as expected, the receivers near the secondary transmitter experience high overloading probabilities. Note that the receivers near the primary transmitter do also suffer from overloading, but it is usual that the receiver infrastructure in that zone is prepared to deal with this problem (for example, by the insertion of a attenuator after the receive antenna, and before the power amplifier). The case of the zone near the secondary transmitter is different, as the receivers are not prepared for such a high signal level, and they will require to modify the receive

chain, thus disabling the possibility of unplanned deployment of secondary networks. Note that the overloading caused by the secondary transmitter is not isotropic due to the directivity of the receive antennas.

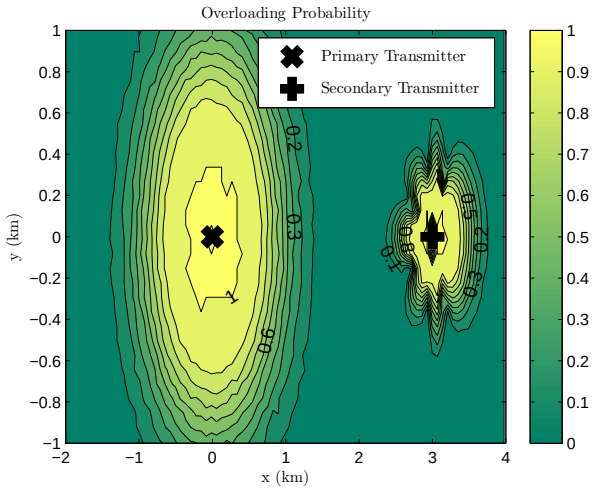


Figure 4.12: Overloading probability of the primary receivers. The overloading area around the secondary transmitter is not circular due to the directivity of the receive antennas.

Finally, in Figure 4.13 the probability of coverage is shown, taken into account both the SINR and the overloading probability. Not surprisingly, the zones near the transmitters appear as *coverage holes* due to the receive overloading. Moreover, it can be seen that the insertion of the secondary transmitter is beneficial for the primary system, as the (approximately probability one, lightest zone) coverage area reaches more than 5 kilometers to the east of the primary transmitter (where the secondary transmitter is located) and only about 3 kilometers to the west, north and south.

4.7 Concluding remarks

In this chapter we analyzed the problem of overlay cognitive access to the spectrum licensed to a primary broadcast user. We show that, in some cases the insertion of the secondary network is beneficial for the primary system, as it expands its coverage area. This kind of overlay operation could result of special interest in those cases where the primary service provider outsources the extension of his coverage area to a secondary operator. We also studied the problems arising from a high power secondary transmitter, which can overload the primary receivers.

The content in this chapter is an extended version of a joint paper with Prof. Carlos

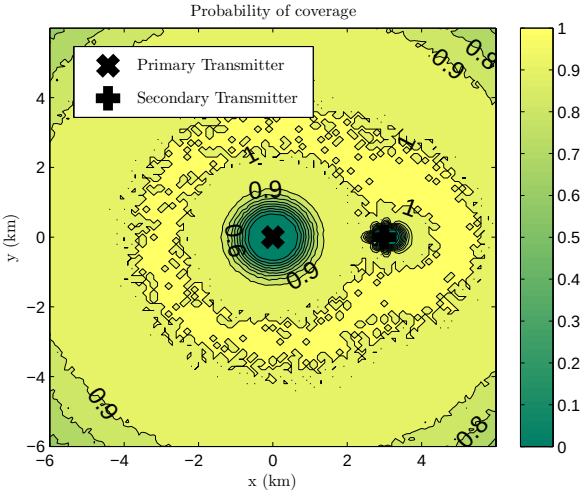


Figure 4.13: Coverage probability of the primary system. The probability of correct reception is calculated by taking into account the SINR and overloading probabilities.

Mosquera presented in DySPAN 2012 [97].

Chapter 5

Point to Point Transmission Exploiting Primary Feedback

Contents

5.1 Introduction

In the previous chapters we analyzed the insertion of a secondary transmitter in a primary broadcast network. One of the main problems we dealt with is the absence of channel state information (CSI) at the transmitter. This information cannot be acquired due to the broadcast nature of the primary user, as the primary receivers are not able to feed any information back even to the primary transmitter. In point to point systems, however, the secondary transmitter can exploit some degree of CSI to access the spectrum. In this chapter, we explore simple techniques that allow gaining access to CSI, and show how a properly designed secondary transmitter can improve the rate of the primary system to gain access to a fraction of the transmit resources.

In point to point systems, some CSI information can be obtained by analyzing the message exchange between primary transmitter and receiver. Depending on the nature of the system, and especially on the existence of feedback channels and adaptive transmission, the primary transmitter can change its behavior in presence of a secondary transmitter that conveys the primary information. For example, in the case of point to point communications it is usual to have a feedback channel from the receiver to the transmitter to perform some tasks such as adaptive modulation and coding (AMC), power and bit loading, etc. In this case the secondary transmitter can obtain additional information about the primary link if it is able to demodulate the feedback signal. In these systems, the primary transmitter can operate in different modes, trying to maximize the spectral efficiency for a given channel quality. Thus, the correct metric for the primary user communication is the resulting bit rate. If the bit rate is larger than the one the primary communication needs, the primary transmitter will be able to free some transmission resources (in the frequency or time domain) [16,17]. If this is the case, the secondary information can be transmitted in the released resources due to the increased capacity of the primary link.

It is usually assumed, however, that the secondary transmitter has full CSI, or, if measured in capacity terms, that the contributions from the primary and secondary transmitters are coherently added at the primary receiver location [16]. As opposed to this, in this chapter we focus on a point to point scenario where the secondary transmitter maximizes the rate of the primary link based on partial CSI. This partial CSI is obtained by means of the primary feedback containing the measured signal to noise ratio (SNR). We show how to exploit this information to obtain CSI.

As we will show, it is possible to acquire CSI based on SNR information, but the analysis and generalization of this method is somehow complicated. To overcome this problem, we extend the analysis to the case of having a primary receiver feeding back a complex channel estimate,

following ideas taken from [15]. Under this setting, we develop CSI acquisition strategies for multiple input multiple output (MIMO) and time varying channels.

This chapter is organized as follows: in Section 5.2 the proposed scenario is presented; in Section 5.3 a general optimization problem, based on effective SNR metrics, is introduced; in Section 5.4 different approximations for the mutual information effective SNR metric are presented, and the optimization problem is solved assuming perfect CSI; in Section 5.5 the problem of obtaining CSI by exploiting the primary feedback channel is stated and the effects of imperfect CSI in the previously solved optimization problem are presented; Section 5.6 presents the CSI acquisition problem with complex channel feedback, and the extension to MIMO and time varying channels; finally, Section 5.7 concludes the chapter.

5.2 Proposed scenario

Consider an orthogonal frequency division multiplexing (OFDM) point to point communication system where a primary transmitter (PT) is communicating with a single primary receiver (PR). A secondary transmitter (ST) tries to exploit the knowledge of the primary signal in order to communicate with a secondary receiver (SR). The PR conveys CSI to the PT, so the latter can use AMC to maximize the link throughput, or minimize the communication time. Assume that the communication needs of the primary system can be set to R_p bits per time unit. For the sake of simplicity, we assume that the OFDM symbols have constant length, and use the OFDM block as the time unit. In this case, R_p denotes the necessary bit rate for the primary system, measured in terms of bits per OFDM symbol.

Depending on the quality of the link, the resulting AMC mode will set the transmission rate to R_s bits per OFDM block. It is clear that if $R_s < R_p$ the link does not provide enough quality for the transmission, but if $R_s \geq R_p$ only a fraction $\rho \triangleq \frac{R_p}{R_s}$ of the transmission resources will be used, and the remaining $(1-\rho)$ could be used by the secondary transmitter. For a system using adaptive repeat request (ARQ), another convenient figure of merit is the throughput or the goodput, metrics that include the performance loss due to the presence of message errors and the corresponding retransmissions. In this case, the fraction of used resources can be written as $\rho = \frac{\mu_p}{\mu_s}$, where μ_p denotes the throughput required by the primary system, and μ_s the total throughput after the insertion of the secondary transmitter. Thus, the maximization of the secondary user rate is equivalent to the minimization of ρ or, equivalently, the maximization of R_s or μ_s.

In general, choosing the correct AMC mode for a given channel state is not a trivial problem when facing frequency selective channels in OFDM communications, as the mean SNR is not

a good indicator of the channel quality. In order to face this problem, different effective SNR metrics (ESM) were recently developed [39]. These metrics can be expressed as a generalized mean, parametrized[1] by the function $\Phi(\cdot)$, of the SNR at each carrier

$$\psi_e = \Phi^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\Phi(\psi_i)\right) \tag{5.1}$$

where ψ_e denotes the effective SNR, N denotes the number of carriers, and ψ_i denotes the SNR of the i-th carrier. The AMC mode (also known as modulation and coding scheme - MCS) will be selected depending on the value ψ_e from a set $\mathcal{M} = \{m_1, ..., m_M\}$ of M different modes, each one with an associated rate of R_i $i = 1 ... M$ bits per OFDM block. Without loss of generality, we will order the modes in such a way that $R_1 < R_2 < ... < R_M$, with associated mode thresholds $0 = t_0 < t_1 < ... < t_{M-1} < t_M = +\infty$ such that the mode m_i is selected if $t_{i-1} < \psi_e < t_i$. In general, the throughput function is more difficult to approximate, as it must take into account packet errors and retransmissions. In this chapter, we approximate the throughput $\mu(\psi_e)$ by the linear interpolation of the rate at the MCS threshold values, i.e. $\mu(\psi_e) = \frac{R_i - R_{i-1}}{t_i - t_{i-1}}(\psi_e - t_{i-1}) + R_{i-1}$, with $t_{i-1} \leq \psi_e < t_i$.

Let us denote by η_i the complex channel coefficient on the i-th carrier of the PT to PR link, and by σ^2 the noise power at the PR, assumed to be constant through all carreirs. Thus, in absence of the ST, and assuming a unit power primary signal, we can write $\psi_i = \frac{|\eta_i|^2}{\sigma^2}$. With the insertion of the ST, the resulting SNR can be written as

$$\psi_i = \frac{|\alpha_i\gamma_i + \eta_i|^2}{\sigma^2} \tag{5.2}$$

where α_i is the complex channel coefficient of the i-th carrier of the ST to PR link, and γ_i is a one tap pre-equalizer at the ST that allows us to change the amplitude and phase of the transmitted symbols. If the ST had access to perfect CSI, then the role of γ_i would be to perform power allocation and distributed beamforming, so that the replicas from the PT and ST were coherently added at the PR. This scenario is depicted in Figure 5.1.

5.3 Problem statement

The objective of the secondary user is to maximize the quality of the primary link (measured in terms of ESM) to obtain a fraction of released resources $1 - \rho$ as large as possible. In our first

[1]For example, following expression (5.1), the arithmetic mean is parametrized by the function $\Phi(x) = x$, and the geometric mean by $\Phi(x) = \log(x)$.

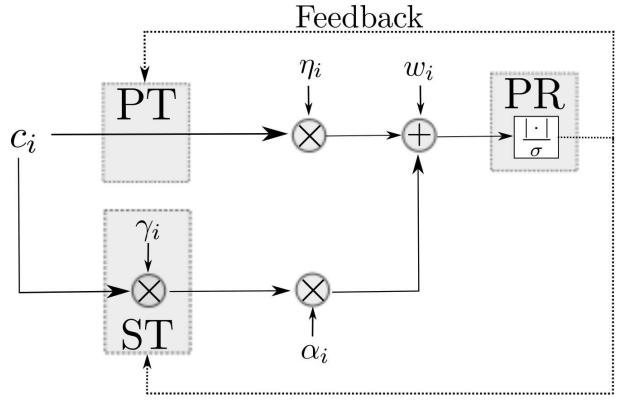


Figure 5.1: Diagram of the proposed scenario. The ST uses a one-tap per carrier pre-equalizer to maximize the primary link ESM. The PR feeds back information related to the SNR of each carrier (see Section 5.5). c_i is the symbol to be transmitted in the i-th carrier in a given OFDM block, assumed to have unit variance, and $w_i \sim \mathcal{CN}(0, \sigma^2)$.

approach, we assume that the channel coefficients α_i, η_i are perfectly known at the secondary transmitter, and in Section 5.5 a method that estimates these parameters exploiting the feedback channel will be described.

The design variables in our optimization problem are the complex values γ_i for a total transmit power below a given value P. For a given power allocated to the i-th carrier $|\gamma_i|^2$, the optimum value for γ_i is $\gamma_i = |\gamma_i| e^{j(\angle\eta_i - \angle\alpha_i)}$, so the signals are coherently combined at the receiver, and $\psi_i = \left(|\alpha_i\gamma_i|^2 + |\eta_i|^2 + 2|\alpha_i\gamma_i\eta_i|\right)/\sigma^2$. Therefore, and without loss of generality, we assume that α_i, γ_i and η_i are real and non-negative values (just by taking the modulus of the complex coefficients), so the optimization problem can be stated as

$$\begin{array}{ll} \text{minimize} & -\Phi^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\Phi\left(\psi_i\right)\right) \\ \text{subject to} & \frac{1}{N}\sum_{i=1}^{N}\gamma_i^2 \leq P \\ & -\boldsymbol{\gamma} \preceq 0 \end{array} \tag{5.3}$$

with $\psi_i = \frac{(\alpha_i\gamma_i+\eta_i)^2}{\sigma^2}$, and $\boldsymbol{\gamma} = [\gamma_1, \ldots \gamma_N]^T$. Obviously, the result of the optimization problem will vary depending on the function Φ. In the following section, we will study the optimum power allocation corresponding to the mutual information ESM (MIESM).

5.4 Power allocation for MIESM metric

The MIESM is based on the mutual information per bit. The expression for Φ, taken from [73], is

$$\Phi\left(\psi\right) = 1 - \frac{1}{M \log_2 M} \sum_{m=1}^{M} \mathbb{E}_U \left[\log_2 \left(\sum_{k=1}^{M} e^{-\frac{|X_m - X_k + U| - |U|^2}{1/\psi}} \right) \right] \tag{5.4}$$

where M is the number of symbols in the constellation, and U is a complex Gaussian random variable of zero mean and variance $1/\psi$ ($1/(2\psi)$ per real dimension). We also denote by X_m, $m = 1, ..., M$ the M complex constellation points. Note that this metric does not depend on the code rate being used, but only on the constellation.

As there is not a closed-form expression for (5.4), we will approximate Φ by two different functions in order to obtain analytical results of interest, although these results will be evaluated using the actual value of Φ, obtained by Monte Carlo integration.

On a first approach, we will approximate Φ by a parametrized exponential function, similarly to [41],

$$\Phi(\psi) = 1 - \sum_{l=1}^{L} \phi_l e^{-\beta_l \psi} \tag{5.5}$$

where $\sum_{l=1}^{L} \phi_l = 1$, and $\phi_l \geq 0$ and $\beta_l \geq 0$ are parameters that have to be properly chosen in order to fit the actual value of (5.4). Note that the approximation with $L = 1$ makes this metric equivalent to the exponential ESM (EESM) [39], so we can think of this approximation as a generalized exponential ESM of degree L (L-GEESM). Therefore, the results for this approximation can be directly applied to the EESM metric just by setting $L = 1$.

Additionally, we propose to approximate the function Φ by a piecewise linear function (PLF) in the logarithmic domain

$$\Phi(\psi) = \begin{cases} 0 & \psi < \psi_0 \\ \frac{\log_{10}(\psi) - \log_{10}(\psi_0)}{\log_{10}(\psi_1) - \log_{10}(\psi_0)} & \psi_0 \leq \psi \leq \psi_1 \\ 1 & \psi > \psi_1 \end{cases} \tag{5.6}$$

where ψ_0 and ψ_1 have to be adjusted to approximate (5.4). Note that this PLF approximation makes the ESM ψ_e equal to the arithmetic mean of the SNR values ψ_i in the logarithmic domain, where extreme values for ψ_i are not taken into account, as values of $\psi_i > \psi_1$ and $\psi_i < \psi_0$ are *clipped* to ψ_1 and ψ_0 respectively.

The results of the fitting, performed with the MATLAB *Curve Fitting* toolbox, are shown in Figure 7.2. The approximation with the L-GEESM is only shown for values of $L = 1, 2, 3$, as the benefit of using higher order approximation is almost unnoticeable. In the following, we will try to maximize the MIESM by using these two approximations in (5.3).

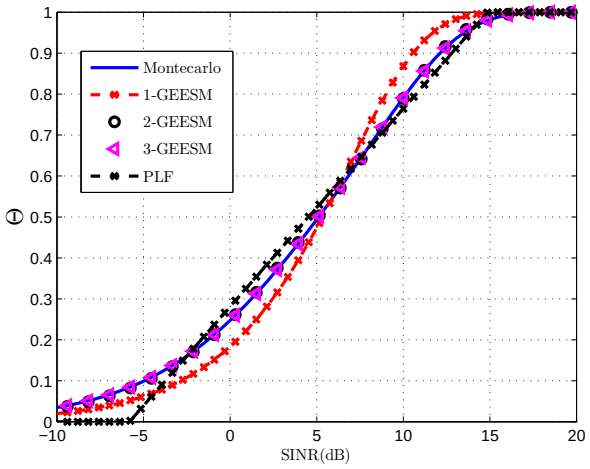


Figure 5.2: Approximations for the mutual information Φ for a 16-QAM constellation.

5.4.1 L-GEESM approximation

The maximization of the L-GEESM can be seen to be equivalent to maximizing $\Phi(\psi_e) = \frac{1}{N}\sum_{i=1}^{N}\Phi(\psi_i)$, as in this case Φ is given by (5.5), which is a monotonic increasing function. Therefore, by removing constant terms in the objective function and adding a power constraint on $\boldsymbol{\gamma}$, we arrive to

$$\begin{aligned} \text{minimize} \quad & f_0(\boldsymbol{\gamma}) \triangleq \frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{L}\phi_l e^{-\beta_l\frac{(\alpha_i\gamma_i+\eta_i)^2}{\sigma^2}} \\ \text{subject to} \quad & f_1(\boldsymbol{\gamma}) \triangleq \frac{1}{N}\sum_{i=1}^{N}\gamma_i^2 - P \leq 0 \\ & \mathbf{f}_2(\boldsymbol{\gamma}) \triangleq -\frac{1}{N}\boldsymbol{\gamma} \preceq 0 \end{aligned} \tag{5.7}$$

where the factor $1/N$ has been introduced in the last constraint in order to simplify the upcoming expressions, and $\mathbf{f}_2(\boldsymbol{\gamma}) = [f_{2,1}(\boldsymbol{\gamma}), ..., f_{2,N}(\boldsymbol{\gamma})]^T = -\frac{1}{N}[\gamma_1, ..., \gamma_N]^T$.

The Karush-Kuhn-Tucker (KKT) conditions for the problem (5.7) are

$$-\sum_{l=1}^{L}\phi_l\frac{2\beta_l}{\sigma^2}\alpha_i\left(\alpha_i\gamma_i+\eta_i\right)e^{-\frac{\beta_l\left(\alpha_i\gamma_i+\eta_i\right)^2}{\sigma^2}}+2\lambda_1\gamma_i-\lambda_{2,i} = 0 \tag{5.8}$$

$$\lambda_1 f_1\left(\gamma\right) = 0 \tag{5.9}$$

$$\lambda_1 \geq 0 \tag{5.10}$$

$$\boldsymbol{\lambda}_2^T\mathbf{f}_2\left(\gamma\right) = 0 \tag{5.11}$$

$$\boldsymbol{\lambda}_2 \succeq 0 \tag{5.12}$$

where $\boldsymbol{\gamma}$ is a feasible point of (5.7). We will study the conditions by making assumptions about the different constraints being active or not.

If f_1 is not active ($f_1\left(\boldsymbol{\gamma}\right)<0$), then we have from (5.9) that $\lambda_1=0$, so (5.8) reads as

$$-\sum_{l=1}^{L}\phi_l\frac{2\beta_l}{\sigma^2}\alpha_i\left(\alpha_i\gamma_i+\eta_i\right)e^{-\frac{\beta_l\left(\alpha_i\gamma_i+\eta_i\right)^2}{\sigma^2}}-\lambda_{2,i}=0\,\forall\, i=1\ldots,\,N. \tag{5.13}$$

As the values α_i,γ_i, η_i, β_l, σ^2 and $\lambda_{2,i}$ are non-negative, the N equalities in (5.13) will never be met, except in some degenerate cases, such as $\alpha_i=0\ \forall i$, which are not of interest. This means that if the point $\boldsymbol{\gamma}$ is optimum, $f_1\left(\gamma\right)=0$. This fact can be easily seen in problem (5.7), where the terms in the sum of the objective function are decreasing functions of γ_i, so allocating the remaining power to any of the terms will make the objective function decrease and, therefore, a point with non-active f_1 cannot be optimum.

If $f_{2,i}$ and f_1 are active, i.e., $\gamma_i=0$, then (5.8) reads as $-\sum_{l=1}^{L}\phi_l\frac{2\beta_l}{\sigma^2}\alpha_i\eta_i e^{-\frac{\eta_i\beta_l}{\sigma^2}}=\lambda_{2,i}$, or, as the left part of the equation is non-positive, and condition (5.12) constraints $\lambda_{2,i}$ to be non-negative, $\sum_{l=1}^{L}\phi_l\frac{2\beta_l}{\sigma^2}\alpha_i\eta_i e^{-\frac{\eta_i\beta_l}{\sigma^2}}=0$, so the condition is only met if α_i or η_i are equal to zero. Note that in the case $\alpha_i=0$ it is clear that allocating power to the i-th carrier is not going to change the objective function value, so that power consumption is useless. Since these are again degenerate cases, we can state that for a non-degenerate problem (α_i and η_i being strictly positive), the power constraint f_1 is going to be active ($f_1\left(\boldsymbol{\gamma}\right)=0$), and the N constraints $\mathbf{f}_2$ are going to be inactive ($\gamma_i>0\ \forall i$).

Therefore, the condition (5.8) for a non-degenerate problem is

$$\sum_{l=1}^{L}\phi_l\frac{\beta_l}{\sigma^2}\alpha_i\left(\alpha_i+\frac{\eta_i}{\gamma_i}\right)e^{-\frac{\beta_l\left(\alpha_i\gamma_i+\eta_i\right)^2}{\sigma^2}}=\lambda_1, \tag{5.14}$$

with a value of λ_1 such that the power constraint is met with equality. In order to obtain a

solution, we define the following function

$$h_i(\gamma_i) = \sum_{l=1}^{L} \phi_l \frac{\beta_l}{\sigma^2} \alpha_i \left(\alpha_i + \frac{\eta_i}{\gamma_i} \right) e^{-\frac{\beta_l (\alpha_i \gamma_i + \eta_i)^2}{\sigma^2}} \tag{5.15}$$

which is the sum of products of two strictly decreasing functions of γ_i and, therefore, is a strictly decreasing function of γ_i. Taking this fact into account, we can state that the function h is injective, so the inverse h^{-1} is unique.

From all the above, the optimum ST power distribution based on the L-GEESM approximation is computed in two steps:

- Obtain λ_1 as the root for $\frac{1}{N} \sum_{i=1}^{N} h_i^{-2}(\lambda_1) = P$
- Obtain γ_i as $h_i^{-1}(\lambda_1)$.

The inversion of h is a computationally expensive operation, and although its values could be stored in a lookup table in order to speed up the optimization, it is convenient to have an alternative computationally efficient approximation. In the following section we derive the optimum power allocation for the PLF approximation, which is much simpler than the previous approach.

5.4.2 PLF approximation

In this approximation, we have the problem that the objective function is not differentiable. In a first approach, we will only take into account the logarithmic part of the piecewise function, and afterwards we will add the upper part $\psi > \psi_1$. The lower clipping $\psi < \psi_0$ will be omitted for convenience, as its effect in the final results was found to be negligible. For $\psi_0 < \psi < \psi_1$, the optimum value of $\boldsymbol{\gamma}$ is obtained by solving

$$\begin{aligned} \text{minimize} \quad & f_0(\boldsymbol{\gamma}) \triangleq -\frac{1}{N} \sum_{l=1}^{N} 2 \log(\eta_i + \alpha_i \gamma_i) \\ \text{subject to} \quad & f_1(\boldsymbol{\gamma}) \triangleq \frac{1}{N} \sum_{i=1}^{N} \gamma_i^2 - P \leq 0 \\ & \mathbf{f}_2(\boldsymbol{\gamma}) \triangleq -\frac{2}{N} \boldsymbol{\gamma} \preceq 0 \end{aligned} \tag{5.16}$$

which is a convex problem. The optimality conditions for this problem read as

$$-\frac{2}{\gamma_i + \eta_i/\alpha_i} + 2\lambda_1\gamma_i - 2\lambda_{2,i} = 0 \tag{5.17}$$

$$\lambda_1 f_1(\boldsymbol{\gamma}) = 0 \tag{5.18}$$

$$\lambda_1 \geq 0 \tag{5.19}$$

$$\boldsymbol{\lambda}_2^T \mathbf{f}_2(\gamma) = 0 \tag{5.20}$$

$$\boldsymbol{\lambda}_2 \succeq 0. \tag{5.21}$$

If the power constraint is not active ($f_1(\boldsymbol{\gamma}) < 0$), following (5.18) we have that $\lambda_1 = 0$. Thus, we arrive to condition $-\frac{1}{\gamma_i+\eta_i/\alpha_i} - \lambda_{2,i} = 0$, that will only be met in the case $\alpha_i = 0 \; \forall\, i$. Similarly, if $\gamma_i = 0$ for some values of i, then $-\frac{\alpha_i}{\eta_i} - \lambda_{2,i} = 0$, a condition that will only be met in the case $\alpha_i = 0$. Therefore, for the non-degenerate cases we have that the power constraint is met with equality, and the non-negativity constraint with inequality.

With the previous conditions, and for non-degenerate cases, we have that

$$\gamma_i = \frac{-\eta_i + \sqrt{4\alpha_i^2 + \eta_i^2\lambda_1}/\sqrt{\lambda_1}}{2\alpha_i} \tag{5.22}$$

with a value of λ_1 such that the power constraint is met with equality, i.e.,

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{-\eta_i + \sqrt{4\alpha_i^2 + \eta_i^2\lambda_1}/\sqrt{\lambda_1}}{2\alpha_i}\right)^2 = P. \tag{5.23}$$

If we add the upper *clipping* to the problem, it is clear that if $(\eta_i + \alpha_i\gamma_i)^2/\sigma^2 > \psi_1$ some power is being wasted on the i-th carrier, as a value of

$$\gamma_i = \frac{\sqrt{\psi_1}\sigma - \eta_i}{\alpha_i} \tag{5.24}$$

will lead to the same objective function value with less power consumption. Therefore, we propose to solve the optimization problem iteratively by clipping the values of γ_i with $\psi_i > \psi_1$ according to (5.24), removing those γ_i from the optimization, and running the algorithm once again. Algorithm 1 describes this iterative approximation.

Algorithm 1: Iterative approximation for the upper-clipped PLF optimization

$\mathcal{O} = \{i|\psi_i < \psi_1\}$;
$end \leftarrow$ false
while not end **do**
  Solve Problem (5.16) over $\gamma_i, i \in \mathcal{O}$
  $\mathcal{A} \leftarrow \{i|\psi_i > \psi_1\}$
  if $\mathcal{A} = \emptyset$ **then**
    $end \leftarrow$ true
  else
    $\gamma_i = \frac{\sqrt{\psi_1}\sigma - \eta_i}{\alpha_i} \forall\, i \in \mathcal{A}$
    $\mathcal{O} \leftarrow \mathcal{O} \setminus \mathcal{A}$
  end if
end while

5.4.3 Results

We analyzed the obtained MIESM values for the PLF and 2-GEESM approximations for a 16-QAM constellation. The noise variance was set to $\sigma^2 = 0.2$, and the number of carriers was $N = 128$. The optimization was run for different values of P. In Figure 5.3 we show the coefficients of the simulated channel (recall that α_i and η_i were assumed, without loss of generality, to be non-negative values) and the resulting $\boldsymbol{\gamma}$ for different values of P for the 2-GEESM approximation. It can be seen that the secondary power allocation concentrates on those carriers with a weaker primary channel.

In Figure 5.4 we show the fraction of released resources $1 - \rho$ as a function of the available secondary power P. It can be seen that the PLF approximation offers a performance that is comparable with the one offered by the GEESM approximation, and outperforms a uniform power allocation policy (i.e., $\gamma_i = \sqrt{P}\ \forall i$). Thus, this PLF approximation can be of special interest because of its reduced complexity. The fraction of released resources $1 - \rho$ is shown to increase with P, reaching near a 40% of released resources for values of P near 1. The thresholds t_i that conform the mapping from effective SNR to MCS rate and throughput were taken from the LTE performance study in [98]. In our simulations, the MCS evolves from 16-QAM 1/3 to 16-QAM 4/5.

5.5 Channel estimation with SNR feedback

In the previous sections we assumed perfect knowledge of the channel coefficients α_i and η_i. In a practical scenario this perfect knowledge is not possible, so CSI has to be obtained by

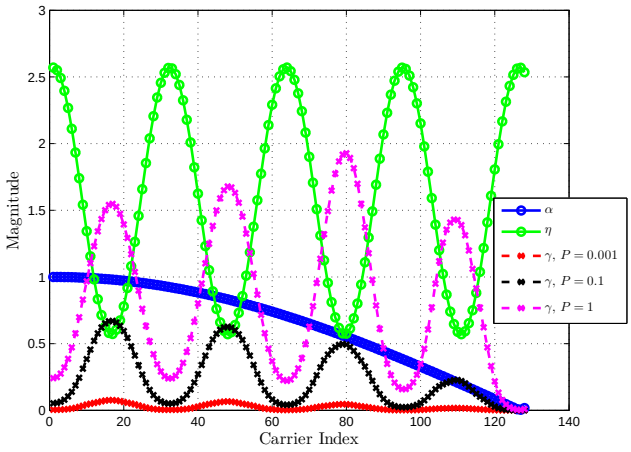


Figure 5.3: Channel under study, and optimum power allocation for different values of P. 2-GEESM approximation.

exploiting the information exchanged between the primary transmitter and receiver. We consider a simplified scenario where the channel coefficients α_i, η_i are time invariant, and the noise power σ^2 is known at both PR and ST. After a group of OFDM symbols, in the n-th feedback message, the PR conveys the square root of the SNR measured at each carrier, that we will model as

$$f_{i,n} = \frac{|\eta_i + \gamma_{i,n}\alpha_i + w_{i,n}|}{\sqrt{(2\sigma^2)}} \tag{5.25}$$

where the terms $w_{i,n}$ are independent and identically distributed random variables[2] $w_{i,n} \sim \mathcal{CN}\left(0, 2\sigma^2\right)$ that account for the SNR estimation error. This could be the case of an OFDM system that uses a pilot-based estimation scheme where the pilot symbols have unit power, and the PR just feeds back the modulus of the received pilot. η_i and α_i are modeled as deterministic but unknown parameters, so we will follow a maximum likelihood (ML) estimation approach.

As the SNR is fed back separately for each carrier, the estimation can be carried out independently for every carrier, so we will drop the carrier index i in the following expressions. For convenience, we will consider our observations x_n to be

$$x_n = f_n\sqrt{2\sigma^2} = |\eta + \gamma_n\alpha + w_n|\,. \tag{5.26}$$

Let us denote $\mathbf{x}_J \triangleq [x_1, ..., x_J]^T$ as the result of stacking J observations into a vector. It can be seen that the observations are independently Rician distributed, so the probability density

[2]In this case, the noise variance was set to $2\sigma^2$ to keep the notation consistent with the usual representation of a Rician random variable, where σ^2 denotes the noise variance per dimension.

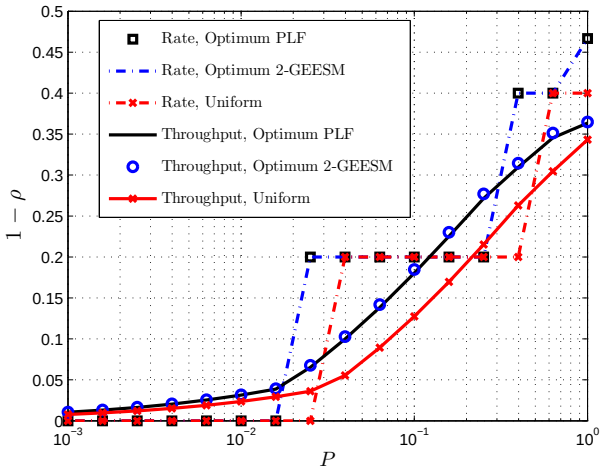


Figure 5.4: Fraction of released resources measured in terms of transmit rate and throughput.

function of $\mathbf{x}_J$, parametrized by the unknown parameters η and α, is given by

$$p\left(\mathbf{x}_J;\,\eta,\alpha\right)=\prod_{n=1}^{J}\frac{x_n}{\sigma^2}\exp\left(-\frac{x_n^2+\nu_n^2}{2\sigma^2}\right)I_0\left(\frac{x_n\nu_n}{\sigma^2}\right) \tag{5.27}$$

where $\nu_n=|\eta+\gamma_n\alpha|$. In this case, γ_n can be seen as a training sequence for the estimation procedure. With this, we can write the log-likelihood function of (η,α) as

$$\mathcal{L}\left(\eta,\alpha\right)=\sum_{n=1}^{J}\frac{-\nu_n^2}{2\sigma^2}+\log\left(I_0\left(\frac{x_n\nu_n}{\sigma^2}\right)\right) \tag{5.28}$$

where $I_0\left(\cdot\right)$ is the zeroth order modified Bessel function of the first kind, and a constant term that is independent of the parameters (η,α) has been omitted.

It is important to remark that the parameter ν_n is not constant with n, as γ_n can change with time. In fact, if we try to simplify the log-likelihood function by making γ_n constant with n an ambiguity will appear in the estimation procedure, as the likelihood function will have an infinite number of maxima. This fact is illustrated in Figure 5.5, where the existence of multiple (or even infinite) global and local maxima complicates the problem, even for a simple case with real parameters $(\eta=3,\alpha=1,\Phi=0)$. In fact, there exists an additional ambiguity that cannot be removed, as the points (η,α) and $\left(\eta e^{j\Phi_0},\alpha e^{j\Phi_0}\right)$ lead to the same likelihood value. However, this ambiguity does not affect our optimization procedure, as it only depends on the modulus of the channels ($|\eta_i|$ and $|\alpha_i|$) and the difference in its phase (Φ), as presented in Section 5.3.

With this, we can rewrite $\mathcal{L}$ as a function of the parameters of interest

$$\mathcal{L}\left(|\eta|,|\alpha|,\Phi\right)=\sum_{n=1}^{J}\frac{-\left||\eta|\,e^{j\Phi}+\gamma_n\,|\alpha|\right|^2}{2\sigma^2} \\ +\log\left(I_0\left(\frac{x_n\left||\eta|\,e^{j\Phi}+\gamma_n\,|\alpha|\right|}{\sigma^2}\right)\right). \tag{5.29}$$

5.5.1 Results

The log-likelihood function was maximized using a gradient based algorithm with two initial points $(\alpha,\eta,\Phi)=(1,1,\pi/2)$ and $(1,1,-\pi/2)$, selecting afterwards the one that led to a higher value of $\mathcal{L}$ in order to cope with the presence of local maximum. The training sequence γ was selected randomly following a complex Gaussian distribution, and normalized afterwards to meet the power constraint P. The studied channel is the one in Figure 5.3 with additional phase terms[3] multiplying the coefficients η_i with $e^{j\Phi_i}$, $\Phi_i=4\pi i/N$. The estimation procedure was run for different training sequence lengths J, and the PLF-based Algorithm 1 was run taking as input the estimated $(|\alpha_i|,|\eta_i|)$. The knowledge of Φ_i was used to make the primary and secondary contributions to be coherently added at the primary receiver, as explained in Section 5.3. The conditions of the simulation are the same as in Section 5.4, and the estimation noise was set to $w_k\sim\mathcal{CN}(0,0.2)$. The obtained results are shown in Figure 5.6, where it can be seen that even the scenario with a short training sequence ($J=5$) outperforms the uniform allocation[4], especially for large values of P. For smaller values of P the estimation error is much larger, so the fraction of released resources can be increased by the use of longer training sequences.

5.6 Other CSI acquisition schemes

5.6.1 Introduction

In the previous sections we showed how to exploit the primary feedback to obtain CSI. This CSI acquisition scheme, however, is difficult to analyze due to the lack of phase information on the channel. For example, it is difficult to obtain optimum training sequences γ_i, or a closed form expression for the estimation mean squared error. In this section, we follow a different approach, taken from [15]. We consider single carrier schemes, that can be extended to the OFDM case

[3] The obtained results were similar when a random phase component was applied to each carrier separately.

[4] In Section 5.4 the uniform power allocation assumed phase knowledge, so the primary and secondary contributions were coherently added. In this case, no phase knowledge is assumed, so some of the carriers can experience a lower SNR than that in absence of the ST, so the gain is much smaller.

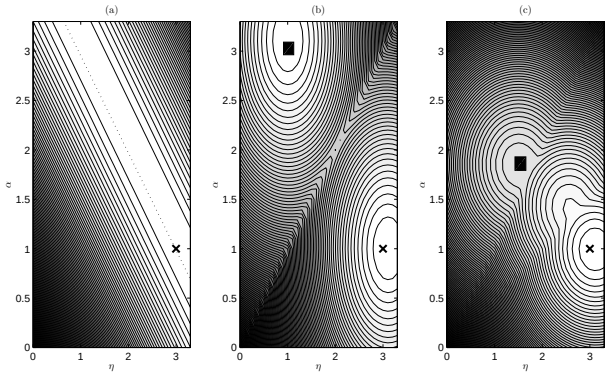


Figure 5.5: Level curves of the likelihood function for different training sequences. The point $\times = (3, 1)$ represents the true value of (η, α) In (a) $\boldsymbol{\gamma} = [1,\ 1,\ 1,\ 1]$, and the dashed line represents the points with the same value as $\times$; in (b) ($\boldsymbol{\gamma} = [1,\ -1,\ 1,\ -1]$) the point $\blacksquare = (1, 3)$ has the same value as $\times$; in (c) ($\boldsymbol{\gamma} = [1,\ -1,\ 2,\ -2]$) the only global maximum is $\times$, but a local maximum $\blacksquare$ appears.

by repeating the process for each carrier. If we assume that the PR estimates the *equivalent channel* (the channel resulting from the addition of the contributions coming from the ST and the PT) and broadcasts its value, the ST might gain access to CSI in the following way: in the first time slot, when the ST joins the network, it does not transmit the primary message, so the received waveform at the PR is

$$r_i = \eta x_i + z_i, \tag{5.30}$$

with x_i the primary codeword, and z_i a sample of Gaussian noise. At that time slot, the PR broadcasts the estimated channel $h_i \approx \eta$. In the second time slot, the ST allocates $|\gamma|^2$ units of power to the primary message, so the received waveform is

$$r_{i+1} = (\eta + \gamma\alpha)\, x_{i+1} + z_{i+1}. \tag{5.31}$$

At this time, the PR broadcasts the estimated channel $h_{i+1} \approx \eta + \gamma\alpha$, so the channel values α and η can be approximately obtained as $\alpha \approx \frac{h_{i+1} - h_i}{\gamma}$, $\eta \approx h_i$. Note that the one-tap precoding sequence $\boldsymbol{\gamma} = [0, \gamma]$ can be considered as a **training sequence** for the channel estimation problem.

Although this simple interaction framework has been cited in other works like [99], and even for a multiple antenna channel in [100], to the best of our knowledge this CSI acquisition technique has not been sufficiently studied in the literature. In this section, we derive closed-form expressions for the mean squared error (MSE) of the channel estimates in a single input single output (SISO) channel, in a multiple input multiple output (MIMO) channel with transmit

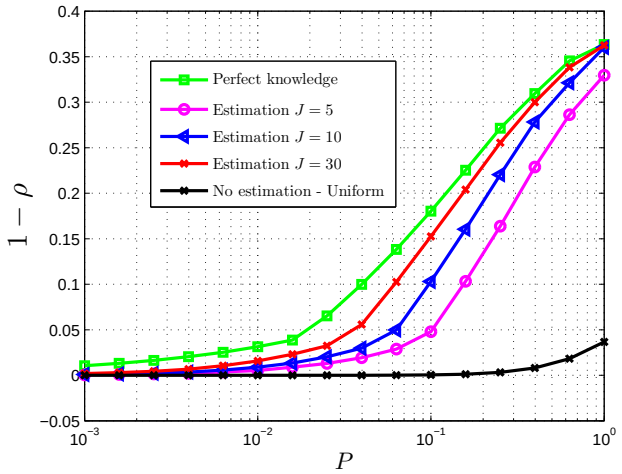


Figure 5.6: Fraction of released resources, measured in terms of throughput, for different values of training sequence length. The optimization was run with the PLF approximation.

beamforming and in a SISO time varying channel.

5.6.2 SISO Channel

We assume α, $\eta \in \mathbb{C}$ are the time-invariant ST to PR and PT to PR channels, respectively, and model the i-th feedback message as

$$y_i = (\alpha\gamma_i + \eta) + n_i,\ n_i \sim \mathcal{CN}\left(0, \sigma^2\right) \tag{5.32}$$

where n_i is a zero-mean Gaussian random variable that accounts for the estimation error and γ_i is the one-tap pre-equalizer previously introduced. We assume that feedback is error-free.

Note that if we remove the term γ_i, we have that the parameters α and η are not identifiable. We define the vector observation $\mathbf{y} = [y_1, ..., y_M]^T$ resulting from stacking M scalar observations as

$$\mathbf{y} = \boldsymbol{\gamma}\alpha + \mathbf{1}_M\eta + \mathbf{n} \tag{5.33}$$

with $\boldsymbol{\gamma} = [\gamma_1, ... \gamma_M]^T$, $\mathbf{1}_M$ a column vector with its M entries equal to one and $\mathbf{n} \sim \mathcal{N}\left(0, \sigma^2\mathbf{I}_M\right)$. The probability density function of the observations $\mathbf{y}$, conditioned on the parameters to estimate (α, η), is

$$p\left(\mathbf{y}; \alpha, \eta\right) = \left(\pi\sigma^2\right)^{-M} \exp\left(-\frac{\|\mathbf{y} - \alpha\boldsymbol{\gamma} - \eta\mathbf{1}_M\|^2}{\sigma^2}\right). \tag{5.34}$$

In this case, the Fisher Information Matrix (FIM) $\boldsymbol{\mathcal{I}}$ reads as

$$\boldsymbol{\mathcal{I}} = \frac{1}{\sigma^2} \begin{bmatrix} \|\boldsymbol{\gamma}\|^2 & \boldsymbol{\gamma}^* \mathbf{1}_M^T \\ \mathbf{1}_M^T \boldsymbol{\gamma} & M \end{bmatrix}. \tag{5.35}$$

Note that if γ_i is constant $\forall\, i$ then the FIM is not invertible, thus leading to unidentifiable parameters [101].

The Cramér-Rao bound (CRB) for the variance of any unbiased estimator $\hat{\alpha}$ and $\hat{\eta}$ can be obtained from (5.35) as

$$\mathbb{V}\mathrm{ar}\left[\hat{\alpha}\right] \geq \frac{M}{\sigma^2 \det \boldsymbol{\mathcal{I}}} \tag{5.36}$$

$$\mathbb{V}\mathrm{ar}\left[\hat{\eta}\right] \geq \frac{\|\boldsymbol{\gamma}\|^2}{\sigma^2 \det \boldsymbol{\mathcal{I}}}. \tag{5.37}$$

The determinant of $\boldsymbol{\mathcal{I}}$ can be written as

$$\det \boldsymbol{\mathcal{I}} = \sigma^{-4} \left(M \|\boldsymbol{\gamma}\|^2 - \|\boldsymbol{\gamma}^* \mathbf{1}_M\|^2 \right). \tag{5.38}$$

For a given total power $\|\boldsymbol{\gamma}\|^2 \leq P$, the values of $\boldsymbol{\gamma}$ that maximize the determinant of $\boldsymbol{\mathcal{I}}$, and, therefore, minimize the CRB, are those with $\boldsymbol{\gamma}^* \mathbf{1}_M = 0$ and $\|\boldsymbol{\gamma}\|^2 = P$. Just by taking any vector of this family, we arrive to

$$\mathbb{V}\mathrm{ar}\left[\hat{\alpha}\right] \geq \frac{\sigma^2}{P} \tag{5.39}$$

$$\mathbb{V}\mathrm{ar}\left[\hat{\eta}\right] \geq \frac{\sigma^2}{M}. \tag{5.40}$$

Moreover, it can be easily seen that, in this case, an efficient estimator exists [102], and is given by

$$[\hat{\alpha}, \hat{\eta}]^T = \begin{bmatrix} \frac{M\boldsymbol{\gamma}^* \mathbf{y} - \boldsymbol{\gamma}^* \mathbf{1}_M \mathbf{1}_M^T \mathbf{y}}{M\|\boldsymbol{\gamma}\|^2 - \|\mathbf{1}_M \boldsymbol{\gamma}\|^2} \\ \frac{\|\boldsymbol{\gamma}\|^2 \mathbf{1}_M^T \mathbf{y} - \mathbf{1}_M^T \boldsymbol{\gamma}\boldsymbol{\gamma}^* \mathbf{y}}{M\|\boldsymbol{\gamma}\|^2 - \|\mathbf{1}_M \boldsymbol{\gamma}\|^2} \end{bmatrix} \tag{5.41}$$

that for any optimum sequence $\boldsymbol{\gamma}$ reads as

$$[\hat{\alpha}, \hat{\eta}]^T = \begin{bmatrix} \frac{\boldsymbol{\gamma}^* \mathbf{y}}{P} \\ \frac{\mathbf{1}_M^T \mathbf{y}}{M} \end{bmatrix}. \tag{5.42}$$

Note that this estimator, with the optimum training sequence, performs the following operations:

- For the estimation of α, it correlates the received signal with $\boldsymbol{\gamma}$. As $\boldsymbol{\gamma}$ is zero mean, it completely removes the parameter η. A power normalization is performed afterwards.

- For the estimation of η, it averages the received signal, thus removing the parameter α, as it is multiplied by a zero-mean sequence. A normalization is performed afterwards.

Thus, the mean and the correlation are sufficient statistics for the estimation.

At the view of these results, we conclude that the *training sequence* $\boldsymbol{\gamma} = [0, \gamma]$ introduced in [15] is not optimum in the sense of minimum estimation variance for a given total power, as $\boldsymbol{\gamma}^* \mathbf{1}_2 = \gamma \neq 0$.

5.6.3 MIMO channel

In this section we extend the results to the case of transmit MIMO channels with transmit beamforming.

Let us denote by $\mathbf{H}_s \in \mathbb{C}_{N_r \times N_s}$ the (assumed to be flat fading) MIMO channel from the ST to the PR, and as $\mathbf{H}_p \in \mathbb{C}_{N_r \times N_p}$ the MIMO channel from the PT to the PR. We will assume that the PT is transmitting only one data layer at a time by applying a fixed beamforming vector $\mathbf{w}_p \in \mathbb{C}_p^N$, and the ST is also conveying the primary information by using a beamforming vector $\mathbf{w}_{s,i} \in \mathbb{C}^{N_s}$, which can be time-varying. We will also assume that the PR is estimating the SIMO channel (combination of the beamforming + MIMO channel) such that the SIMO channel estimate $\mathbf{y}_i \in \mathbb{C}_{N_r}$ can be written as

$$\mathbf{y}_i = \mathbf{H}_p \mathbf{w}_p + \mathbf{H}_s \mathbf{w}_{s,i} + \mathbf{n}_i,\ \mathbf{n}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_n) \tag{5.43}$$

where $\mathbf{n}_i$ accounts for the estimation error.

If the PR broadcast these channel estimates, then the PT would be able to acquire some CSI. Note that the assumption of fixed beamforming $\mathbf{w}_p$ makes the separate estimation of $\mathbf{H}_p$ and $\mathbf{w}_p$ impossible, so we will denote by $\mathbf{g} \triangleq \mathbf{H}_p \mathbf{w}_p$ the SIMO channel comprising the combination of beamforming and MIMO channel from the PT to the PR. The objective of the ST is to estimate both $\mathbf{g}$ and $\mathbf{H}_s$ from the observations $\mathbf{y}_i$, by treating the sequence of beamforming vectors $\mathbf{w}_{s.i}$ as a training sequence.

Estimation problem

For notational simplicity we will denote $\mathbf{H} \triangleq \mathbf{H}_s$ and $\mathbf{w}_i \triangleq \mathbf{w}_{s,i}$, as the primary MIMO channel and beamforming vectors are included in the SIMO channel $\mathbf{g}$. At a given time instant i, our observation will be

$$\mathbf{y}_i = \mathbf{g} + \mathbf{H}\mathbf{w}_i + \mathbf{n}_i. \tag{5.44}$$

If we stack M observations into a column vector we obtain the $MN_r \times 1$ vector $\mathbf{y}$ as follows:

$$\mathbf{y} = \begin{bmatrix} \mathbf{g} \\ ... \\ \mathbf{g} \end{bmatrix} + \begin{bmatrix} \mathbf{H}\mathbf{w}_1 \\ ... \\ \mathbf{H}\mathbf{w}_M \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ ... \\ \mathbf{n}_M \end{bmatrix}. \tag{5.45}$$

As we are interested in estimating both the vector $\mathbf{g}$ and the matrix $\mathbf{H}$, we will rewrite (5.45) as

$$\mathbf{y} = (\mathbf{1}_M \otimes \mathbf{I}_{N_r})\,\mathbf{g} + (\mathbf{W} \otimes \mathbf{I}_{N_r})\,\mathbf{h} + \mathbf{n} \tag{5.46}$$

with $\mathbf{I}_M$ the $M \times M$ identity matrix, $\otimes$ the Kronecker product operator, and $\mathbf{h} = \operatorname{vec}\mathbf{H}$ is the result of stacking the columns of $\mathbf{H}$ into a vector, so $\mathbf{h} \in \mathbb{C}^{N_s N_r}$. The training sequence matrix is the result of stacking into a matrix the training sequence $\mathbf{w}_i^T$: $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_M]^T \in \mathbb{C}_{M \times N_s}$. The vector $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{C})$, with $\mathbf{C} = \mathbf{C}_n \otimes \mathbf{I}_M$, is the result of stacking the M noise vectors $\mathbf{n}_i$. With this, we can rewrite (5.46) as

$$\mathbf{y} = \mathbf{K}\mathbf{g} + \mathbf{R}\mathbf{h} + \mathbf{n} \tag{5.47}$$

with $\mathbf{K} \triangleq (\mathbf{1}_M \otimes \mathbf{I}_{N_r}) \in \mathbb{C}_{N_r M \times N_r}$ and $\mathbf{R} \triangleq (\mathbf{W} \otimes \mathbf{I}_{N_r}) \in \mathbb{C}_{N_r M \times N_r N_s}$. It can be easily seen that (5.47) is a Gaussian linear model [102], so if we define $\mathbf{A} = [\mathbf{K}\,\mathbf{R}]$ and $\mathbf{b} = \left[\mathbf{g}^T\,\mathbf{h}^T\right]^T$ the minimum variance unbiased (MVU) estimator (which is efficient) is given by

$$\hat{\mathbf{b}} = \left(\mathbf{A}^*\mathbf{C}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^*\mathbf{C}^{-1}\mathbf{y} \tag{5.48}$$

which is distributed according to

$$\hat{\mathbf{b}} \sim \mathcal{CN}\left(\mathbf{b}, \left(\mathbf{A}^*\mathbf{C}^{-1}\mathbf{A}\right)^{-1}\right). \tag{5.49}$$

Note that in the previous equations we have assumed $M \geq N_s + 1$ so the provided inverse matrices exist.

Training sequence design

We will design our training sequence $\mathbf{W}$ in order to minimize the total estimation variance, subject to a total power constraint P:

$$\begin{aligned} &\text{minimize} && \operatorname{tr}\left(\mathbf{A}^*\mathbf{C}^{-1}\mathbf{A}\right)^{-1} \\ &\text{subject to} && \operatorname{tr}\left(\mathbf{W}^*\mathbf{W}\right) \leq P. \end{aligned} \tag{5.50}$$

In the following, we will assume that $\mathbf{C} = \sigma^2 \mathbf{I}_{MN_r}$, so the matrix in the objective function reads as

$$\mathbf{A}^* \mathbf{C}^{-1} \mathbf{A} \;=\; \frac{1}{\sigma^2} \mathbf{M} \otimes \mathbf{I}_{N_r} \tag{5.51}$$

with

$$\mathbf{M} \triangleq \begin{bmatrix} \mathbf{1}_M^T \\ \mathbf{W}^* \end{bmatrix} [\mathbf{1}_M,\, \mathbf{W}]\,. \tag{5.52}$$

Therefore, the objective function in (5.50) reads as

$$\operatorname{tr}\left(\mathbf{A}^* \mathbf{C}^{-1} \mathbf{A}\right)^{-1} = \sigma^2 N_r \operatorname{tr} \mathbf{M}^{-1}. \tag{5.53}$$

As N_r and σ^2 do not depend on $\mathbf{W}$, we can rewrite (5.50) as

$$\begin{array}{ll} \text{minimize} & \operatorname{tr} \mathbf{M}^{-1} \\ \text{subject to} & \operatorname{tr}(\mathbf{W}^* \mathbf{W}) \leq P. \end{array} \tag{5.54}$$

Note that

$$\mathbf{M} = \begin{bmatrix} M & \mathbf{1}_M^T \mathbf{W} \\ \mathbf{W}^* \mathbf{1}_M & \mathbf{W}^* \mathbf{W} \end{bmatrix}, \tag{5.55}$$

so the trace of the inverse matrix of $\mathbf{M}$ can be written as a function of the Schur complements of the submatrices in $\mathbf{M}$ as $\operatorname{tr} \mathbf{M}^{-1} = \operatorname{tr} \mathbf{S}_M^{-1} + \operatorname{tr} S^{-1}_{\mathbf{W}^M \mathbf{W}}$, with $\mathbf{S}_M$ and $S_{\mathbf{W}^* \mathbf{W}}$ the Schur complements of M and $\mathbf{W}^* \mathbf{W}$ in $\mathbf{M}$. The inverse of the latter can be expanded by using the Sherman-Morrison formula:

$$\mathbf{S}^{-1}_{\mathbf{W}^* \mathbf{W}} = (\mathbf{W}^* \mathbf{W})^{-1} + \frac{1}{M} \frac{(\mathbf{W}^* \mathbf{W})^{-1} \mathbf{W}^* \mathbf{1}_M \mathbf{1}_M^T \mathbf{W} (\mathbf{W}^* \mathbf{W})^{-1}}{M - \mathbf{1}_M^T \mathbf{W} (\mathbf{W}^* \mathbf{W})^{-1} \mathbf{W}^* \mathbf{1}_M} \tag{5.56}$$

so the objective function is

$$\operatorname{tr} \mathbf{M}^{-1} = c \left(1 + \frac{\operatorname{tr} \mathbf{Q}}{M}\right) + \operatorname{tr} (\mathbf{W}^* \mathbf{W})^{-1} \tag{5.57}$$

where

$$\mathbf{Q} \triangleq (\mathbf{W}^* \mathbf{W})^{-1} \mathbf{W}^* \mathbf{1}_M \mathbf{1}_M^T \mathbf{W} (\mathbf{W}^* \mathbf{W})^{-1} \tag{5.58}$$

and

$$c \triangleq S^{-1}_{\mathbf{W}^* \mathbf{W}} = \left(M - \mathbf{1}_M^T \mathbf{W} (\mathbf{W}^* \mathbf{W})^{-1} \mathbf{W}^* \mathbf{1}_M\right)^{-1} \tag{5.59}$$

In the following, we will minimize separately the two terms in the sum (5.57).

Minimization of $c\left(1+\frac{\operatorname{tr}\mathbf{Q}}{M}\right)$:

If we define $\mathbf{j}=\left(\mathbf{W}^{*}\mathbf{W}\right)^{-1}\mathbf{W}^{*}\mathbf{1}_{M}$ then we have that $\operatorname{tr}\mathbf{Q}=\operatorname{tr}\mathbf{j}\mathbf{j}^{*}=\|\mathbf{j}\|^{2}$ so we can write

$$c\left(1+\frac{\operatorname{tr}\mathbf{Q}}{M}\right)=\frac{1+\frac{1}{M}\|\mathbf{j}\|^{2}}{M-\mathbf{1}_{M}^{T}\mathbf{W}\mathbf{j}}. \tag{5.60}$$

The denominator is always positive since

$$\mathbf{1}_{M}^{T}\mathbf{W}\mathbf{j}=\mathbf{1}_{M}^{T}\mathbf{P}_{\mathbf{W}}\mathbf{1}_{M}\leq M \tag{5.61}$$

with $\mathbf{P}_{\mathbf{W}}=\mathbf{W}\left(\mathbf{W}^{*}\mathbf{W}\right)^{-1}\mathbf{W}^{*}$ the projection matrix into the subspace spanned by the columns of $\mathbf{W}$, so (5.60) is minimized when $\mathbf{j}=\mathbf{0}$ or, equivalently $\mathbf{W}^{*}\mathbf{1}_{M}=0$. Note that this minimization is not affected by the power constraint.

Minimization of $\operatorname{tr}\left(\mathbf{W}^{*}\mathbf{W}\right)^{-1}$:

This minimization is affected by the power constraint $\operatorname{tr}\mathbf{W}^{*}\mathbf{W}\leq P$. As $\operatorname{tr}\mathbf{A}=\sum_{i=1}^{N}\lambda_{i}\left(\mathbf{A}\right)$ and $\operatorname{tr}\mathbf{A}^{-1}=\sum_{i=1}^{N}\lambda_{i}\left(\mathbf{A}^{-1}\right)=\sum_{i=1}^{N}\lambda_{i}^{-1}\left(\mathbf{A}\right)$ for $\mathbf{A}\in\mathbb{C}_{N\times N}$ we can state our optimization problem as

$$\begin{aligned}&\text{minimize} && \sum_{i=1}^{N_s}\frac{1}{\lambda_i}\\ &\text{subject to} && \sum_{i=1}^{N_s}\lambda_i\leq P,\quad -\lambda_i\leq 0\end{aligned} \tag{5.62}$$

which is convex. If we define $\boldsymbol{\lambda}=[\lambda_1,\ ...,\ \lambda_{N_s}]$, it can be proved by using the KKT conditions that the optimum value is given by $\boldsymbol{\lambda}=\frac{P}{N_s}\mathbf{1}_{N_s}$, leading to an objective function value of $\operatorname{tr}\left(\mathbf{W}^{*}\mathbf{W}\right)^{-1}=\frac{N_s^2}{P}$.

Putting all together:

From the previous results, if we can find a matrix that meets the following properties

1. The N_s eigenvalues of $\mathbf{W}^{*}\mathbf{W}$ are all equal to $\frac{P}{N_s}$, and
2. $\mathbf{W}^{*}\mathbf{1}_{M}=\mathbf{0}$,

then the optimum training sequence will be given by $\mathbf{W}$. We can write the Singular Value Decomposition (SVD) of the matrix $\mathbf{W}$ as

$$\mathbf{W}=\mathbf{U}\begin{bmatrix}\boldsymbol{\Sigma}\\ \mathbf{0}_{M-N_s\times N_s}\end{bmatrix}\mathbf{V}^{*} \tag{5.63}$$

with $\mathbf{U} \in \mathbb{C}_{M\times M}$ and $\mathbf{V} \in \mathbb{C}_{N_s\times N_s}$ unitary matrices, and $\mathbf{\Sigma} = \text{diag}\left(\sigma_1 ... \sigma_{N_s}\right)$ is a diagonal matrix containing the nonzero singular values of $\mathbf{W}$.

Now, we have that $\lambda_i\left(\mathbf{W}^*\mathbf{W}\right) = \sigma_i^2$, so property 1 does only depend on the values of the matrix $\mathbf{\Sigma}$.

In order to characterize the second property, we can rewrite (5.63) as the *thin SVD*

$$\mathbf{W} = [\mathbf{U}_1\, \mathbf{U}_2] \begin{bmatrix} \mathbf{\Sigma} \\ \mathbf{0}_{M-N_s\times N_s} \end{bmatrix} \mathbf{V}^* = \mathbf{U}_1\mathbf{\Sigma}\mathbf{V}^* \tag{5.64}$$

with $\mathbf{U}_1 \in \mathbb{C}_{M\times N_s}$ the matrix containing the first N_s columns of $\mathbf{U}$. Condition 2 can be rewritten as $\mathbf{V}^*\mathbf{\Sigma}\mathbf{U}_1^*\mathbf{1}_M = \mathbf{0}$. Note that $\mathbf{V}$ and $\mathbf{\Sigma}$ are invertible, so the previous condition is equivalent to $\mathbf{U}_1^*\mathbf{1}_M = \mathbf{0}$, that only depends on the submatrix $\mathbf{U}_1$. Therefore, it is possible to find a matrix $\mathbf{W}$ that meets the two conditions at the same time by means of the following procedure:

1. Let $\mathbf{U}$ be an orthonormal base of $\mathbb{C}^M$ with $\frac{1}{\sqrt{M}}\mathbf{1}_M$ as a vector.
2. Choose N_s of the vectors in $\mathbf{U}$ except $\frac{1}{\sqrt{M}}\mathbf{1}_M$. Put them into the matrix $\mathbf{U}_1$.
3. Set $\mathbf{\Sigma} = \sqrt{P/N_s}\mathbf{I}_{N_s}$.
4. Let $\mathbf{V}$ be an orthonormal base of $\mathbb{C}^{N_s}$.
5. Obtain the matrix training sequence as $\mathbf{W} = \mathbf{U}_1\mathbf{\Sigma}\mathbf{V}^*$.

With this family of training sequences, the matrix $\mathbf{M}$ in (5.55) is block-diagonal, and the optimization problem in (5.50) is solved with a value of

$$\text{tr}\left(\mathbf{A}^*\mathbf{C}^{-1}\mathbf{A}\right)^{-1} = \sigma^2 N_r \left(\frac{1}{M} + \frac{N_s^2}{P}\right). \tag{5.65}$$

5.6.4 Time varying channel

In this section we will study a scenario with time varying (TV) channels. We assume that for a given observation period of M samples, the TV channels $\boldsymbol{\alpha} = [\alpha_1 ..., \alpha_M]^T$, $\boldsymbol{\eta} = [\eta_1, ..., \eta_M]^T$ can be written following a basis expansion model (BEM) as

$$\boldsymbol{\alpha} = \mathbf{F}_\alpha \mathbf{b}_\alpha,\ \boldsymbol{\eta} = \mathbf{F}_\eta \mathbf{b}_\eta, \tag{5.66}$$

with $\mathbf{b}_\alpha \in \mathbb{C}^{K_\alpha}$ and $\mathbf{b}_\eta \in \mathbb{C}^{K_\eta}$ the BEM coefficients, and $\mathbf{F}_\alpha$ and $\mathbf{F}_\eta$ the matrices containing the K_α and K_η first columns of the $M \times M$ Discrete Fourier Transform (DFT) matrix $\mathbf{F}$:

$$\mathbf{F} = \frac{1}{\sqrt{M}} \left[\begin{array}{cccc} \mathbf{f}_0 & \mathbf{f}_1 & ... & \mathbf{f}_{M-1} \end{array} \right] \tag{5.67}$$

with $\mathbf{f}_i = \left[1,\, e^{-j\frac{2\pi i}{M}},\, e^{-j2\frac{2\pi i}{M}},\, ... e^{-j(M-1)\frac{2\pi i}{M}}\right]^T$.

Note that $\mathbf{F}_x^* \mathbf{F}_x = \mathbf{I}_{K_x}$ but $\mathbf{F}_x \mathbf{F}_x^* \neq \mathbf{I}_M$ in general, for $x \in \{\eta, \alpha\}$.

Let $\boldsymbol{\gamma} = [\gamma_1,\, ...,\, \gamma_M]^T$ be the sequence of pre-equalizers applied at the ST, and $\boldsymbol{\Gamma} = \mathrm{diag}\,(\boldsymbol{\gamma})$. We can write our observation vector $\mathbf{y} = [y_1,\, ...,\, y_M]^T$ as

$$\mathbf{y} = \boldsymbol{\Gamma}\mathbf{F}_\alpha \mathbf{b}_\alpha + \mathbf{F}_\eta \mathbf{b}_\eta + \mathbf{n} \tag{5.68}$$

with $\mathbf{n} \sim \mathcal{CN}\left(0, \frac{\sigma^2}{M}\mathbf{I}_M\right)$. The factor $1/M$ is introduced in the noise variance to have a constant *signal to noise ratio per observation*, independently from the value of M.

The objective of the ST is to estimate the parameters $\mathbf{b}_\alpha$ and $\mathbf{b}_\eta$ in order to predict the channel. Since (5.68) is a linear model, an efficient estimator $\hat{\mathbf{b}}$ of $\mathbf{b} \triangleq \left[\mathbf{b}_\alpha^T, \mathbf{b}_\eta^T\right]^T$ can be found with a distribution $\hat{\mathbf{b}} \sim \mathcal{CN}\,(\mathbf{b}, \mathbf{C}_b)$ where

$$\mathbf{C}_b = \frac{\sigma^2}{M}\,(\mathbf{H}^*\mathbf{H})^{-1} \tag{5.69}$$

and $\mathbf{H} = \left[\begin{array}{cc} \boldsymbol{\Gamma}\mathbf{F}_\alpha & \mathbf{F}_\eta \end{array} \right]$.

In the following we will assume that $M \geq K_\alpha + K_\eta$ so the inverse in (5.69) exists. The total estimation variance can be written as

$$\frac{\sigma^2}{M}\,\mathrm{tr}\,(\mathbf{H}^*\mathbf{H})^{-1} \;=\; \frac{\sigma^2}{M}\,\mathrm{tr}\left[\begin{array}{cc} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{I}_{K_\alpha} \end{array} \right]^{-1} \tag{5.70}$$

with $\mathbf{A} = \mathbf{F}_\alpha^* \boldsymbol{\Gamma}^* \boldsymbol{\Gamma} \mathbf{F}_\alpha$ and $\mathbf{B} = \mathbf{F}_\alpha^* \boldsymbol{\Gamma}^* \mathbf{F}_\eta$. Note that the trace can be written as a function of the Schur complements of the block matrices as

$$\mathrm{tr}\left[\begin{array}{cc} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{I}_{K_\alpha} \end{array} \right]^{-1} = \mathrm{tr}\,\mathbf{S}_{\mathbf{I}_{K_\alpha}}^{-1} + \mathrm{tr}\,\mathbf{S}_{\mathbf{A}}^{-1} \tag{5.71}$$

that can be written, following the Woodbury matrix identity, as

$$\mathbf{S}_{\mathbf{I}_{K_\alpha}}^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}^*\left(\mathbf{I}_{K_\alpha} + \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^*\right)^{-1}\mathbf{B}\mathbf{A}^{-1} \tag{5.72}$$

and

$$\mathbf{S}_{\mathbf{A}}^{-1} = \mathbf{I}_{K_\alpha} + \mathbf{B}\left(\mathbf{A}^{-1} + \mathbf{B}^*\mathbf{B}\right)^{-1}\mathbf{B}^*. \tag{5.73}$$

Now, we proceed to separately minimize the traces of the inverses of the Schur complements.

Minimization of $\operatorname{tr}\mathbf{S}_{\mathbf{A}}^{-1}$

Note that

$$\operatorname{tr}\mathbf{S}_{\mathbf{A}}^{-1} = K_\alpha + \operatorname{tr}\mathbf{B}\left(\mathbf{A}^{-1} + \mathbf{B}^*\mathbf{B}\right)^{-1}\mathbf{B}^* \geq K_\alpha \tag{5.74}$$

since the second term in the sum is a positive semidefinite matrix. Recall that $\mathbf{B} = \mathbf{F}_\eta^*\boldsymbol{\Gamma}\mathbf{F}_\alpha$, so (5.74) will be minimized when $\mathbf{B} = \mathbf{0}$. If we set

$$\gamma_i = \sqrt{r}e^{-j\frac{2\pi i K_\eta}{M}},\ i = 1,\ ...,\ M \tag{5.75}$$

with r a positive number, we have that

$$\boldsymbol{\Gamma}\mathbf{F}_\alpha = \sqrt{\frac{r}{M}}\left[\begin{array}{cccc}\mathbf{f}_{K_\eta} & \mathbf{f}_{K_\eta+1} & ... & \mathbf{f}_{K_\eta+K_\alpha-1}\end{array}\right] \tag{5.76}$$

so $\mathbf{B} = \mathbf{F}_\eta^*\boldsymbol{\Gamma}\mathbf{F}_\alpha = \mathbf{0}$. In this case, the training sequence $\boldsymbol{\gamma}$ is *modulating* the channel, thus artificially introducing a faster time variation in the channel from the ST to the PR. In this case, the estimator is able to distinguish the lowpass variations of the channel, which correspond to $\boldsymbol{\eta}$, from those variations at higher frequencies, which correspond to $\boldsymbol{\alpha}$. With this training sequence we have that $\operatorname{tr}\mathbf{S}_{\mathbf{A}}^{-1} = K_\alpha$.

Minimization of $\operatorname{tr}\mathbf{S}_{\mathbf{I}_{K_\alpha}}^{-1}$

Since

$$\operatorname{tr}\mathbf{S}_{\mathbf{I}_{K_\alpha}}^{-1} = \operatorname{tr}\mathbf{A}^{-1} + \operatorname{tr}\mathbf{A}^{-1}\mathbf{B}^*\left(\mathbf{I}_{K_\alpha} + \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^*\right)^{-1}\mathbf{B}\mathbf{A}^{-1} \geq \operatorname{tr}\mathbf{A}^{-1}, \tag{5.77}$$

the second term of the sum is also minimized when $\mathbf{B} = \mathbf{0}$.

We can find an optimum training sequence for the first term $\operatorname{tr}\mathbf{A}^{-1} = \operatorname{tr}\left(\mathbf{F}^*\boldsymbol{\Gamma}^*\boldsymbol{\Gamma}\mathbf{F}\right)^{-1}$ as follows. Let $\boldsymbol{\Sigma} \triangleq \boldsymbol{\Gamma}^*\boldsymbol{\Gamma}$. Since $\mathbf{A}$ is positive definite we have that (see Appendix 5.A for a proof)

$$\operatorname{tr}\mathbf{A}^{-1} \geq \sum_{i=1}^{K_\alpha}\frac{1}{a_i} \tag{5.78}$$

with equality if and only if $\mathbf{A}$ is diagonal, and where

$$a_i = \mathbf{f}_{i-1}^* \mathbf{\Sigma} \mathbf{f}_{i-1} = \frac{1}{M} \operatorname{tr} \mathbf{\Sigma} \tag{5.79}$$

is the i-th element of the diagonal of $\mathbf{A}$. Therefore, the optimum value of $\mathbf{\Sigma}$, subject to $\operatorname{tr} \mathbf{\Sigma} \leq P$ is

$$\mathbf{\Sigma} = \frac{P}{M} \mathbf{I}_M \tag{5.80}$$

so that $\mathbf{A}$ is diagonal and $\operatorname{tr} \mathbf{A}^{-1} = \frac{K_\alpha M}{P}$.

Putting all together

It can be easily seen that the optimality conditions obtained in (5.6.4) and (5.6.4) are compatible, since a training sequence

$$\boldsymbol{\gamma} = \sqrt{\frac{P}{M}} \left[1,\, e^{-j\frac{2\pi K_\eta}{M}},\, \cdots e^{-j(M-1)\frac{2\pi K_\eta}{M}}\right] \tag{5.81}$$

meets (5.80) and (5.75). Therefore, the total estimation variance is $\sigma^2 \left(\frac{K_\eta}{M} + \frac{K_\alpha}{P}\right)$.

5.6.5 Results

In Table 5.1 we summarize the results obtained in this section. In Figures 5.7a-5.7c the estimation variance results for the SISO, MIMO and TV channels are shown. In all cases, dashed lines represent the analytical expression for the MSE with the *On-Off*[5] procedure described in [15], solid lines those of the optimum training sequence, and squares and circles the results of the Monte Carlo simulations in the *On-Off* and optimum cases, respectively. As expected, the optimum training clearly outperforms the *On-Off* one, being this difference specially noticeable in the TV channel.

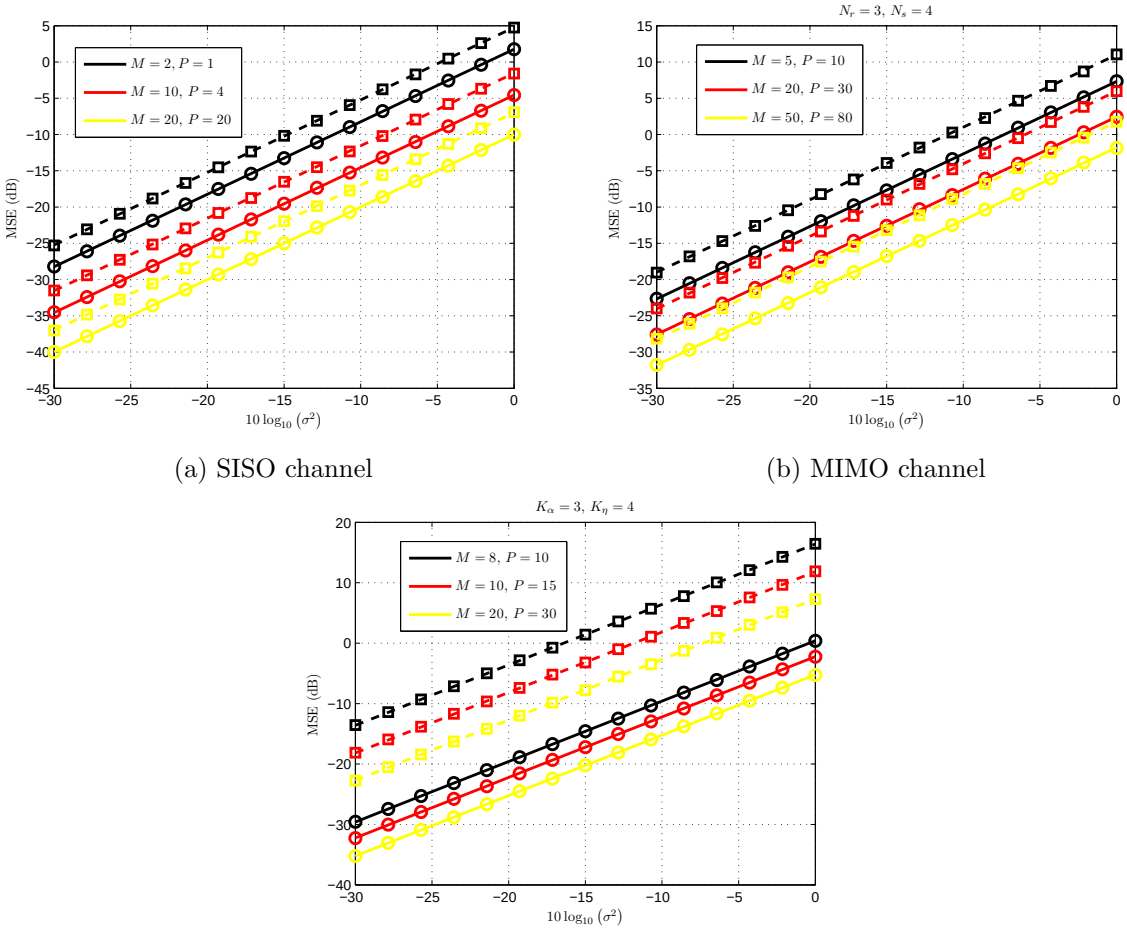
5.7 Conclusions

In this chapter we presented a scenario where a secondary transmitter is aware of the primary message, and exploits this knowledge to free some primary transmission resources to convey a secondary message. Channel knowledge is obtained by exploiting the SNR-based feedback

[5]For the MIMO case, the *On-Off* approach uses only one transmit antenna in each time slot. For the SISO case, when more than two time slots are used, the *On-Off* sequence is of the form 0, γ, 0, $\gamma \ldots$.

Channel	SISO	MIMO	SISO-TV
Signal Model	$\mathbf{y} = \boldsymbol{\gamma}\alpha + \mathbf{1}_M\eta + \mathbf{n}$	$\mathbf{y} = (\mathbf{1}_M \otimes \mathbf{I}_{N_r})\,\mathbf{g} + (\mathbf{W} \otimes \mathbf{I}_{N_r})\,\mathbf{h} + \mathbf{n}$	$\mathbf{y} = \boldsymbol{\Gamma}\mathbf{F}_\alpha\mathbf{b}_\alpha + \mathbf{F}_\eta\mathbf{b}_\eta + \mathbf{n}$
Signal Model Parameters	M: Number of Observations. σ^2: Noise Variance. P: Power of the training sequence.	M: Number of Observations. σ^2: Noise Variance. P: Power of the training sequence. N_r: Number of receive antennas. N_s: Number of transmit antennas at the ST.	M: Number of Observations. σ^2: Noise Variance. P: Power of the training sequence. K_α: Number of BEM coefficients of the ST to PR channel. K_η: Number of BEM coefficients of the PT to PR channel.
Parameters to estimate	α, η	$\mathbf{g}$, $\mathbf{h}$	$\mathbf{b}_\alpha$, $\mathbf{b}_\eta$
Optimum Training Sequence	$\boldsymbol{\gamma}\,\vert\boldsymbol{\gamma}^T\mathbf{1} = 0,\ \Vert\gamma\Vert^2 = P$	$\mathbf{W} = \mathbf{U}_1\boldsymbol{\Sigma}\mathbf{V}^H$ with $\boldsymbol{\Sigma} = \sqrt{\frac{P}{N_s}}\mathbf{I}$, $\mathbf{U}_1^T\mathbf{1} = \mathbf{0}$	$\boldsymbol{\gamma} = \sqrt{\frac{P}{M}}\begin{bmatrix} e^{-j0\frac{2\pi K_\eta}{M}} \\ \vdots \\ e^{-j(M-1)\frac{2\pi K_\eta}{M}} \end{bmatrix}$
Estimation Variance	$\sigma^2\left(\frac{1}{P} + \frac{1}{M}\right)$	$\sigma^2 N_r\left(\frac{N_s^2}{P} + \frac{1}{M}\right)$	$\sigma^2\left(\frac{K_\alpha}{P} + \frac{K_\eta}{M}\right)$

Table 5.1: Summary of results



(a) SISO channel

(b) MIMO channel

(c) Time-Varying channel

Figure 5.7: Channel estimation error for SISO, MIMO and time-varying channels. Solid lines represent the MSE obtained by using the training sequences in Table 5.1. Lines represent the analytical MSE, and circles/squares the simulation results.

from the primary receiver. This knowledge is shown to dramatically increase the fraction of released resources with respect to a non-CSI aware secondary transmitter. We also analyzed the CSI acquisition problem as presented in [15], and extended its results to the case of MIMO and time-varying channels. We derived closed form expressions for the variance of the channel estimators and designed optimum training sequences that minimize such a variance.

The content of this chapter was partially published in the conference proceedings of CIP 2012 and Asilomar 2012 [103, 104] with Prof. Carlos Mosquera as a co-author.

Appendix 5.A Trace inverse inequality

We will prove that if $\mathbf{A} \succ \mathbf{0}$, $A \in \mathbb{C}_{N\times N}$ then

$$\operatorname{tr} \mathbf{A}^{-1} \geq \sum_{i=1}^{N} \frac{1}{a_i} \tag{5.82}$$

with a_i the i-th element of the diagonal of $\mathbf{A}$. Note that

$$\operatorname{tr} \mathbf{A}^{-1} = \operatorname{tr} \frac{\operatorname{adj}\mathbf{A}}{\det \mathbf{A}} = \frac{1}{\det \mathbf{A}} \sum_{i=1}^{N} \det \mathbf{A}_i \tag{5.83}$$

with $\mathbf{A}_i$ the matrix resulting from the removal of the i-th row and column of matrix $\mathbf{A}$.

It suffices to prove that

$$\det \mathbf{A} \leq a_i \det \mathbf{A}_i \, \forall \, i. \tag{5.84}$$

More specifically, it suffices to prove (5.84) for $i = 1$ (for other values of i just apply a permutation). If we write

$$\mathbf{A} = \begin{bmatrix} a_1 & \mathbf{x}^T \\ \mathbf{x} & \mathbf{A}_1 \end{bmatrix} \tag{5.85}$$

we have that

$$\det \mathbf{A} = \det \mathbf{A}_1 \left(a_1 - \mathbf{x}^T \mathbf{A}_1^{-1} \mathbf{x}\right) \leq \det \mathbf{A}_1 a_1 \tag{5.86}$$

since $A^{-1} \succ \mathbf{0}$.

Part II

Link Adaptation

Chapter 6

Link Adaptation for Mobile Satellite Channels

Contents

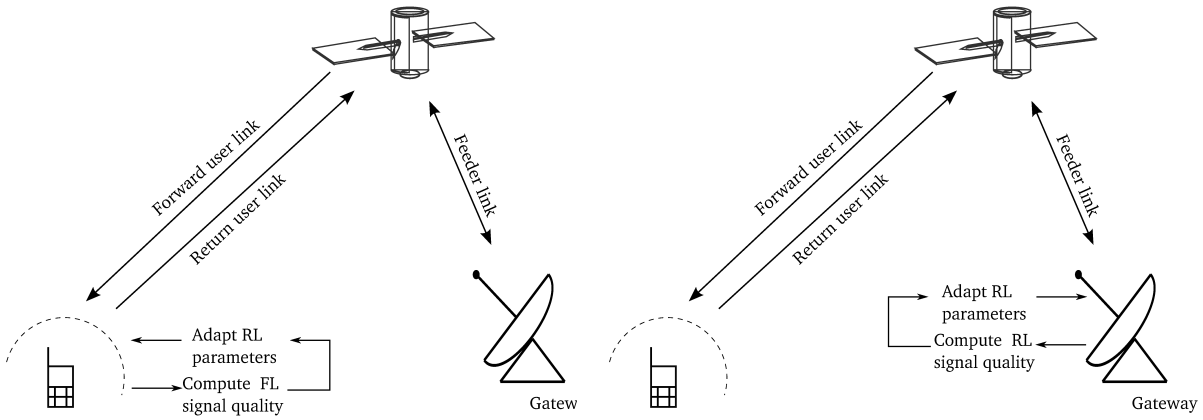
6.1 Introduction

In mobile satellite communications, there is an increasing need for more efficient transmission techniques that enable higher bit-rates at an affordable cost. To this extent, adaptive modulation and coding (AMC) allows the provision of broadband services to large user populations at lower costs, since it makes it possible to operate the links more efficiently by selecting the most suitable modulation and coding scheme (MCS) at each time [36]. However, the use of AMC for mobile links operating at S-band is hindered by the behavior of the land mobile satellite (LMS) channel [105]. This channel is usually modeled by a fast fading component, whose spectrum is related to the mobile speed by the Doppler effect, superimposed on a slow shadowing component; the parameters of both fading and shadowing depend on the environment in which the receiver happens to be. In short, the mobility of the user terminal will cause fast, difficult to predict channel variations, which will pose additional difficulties on the design of both forward and return link strategies.

Elaborating more on this issue, adaptation can be performed in *open-loop* or *closed loop* mode. In an open-loop scheme, the transmitter directly measures the signal quality from the reverse link and changes the parameters accordingly; it usually enjoys negligible delays, although at the price of having only partial information when both links are not perfectly correlated. On the other hand, closed-loop strategies wait for the other end to process their data and operate upon receiving some information about its reception. This makes them more accurate, although the experienced delay is much higher. In Figure 6.1 we show a diagram with the difference between closed loop and open loop adaptation.

Another key problem is that it is difficult to relate the channel statistics -even if we knew them- to the performance of the link because the channel is time varying. For example, it is difficult to tell what is best, a fixed channel with low quality or a fast varying channel with sharp transitions and a higher average quality. In an LMS channel, where transmission usually entails the use of complicated channel codes, if is difficult to do this without resorting to extensive simulations. This, however, is highly undesired for the design of an AMC strategy. In order to overcome this problem, we propose to use physical layer abstraction techniques. Thus, we will use effective SNR metrics (ESM) that have been reported to map complicated channel profiles with their actual performance in a one-to-one manner [106]. Therefore, our key assumption will be the following: a codeword will be correctly transmitted if the ESM of the channel that it undergoes is higher than or equal to the decoding threshold specified for the MCS in use in a static channel.

Focusing firstly on the forward link, the fading is attached to the receiver and, as a conse-



(a) Open loop adaptation.

(b) Closed loop adaptation.

Figure 6.1: Example of link adaptation in the return link. In open loop mode, the terminal observes the forward link signal and selects the MCS based on this value. In closed loop mode, the gateway measures the return link channel and feeds back the optimum MCS value to the terminal.

quence, the delay experienced by the CSI will be much longer than the channel coherence time for most speeds of practical interest. Therefore, the received CSI will be completely outdated -even if considering an open-loop scheme [107]- and rate adaptation will be of no use. However, if frames can be retransmitted or, at least, some additional redundancy or parity bits can be sent, then higher throughputs can be achieved at the cost of some latency. The underlying idea would be to transmit at a higher rate during good states of the channel, while somehow keeping the outage probability low. To achieve this, we propose two different strategies. The first one suggest resorting to superposition or multi-layer coding (MLC) with retransmissions, in a spirit close to hybrid-ARQ systems or layered rateless codes [108]. Particularly, we consider the use of two layers of information, each of them modulated at a different power. Upon reception, a decoding performance indicator will be sent and those bits which were not decoded by the receiver will be resent in a robust way (that is, in the high power layer) in order to keep the latency low. The second strategy for the forward link relies on the use of different MCS for different retransmission indexes. A high throughput is achieved by using high rate MCS in the first transmissions, and an outage constraint is met by reducing the rate when the retransmission index is high. A similar idea was presented in [108], where MLC and MCS that change with the retransmissions are used.

On the other hand, the return link enjoys timely channel information when operating in open-loop mode, but at the cost of having only partial CSI. This partial CSI knowledge is caused

by the operation of forward and return links on different frequency bands, so the multipath fading is uncorrelated [109]. The shadowing component, however, is the same for both channels. In some cases, open loop adaptation might offer a better performance than closed loop, but selecting the optimum operating regime is not a simple task. We propose to adaptively weight the open loop and closed loop CSI using only the ACK/NAK interchange.

The remaining of the chapter is structured as follows. Section 6.2 describes the system model; Section 6.3 contains the multi-layer coding approach to adaptation in the forward link; Section 6.4 describes the variable MCS adaptation in the forward link; Section 6.5 explains the adaptive CSI balancing for adaptation in the return link; Section 6.6 concludes the chapter.

6.2 System model

Consider a mobile satellite link operating at the L-band; the forward link band is centered at a frequency of 1550 MHz and the return link at a frequency of 1650 MHz. In the following we describe the signal model in detail, along with some key system assumptions.

6.2.1 Signal model

The signal model at a given time instant i, for both the forward and return link, is

$$y_i = \sqrt{\mathrm{snr}} \cdot h_i^{\mathrm{xl}} x_i + w_i \tag{6.1}$$

with y_i the received symbol, s_i the transmitted symbol, h_i^{xl} the channel coefficient, with $\mathrm{xl} \in \{\mathrm{fl}, \mathrm{rl}\}$ the channel in the forward or return link[1], and snr the signal to noise ratio; we also define $\sigma^2 \triangleq \mathrm{snr}^{-1}$. Accordingly, w_i is the unit-power noise contribution.

In the following, we describe the channel model and present some assumptions on the coding of the system under study. The methods developed in this chapter do not depend in the specific channel model, that will be abstracted by the use of ESM.

Channel model

We assume h_i^{xl} follows a Loo distribution [110]: slow variations in the LOS component (*shadowing*) are described by a log-normal distribution, whereas fast fluctuations of the signal amplitude (*fading*) are given by a Rician distribution. The PDF of the signal amplitude at a given time

[1] Throughout the chapter, the subscript fl or rl might be removed if the context is clear.

instant would be given by

$$f_r(x) = \frac{x}{b_0\sqrt{2\pi d_0}} \int_0^\infty \frac{1}{z} \exp\left(-\frac{(\log z - \mu)^2}{2d_0} - \frac{x^2+z^2}{2b_0}\right) I_0\left(\frac{x\cdot z}{b_0}\right) \, \mathrm{d}z \tag{6.2}$$

where d_0 and μ are the scale parameter and the location parameter of the log-normal distribution, respectively, and b_0 is the variance of the Rician distribution; to determine these parameters we follow the Fontan 3-state model [105]. If we average over the different states of the Fontan model, the resulting channel amplitude has the form of a Loo mixture, with PDF

$$f_r(x) = \sum_{i=1}^{N} p_i \frac{x}{b_0\sqrt{2\pi d_{0,i}}} \int_0^\infty \frac{1}{z} \exp\left(-\frac{(\log z - \mu_i)^2}{2d_{0,i}} - \frac{x^2+z^2}{2b_{0,i}}\right) I_0\left(\frac{x\cdot z}{b_{0,i}}\right) \, \partial z \tag{6.3}$$

From an implementation point of view, [105] advocates for generating the LOS component by first obtaining independent Gaussian samples n, exponentiating them $10^{(n/20)}$ to obtain log-normally distributed numbers, and finally interpolating them to obtain the correlation properties specified by the model. This procedure, followed by many more references afterwards, has two main drawbacks: the resulting sequence is not log-normal (interpolating a log-normal sequence does not preserve the original distribution), and the resulting process is not stationary.

A different procedure that preserves log-normality and ensures stationarity is proposed here. Starting from i.i.d. Gaussian samples, the correlation properties are introduced before the exponentiation [111] by a low pass filter whose cutoff frequency is given by $f_{\mathrm{LOS}} = v \cdot T_{\mathrm{symb}}/d_{\mathrm{c}}$ where v is the terminal speed, T_{symb} is the symbol period and d_{c} is the measured correlation distance of the LOS component, which we have obtained from [112]. We assume that the bandwidth of the system in both forward and return links is of 33.6 KHz, which leads to $T_{\mathrm{symb}} = 1/(33.6\,10^3)$ s.

The NLOS component, on the other hand, is obtained by filtering complex Gaussian samples with a low-pass filter whose cutoff frequency is given by the Doppler spread. Through this work we will assume that forward and return link channels experience the exact same LOS realization, but independent NLOS components with the same b_0 parameter and equal Doppler frequencies[2].

To sum up, the forward and return link channels are generated by

$$h_i^{\mathrm{rl}} = h_i^{\mathrm{LOS}} + h_i^{\mathrm{NLOS,rl}}, \;\; h_i^{\mathrm{fl}} = h_i^{\mathrm{LOS}} + h_i^{\mathrm{NLOS,fl}}. \tag{6.4}$$

[2]This is a simplification, as return and forward link operate in different frequencies. The difference between them, however, is less than 7%.

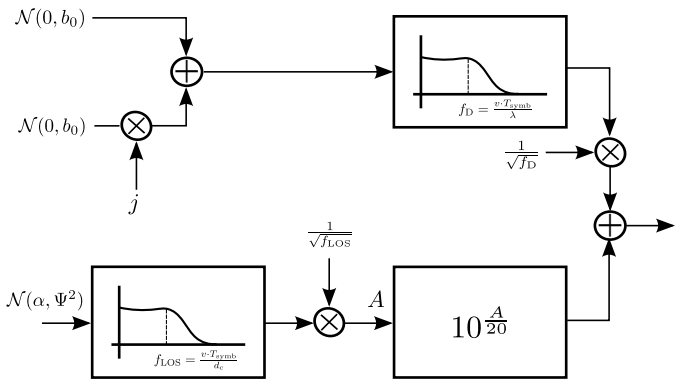


Figure 6.2: Diagram of the channel generation process. α and Ψ denote the mean and standard deviation of the lognormal shadowing when expressed in dB.

MCS	Coding rate	γ_{th} (dB)
QPSK 1/4	0.357	-1.5
QPSK 1/3	0.616	-0.3
QPSK 2/5	0.745	0.6
QPSK 1/2	0.831	1.9
QPSK 3/5	1.132	3.1
QPSK 2/3	1.261	4
QPSK 3/4	1.390	4.9
QPSK 4/5	1.476	5.6
QPSK 5/6	1.562	6.1
QPSK 8/9	1.691	7.1

MCS	Coding rate	γ_{th} (dB)
L8 QPSK	0.34	-2.1534
L7 QPSK	0.39	-1.3663
L6 QPSK	0.46	-0.3605
L5 QPSK	0.53	0.5733
L4 QPSK	0.61	1.5948
L3 QPSK	0.69	2.6092
L2 QPSK	0.73	3.1288
L1 QPSK	0.77	3.6674
R QPSK	0.81	4.2367

Table 6.1: MCS for forward (left) and return links (right). The MCS for the forward link are the ones used ind DVB-S2 [24], and the return link ones are the used in BGAN [61].

A block diagram of the channel generation process is shown in Figure 6.2.

Transmitted and received symbols

The transmitted symbols are the result of applying forward error correction coding and constellation mapping to a stream of bits; we consider a finite set of available codes, as described in Table 6.1 for the forward and return links. Throughout the chapter we will denote as $\gamma_{\mathrm{th},j}$ the threshold SNR for the i-th MCS, and drop the index i when its meaning is clear. We denote as r_j the spectral efficiency of the j-th MCS, that is obtained by multiplying the code rate with the $\log_2$ of the constellation size (2 for QPSK). In the following, we describe the input-output relationship of a channel using single layer coding. We denote the number of MCS as K. The definitions for multi-layer coding are similar, and will be introduced in Section 6.3.

Symbols form codewords $\mathbf{x}_i = \left[x_{iN},\, x_{iN+1},\, \ldots,\, x_{(i+1)N-1}\right]$ of constant length N, such that they see the channel samples

$$\mathbf{h}_i^{\mathrm{xl}} \triangleq \left[h_{iN}^{\mathrm{xl}}, h_{iN+1}^{\mathrm{xl}},\, \ldots,\, h_{(i+1)N-1}^{\mathrm{xl}}\right] \tag{6.5}$$

with $\mathrm{xl} \in \{\mathrm{fl}, \mathrm{rl}\}$. For each codeword sent, we assume that the other end feeds back an ACK if decoding was possible, and a NAK otherwise. We assume that $N = 2700$, which gives codewords of approximately $80ms$. We neglect the effect of headers or other sort of overhead.

Determining whether a codeword seeing different channel samples will be correctly decoded or not is a tough task. In general, the average SNR (possibly estimated from pilot symbols scattered through the codeword) is a poor indicator, as very different channel realizations could share the same value. To ease the simulation part, we use the effective SNR metrics (ESM) instead, which is given by

$$\gamma_{\mathrm{eff},i}^{\mathrm{xl}} \triangleq \Phi^{-1}\left(\frac{1}{N}\sum_{k=iN}^{(i+1)N-1}\Phi\left(\mathrm{snr}\cdot\left|h_k^{\mathrm{xl}}\right|^2\right)\right), \tag{6.6}$$

that is the SNR of an additive white Gaussian noise channel with the same mutual information as the faded channel $\mathbf{h}_i^{\mathrm{xl}}$, and with $\Phi(\gamma)$ the mutual information over a Gaussian channel with SNR γ and input restricted to a certain constellation $\{X_1, \ldots, X_L\}$

$$\Phi(\gamma) = 1 - \frac{1}{L\log_2 L}\times\sum_{\ell=1}^{L}\mathbb{E}_w\left[\log_2\left(\sum_{k=1}^{M}e^{-\frac{|X_\ell - X_k + w| - |w|^2}{1/\gamma}}\right)\right] \tag{6.7}$$

with $w \sim \mathcal{CN}\left(0, \frac{1}{\gamma}\right)$.

Using this metric, we assume that the transmission of the i-th codeword fails when $\gamma_{\mathrm{eff},k}$ is below the threshold SNR of the MCS used, and that it succeeds otherwise. In the case of multi-layer coding, each layer has a different SNR value. Define $\epsilon_i \in \{0, 1\}$ as the error event of the i-th codeword, then

$$\epsilon_i = \begin{cases} 1 & \text{if } \gamma_{\mathrm{eff},i} < \gamma_{\mathrm{th}} \\ 0 & \text{otherwise.} \end{cases} \tag{6.8}$$

6.3 Forward link: multi-layer coding

6.3.1 Introduction

If the transmitter uses Multi-Level Coding (MLC), the received signal can be written for the two-layer case as

$$y_i = \sqrt{\text{snr}} \cdot h_i \left(\sqrt{\alpha} x_i^{\mathcal{H}} + \sqrt{1-\alpha} x_i^{\mathcal{L}} \right) + w_i \tag{6.9}$$

with $x_i^{\mathcal{H}}$ and $x_i^{\mathcal{L}}$ the i-th symbols of the $\mathcal{H}$ (High priority) and $\mathcal{L}$ (Low priority) levels, and α a parameter that weights the power sharing between the two layers. Note that for $\alpha = 1$ we have a single-layer transmission. This MLC scheme is depicted in Figure 6.3.

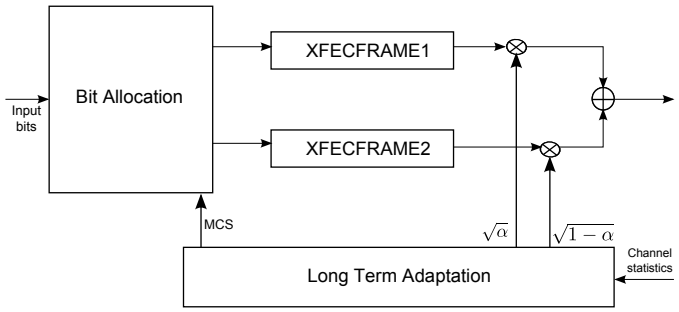


Figure 6.3: Schematic of the MLC scheme: the input bits are divided into two different layers, and the transmission parameters are selected according to long-term channel statistics. Although two different DVB-S2 XFECFRAME are transmitted, they are restricted to be coded with the same MCS to ease implementation.

For the sake of clarity, we define the SNR of the i-th symbol of the $\mathcal{H}$ and $\mathcal{L}$ levels as

$$\gamma_i^{\mathcal{H}} = \frac{|h_i|^2 \alpha}{\sigma^2 + (1-\alpha)|h_i|^2} \tag{6.10}$$

$$\gamma_i^{\mathcal{L}} = \frac{|h_i|^2 (1-\alpha)}{\sigma^2}. \tag{6.11}$$

Note that this approach is conservative, as the interference caused by the $\mathcal{L}$ layer to the $\mathcal{H}$ layer is treated as Gaussian noise, when it is clear that it has a lower entropy, thus leading to a higher mutual information. The effective snr values $\gamma_{\text{eff},i}^{\mathcal{L}}$ and $\gamma_{\text{eff},i}^{\mathcal{H}}$ are defined by combining (6.10-6.11) and (6.6).

The $\mathcal{L}$ layer will be correctly decoded if the $\mathcal{H}$ layer was correctly decoded (i.e., if $\gamma_{\text{eff},i}^{\mathcal{H}} > \gamma_{\text{th}}$)

and if $\gamma_{\text{eff},i}^{\mathcal{L}} > \gamma_{\text{th}}$, or, equivalently, if the following condition is met:

$$\gamma_{\text{eff},i}^{\mathcal{H}} \geq \gamma_{\text{th}} \cap \gamma_{\text{eff},i}^{\mathcal{L}} \geq \gamma_{\text{th}}. \tag{6.12}$$

Note that we are restricting both layers to have the same constellation and code rates, which is desirable for implementation issues. In the following, we will restrict our analysis to the first codeword $i = 0$, without loss of generality, and drop the indexes that take into account the codeword number.

Our objective function will be the average spectral efficiency (ASE), which is defined for an MCS index j and power weighting α as

$$\eta\left(j,\alpha\right) \triangleq r_j \cdot \left(\mathbb{P}\left[\gamma_{\text{eff}}^{\mathcal{H}} \geq \gamma_{\text{th},j}\right] + \mathbb{P}\left[\gamma_{\text{eff}}^{\mathcal{H}} \geq \gamma_{\text{th},j} \cap \gamma_{\text{eff}}^{\mathcal{L}} \geq \gamma_{\text{th},j}\right]\right) \tag{6.13}$$

with r_j the spectral efficiency of the j-th MCS. The outage probability constraint is defined as

$$g\left(j,\alpha\right) \triangleq \mathbb{P}\left[\gamma_{\text{eff}}^{\mathcal{H}} < \gamma_{\text{th},j}\right] \leq p_0. \tag{6.14}$$

Therefore, our design problem is stated as

$$\begin{array}{ll} \text{maximize} & \eta\left(j,\alpha\right) \\ \text{subject to} & g\left(j,\alpha\right) \leq p_0 \\ & 0 \leq \alpha \leq 1 \end{array} \tag{6.15}$$

where the maximization is performed over the MCS index j and $\alpha \in \mathbb{R}$. The MCS choice is assuming the knowledge of the channel statistics, which we consider to be available at the gateway.

In the following, we analyze the problem by assuming a block fading channel, which is a realistic assumption if the mobile speed is small enough, and makes the problem analytically tractable, and afterwards evaluate the evolution of throughput and outage probability as a function of the mobile speed.

6.3.2 Block fading

The histogram of the effective SNR in a Fontan channel has been obtained for different mobile speeds and for different *average SNR* values. In Figure 6.4 we can see the variation of the effective SNR with speed. Clearly, as the speed increases the effective SNR variance diminishes as the result of averaging more *channel states* in the same codeword.

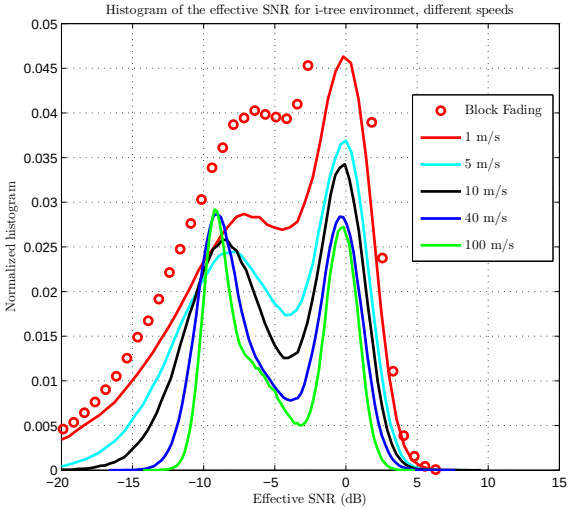


Figure 6.4: Histogram of the effective SNR for different speed values, ITS environment, very low SNR.

Unfortunately, obtaining the pdf of the effective SNR for a mobile seems to be analytically intractable, so we will assume that all the channel states seen by a codeword are approximately the same (*block fading*):

$$h_1 \approx h_i,\ i = 2,\ ...,\ N \tag{6.16}$$

so we can approximate

$$\gamma_{\text{eff}}^{\mathcal{H}} \approx \gamma^{\mathcal{H}} \tag{6.17}$$

with $\gamma^{\mathcal{H}} \approx \gamma_1^{\mathcal{H}},\ ...,\ \gamma_N^{\mathcal{H}}$. If this is the case we can define the event of decodability of the $\mathcal{H}$ layer as

$$\gamma^{\mathcal{H}} \geq \gamma_{\text{th},j} \tag{6.18}$$

Analogously, we can rewrite the event of decodability of the $\mathcal{L}$ layer as

$$\left(\gamma^{\mathcal{H}} \geq \gamma_{\text{th},j}\right) \cap \left(\gamma^{\mathcal{L}} \geq \gamma_{\text{th},j}\right). \tag{6.19}$$

Following (6.10) we can rewrite (6.18) as

$$|h|^2 \geq \frac{\sigma^2 \gamma_{\text{th},j}}{\alpha - (1-\alpha)\,\gamma_{\text{th},j}}, \tag{6.20}$$

and (6.19) according to (6.11) as

$$|h|^2 \geq \max\left\{\frac{\sigma^2\gamma_{\mathrm{th},j}}{\alpha-(1-\alpha)\,\gamma_{\mathrm{th},j}}, \frac{\gamma_{\mathrm{th},j}\sigma^2}{1-\alpha}\right\}. \tag{6.21}$$

Note that if we choose a value of α such that

$$\frac{\sigma^2\gamma_{\mathrm{th},j}}{\alpha-(1-\alpha)\,\gamma_{\mathrm{th},j}} > \frac{\gamma_{\mathrm{th},j}\sigma^2}{1-\alpha} \tag{6.22}$$

then the $\mathcal{L}$ layer will be only limited by the decodability of the $\mathcal{H}$ layer, which is clearly not optimum. Therefore, we might want to choose a value of α such that

$$\alpha \geq \frac{\gamma_{\mathrm{th},j}+1}{\gamma_{\mathrm{th},j}+2}. \tag{6.23}$$

By adding this constraint to α, the ASE can be written as

$$\eta\,(j,\alpha) = r_j\ \left(\mathbb{P}\left[|h|^2 \geq \frac{\sigma^2\gamma_{\mathrm{th},j}}{\alpha-(1-\alpha)\,\gamma_{\mathrm{th},j}}\right] + \mathbb{P}\left[|h|^2 \geq \frac{\gamma_{\mathrm{th},j}\sigma^2}{1-\alpha}\right]\right) \tag{6.24}$$

and the outage probability

$$g\,(j,\alpha) = P\left[|h|^2 \leq \frac{\sigma^2\gamma_{\mathrm{th},j}}{\alpha-(1-\alpha)\,\gamma_{\mathrm{th},j}}\right] \leq p_0. \tag{6.25}$$

Note that the outage probability applies to the $\mathcal{H}$ layer and, in consequence, $g\,(j,\alpha)$ is a monotonic decreasing function of α, as it is clear that allocating more power to the $\mathcal{H}$ layer will decrease the outage probability. Therefore, if we assume that $g\,(j,1) \leq p_0$ (otherwise the problem will be infeasible for the MCS j), constraint (6.25) is equivalent to

$$\alpha \geq \alpha_{0,j} \tag{6.26}$$

with $\alpha_{0,j}$ such that $g\,(j,\alpha_{0,j}) = p_0$. Therefore, for a given j such that the problem is feasible, the optimum value $\eta_j^\star$ is obtained as

$$\eta_j^\star = \max_{\alpha_{\min,j}\leq\alpha\leq 1}\{\eta\,(j,\alpha)\} \tag{6.27}$$

with

$$\alpha_j^{min} \triangleq \max\left\{\alpha_{0,j}, \frac{\gamma_{\mathrm{th},j}+1}{\gamma_{\mathrm{th},j}+2}\right\}, \tag{6.28}$$

$$\eta^\star = \max\left\{\eta_j^\star\right\}_{j=1}^{K}. \tag{6.29}$$

Note that the previous expressions involving probabilities of the channel power $|h|^2$ being smaller than a given threshold k can be easily rewritten as

$$\mathbb{P}\left[|h|^2 \leq a\right] = F_r\left(\sqrt{a}\right) \tag{6.30}$$

where

$$F_r\left(\sqrt{a}\right) = \int_0^{\sqrt{a}} f_r\left(x\right) \partial x, \tag{6.31}$$

being f_r the pdf of the Loo mixture (6.3).

6.3.3 Evolution of outage probability with speed and environment

It is clear that for a given line of sight SNR (the SNR a receiver would experience without shadowing and fading), the performance of a receiver (in terms of ASE or outage probability) is going to be heavily dependent on the environment, which determines the fading and shadowing statistical characterization. Moreover, as we have already seen, an increasing mobile speed causes different channel states to be *averaged* during the same codeword, thus reducing the variance of the effective SNR and, as a consequence, reducing the outage probability.

Therefore, as we are constraining the outage probability to lie below a given threshold (according to (6.15)), it would be useful to analyze the outage probability performance of the most protected MCS for different speeds and environments. Note that if the most protected MCS does not provide the desired outage probability, then it is clear that the optimization problem (6.15) is unfeasible.

In order to anticipate the performance of the MLC strategy for the LMS channel, we have simulated its performance for different values of α and speed, as we can see in Figure 6.5. Obviously, the larger α is, the less interference we receive from the $\mathcal{L}$ layer, so the outage probability decreases as α increases. Moreover, we have seen in Figure 6.4 that the effect of incrementing the speed is to reduce the variance of the ESM, as more states are averaged for the same codeword. As an outage event occurs when the ESM of a codeword is very small, the variance reduction implies that this event is less likely. As a consequence, the outage probability decreases with the speed[3].

Results for the intermediate-tree shadowing, heavy-tree shadowing, open and suburban

[3] A different effect is observed in [113, 114] where the ergodic capacity is shown to decrease with speed if pilots are used to estimate the channel. In our case, we assume perfect CSI and measure the outage capacity. The analysis of outage capacity in fading channels is quite involved even for the simple case of correlated Rayleigh taps with perfect CSI [115]. Therefore, the evolution of outage capacity with imperfect CSI does not seem to be tractable.

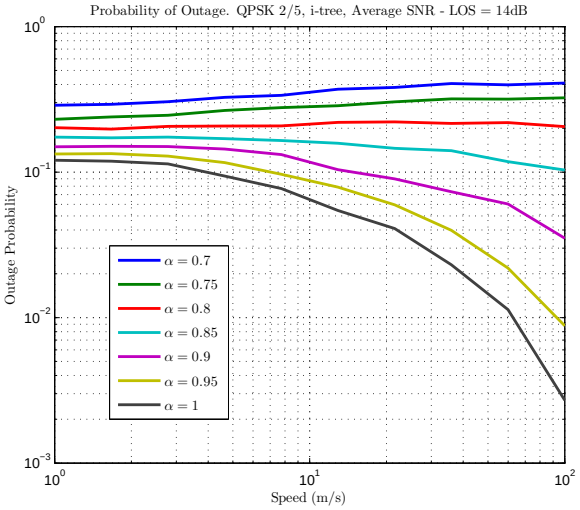
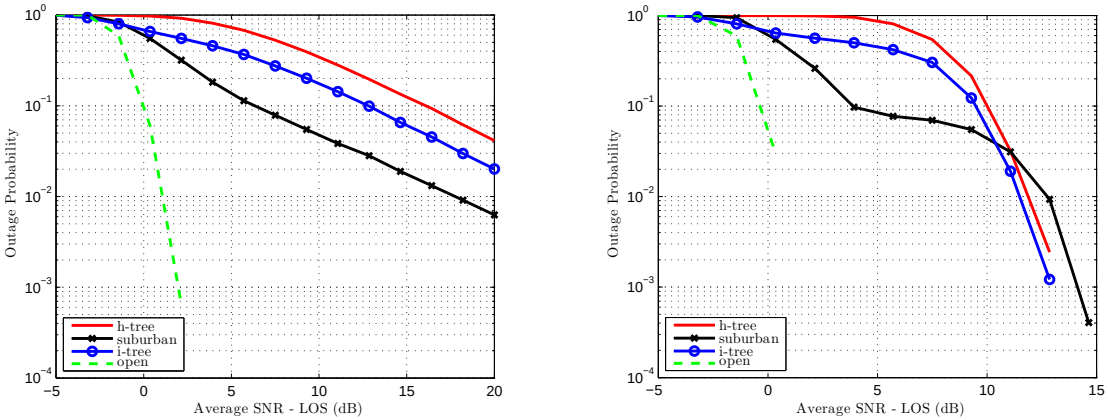


Figure 6.5: Outage probability for different values of α and speed.

environments are shown for mobile speeds of 0.1 m/s and 40 m/s on Figure 6.6. It is easy to conclude that higher speeds dramatically reduce the required SNR for a given outage probability, thus allowing the system to operate with lower SNR values. As an example, note that enforcing an outage probability of 0.1 in the 0.1 m/s case under the heavy-tree shadowing environment requires the received LOS SNR to be larger than 16 dB; this value is reduced to 10 dB if the speed is increased up to 40 m/s.



(a) 0.1 m/s (b) 40m/s

Figure 6.6: Outage probability for the most protected MCS in different environments.

6.3.4 Performance of the proposed optimized MCS

We have evaluated the outage and throughput of the forward link with the optimized MCSs, assuming that the gateway is aware of the statistics of the channel. First of all, we have solved the optimization for a constraint on the outage probability of $p_0 = 2 \cdot 10^{-2}$. The LOS SNR was set to a relatively high –and maybe unrealistic– value, in order to be able to use some higher rate MCSs. In Figure 6.7, it can be seen that the outage probability constraint is met for all speeds, and specially for the larger ones. Therefore, the block fading approach turns out to be quite conservative.

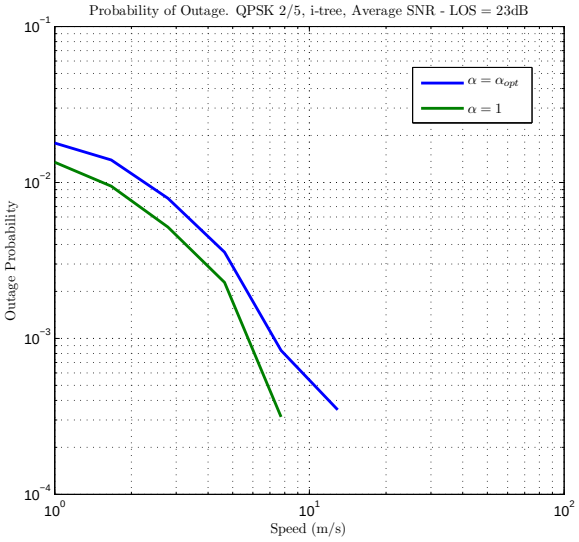


Figure 6.7: Outage probability with the parameters obtained from the optimization (assuming block fading), compared with the SLC case ($\alpha = 1$). Target $p_0 = 0.02$.

Similarly to the results on the outage probability presented in the previous section, we will evaluate the performance of the proposed scheme as a function of the LOS SNR. In this case we have set the outage probability constraint to 0.1 in order to have a feasible problem for realistic SNR values. In Figure 6.8 there is a plot of the ASE as a function of the LOS SNR for the heavy-tree, suburban and intermediate-tree scenarios, respectively, for both MLC and SLC ($\alpha = 1$). It can be seen that the MLC outperforms SLC for almost every SNR value, while meeting the outage probability constraint due to the selection of the parameter α. It is worth remarking that the throughput does not change with the speed in most of the cases, although higher speeds obviously attain a lower outage probability. Note that an ASE equal to zero indicates that the problem is unfeasible for the selected SNR value. The obtained values of

outage probability stayed below the target value in both cases, with lower values in the case of $v = 20\,\mathrm{m/s}$.

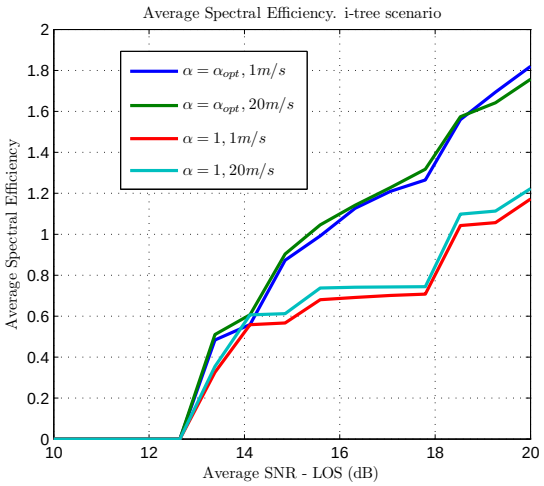


Figure 6.8: Average spectral efficiency with SLC and MLC in ITS environment. Target $p_0 = 0.1$.

6.4 Forward link: ARQ with different MCS

6.4.1 Introduction

In the previous section we used multi-layer coding to increase the throughput in the forward link of a mobile satellite scenario. Although not explicitly noted, the proposed retransmission strategy, together with the outage constraint, introduces a constraint in the maximum allowed delay.

The use of multi-layer coding allows to increase the spectral efficiency, but the receivers are more complex (they have to perform inter-layer interference cancellation) and the statistics of the channel are needed to design the optimum power weighting and MCS index. In this section, we propose to use a simpler single-layer transmission, but with the possibility of using different MCS in each retransmission index. The outage probability is defined in terms of the post-ARQ packet error probability, and the objective is once again to maximize the throughput subject to a packet error rate constraint.

6.4.2 Link adaptation

Problem statement

The objective of link adaptation is to maximize the throughput guaranteeing the data-link layer QoS, expressed in terms of outage or packet error rate and maximum delay. For a maximum delay or, equivalently, for a maximum number of transmissions t, our design variables are the MCS indexes to use on each attempt $\{s_1, \ldots, s_t\} \in [1, \ldots, K]^t$. These values must satisfy an outage constraint given by p_0, so

$$\prod_{i=1}^{t} P_{s_i} \leq p. \tag{6.32}$$

Each bit waiting to be transmitted can be in one of t different states denoted by M_i, $i = 1, \ldots, t$, with i the number of maximum attempts before being discarded. In state M_i the data still has i transmissions available. Decoding will fail with a probability P_{s_i}, and data will move onto the next state, M_{i-1}, with $i-1$ transmissions left; on the other hand, if the transmission is successful -with probability $(1 - P_{s_i})$- the system moves back to the first state. The Markov chain representing this transmission procedure is depicted in Figure 6.9.

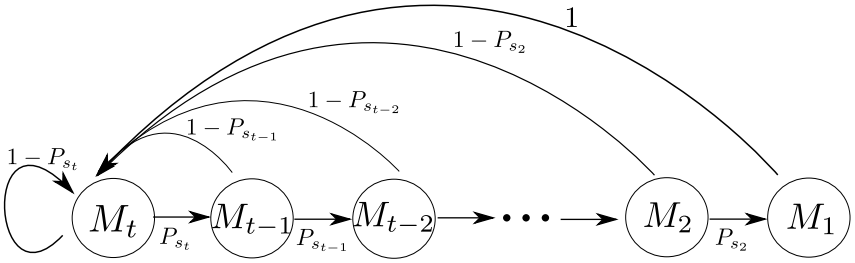


Figure 6.9: Markov model representing the transitions between states.

With this, we have that the ASE is given by

$$\eta = \sum_{i=1}^{t} \mathbb{P}\left[M_i\right] r_{s_i} \left(1 - P_{s_i}\right) \tag{6.33}$$

where the probability for a given bit to belong to state i is easily obtained from the diagram in

Figure 6.9 as

$$\begin{aligned}
\mathbb{P}[M_t] &= 1/P \\
\mathbb{P}[M_{t-1}] &= P_{s_t}/P \\
\mathbb{P}[M_{t-2}] &= P_{s_t} \cdot P_{s_{t-1}}/P \\
&\dots \\
\mathbb{P}[M_1] &= P_{s_t} \cdot P_{s_{t-1}} \cdot \ldots \cdot P_{s_2}/P.
\end{aligned} \tag{6.34}$$

For notational convenience, we have defined $P \triangleq 1 + P_{s_t} + P_{s_{t-1}} \cdot P_{s_{t-1}} + \cdots + P_{s_t} \cdot P_{s_{t-1}} \cdot \ldots \cdot P_{s_2}$. Note that the outage probability of the last state P_{s_1} does not affect the probability of the different states, but only the average spectral efficiency of the last retransmission and the overall outage probability. The input bit rate is supposed to be such that these probabilities remain stationary with time.

Based on the previous considerations, the choice of the MCS indexes can be recast as the following optimization problem:

$$\begin{array}{ll}
\text{maximize} & \sum_{i=1}^{t} \mathbb{P}[M_i]\, r_{s_i}(1 - P_{s_i}) \\
\text{subject to} & \prod_{i=1}^{t} P_{s_i} \leq p_0.
\end{array} \tag{6.35}$$

This problem seems hard to solve, at least for the general case of different rates being assigned to subsequent retransmissions. We will now present some simplifications to this problem.

Simplifying the optimization problem

There are some optimality properties that can be exploited to avoid the need for an exhaustive search over the whole set of codes. Intuitively, for example, it seems plausible to expect higher rates at earlier attempts and more robust MCS for the subsequent retransmissions. Next we will prove this for $t = 2$, with the conjecture, supported by practical optimizations, that this is also the case for any t.

Proposition 6.1. *If t=2, then the solution to (6.35) meets the following two propositions:*

1. *if* $\exists\, j \neq s_i$ *such that* $r_{s_i}(1 - P_{s_i}) < r_j(1 - P_j)$ *then* $P_{s_i} \leq P_j$ $\forall i$*=1, 2*
2. $r_{s_2} \geq r_{s_1}$ *(or, equivalently,* $P_{s_2} \geq P_{s_1}$*).*

Proof. The first proposition is easy to prove for $i = 1$. Let us assume (s_2^*, s_1^*) is optimum, and

denote $T_x \triangleq r_x\,(1-P_x)$, so we can write the objective function as

$$r = \frac{T_{s_2^*} + P_{s_2^*} T_{s_1^*}}{1 + P_{s_2^*}}. \tag{6.36}$$

It is easy to see that if there exists j such that $T_j > T_{s_1^*}$ and $P_j < P_{s_1^*}$ the outage constraint will be met if it was met for (s_2^*, s_1^*) and the objective function will increase, so (s_2^*, s_1^*) cannot be optimum.

Next, assume that (s_2^*, s_1^*) is optimum with $P_{s_1^*} > P_{s_2^*}$. This implies that $T_{s_1^*} > T_{s_2^*}$ from the previous property. Also, we have that the outage probability of (s_2^*, s_1^*) is the same as that of (s_1^*, s_2^*). Therefore, to prove that (s_2^*, s_1^*) is not optimum it suffices to prove that

$$T_{s_1^*}\theta_{s_1^*} + T_{s_2^*}(1-\theta_{s_1^*}) \geq T_{s_2^*}\theta_{s_2^*} + T_{s_1^*}(1-\theta_{s_2^*}) \tag{6.37}$$

where $\theta_{s_2^*} \triangleq 1/(1+P_{s_2^*})$ is the probability of a bit being transmitted in the first state if the order is (s_2^*, s_1^*), and $\theta_{s_1^*} \triangleq 1/(1+P_{s_1^*})$ is the probability of a bit being transmitted in the first state if the order is (s_1^*, s_2^*). Note also that $\theta_i \geq (1-\theta_i)$, and $\theta_{s_2^*} \geq \theta_{s_1^*}$. With this, we have that

$$\begin{aligned} T_{s_1^*}\theta_{s_1^*} + T_{s_2^*}(1-\theta_{s_1^*}) &= \theta_{s_1^*}\left(T_{s_1^*} - T_{s_2^*}\right) + T_{s_2^*} \\ &\overset{(i)}{\geq} \left(1-\theta_{s_1^*}\right)\left(T_{s_1^*} - T_{s_2^*}\right) + T_{s_2^*} \\ &= T_{s_1^*} + \theta_{s_1^*}\left(T_{s_2^*} - T_{s_1^*}\right) \\ &\overset{(ii)}{\geq} T_{s_1^*} + \theta_{s_2^*}\left(T_{s_2^*} - T_{s_1^*}\right) \\ &= T_{s_1^*}\left(1-\theta_{s_2^*}\right) + T_{s_2^*}\theta_{s_2^*} \end{aligned} \tag{6.38}$$

where (i) is due to $T_{s_1^*} > T_{s_2^*}$ and $\theta_{s_1^*} > (1-\theta_{s_1^*})$, and (ii) is due to $T_{s_1^*} > T_{s_2^*}$ and $\theta_{s_1^*} > \theta_{s_2^*}$. With this, we prove that it is better to transmit first with the higher rate code, so (s_2^*, s_1^*) is not optimum if $P_{s_1^*} > P_{s_2^*}$. Now, we have to prove proposition I for $i = 2$. This is easy to prove since

$$\frac{\mathrm{d}r}{\mathrm{d}P_{s_2}} = \frac{T_{s_1} - T_{s_2}}{(1+P_{s_2})} \leq 0 \tag{6.39}$$

as $T_{s_1} \leq T_{s_2}$, and

$$\frac{\mathrm{d}r}{\mathrm{d}T_{s_2}} = \frac{1}{(1+P_{s_2})} \geq 0 \tag{6.40}$$

so decreasing the error probability of s_2 while increasing T_{s_2} will increase the objective function.

This lemma proves that for $t = 2$ we only have to take into account those MCS where an increment in the outage probability implies an increment in the throughput, and that the MCS are of non-increasing rate. Note that this result is quite general, since the only assumption for

the relationship between code rate and error probability is that higher rates imply higher error probabilities. □

Conjecture 6.1. *For all t, the solution to (6.35) meets the following two propositions:*

1. *if $\exists\, j \neq s_i$ such that $r_{s_i}(1-P_{s_i}) < r_j(1-P_j)$ then $P_{s_i} \leq P_j$ $\forall i = 1, \ldots, t$*
2. *$r_{s_{i+1}} \geq r_{s_i}$ (or, equivalently, $P_{s_{i+1}} \geq P_{s_i}$) $\forall i = 1, \ldots, t-1$.*

6.4.3 Optimization algorithm

With the above conjecture, the optimization algorithm could be written as follows: in a first step, we will obtain the subset of K' MCS that meet Condition 1; then, we will perform the optimization over them using Condition 2.

Require: MCS set sorted by descending rate, $i < j \Leftrightarrow r_j \leq r_i \quad \forall r_i, r_j \in \{r_i\}_{i=1}^{M}$

```
T_0 ← 0
for i = 1 to K do
  if r_i(1 − P_i) < T_{i−1} then
    Delete MCS i
  else
    T_i = r_i (1 − P_i)
  end if
end for
```

We now compute all the valid combinations of MCS:

```
i ← 0
for a_1 = 1 to K' do
  for a_2 = 1 to a_1 do
    ⋮
    for a_t = 1 to a_{t−1} do
      F_i = {a_1, a_2, ··· , a_t}
      if ∏_{j=1}^{t} P_{a_j} ≤ p_0 then
        G_i = Σ_{j=1}^{t} ℙ[M_{a_j}] r_{a_j}(1 − P_{a_j})
      else
        G_i = ∞
      end if
    end for
  end for
```

end for

return $F_{\text{opt}} = \arg\max_{F_i} G_i$

The complexity of this algorithm is less than that of a brute-force search over all the possible combinations, which would result into K^t combinations; the complexity is now given by the following proposition.

Lemma 6.1. *The modified algorithm only needs to check* $\binom{K'-1+t}{t}$ *combinations.*

Proof. From the algorithm description, we have that the number of combinations will be given by

$$\begin{aligned} N &= \sum_{a_1=0}^{K'-1} \sum_{a_2=0}^{a_1} \cdots \sum_{a_{t-2}=0}^{a_{t-3}} \sum_{a_{t-1}=0}^{a_{t-2}} \sum_{a_t=0}^{a_{t-1}} 1 \\ &= \sum_{a_1=0}^{K'-1} \sum_{a_2=0}^{a_1} \cdots \sum_{a_{t-2}=0}^{a_{t-3}} \sum_{a_{t-1}=0}^{a_{t-2}} \binom{a_{t-1}+1}{1} \\ &= \sum_{a_1=0}^{K'-1} \sum_{a_2=0}^{a_1} \cdots \sum_{a_{t-2}=0}^{a_{t-3}} \binom{a_{t-2}+2}{2} \\ &= \sum_{a_1=0}^{K'-1} \binom{a_1+(t-1)}{t-1} = \binom{K'-1+t}{t} \end{aligned} \tag{6.41}$$

where we have used $\sum_{j=0}^{k} \binom{n+j}{n} = \binom{k+n+1}{n+1}$. □

The reduction in complexity with respect to the brute force solution is quite remarkable, reaching almost an order of magnitude for $t = 3$ and $K' = 20$.

Knowledge of P_i

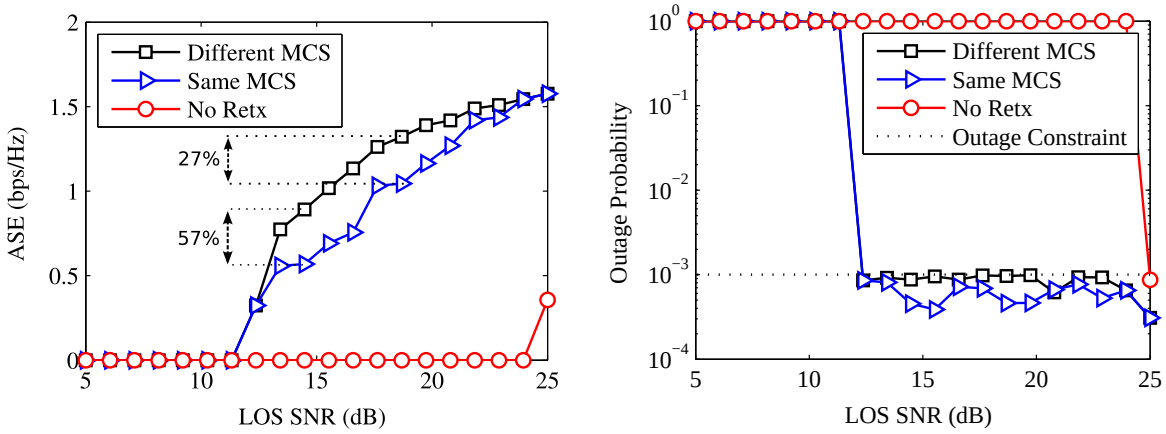
So far we have been implicitly assuming that the transmitter will know the pdf of γ_{eff} in order to obtain P_k perfectly. But, indeed, the proposed link adaptation algorithm can be easily implemented without this knowledge, in an online fashion. A similar idea was introduced in [116] for link adaptation in MIMO-OFDM systems. Here, the transmit side would estimate directly P_k by just observing the ACK/NAK of the ARQ protocol:

$$\hat{P}_{k,n_{k+1}} = \frac{n_k}{n_k+1}\hat{P}_{k,n_k} + \frac{1}{n_k+1}a_k \tag{6.42}$$

with n_k the number of ACK/NAK observations for MCS c_k, $a_k = 1(0)$ if an ACK (NAK) is received, and $\hat{P}_{k,n_k}$ the estimation of P_k after n_k packets are received.

A transmitter selecting the MCS following the adaptation algorithm and updating the PER estimations following (6.42) can be easily stuck in a suboptimal solution. If some of the optimum MCS have not been *explored*, the corresponding PER estimation will have a large variance and it can cause the adaptation algorithm not to select that MCS, thus disabling the possibility of improving the estimation. To overcome this issue, we set a fraction ϵ of the packets to be scheduled with a random MCS, so we can improve the estimation of the PER in case an MCS is seldom selected. In our simulations we selected $\epsilon = 0.01$.

6.4.4 Simulation results



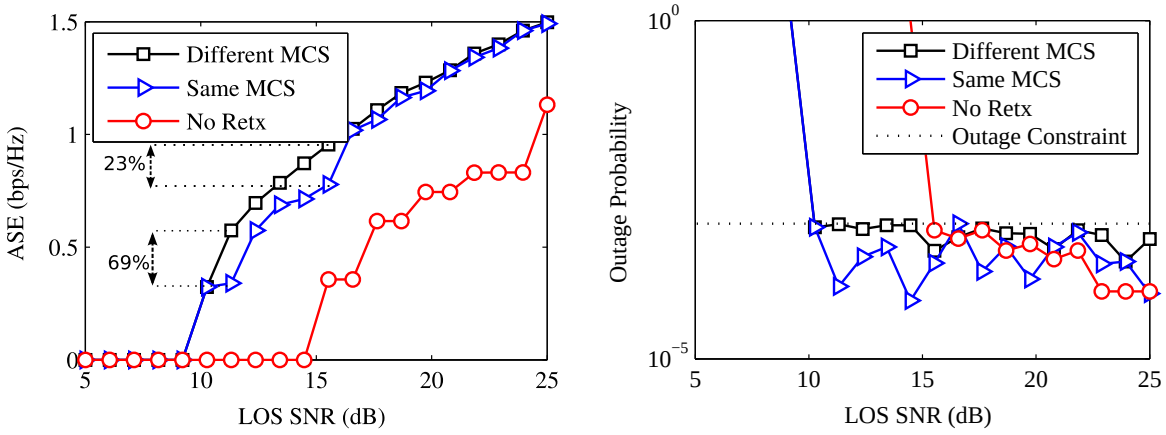
(a) Average Spectral Efficiency.

(b) Frame error rate.

Figure 6.10: Average spectral efficiency and outage probability. Receiver speed: 1 m/s.

We have simulated the performance of the proposed ARQ scheme; we simulated a 3-transmission strategy (2 retransmissions) with an outage probability of 10^{-3}. In this first approximation, we assume perfect empirical knowledge of the probability density function of γ_{eff}. We compare the results with the case of not allowing retransmissions at all ($t = 1$) with the same outage probability constraint, that could be the constraint in interactive applications, for example.

Results are shown on Figures 6.10-6.11 for speeds of 1 m/s and 5 m/s. The results are compared with the case of using the same MCS for every retransmission (i.e., an outage probability of 0.1 for each transmission). It can be seen that our approach outperforms the fixed MCS in all scenarios, specially in the low speed ones due to the lower time diversity. Note that in some



(a) Average Spectral Efficiency.

(b) Frame error rate.

Figure 6.11: Average spectral efficiency and outage probability. Receiver speed: 5 m/s.

cases the fixed MCS is the solution to our problem, leading to the same throughput and outage probability. In the case of not allowing retransmissions the throughput is severely reduced due to the smaller diversity. In fact, for the low speed case, the small diversity introduced by the channel variations causes the system to be unable to meet the outage constraint for almost every SNR.

The solution for 1 m/s for the different points is shown on Table 6.2, where it can be seen that the MCS rate decreases with the transmission index, and that there can be a huge variation between the first and last transmissions. In the simulations we solved the problem by brute force, so these results agree with Conjecture 6.1.

In Figure 6.12 we show the evolution of the spectral efficiency with time (averaged over a window of 700 packets) when online adaptation is used following (6.42). We can see that in the high SNR case the offline optimization outperforms the online approach due to the perfect knowledge of the error probabilities P_k. For the lower SNR case both approaches offer approximately the same performance. For the sake of simplicity, we used *Stop-and-Wait* ARQ for the simulations, so only one packet is transmitted every RTT, although the method can be also applied for more efficient protocols like *Selective Repeat ARQ* [117].

Table 6.2: Optimum MCS evolution.

LOS SNR (dB)	s_3	s_2	s_1	Same MCS
23.9474	QPSK 8/9	QPSK 8/9	QPSK 8/9	QPSK 8/9
22.8947	QPSK 8/9	QPSK 8/9	QPSK 4/5	QPSK 5/6
21.8421	QPSK 8/9	QPSK 8/9	QPSK 3/4	QPSK 5/6
20.7895	QPSK 8/9	QPSK 8/9	QPSK 3/5	QPSK 3/4
19.7368	QPSK 8/9	QPSK 8/9	QPSK 2/5	QPSK 2/3
18.6842	QPSK 8/9	QPSK 5/6	QPSK 1/4	QPSK 2/3
17.6316	QPSK 8/9	QPSK 5/6	QPSK 1/4	QPSK 3/5
16.5789	QPSK 5/6	QPSK 2/3	QPSK 1/4	QPSK 1/2
15.5263	QPSK 3/4	QPSK 3/5	QPSK 1/4	QPSK 2/5
14.4737	QPSK 3/4	QPSK 1/3	QPSK 1/4	QPSK 2/5
13.4211	QPSK 3/5	QPSK 1/4	QPSK 1/4	QPSK 1/3
12.3684	QPSK 1/4	QPSK 1/4	QPSK 1/4	QPSK 1/4

6.5 Return link: automatic CSI balancing

6.5.1 Introduction

If the return link is operating in closed loop mode, then the previously described techniques for the forward link can be applied without any change. The closed loop operation makes the CSI delay be equal to one round-trip time and, therefore, strategies based on statistical CSI might offer a good performance.

In some cases, open loop operation can offer advantages to closed loop, since the CSI is timely (shadowing events occur next to the mobile terminal). The CSI accuracy, however, is reduced due to uncorrelated multipath fading in return and forward link. Thus, scenarios with strong multipath and low speed seem more favorable for closed loop operation, whereas open loop would be a better choice for high speed and strong line of sight scenarios. Switching between open and closed loop modes depending on the scenario is not a trivial task, and different methods could be proposed. One possible option could be the use of a look-up table (LUT), built from simulation results, and containing the optimum operating mode for different parameters (speed, multipath, shadowing coherence distance, etc.). This option, however, requires the receiver to estimate these parameters in an accurate way, and is not robust against practical impairments not considered in the simulation step.

In this section we present a method to automatically balance the open loop and closed loop CSI. This method relies only on the ACK/NAK interchange between transmitter and receiver, and is able to exploit both CSI values at the same time. An automatic backoff margin is also

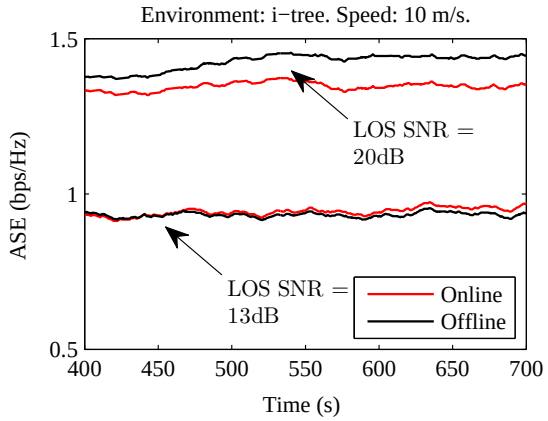


Figure 6.12: Evolution of the average spectral efficiency with online adaptation, and comparison with the offline approach.

designed as a byproduct of the proposed method.

6.5.2 Problem statement

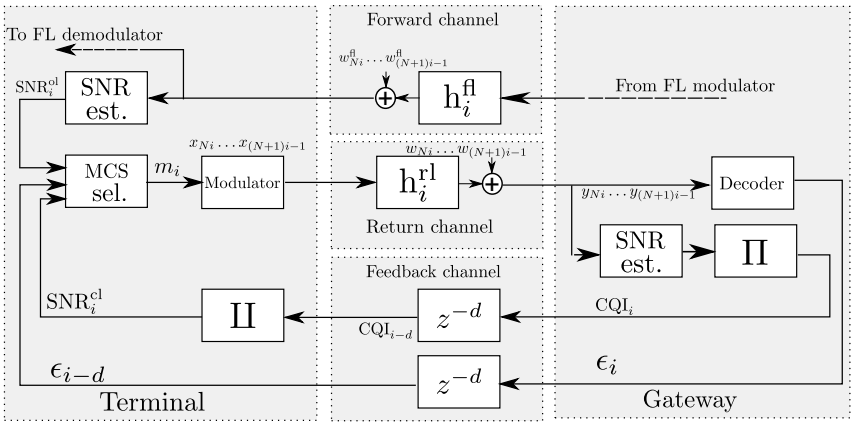


Figure 6.13: Diagram of the information interchange and link adaptation procedure.

Let us denote by $m_i \in \{1, \ldots, M\}$ the index of the MCS selected in time instant i. We also assume that the error probability $\mathbb{P}[\epsilon_k = 1]$ depends only on m_k and $\mathbf{h}_k$, and thus is independent of the transmitted message. This assumption is compatible with the proposed error function (6.8), but includes other error functions that characterize coded transmission. We denote the probability of error of the j-th MCS under channel $\mathbf{h}$ as $E(j, \mathbf{h})$, and assume that $E(j+1, \mathbf{h}) \geq E(j, \mathbf{h})$, i.e., higher transmission rates imply higher error probabilities.

We consider two types of feedback: first, the receiver acknowledges the correct decoding of

the $(i-d)$-th codeword, so the values $\epsilon_0, \ldots, \epsilon_{i-d}$ are available at the transmitter at time instant i, with d the feedback delay; second, the receiver estimates the channel quality in time instant $i-d$ and includes a channel quality indicator (CQI) in the feedback message. Although there are different ways to calculate the CQI, we consider the index of the highest MCS supported by channel $\mathbf{h}_i^{\mathrm{rl}}$, i.e., $\mathrm{CQI}_i = \arg\max_j \left[j | E\left(j, \mathbf{h}_i^{\mathrm{rl}}\right) \leq p_{\mathrm{CQI}} \right]$. If no MCS is supported, then $\mathrm{CQI}_i = 1$. p_{CQI} is a value that determines a limit on the error probability. In the case of using Gaussian coding with a sufficiently large code size, for example, p_{CQI} can be set to a number as close to 0 as desired. With the proposed error function (6.8), the error probability can be set to zero if the channel is sufficiently favorable, so we assume $p_{\mathrm{CQI}} = 0$.

The CQI value can be obtained from the effective SNR value by means of a function Π. The function $\Pi(\mathrm{snr})$ is an LUT that maps SNR intervals to values. Throughout this section, we assume that the SNR values are in decibels for convenience. For M MCS values, the function Π can be parametrized by $M-1$ thresholds t_i, $i = 1, \ldots, M-1$, such that $\Pi(\mathrm{snr}) = j \iff t_{j-1} \leq \mathrm{snr} < t_j$, where the higher and lower thresholds are defined as $t_0 = -\infty$ and $t_M = +\infty$. The function Π is usually referred to as *inner loop* for link adaptation, and we assume the thresholds to be the γ_{th} values for each MCS[4]. If perfect SNR information was available at the transmitter, the link adaptation procedure would be trivial for a calibrated receiver. In practice, however, this information may not be available, so a correction has to be made to the estimated SNR value.

We define the function $\amalg$ as a function that maps MCS indexes to SNR values, such that $\Pi\left(\amalg\left(j\right)\right) = j$. For example, a function that maps every MCS index to its SNR threshold γ_{th} meets this requirement, and is the one we will use throughout this section. The objective of $\amalg$ is to map CQI to SNR values. We also define $\mathrm{snr}_i^{\mathrm{cl}} \triangleq \amalg\left(\mathrm{CQI}_{i-d}\right)$ as the SNR that the transmit side gets to know about the received quality d frames earlier in closed loop mode. Note that $\mathrm{snr}_i^{\mathrm{cl}}$ denotes an estimate of the effective SNR performed in time instant $i-d$, but used at the terminal in instant i. In absence of feedback delay ($d = 0$), channel estimation error and other impairments, the optimal MCS selection would be $m_i = \Pi\left(\amalg\left(\mathrm{CQI}_i\right)\right) = \mathrm{CQI}_i$. A common practice to accommodate these impairments is the application of margins to the received CQI value, so

$$m_i = \Pi\left(\mathrm{snr}_i^{\mathrm{cl}} + c^{\mathrm{cl}}\right) \tag{6.43}$$

with c^{cl} the SNR margin in dB. A possible approach to select c is by means of an LUT that stores values of c for different scenarios, where parameters like channel distribution, Doppler, detection complexity, etc. have to be taken into account. This approach has some drawbacks that limit its application to practical settings. First, filling the LUT requires running exhaustive simulations

[4]Note that the γ_{th} for the first MCS is not used as a threshold.

under many different settings to be applicable to practical scenarios, and its behavior will be unpredictable under conditions that differ from the stored ones. Second, the receiver has to estimate the required parameters, which can be computationally expensive, and errors in the estimation of these parameters might lead to unexpected behavior. Thus, an adaptive method to adjust c^{cl} is required in many cases. An adaptation of c^{cl} based on ACK/NAK reception was proposed in [118], and applied to the satellite scenario in [36].

On top of the feedback information, the terminal is also observing the channel in the forward link. If the duplexing scheme was TDD, the terminal might gain access to timely and accurate CSI just by measuring the forward link channel. This sort of CSI is called *open loop CSI*. In our setting, duplexing is performed by means of frequency separation, so this assumption does not hold. Under our model, however, there is some degree of correlation between the forward and return link, as the LOS component is the same for both links. Therefore, depending on the scenario, the open loop CSI would be more or less accurate. Let us define $\mathrm{snr}^{\mathrm{ol},i}$ as the most recent SNR estimation on the forward link. We assume that this SNR estimation is perfect, and equal to the effective SNR of the previous codeword, i.e., $\mathrm{snr}^{\mathrm{ol},i} = \gamma^{\mathrm{fl}}_{\mathrm{eff},i-1}$. This assumption does not affect the design of the method, and is made for the sole purpose of simplifying the simulations. We might think of performing a similar adaptation as in the closed loop case (6.43)

$$m_i = \Pi\left(\mathrm{snr}_i^{\mathrm{ol}} + c^{\mathrm{ol}}\right). \tag{6.44}$$

One again, the margin c^{ol} should be obtained adaptively or by means of an LUT. In Figure 6.13 we show a diagram containing the main variables of the system model.

A further question is how to determine the scenarios where (6.44) or (6.43) are more appropriate to be used. It is expected that in scenarios with relatively low speed or strong multipath the closed loop approach would perform better, while strong LOS and high speed scenarios are more suitable for the open loop one. A possible approach is to perform parameter estimation (speed, multipath, etc.) and obtain the optimum strategy from an LUT, which had to be previously filled according to exhaustive simulation results. In the following section we present an adaptive approach that does not require to perform this parameter estimation, and solely relies on the feedback of CQI and ACK/NAK, as well as in the open loop SNR estimation.

6.5.3 Adaptive CSI balancing

A key observation in (6.43)-(6.44) is that they can be jointly described by

$$m_i = \Pi\left(\xi^{\mathrm{ol}}\mathrm{snr}_i^{\mathrm{ol}} + \xi^{\mathrm{cl}}\mathrm{snr}_i^{\mathrm{cl}} + c\right). \tag{6.45}$$

If we set $\xi^{\text{ol}} = 0$, $\xi^{\text{cl}} = 1$ we arrive to (6.43), and $\xi^{\text{cl}} = 0$, $\xi^{\text{ol}} = 1$ leads to (6.44). Note that (6.45) includes any affine combination of snr^{ol} and snr^{cl}, so it generalizes the open loop and closed loop strategies. We now derive an adaptation method for general values of ξ^{cl} and ξ^{ol}, and in Section 6.5.4 we introduce an alternative formulation where their value is constrained to sum one, i.e., $\xi^{\text{cl}} + \xi^{\text{ol}} = 1$.

For simplicity, we denote $\mathbf{snr}_i \triangleq \left[\text{snr}_i^{\text{cl}}\ \text{snr}_i^{\text{ol}}\right]^T$ and $\boldsymbol{\xi} \triangleq \left[\xi^{\text{cl}}\ \xi^{\text{ol}}\right]^T$; the derivations from now on could be generalized for vectors $\mathbf{snr}$ and $\boldsymbol{\xi}$ of any size, so we could include channel prediction in this framework, for example.

Following a similar approach as [119], we state the problem of finding the margin c and SNR balancing weights $\boldsymbol{\xi}$ such that the PER converges to a fixed *target PER* p_0. The desired values can be obtained as the solution to the following optimization problem

$$\min_{c,\boldsymbol{\xi}} J(c, \boldsymbol{\xi}) = |\mathbb{E}[\epsilon] - p_0|^2. \tag{6.46}$$

Note that (6.46) does not have any optimality properties in terms of throughput, but just sets the mean packet error rate to the desired value p_0. In practice, nevertheless, it is expected that high SNR values will lead to the use of higher rate MCS to meet the target PER p_0, so the throughput is implicitly increased.

Problem (6.46) can be solved by performing a gradient descent on $J(c, \boldsymbol{\xi})$. The gradient of $J(c, \boldsymbol{\xi})$ can be worked out as

$$\nabla J(c, \boldsymbol{\xi}) = 2\left(\mathbb{E}[\epsilon] - p_0\right) \nabla \mathbb{E}[\epsilon]. \tag{6.47}$$

A gradient descent iteration reads as

$$\begin{bmatrix} c_{i+1} \\ \boldsymbol{\xi}_{i+1} \end{bmatrix} = \begin{bmatrix} c_i \\ \boldsymbol{\xi}_i \end{bmatrix} - \mu_i \cdot \nabla\, J(c, \boldsymbol{\xi})|_{c_i, \boldsymbol{\xi}_i}. \tag{6.48}$$

Obtaining a numerical expression for the gradient $J(c, \boldsymbol{\xi})|_{c_i, \boldsymbol{\xi}_i}$ is not possible: the expectation of ϵ depends on the PDF of the channel, which we assume unknown at the transmitter. On top of this, the PDF of the channel might change over time. Instead, we propose a stochastic gradient approach, where the expectations are substituted by instantaneous observations.

Let us define

$$\Omega \triangleq \boldsymbol{\xi}^T \mathbf{snr} + c, \tag{6.49}$$

as the indicator SNR with which the MCS m_i is selected in (6.45). Ω is a function of c and $\boldsymbol{\xi}$

whose gradient is trivial to compute, so that applying the chain rule of differentiation in (6.47) we arrive at

$$\begin{aligned}\nabla J(c,\boldsymbol{\xi}) &= 2\left(\mathbb{E}\left[\epsilon\right]-p_0\right)\nabla\mathbb{E}\left[\epsilon\right]\\ &= 2\left(\mathbb{E}\left[\epsilon\right]-p_0\right)\mathbb{E}\left[\frac{\partial\epsilon}{\partial\Omega}\nabla\Omega\right]\\ &= 2\mathbb{E}\left[\frac{\partial\epsilon}{\partial\Omega}\right]\left(\mathbb{E}\left[\epsilon\right]-p_0\right)\begin{bmatrix}1\\ \mathbf{snr}\end{bmatrix}\end{aligned} \tag{6.50}$$

Following the stochastic gradient approach, we substitute $\mathbb{E}\left[\epsilon\right]$ by the instantaneous value ϵ_{i-d}; also, and since $2\partial\epsilon/\partial\Omega$ is positive[5], we embed $2\mathbb{E}\left[\partial\epsilon/\partial\Omega\right]$ into the positive adaptation constant μ_i. The resulting expression for the update of c and $\boldsymbol{\xi}$ reads as (see Figure 6.14)

$$\begin{bmatrix}c_{i+1}\\ \boldsymbol{\xi}_{i+1}\end{bmatrix} = \begin{bmatrix}c_i\\ \boldsymbol{\xi}_i\end{bmatrix} - \mu\left(\epsilon_{i-d}-p_0\right)\begin{bmatrix}1\\ \mathbf{snr}_{i-d}\end{bmatrix}. \tag{6.51}$$

where we removed the dependence of μ with time, so we are using a constant stepsize.

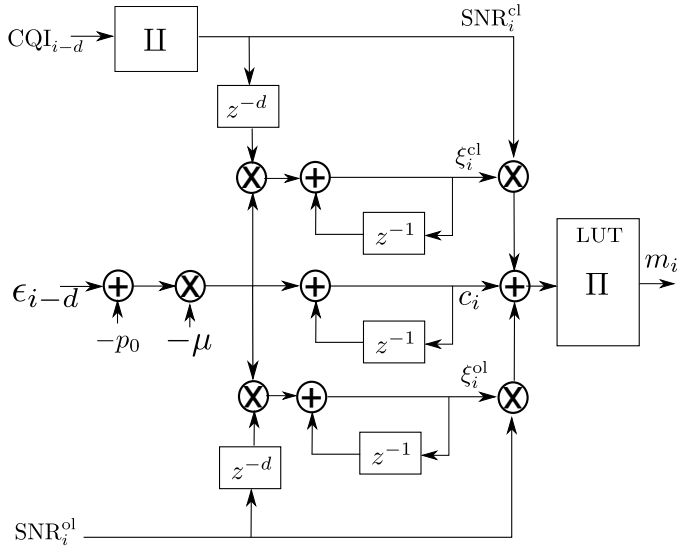


Figure 6.14: Diagram of the adaptation process

Note that in time instant i the last received feedback is the one corresponding to the information transmitted in time $i-d$. The SNR values used for adaptation in (6.51) have to

[5]This can be proved by writing probability of error, averaged over all channel states, in integral form, and using the assumption that for a channel state, the probability of error is higher for higher MCS. Higher values of Ω lead to higher MCS values, which increase the probability of error.

be the ones used for the MCS selection of the packet the ACK/NAK is referred to. If the transmitter knows the delay introduced by the channel, a delay of z^{-d} has to be introduced in the adaptation algorithm, as shown in Figure 6.14. If the delay value is not known or is variable (in case of ACK/NAK grouping, for example), then the transmitter should store in memory the SNR values used for adaptation of every packet, indexed by a packet ID. When the ACK/NAK for a packet ID is received, the parameter update is performed by recovering the corresponding SNR values from memory. Note that the closed loop SNR value snr_i^{cl} used for MCS selection is the one generated by CQI_{i-d}, but the one used for the adaptation of ξ_i^{cl} is $\text{snr}_{i-d}^{\text{cl}}$, which is generated by CQI_{i-2d}.

Remark 6.1. Note that the adaptive margin algorithm proposed in [36, 118] is equivalent to the adaptation of c in (6.51). This adaptation is also the one described in [119]. More precisely, the algorithm for update described in [36, 118] is

$$c_{i+1} = \begin{cases} c_i + \delta_{\text{up}} & \text{if } \epsilon_{i-d} = 0 \\ c_i - \delta_{\text{down}} & \text{if } \epsilon_{i-d} = 1 \end{cases} \tag{6.52}$$

with δ_{up} and δ_{down} values such that[6]

$$\delta_{\text{down}} = \delta_{\text{up}} \frac{p_0}{1 - p_0}. \tag{6.53}$$

It can be seen that (6.51) and (6.52) describe the same adaptation, provided that $\delta_{\text{up}} = \mu p_0$ and $\delta_{\text{down}} = \mu (1 - p_0)$.

6.5.4 Convergence enhancements

We observed that the adaptation method described by (6.51) offers a noisy behavior in convergence, thus needing small values of μ. This affects the convergence speed of the algorithm, dramatically decreasing it. Note that (6.51) resembles a least mean squares (LMS) adaptation with input $[1 \ \mathbf{snr}_{i-d}]^T$ and error $\epsilon_{i-d} - p_0$. Normalized LMS (NLMS) [120] is well known to outperform LMS in convergence speed. If the step-size is normalized in (6.51), the NLMS-like version reads as

$$\begin{bmatrix} c_{i+1} \\ \boldsymbol{\xi}_{i+1} \end{bmatrix} = \begin{bmatrix} c_i \\ \boldsymbol{\xi}_i \end{bmatrix} - \frac{\mu}{1 + \|\mathbf{snr}_{i-d}\|^2} (\epsilon_{i-d} - \tilde{p}_{0,i}) \begin{bmatrix} 1 \\ \mathbf{snr}_{i-d} \end{bmatrix}. \tag{6.54}$$

Note that we substituted p_0 by $\tilde{p}_{0,i}$. In general, (6.54) does not converge to a PER of $\hat{p}_{0,i}$,

[6] In [36] the steps are selected to meet $\delta_{\text{down}} = \delta_{\text{up}} p_0$ instead of (6.53). Both formulations, however, are equivalent for low values of p_0.

since the first component of a stationary point meets

$$\mathbb{E}\left[\frac{\epsilon_i - \tilde{p}_{0,i}}{1 + \|\mathbf{snr}_{i-d}\|^2}\right] = 0 \tag{6.55}$$

which does no necessarily imply $\mathbb{E}[\epsilon_i] = \tilde{p}_{0,i}$. It is expected, however, that an appropriate choice of $\tilde{p}_{0,i}$ (not equal to p_0) led to a PER of p_0. We propose to adjust $\tilde{p}_{0,i}$ following a recursion

$$\tilde{p}_{0,i+1} = \tilde{p}_{0,i} - \lambda\,(\epsilon_{i-d} - p_0) \tag{6.56}$$

It is clear that $\mathbb{E}[\epsilon_i] = p_0$ is a stationary point of (6.56), thus leading to the desired PER. A block diagram of this NLMS adaptation is shown in Figure 6.15.

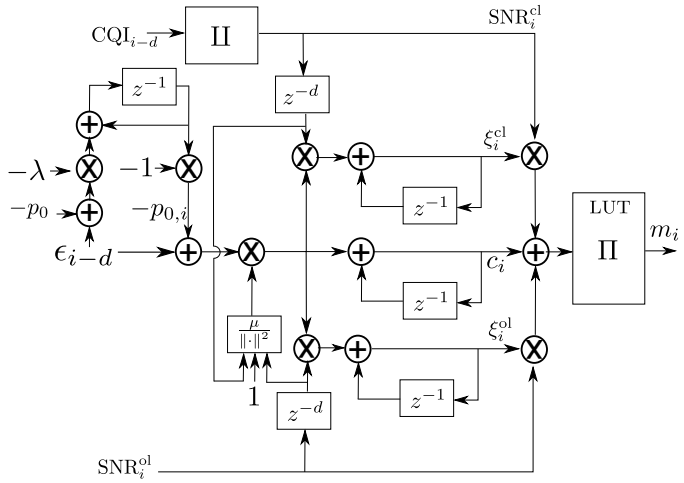


Figure 6.15: Diagram of NLMS adaptation

It has been also observed that the NLMS adaptation offers a good convergence performance for the terms ξ, but not for the margin c. This might be caused because the corrections to this term are smaller in absolute value than those for ξ because $\mathrm{snr}^{\mathrm{ol}}$ and $\mathrm{snr}^{\mathrm{cl}}$ are usually bigger than 1. To overcome this problem, we propose an alternative formulation that increases the speed of convergence of c

$$\begin{bmatrix} c_{i+1} \\ \boldsymbol{\xi}_{i+1} \end{bmatrix} = \begin{bmatrix} c_i \\ \boldsymbol{\xi}_i \end{bmatrix} - \frac{\mu}{\theta^2 + \|\mathbf{snr}_{i-d}\|^2}\,(\epsilon_{i-d} - \tilde{p}_{0,i}) \begin{bmatrix} \theta \\ \mathbf{snr}_{i-d} \end{bmatrix}. \tag{6.57}$$

We also performed experiments with only one weight ξ instead of two. In this case, we defined the MCS selection rule as

$$m_i = \Pi\left(\left(1-\xi^{\mathrm{cl}}\right)\mathrm{snr}_i^{\mathrm{ol}} + \xi^{\mathrm{cl}}\mathrm{snr}_i^{\mathrm{cl}} + c\right), \tag{6.58}$$

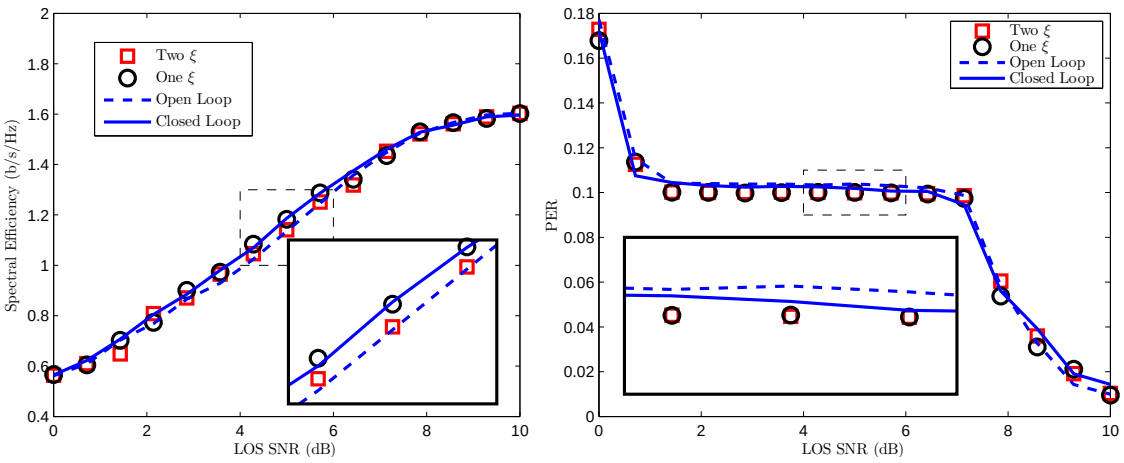
and the corresponding adaptation rule as

$$\begin{bmatrix} c_{i+1} \\ \xi_{i+1}^{\mathrm{cl}} \end{bmatrix} = \begin{bmatrix} c_i \\ \xi_i^{\mathrm{cl}} \end{bmatrix} - \frac{\mu}{\theta^2 + \left(\mathrm{snr}_{i-d}^{\mathrm{cl}} - \mathrm{snr}_{i-d}^{\mathrm{ol}}\right)^2} \times \left(\epsilon_{i-d} - \tilde{p}_{0,i}\right) \begin{bmatrix} \theta \\ \mathrm{snr}_{i-d}^{\mathrm{cl}} - \mathrm{snr}_{i-d}^{\mathrm{ol}} \end{bmatrix}. \tag{6.59}$$

and $\tilde{p}_{o,i}$ following the recursion in (6.56).

The convergence properties of the methods described in this section are object of future research. Nevertheless, their convergence to the desired PER value has been empirically observed, as described the next section.

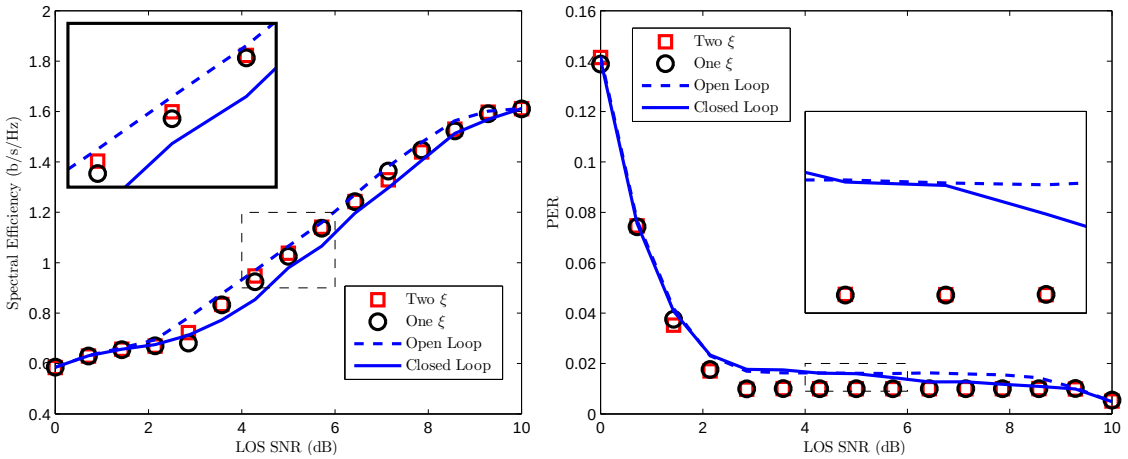
6.5.5 Simulation results



(a) Average Spectral Efficiency. (b) Frame error rate.

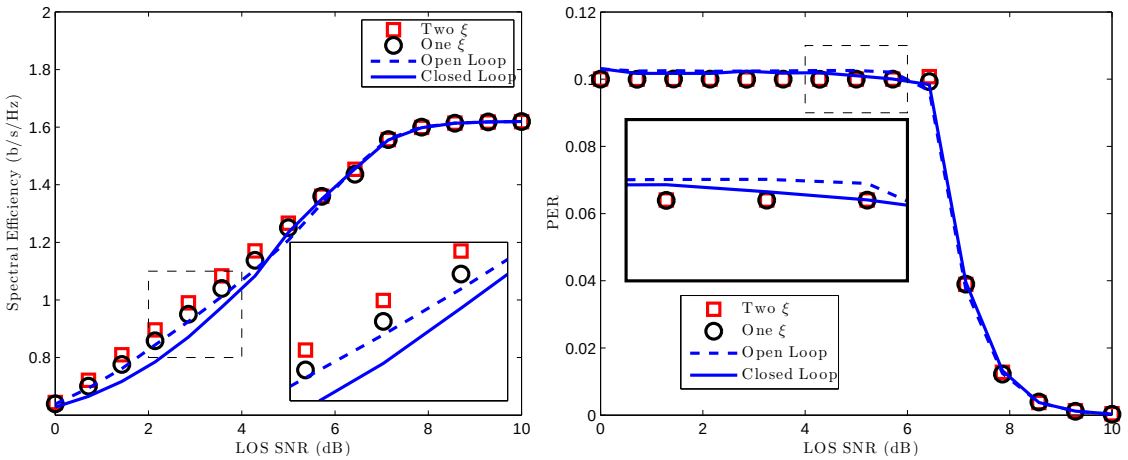
Figure 6.16: PER and throughput for different methods in intermediate tree shadowed environment, state 1, 0.3 m/s, $p_0 = 0.1$

We performed simulations of the proposed methods, given by equations (6.59) and (6.57), and compare them with open loop and closed loop with automatic margin adaptation, given by (6.52). The adaptation was performed with $\theta = 10$, $\lambda = 10^{-3}$ and $\mu = 1$ for the case of the proposed methods, and $\mu = 10^{-2}$ for the open and closed loop cases, which correspond to $\delta_{\mathrm{up}} = 0.001$ and $\delta_{\mathrm{down}} = 0.009$ for a target PER of 10^{-1}. The parameters were initialized as



(a) Average Spectral Efficiency. (b) Frame error rate.

Figure 6.17: PER and throughput for different methods in intermediate tree shadowed environment, state 1, 3 m/s, $p_0 = 0.01$



(a) Average Spectral Efficiency. (b) Frame error rate.

Figure 6.18: PER and throughput for different methods in intermediate tree shadowed environment, state 1, 15 m/s, $p_0 = 0.1$

$c_0 = 0$, $\xi_0^{\text{cl}} = 0.5$, and $\xi_0^{\text{ol}} = 0.5$. We set the feedback delay to $d = 5$ codewords to model the round trip time in a GEO satellite.

We performed simulations over a Loo channel with the parameters for intermediate tree shadowed environment, state 1. Simulations were also carried out for other settings, with similar results observed. We simulated three different terminal speeds, corresponding to 0.3 m/s, 3 m/s and 15 m/s. The target PER was set to $p_0 = 0.1$ in the first and last speeds, and to $p_0 = 0.01$ for the 3 m/s case. Average spectral efficiency and throughput results were averaged over the transmission of $6 \cdot 10^4$ packets. Results are shown in Figures 6.16-6.18. Spectral efficiency is defined as $\frac{1}{N}\sum_{i=1}^{N} \epsilon_i r_{m_i}$, with r_j the rate of the j-th MCS.

We see that the lower speed scenario shows a better performance of closed loop with respect to open loop, and that the adaptation with two ξ performs worse than closed loop, and similarly to open loop. Adaptation with one ξ offers approximately the same performance as closed loop. For higher speeds, open loop performs better than closed loop, and so do the proposed methods. For the $15m/s$ case, adaptation with one and two ξ outperform both open loop and closed loop adaptation. It is also noticeable that the proposed methods are more accurate in converging to the desired PER p_0 than the open and closed loop ones. A PER p_0 might not be achievable in very high or very low SNR scenarios. In these cases, the PER converges to the minimum or maximum possible PER values, respectively. The method with one ξ seems to be more robust than the one with two ξ, and offers a better or at least the same performance than closed loop link adaptation in all the cases, even beating the open loop adaptation for high speeds.

6.5.6 Implementation aspects

In this chapter we made some simplifications to make the link adaptation problem more tractable. In the following, we make comments about some possible implementation aspects of the proposed algorithm

- **CQI feedback** Throughout the section, it was assumed that a CQI value was fed back for every packet. In modern communication standards this is not usually the case, so some packets would have a more outdated CQI than others. Assume that a CQI value is transmitted every K packets. Applying the same weight ξ^{cl} in all K time instants could be suboptimal, since the first packet in every period has much more precise CSI than the last one. A possible approach to overcome this problem is to have a different adaptation loop for each of the possible CQI delays.
- **Use of effective SNR** Although in this section we used effective SNR for convenience, the proposed method is expected to work with other CSI metrics, such as average SNR or

RSSI. This can be convenient in case the effective SNR calculation could not be performed by the receiver because of computational complexity.

- **Interference in return link** We neglected the possible interference in the return link, which is impossible to estimate from open loop observations [109]. In this case, the proposed method is expected to converge to weights ξ that reduce the impact of errors in open loop CSI.

- **Estimation errors and uncalibrated receivers** We assumed in the simulations that CQI and SNR estimates in the forward link were perfect, as well as the knowledge of the SNR thresholds for decoding. In practice, there might be some non-negligible errors in the SNR estimation, and the performance of a receiver might not be known a priori. In these cases, the proposed adaptive method should be able to adjust the parameters to meet the PER constraint, although possibly reducing the throughput with respect to the ideal case.

- **Divergence of parameters** In cases of very high or very low SNR, the adaptation parameters ξ and c diverge, as it is not possible to converge to the desired PER p_0. A threshold should be included in the adaptation to prevent this behavior.

6.6 Conclusions

In this chapter we proposed different link adaptation techniques for mobile satellite channels. We focused first on the return link, where timely CSI is difficult to obtain, and proposed adaptation techniques based only on statistical CSI. The first technique uses multi-layer coding to increase the spectral efficiency while controlling the outage of the system, and relies on statistical information on the channel to perform adaptation. The second technique uses different MCS in different retransmission index, and the needed statistical CSI can be acquired online by observing the ACK/NAK interchange.

The return link is studied in a different way, and open loop CSI is proposed to be used. We focus on the problematic of switching between open loop and closed loop in an automatic manner, just from the observation of ACK/NAK feeds.

The content of this chapter is the result of a collaboration with Jesus Arnau and Prof. Carlos Mosquera, and was partially published in ICSSC 2012 and Globecom 2013 [121,122], and submitted to ASMS 2014 [123].

Chapter 7

Learning-Based Link Adaptation for Limited Feedback Multiuser MIMO-OFDM

Contents

7.1 Introduction

Dynamically adapting the transmitter in response to changing channel conditions is key to achieving both throughput and reliability on wireless communication links. Reconfiguring the link requires adjusting several transmission parameters: the modulation and coding scheme (MCS), the multiple input multiple output (MIMO) precoding matrices in multiple antenna systems, the spatial mode (number of spatial data streams for each user), and the assignment of transmit resources for the different users, among other parameters. In this chapter we consider a learning-based approach to link adaptation in multiuser (MU) MIMO orthogonal frequency division multiplexing (OFDM) systems.

We present a data-driven link adaptation method that includes user selection, mode selection, precoding, MCS selection, and limited feedback. The main objective of this chapter is to give insight into the effect of limited feedback in MU-MIMO systems with link adaptation. To do so, our solution assumes perfect channel state information (CSI) at the receivers, and different degrees of CSI at the transmitter. With this information, and taking into account the degradation due to CSI inaccuracy, the transmitter is able to perform link adaptation. Our work is different from previous approaches in several ways. The focus of [124, 125] is to study fairness in a multiuser setting, with limited feedback information not taken into account, and MCS selection is performed by means of effective signal to noise ratio (SNR). In [126], only a single spatial stream is allowed per user, and the effective SNR approach is followed for MCS selection. In [127], adaptation is performed following a data-driven approach, and limited feedback is taken into account, but the communication scenario is multicast (i.e., there is only a common message for all receivers). This chapter studies the broadcast scenario, takes into account limited feedback, includes the possibility of transmitting more than one spatial stream to each user, and performs MCS selection following a data driven approach.

We focus on the multiuser capabilities of IEEE 802.11ac [62] to develop our link adaptation method. We consider this standard due to its novelty and the challenging constraint of not allowing user allocation among subcarriers, like in LTE or WIMAX. The adaptation is performed at the access point (AP), thus requiring a minimum communication overhead with the user stations (STA) to obtain CSI. This AP-centric approach is specially suitable for link adaptation in scenarios with cheap STAs, as they might not implement some of the optional features in IEEE 802.11ac to perform MCS selection.

The main contributions of this chapter are summarized as follows.

- We derive a closed form approximation of the interuser interference leakage due to the use of block diagonalization precoding with zero forcing (ZF) receivers. This approximation,

unlike previous approaches, is not based on a random vector quantization analysis. We exploit the particular codebook structure used in IEEE 802.11ac, which is induced by the use of Givens decompositions, to derive our approximation. This approximation is shown to be very accurate for different feedback rates. We restrict our analysis to ZF receivers as a *worst-case* scenario, so our interference estimation will be conservative if the receiver employs more advanced algorithms, such as minimum mean squared error (MMSE) or maximum likelihood detectors.

- We apply previous data-driven approaches [47–51], which are limited to the single user, to a multiuser setting with a variable number of spatial streams per user. We show that the MCS selection accuracy of the data-driven approach outperforms unidimensional metrics such as average SNR or effective SNR. The machine learning classifier is shown to be robust to changes in the statistical distribution of the channel, i.e., is able to correctly perform MCS selection even when the training is done with channel samples taken from a different statistical distribution.

- We use a *greedy* algorithm that performs mode and user selection, inspired on previous work like [128–130]. This algorithm exploits information on the feedback rate, the SNR regime and the number of users to perform user and mode selection. The greedy algorithm sequentially adds spatial layers from the different users until the maximum number of spatial streams is reached, or the throughput is no longer increased. This algorithm, with a complexity that is linear in the maximum number of spatial streams and in the number of users, allows solving the user and mode selection problem without resorting to an exhaustive search.

The remaining of the chapter is organized as follows: Section 7.2 describes the mechanisms in IEEE 802.11ac that enable MU-MIMO operation; Section 7.3 presents the system model; Section 7.4 introduces the problem statement. The following sections present the main contributions of the chapter: Section 7.5 describes the MU-MIMO precoding problem, and presents an approximation for the interference leakage due to limited feedback precoding; Section 7.6 describes the data-driven MCS selection; Section 7.7 presents a greedy algorithm for user and mode selection; Section 7.8 presents the simulation results; Section 7.9 concludes the chapter.

7.2 MU-MIMO in IEEE 802.11ac

IEEE 802.11ac is an emerging wireless standard that supports MU-MIMO, single user MIMO, OFDM modulation, and thus needs sophisticated link adaptation algorithms. In this section we

summarize some of the mechanisms that enable MU-MIMO link adaptation in IEEE 802.11ac, and explain how they relate to our system assumptions. Given the functionality required in this chapter, we divide the MU-MIMO operation into three different tasks: *CSI acquisition*, *link adaptation* and *MU-MIMO transmission*.

7.2.1 CSI acquisition

Despite being a time division duplexing system, channel reciprocity is not natively supported by IEEE 802.11ac to obtain CSI at the transmitter (CSIT). IEEE 802.11n [23] describes a procedure to obtain CSI at the AP by the transmission of a training sequence in the uplink [23, Sec. 9.29.2.2] and a calibration procedure to identify the differences between the transmit and receive chains at the STA. These mechanisms are not present in IEEE 802.11ac, so acquisition of CSI based on channel reciprocity cannot rely on any information exchange with the receivers. Although there is some research in estimating the channel from the normal packet exchange in IEEE 802.11ac [131, Sec. 2.3.3], we only consider CSI acquisition at the AP by the mechanism described in the standard. This mechanism comprises sending a sounding sequence in the downlink and feeding back the estimated channel to the AP. In the following, we describe these two tasks.

Sounding

Channel sounding is initiated by an AP by transmitting a very high throughput (VHT - name of the new transmission modes in IEEE 802.11ac) null data packet (NDP) announcement, which is a control frame that includes the identifiers of the set of users which are potentially going to be polled for feedback. Together with the identifiers, the frame carries information on the requested feedback (there are two different types, single user - SU or multiuser - MU) and the number of columns (spatial streams, only for the MU case) in the requested feedback matrix. This frame is described in more detail in [62, Sec. 8.3.1.19].

After the NDP announcement, the AP transmits an NDP, which is used by the receivers to estimate the MIMO channel. After estimating the channel, the first user in the list of the NDP announcement sends feedback information, and the remaining users (if any) transmit their CSI by responding to subsequent beamforming (BF) report polls. This operation is depicted in Figure 7.1.

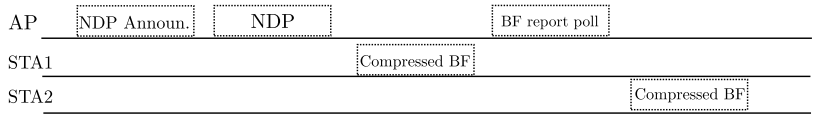


Figure 7.1: Typical message interchange for sounding and feedback.

Quantization and Feedback

The CSI obtained after the training phase is quantized prior to the transmission on the feedback channel. The most relevant parameters for the MU-MIMO operation are the preferred beamforming matrices, which contain the right eigenvectors associated to the largest singular values of the channel matrix. These matrices are represented using a Givens decomposition. The angles resulting from this decomposition are the parameters that are fed back to the AP. The feedback message contains the beamforming matrices plus some additional information. This information depends on the selected feedback mode (MU or SU) indicated by the AP in the feedback request message [62, Sec. 8.4.1.48].

- **Grouping**: Determines if the carriers are grouped in the feedback message, or if feedback is provided for every carrier. We assume no grouping.
- **Codebook information**: Determines the number of bits for the codebook entries. This parameter is selected by the STA. In this chapter, we treat different feedback rates, but the problem of selecting an appropriate feedback rate is out of the scope of this work.
- **Beamforming matrices and SNR information**: Includes preferred beamforming matrix, average SNR for each spatial stream, and SNR for each carrier (or group of carriers) if feedback is in MU mode.

7.2.2 Link adaptation

The objective of link adaptation is to perform user selection, mode selection, MCS selection and MIMO precoding design. This can be done in a centralized way at the transmitter, taking as input the limited feedback described in the previous section, or in a distributed way by requesting an STA to select the preferred mode and MCS for a given MIMO configuration. We follow a centralized approach, since our solution involves joint scheduling and MCS selection. The distributed MCS selection mechanism is an optional feature in IEEE 802.11ac, thus it cannot be expected to be implemented in all devices. The distributed MCS feedback can be used to refine the results when implementing the proposed link adaptation algorithm.

The mechanism for communicating MCS and mode is described in [62, Sec. 9.28.3], and is an optional feature of IEEE 802.11ac. The information is contained in the VHT variant HT Control field, present in the MAC header of all data packets. Roughly speaking, the AP can send a data packet with the bit MRQ (MCS request) set to 1, so the STA *piggybacks* the preferred MCS and number of spatial streams (NSS) in the following data packet. The MCS and NSS can be fed back in an unsolicited way, without being polled by the AP. The content of the link adaptation message is

- Preferred number of spatial streams.
- Preferred MCS.
- Preferred bandwidth.
- SNR, averaged over all carriers and spatial streams.

The way an STA selects the values of the feedback fields is implementation dependent.

7.2.3 MU-MIMO transmission

Once the users, MCS, mode and precoders are selected, MU-MIMO transmission takes place. MU-MIMO in IEEE 802.11ac is *non-transparent*, meaning that the STAs are aware they will be jointly scheduled with other STAs. This allows a given STA to use the training sequences not only to estimate its own MIMO channel, but also the interference [62, Sec. 22.3.11.4], which allows performing some advanced tasks such as the modification of the receive precoders, or the correct calculation of the Viterbi weights for decoding.

7.3 System model

Consider the downlink of an N-carrier OFDM wireless network where an AP equipped with N_{tx} antennas communicates with U STAs, $\mathcal{U} = \{1, \ldots, U\}$ where the u-th user has $N_{\text{rx},u}$ receive antennas. At a given time instant, the transmitter conveys information to a subset of the users $\mathcal{T} \subseteq \mathcal{U}$. In a given time slot all users use all subcarriers.

We assume that the transmitter employs linear precoding, and the receivers use linear equalizers. A single modulation $\mathcal{M}_u = \left\{m_1, \ldots, m_{|\mathcal{M}_u|}\right\} \subset \mathbb{C}$ is selected for the u-th user, constant over all carriers and spatial streams[1]. For a given carrier n, $\mathbf{s}_u[n] \in \mathcal{M}_u^{L_u}$ is the L_u

[1] Although in IEEE 802.11n the use of different modulations in each spatial stream was allowed, it was apparently not implemented in most commercial devices, and finally discarded for 802.11ac.

spatial streams modulated signal containing the information for the u-th user, $\mathbf{F}_u[n] \in \mathbb{C}^{N_{\text{tx}} \times L_u}$ is the transmit precoding matrix for the u-th user, $\mathbf{H}_u[n] \in \mathbb{C}^{N_{\text{rx},u} \times N_{\text{tx}}}$ is the flat-fading MIMO channel from the transmitter to the u-th receiver, $\mathbf{B}_u[n] \in \mathbb{C}^{L_u \times N_{\text{rx},u}}$ is the interference removal matrix, and $\mathbf{G}_u[n] \in \mathbb{C}^{L_u \times L_u}$ is the linear equalizer applied at the u-th receiver. We divide the receive processing into two different matrices for simplicity in the treatment of the multiuser precoding. The objective of $\mathbf{B}_u[n]$ is to reject the interuser interference, while the equalizer $\mathbf{G}_u[n]$ removes the intrauser interference. To limit the transmit power per carrier, we define the power normalization factor $P[n] \triangleq \sum_{u \in \mathcal{T}} \text{tr}\left(\mathbf{F}_u[n]\mathbf{F}_u^*[n]\right)$. We use $\mathbf{n}_u[n] \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ to denote the received noise vector at the u-th receiver. Given these definitions, the post-processed signal at the u-th receiver $\mathbf{y}_u[n] \in \mathbb{C}^{L_u}$ is

$$\mathbf{y}_u[n] = \mathbf{G}_u[n]\mathbf{B}_u[n]\mathbf{H}_u[n] \sum_{i \in \mathcal{T}} \frac{1}{\sqrt{P[n]}} \mathbf{F}_i[n]\mathbf{s}_i[n] + \mathbf{G}_u[n]\mathbf{B}_u[n]\mathbf{n}_u[n] \tag{7.1}$$

with $\mathbb{E}\left(\mathbf{s}_u[n]\mathbf{s}_u^*[n]\right) = \mathbf{I}_{L_u}$. For the sake of clarity, we define $\hat{\mathbf{H}}_{u,i}[n] \triangleq \frac{1}{\sqrt{P[n]}}\mathbf{G}_u[n]\mathbf{B}_u[n]\mathbf{H}_u[n]\mathbf{F}_i[n]$, and $\mathbf{w}_u[n] \triangleq \mathbf{G}_u[n]\mathbf{B}_u[n]\mathbf{n}_u[n]$, so

$$\mathbf{y}_u[n] = \hat{\mathbf{H}}_{u,u}[n]\mathbf{s}_u[n] + \underbrace{\sum_{i \in \mathcal{T},\, i \neq u} \hat{\mathbf{H}}_{u,i}[n]\mathbf{s}_i[n]}_{\text{Interuser interference}} + \underbrace{\mathbf{w}_u[n]}_{\text{Noise}}. \tag{7.2}$$

The transmit signal for each of the scheduled receivers is the result of performing coding, interleaving and constellation mapping operations on a stream of source bits. The MCS for the u-th user c_u is selected from a finite set of MCS $\mathcal{C}$. The selected number of spatial streams and MCS for the u-th user has an associated rate of $\eta\left(c_u, L_u\right)$ bits per second.

In general, the probability that a frame is not correctly decoded at the u-th receiver (i.e., the frame error rate - FER), depends on the transmit power, channel matrices, number of scheduled users, selected MCS for the u-th user, and selected modulation for the interfering users. By treating the noise and residual multi-user interference as Gaussian, and assuming a linear receiver, it is reasonable to write the FER of the u-th user p_u as a function of the the selected MCS c_u and the post-processing SNR values $\boldsymbol{\gamma}_u = \left[\gamma_{u,1}\left[1\right], \ldots, \gamma_{u,1}\left[N\right], \ldots, \gamma_{u,L_u}\left[1\right], \ldots, \gamma_{u,L_u}\left[N\right]\right]^T$, where

$$p_u = \text{FER}\left(\boldsymbol{\gamma}_u, c_u\right). \tag{7.3}$$

The post processing SNR of the u-th user in the i-th spatial stream and n-th carrier is defined as

$$\gamma_{u,i}[n] = \frac{\left|\left[\mathbf{D}_{u,u}\left[n\right]\right]_{ii}\right|^2}{\left[\mathbf{R}_u[n]\right]_{ii}} \quad u = 1 \ldots U, i = 1 \ldots L_u \tag{7.4}$$

with

$$\mathbf{R}_u[n] = \left(\hat{\mathbf{H}}_{u,u}[n] - \mathbf{D}_{u,u}[n]\right)\left(\hat{\mathbf{H}}_{u,u}[n] - \mathbf{D}_{u,u}[n]\right)^* + \sum_{j\in\mathcal{T},\, j\neq u} \hat{\mathbf{H}}_{u,j}[n]\hat{\mathbf{H}}_{u,j}^*[n] + \sigma^2\mathbf{G}_u[n]\mathbf{B}_u[n]\mathbf{B}_u^*[n]\mathbf{G}_u^*[n] \tag{7.5}$$

the covariance matrix of interference plus noise, and $\mathbf{D}_{u,u} \triangleq \hat{\mathbf{H}}_{u,u} \odot \mathbf{I}_{L_u}$. Based on the rate $\eta(c_u, L_u)$ and the FER p_u, we define the throughput of user u as

$$t_u = \eta(c_u, L_u)(1 - p_u). \tag{7.6}$$

7.4 Problem statement

In this section, we formulate the link adaptation problem. The adaptation problem in the multiuser scenario is different from the single user scenario. In the single user case, the usual objective of link adaptation is to maximize the (unique) link throughput subject to a constraint on the FER. In the MU-MIMO case, each user has a different rate, so the objective might be to maximize a function of the rates, subject to a FER constraint $p_0 > 0$ (assumed to be equal for all receivers). We consider the sum rate as the performance objective. If we denote by $\mathbf{t} = [t_1 \dots t_U]$ the vector containing the throughput (7.6) of all users, and by $\nu(\mathbf{t}) \triangleq \sum_{u=1}^{U} t_u$ the sum rate, the LA problem can be stated as

$$\text{maximize} \quad \nu(\mathbf{t}) \quad \text{subject to} \quad p_u \leq p_0 \quad u = 1, \dots, U. \tag{7.7}$$

We will assume $t_u = 0$, $p_u = 0$ if $u \notin \mathcal{T}$ to be consistent with our approach, which involves scheduling a subset of the users. Note that the LA problem can be modified to maximize another utility metric besides the sum rate just by defining $\nu(\mathbf{t})$ accordingly. For example, we can include proportional fairness in this setting by changing the objective function to $\nu(\mathbf{t}) = \sum_{u=1}^{U} \log t_u$ [132].

Trying to solve (7.7) directly is computationally intractable. Besides the difficulty of obtaining a mathematical model that maps the CSI to the FER p_u, the number of design variables is quite large and difficult to handle. The design variables include, among others, the set of active users $\mathcal{T}$, the streams per each active user L_u, the precoding matrices $\mathbf{F}_u[n]$, the interference removal matrices $\mathbf{B}_u[n]$, equalizers $\mathbf{G}_u[n]$ and MCS c_u. Moreover, imperfect CSI at the transmitter creates unknown interference leakage between the different receivers. We propose a procedure for link adaptation that finds a good solution to the sum rate maximization prob-

lem. Our solution has three operational blocks: *precoding and equalization with interference estimation*, *MCS selection* and *user and mode selection*.

7.5 Precoding and equalization with interference estimation

In this section, we present the precoding technique for MU-MIMO transmission and obtain a closed form approximation for the residual interuser interference caused by limited feedback precoding. Given the subset of active users $\mathcal{T}$ and the number of streams per user L_u, the joint problem of selecting optimum precoders[2] $\mathbf{F}_u$, interference rejection matrices $\mathbf{B}_u$, and equalizers $\mathbf{G}_u$ to maximize the sum rate does not have a closed form solution. Several approaches have been proposed, e.g. block diagonalization (BD) [133] or minimizing signal *leakage* [134]. We assume that the precoders and interference rejection matrices are obtained using BD. We choose this precoding algorithm because of its simplicity and its low gap to capacity when used in conjunction with user selection algorithms [135]. The proposed link adaptation framework, however, can be used with other precoding techniques, although we restrict our analysis to BD.

7.5.1 Block diagionalization precoding

BD precoding removes the interference between the different users but not the interference between streams associated to the same user. It is well suited for IEEE 802.11ac because the precoders can be designed based on the feedback information provided. The BD precoder is designed as follows. Let $\mathbf{H}_u = \mathbf{U}_u\mathbf{\Sigma}_u\mathbf{V}_u^*$ be the SVD decomposition of $\mathbf{H}_u$ with the singular values in $\mathbf{\Sigma}_u$ arranged in decreasing order. Note that $\mathbf{U}_u \in \mathbb{C}^{N_{\mathrm{rx},u}\times N_{\mathrm{rx},u}}$, $\mathbf{\Sigma}_u \in \mathbb{R}^{N_{\mathrm{rx},u}\times N_{\mathrm{tx}}}$ and $\mathbf{V}_u \in \mathbb{C}^{N_{\mathrm{tx}}\times N_{\mathrm{tx}}}$. The interference rejection matrix $\mathbf{B}_u$ is formed by taking the first L_u columns of $\mathbf{U}_u$ (corresponding to the left singular vectors associated to the L_u largest singular values). Let us denote $\tilde{\mathbf{H}}_u \triangleq \mathbf{B}_u\mathbf{H}_u$, and

$$\bar{\mathbf{H}}_u \triangleq \begin{bmatrix} \tilde{\mathbf{H}}_1 & \dots & \tilde{\mathbf{H}}_{u-1} & \tilde{\mathbf{H}}_{u+1} & \dots & \tilde{\mathbf{H}}_T \end{bmatrix}^T. \tag{7.8}$$

BD requires that the precoder F satisfies

$$\bar{\mathbf{H}}_u\mathbf{F}_u = \mathbf{0}\,\forall u. \tag{7.9}$$

The set of precoders achieving (7.9) can be written as $\mathbf{N}_u\mathbf{P}_u$, where $\mathbf{N}_u$ is a basis of the nullspace of $\bar{\mathbf{H}}_u$, and $\mathbf{P}_u$ is an arbitrary matrix, that can be used to select the directions of transmission (in

[2] As the design of precoders is independent for each carrier, we will drop the index $[n]$ in this section.

case the nullspace of $\bar{\mathbf{H}}_u$ is of dimension higher than L_u) as well as to perform power allocation. We choose $\mathbf{F}_u$ as the matrix containing the L_u singular vectors associated to the largest singular values of $\tilde{\mathbf{H}}_u\mathbf{N}_u$, i.e., we perform uniform power allocation along the L_u stronger directions of the equivalent channel $\tilde{\mathbf{H}}_u\mathbf{N}_u$. The nullspace of $\mathbf{N}_u$ has dimension $D_u \triangleq N_{\text{tx}} - \sum_{i=1,i\neq u}^{U} L_u$, so $\mathbf{N}_u \in \mathbb{C}^{N_{\text{tx}}\times D_u}$, and $\mathbf{P}_u \in \mathbb{C}^{D_u\times L_u}$. Note that if the system is fully loaded ($\sum_{u=1}^{U} L_u = N_{\text{tx}}$) then the nullspace of $\bar{\mathbf{H}}_u$ will have dimension L_u and, therefore, $\mathbf{P}_u$ will be a square matrix.

The post-processed signal at the u-th receiver is

$$\mathbf{y}_u = \mathbf{G}_u\mathbf{B}_u\mathbf{H}_u\mathbf{F}_u\mathbf{s}_u + \mathbf{w}_u. \tag{7.10}$$

Equation (7.10) can be obtained from (7.2) just by noticing that the BD condition (7.9) is equivalent to $\mathbf{B}_u\mathbf{H}_u\mathbf{F}_i = \mathbf{0}\ \forall u \neq i$ and therefore, $\hat{\mathbf{H}}_{u,i}[n] = \mathbf{0}$, $\forall u \neq i$. Assuming limited feedback of CSI, as anticipated in Section 7.2, full knowledge of $\mathbf{H}_u$ is not possible. We now explain how BD can be performed also with reduced information. Decompose $\mathbf{H}_u$ using the SVD as

$$\mathbf{H}_u = \begin{bmatrix} \tilde{\mathbf{U}}_u & \tilde{\mathbf{U}}_{u,\text{small}} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_{u,0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{u,1} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_u^* \\ \tilde{\mathbf{V}}_{u,\text{small}}^* \end{bmatrix}. \tag{7.11}$$

where $\tilde{\mathbf{U}}_u \in \mathbb{C}^{N_{\text{rx},u}\times L_u}$, $\tilde{\mathbf{U}}_{u,\text{small}} \in \mathbb{C}^{N_{\text{rx},u}\times(N_{\text{rx},u}-L_u)}$ are the matrices containing the left singular vectors associated to the L_u largest singular values and $N_{\text{rx},u} - L_u$ smallest singular values, respectively; $\mathbf{\Sigma}_{u,0} \in \mathbb{C}^{L_u\times L_u}$ is the diagonal matrix containing the L_u largest singular values of $\mathbf{H}_u$; and $\tilde{\mathbf{V}}_u^* \in \mathbb{C}^{L_u\times N_{\text{tx}}}$, $\tilde{\mathbf{V}}_{u,\text{small}}^* \in \mathbb{C}^{(N_{\text{rx},u}-L_u)\times N_{\text{tx}}}$ are the matrices containing the L_u right singular vectors associated with the L_u largest singular values, and the $(N_{\text{rx},u}-L_u)$ right singular vectors associated to the remaining nonzero singular values. Assuming that the receiver uses $\mathbf{B}_u = \tilde{\mathbf{U}}_u^*$, then the *equivalent channel* can be written as $\tilde{\mathbf{H}}_u = \mathbf{\Sigma}_{u,0}\tilde{\mathbf{V}}_u^*$. It can be seen that $\tilde{\mathbf{H}}_u$ can be available at the transmitter (with some quantization error) with the feedback scheme described in Section 7.2 for IEEE 802.11ac, since the matrix $\tilde{\mathbf{V}}_u$ is the beamforming matrix for L_u spatial streams, and the values of $\mathbf{\Sigma}_{u,0}$ can be obtained from the SNR of each subcarrier and spatial stream. Further, since $\mathbf{\Sigma}_{u,0}$ is invertible, the nullspace of $\tilde{\mathbf{H}}_u$ is the same as that of $\tilde{\mathbf{V}}_u^*$, so the SNR information is not necessary to design the precoders. Consequently, the precoders can be found by ensuring that $\tilde{\mathbf{V}}_u^*\mathbf{F}_j = \mathbf{0}\ \forall u \neq j$. Therefore, the set of precoding matrices can be obtained from the preferred beamformers $\tilde{\mathbf{V}}_u$. The presence of limited feedback, however, creates some unknown interference leakage between the different users, which has to be estimated. We present next the quantization scheme in IEEE 802.11ac, and an analytical approximation to this interference leakage.

7.5.2 Quantization

The presence of limited feedback creates unknown interference leakage among the different users. The post-processing SNR depends on the interference leakage, so this value has to be estimated. This estimation can be easily performed at the receive side, and CSIT can be acquired by the use of the feedback link. This procedure, however, is not desirable for various reasons. First, there is a circular dependency between user and mode selection and interference leakage estimation. User and mode selection requires knowing the post processing SNR and, consequently, the interference leakage, but estimating the interference leakage is restricted to a certain user/mode configuration. Second, the amount of overhead and training is roughly doubled with respect to the simple message interchange in Figure 7.1. Therefore, it is desirable to estimate the interference leakage at the transmit side, without additional feedback from the receivers.

Next we derive an approximation for the interference leakage caused by the quantization procedure in IEEE 802.11ac. This approximation can be easily computed at the transmitter by using a statistical characterization of the quantization error. We describe first the quantization method used in IEEE 802.11ac, and then derive an approximation for the interference leakage.

The objective of the quantization task is to provide the transmitter with a quantized version $\hat{\mathbf{V}}_u$ of matrix $\tilde{\mathbf{V}}_u$, which is the preferred beamforming matrix. The quantization process proceeds as follows. The unitary matrix $\tilde{\mathbf{V}} \in \mathbb{C}^{N_{\text{tx}},L}$ is decomposed by using the Givens decomposition [136, Ch. 5] as

$$\tilde{\mathbf{V}} = \left(\prod_{\ell=1}^{L} \mathbf{D}_\ell \left(\phi_{\ell,1} \dots \phi_{\ell,N_{\text{tx}}-\ell+1} \right) \prod_{n=\ell+1}^{N_{\text{tx}}} \mathbf{G}_{n,\ell} \left(\psi_{\ell,n} \right) \right) \tilde{\mathbf{I}} \tag{7.12}$$

where

$$\mathbf{D}_\ell \left(\phi_{\ell,1} \dots \phi_{\ell,N_{\text{tx}}-\ell+1} \right) = \text{diag} \left(\mathbf{1}_{\ell-1}, e^{j\phi_{\ell,1}} \dots e^{j\phi_{\ell,N_{\text{tx}}-\ell+1}} \right) \in \mathbb{C}^{N_{\text{tx}} \times N_{\text{tx}}}. \tag{7.13}$$

$\tilde{\mathbf{I}}$ is a matrix containing the first L columns of an $N_{\text{tx}} \times N_{\text{tx}}$ identity matrix and $\mathbf{G}_{n,\ell} \left(\psi_{\ell,n} \right) \in \mathbb{R}^{N_{\text{tx}} \times N_{\text{tx}}}$ is a rotation matrix operating in the ℓ and n coordinates:

$$\mathbf{G}_{n,\ell} \left(\psi_{\ell,n} \right) = \begin{bmatrix} \mathbf{I}_{\ell-1} & & & & \\ & \cos\psi_{\ell,n} & & \sin\psi_{\ell,n} & \\ & & \mathbf{I}_{n-\ell-1} & & \\ & -\sin\psi_{\ell,n} & & \cos\psi_{\ell,n} & \\ & & & & \mathbf{I}_{N-n} \end{bmatrix}. \tag{7.14}$$

The feedback consists of quantized versions $\hat{\psi}_{\ell,n}$ and $\hat{\phi}_{\ell,n}$ of the angles $\psi_{\ell,n}$ and $\phi_{\ell,n}$. The quantization of $\psi_{\ell,n}$ and $\phi_{\ell,n}$ is performed by the use of b_ψ and b_ϕ bits, respectively. IEEE

802.11ac uses a uniform quantizer. Since $\psi_{\ell,n} \in [0, \pi/2]$ and $\phi_{\ell,n} \in [0, 2\pi]$ [137], the codebook for each of the angles is

$$\hat{\psi}_{\ell,n} \in \left\{ q_{\psi,k} \triangleq \frac{k\pi}{2^{b_\psi+1}} + \frac{\pi}{2^{b_\psi+2}},\ k = 0, 1 \dots 2^{b_\psi} - 1 \right\} \tag{7.15}$$

$$\hat{\phi}_{\ell,n} \in \left\{ q_{\phi,k} \triangleq \frac{k\pi}{2^{b_\phi-1}} + \frac{\pi}{2^{b_\phi}},\ k = 0, 1 \dots 2^{b_\phi} - 1 \right\}. \tag{7.16}$$

The quantization of the angles is performed by finding the minimum distance codeword

$$\hat{\psi}_{\ell,n} = q_{\psi,k} \text{ if } \psi_{l,n} \in \mathcal{Q}_{\psi,k} \triangleq \left[\frac{k\pi}{2^{b_\psi+1}}, \frac{(k+1)\pi}{2^{b_\psi+1}} \right] \tag{7.17}$$

$$\hat{\phi}_{\ell,n} = q_{\phi,k} \text{ if } \phi_{l,n} \in \mathcal{Q}_{\phi,k} \triangleq \left[\frac{k\pi}{2^{b_\phi-1}}, \frac{(k+1)\pi}{2^{b_\phi-1}} \right]. \tag{7.18}$$

For the sake of simplicity, we denote $\delta \triangleq \frac{\pi}{2^{b_\phi}}$ and $\epsilon \triangleq \frac{\pi}{2^{b_\psi+2}}$, so that $\mathcal{Q}_{\psi,k} = [q_{\psi,k} - \epsilon, q_{\psi,k} + \epsilon]$ and $\mathcal{Q}_{\phi,k} = [q_{\phi,k} - \delta, q_{\phi,k} + \delta]$. Note that this quantization scheme is the LBG quantizer [138], optimal in the minimum distortion sense, only if the angles are independent and the distribution of both angles is uniform, which is not the case even in the well-studied independent and identically distributed Gaussian MIMO channel, see e.g. [139].

7.5.3 Interference estimation

Now we characterize the residual interference using only the quantized CSI. The BD precoders are designed using the quantized beamformers $\hat{\mathbf{V}}_u$ to satisfy $\hat{\mathbf{V}}_u^* \mathbf{F}_j = \mathbf{0}\ \forall u \neq j$. Because the beamformers are quantized, the interuser interference cannot be completely removed due to the imperfect CSI, so the total interference plus noise covariance matrix (7.5) assuming BD precoding and ZF equalizer at user u is

$$\mathbf{R}_u = \sum_{j \in \mathcal{T} \setminus \{u\}} \mathbf{R}_{u,j} + \sigma^2 \mathbf{G}_u \mathbf{G}_u^* \tag{7.19}$$

where $\mathbf{R}_{u,j} \triangleq \hat{\mathbf{H}}_{u,j} \hat{\mathbf{H}}_{u,j}^*$ is the covariance matrix of the interference from the message intended to user j. Equation (7.19) follows by assuming perfect CSI at the receiver, so that the ZF equalizer removes all the intra-stream interference, i.e., $\hat{\mathbf{H}}_{u,u} = \mathbf{D}_{u,u}$ in (7.5), and by applying the fact that $\mathbf{B}_u$ is a unitary matrix. If $\mathbf{G}_u$ is a ZF equalizer, and the precoders and interference rejection

matrices are designed following the BD procedure, then

$$\mathbf{G}_u = (\mathbf{B}_u \mathbf{H}_u \mathbf{F}_u)^{-1} = \left(\mathbf{\Sigma}_{u,0} \tilde{\mathbf{V}}_u^* \mathbf{F}_u\right)^{-1}. \tag{7.20}$$

Now, write $\tilde{\mathbf{V}}_u^* = \hat{\mathbf{V}}_u^* + \mathbf{E}_u$, where $\mathbf{E}_u$ is the quantization error matrix. If the quantization error is small, (7.20) can be approximated as

$$\mathbf{G}_u = \left(\mathbf{\Sigma}_{u,0} \hat{\mathbf{V}}_u^* \mathbf{F}_u + \mathbf{\Sigma}_{u,0} \mathbf{E}_u \mathbf{F}_u\right)^{-1} \approx \left(\mathbf{\Sigma}_{u,0} \hat{\mathbf{V}}_u^* \mathbf{F}_u\right)^{-1}. \tag{7.21}$$

Note that for (7.21) to hold it is only necessary that the entries of $\mathbf{E}_u$ are negligible with respect to the entries of $\hat{\mathbf{V}}_u^*$.

The multiuser interference can be written as

$$\mathbf{R}_{u,j} = \frac{1}{P} \mathbf{G}_u \mathbf{\Sigma}_{u,0} \tilde{\mathbf{V}}_u^* \mathbf{F}_j \mathbf{F}_j^* \tilde{\mathbf{V}}_u \mathbf{\Sigma}_{u,0}^* \mathbf{G}_u^* = \frac{1}{P} \mathbf{G}_u \mathbf{\Sigma}_{u,0} \mathbf{E}_u^* \mathbf{F}_u \mathbf{F}_u^* \mathbf{E}_u \mathbf{\Sigma}_{u,0}^* \mathbf{G}_u^* \tag{7.22}$$

where the last equality is due to the BD constraint $\hat{\mathbf{V}}_u^* \mathbf{F}_j = \mathbf{0}$.

First, note that the interference covariance matrices $\mathbf{R}_{u,j}$ in (7.22) are random variables from the transmitter's point-of-view, since the quantization noise $\mathbf{E}_u$ is unknown at the transmitter. We define a new covariance matrix by averaging over the realizations of $\mathbf{E}_u$

$$\bar{\mathbf{R}}_{u,j} \triangleq \mathbb{E}\left[\mathbf{R}_{u,j}\right] = \frac{1}{P} \mathbf{G}_u \mathbf{\Sigma}_{u,0} \mathbf{C}_{u,j} \mathbf{\Sigma}_{u,0}^* \mathbf{G}_u^* \tag{7.23}$$

with

$$\mathbf{C}_{u,j} = \mathbb{E}\left[\tilde{\mathbf{V}}_u^* \mathbf{F}_j \mathbf{F}_j^* \tilde{\mathbf{V}}_u\right] \tag{7.24}$$

and the expectation is taken over the random realization of $\tilde{\mathbf{V}}_u$ given the received feedback $\hat{\mathbf{V}}_u$. Note that we are implicitly averaging over $\mathbf{E}_u$, but the particular structure of the quantization method makes it easier to derive the final result when explicitly averaging over $\tilde{\mathbf{V}}_u | \hat{\mathbf{V}}_u$.

Let us define $\boldsymbol{\phi}_\ell \triangleq (\phi_{\ell,1}, \ldots, \phi_{\ell,N_{\text{tx}}-\ell+1})$ and $\hat{\phi}_\ell \triangleq \left(\hat{\phi}_{\ell,1}, \ldots, \hat{\phi}_{\ell,N_{\text{tx}}-\ell+1}\right)$. With this, we can write $\tilde{\mathbf{V}}_u$ following a Givens decomposition, so that the covariance matrix is

$$\mathbf{C}_{u,j} = \mathbb{E}\left[\tilde{\mathbf{I}}^* \left(\prod_{\ell=1}^{L_u} \mathbf{D}_l\left(\boldsymbol{\phi}_\ell\right) \prod_{n=\ell+1}^{N_{\text{tx}}} \mathbf{G}_{n,\ell}\left(\psi_{\ell,n}\right)\right)^* \mathbf{F}_k \mathbf{F}_k^* \left(\prod_{\ell=1}^{L_u} \mathbf{D}_l\left(\boldsymbol{\phi}_l\right) \prod_{n=\ell+1}^{N_{\text{tx}}} \mathbf{G}_{n,\ell}\left(\psi_{\ell,n}\right)\right) \tilde{\mathbf{I}}\right] \tag{7.25}$$

where we parametrized the matrix random variable $\tilde{\mathbf{V}}_u | \hat{\mathbf{V}}_u$ using the Givens parameters $\phi_{\ell,n} | \hat{\phi}_{\ell,n}$ and $\psi_{\ell,n} | \hat{\psi}_{\ell,n}$. If we assume that all the angles ϕ and ψ are independent, then the expected value over all the angles can be decomposed into several expected values, each one over a different

angle. This is a reasonable assumption for a MIMO channel with zero-mean Gaussian iid entires [139], for example. To simplify the computation of the covariance matrix, we work with a vectorized version of $\mathbf{C}_{u,j}$, $\mathbf{c}_{u,j} \triangleq \operatorname{vec} \mathbf{C}_{u,j}$. Using properties of the Kronecker product [140], we have

$$\mathbb{E}\left[\mathbf{c}_{k,i}\right] = \left(\hat{\mathbf{I}}^{\mathrm{T}} \otimes \hat{\mathbf{I}}^{*}\right) \left(\prod_{\ell=1}^{L_u} \mathbf{R}_{\ell}^{\mathrm{T}} \prod_{n=\ell+1}^{N_{\mathrm{tx}}} \mathbf{W}_{\ell,n}^{\mathrm{T}} \right)^{\mathrm{T}} \operatorname{vec}\left(\mathbf{F}_j \mathbf{F}_j^{*}\right) \tag{7.26}$$

where

$$\mathbf{R}_{\ell} \triangleq \mathbb{E}_{\boldsymbol{\phi}_{\ell}|\hat{\boldsymbol{\phi}}_{\ell}} \left[\mathbf{D}_{\ell}^{\mathrm{T}}\left(\boldsymbol{\phi}_{\ell}\right) \otimes \mathbf{D}_{\ell}^{*}\left(\boldsymbol{\phi}_{\ell}\right) \right] \tag{7.27}$$

and

$$\mathbf{W}_{\ell,n} \triangleq \mathbb{E}_{\psi_{\ell,n}|\hat{\psi}_{\ell,n}} \left[\mathbf{G}_{n,\ell}^{\mathrm{T}}\left(\psi_{\ell,n}\right) \otimes \mathbf{G}_{n,\ell}^{*}\left(\psi_{\ell,n}\right) \right]. \tag{7.28}$$

We now proceed to approximate $\mathbf{R}_{\ell}$ and $\mathbf{W}_{\ell,n}$. We resort to a well known result in high resolution quantization theory [141] and approximate the quantization error by a uniform random variable in the quantization bin. This approximation is exact for the quantization noise $\hat{\phi}_{\ell,i} - \phi_{\ell,i}$ (but not for $\hat{\psi}_{\ell,n} - \psi_{\ell,n}$) if $\mathbf{H}_u$ has Gaussian iid entries, since $\hat{\phi}_{\ell,i}$ is uniformly distributed in $[0, 2\pi]$ [137].

We now calculate a closed form expression for (7.27) and (7.28) assuming uniform quantization noise. For the sake of clarity, we will omit subscripts and matrix angular arguments when their value is clear. Using the definition in (7.13), it is possible to take the Kronecker product and write (7.27) as

$$\operatorname{diag}\left(\mathbf{D}_{\ell}^{*}, \ldots, \mathbf{D}_{\ell}^{*}, e^{j\phi_{\ell,1}}\mathbf{D}_{\ell}^{*}, \ldots e^{j\phi_{\ell,N_{\mathrm{tx}}-\ell+1}}\mathbf{D}_{\ell}^{*}\right). \tag{7.29}$$

Now we compute the expectations of each term. First observe that

$$\mathbb{E}[\mathbf{D}_{\ell}^{*}]_{\ell+i-1,\ell+i-1} \;=\; \mathbb{E}\left[e^{-j\phi_{\ell,i}}\right] = \frac{1}{2\epsilon}\int_{\hat{\phi}_{\ell,i}-\delta}^{\hat{\phi}_{\ell,i}+\delta} e^{-j\phi}\mathrm{d}\phi = \frac{e^{-j\hat{\phi}_{\ell,i}}\sin\delta}{\delta}. \tag{7.30}$$

The expectation of the other diagonal terms can be obtained similarly as

$$\mathbb{E}\left[e^{j\phi_{\ell,j}}[\mathbf{D}_{\ell}^{*}]_{i,i}\right] = \frac{e^{j\hat{\phi}_{\ell,j}}\sin\delta}{\delta},\; i = 1\ldots\ell-1 \tag{7.31}$$

and

$$\mathbb{E}\left[e^{j\phi_{\ell,j}}[\mathbf{D}_{\ell}^{*}]_{\ell+i-1,\ell+i-1}\right] = \begin{cases} 1 & \text{if } i = j \\ \dfrac{e^{j\left(\hat{\phi}_{\ell,j}-\hat{\phi}_{\ell,i}\right)}\sin^2\left(\delta\right)}{\delta^2} & \text{if } i \neq j \end{cases},\; i = 1\ldots N_{\mathrm{tx}}-\ell+1. \tag{7.32}$$

Now, we proceed to obtain a closed form for (7.28) with the uniform error approximation. First,

let us write

$$\mathbf{W}_{\ell,n} = \mathbb{E}\left[\begin{array}{cccc} \mathbf{I}_{\ell-1}\otimes\mathbf{G}_{\ell,n}^* & & & \\ & \cos\psi_{\ell,n}\mathbf{G}_{\ell,n}^* & & -\sin\psi_{\ell,n}\mathbf{G}_{\ell,n}^* \\ & & \mathbf{I}_{n-\ell-1}\otimes\mathbf{G}_{\ell,n}^* & \\ & \sin\psi_{\ell,n}\mathbf{G}_{\ell,n}^* & & \cos\psi_{\ell,n}\mathbf{G}_{\ell,n}^* \\ & & & \mathbf{I}_{N_{\mathrm{tx}}-n}\otimes\mathbf{G}_{\ell,n}^* \end{array}\right]. \tag{7.33}$$

The expectations of the different blocks of the matrix can be obtained by simple integration, similarly to (7.31).

$$\mathbb{E}\left[\mathbf{G}_{\ell,n}^*\right] = \left[\begin{array}{cccc} \mathbf{I}_{\ell-1} & & & \\ & \frac{\sin\epsilon\cos\hat{\psi}_{\ell,n}}{\epsilon} & -\frac{\sin\epsilon\sin\hat{\psi}_{\ell,n}}{\epsilon} & \\ & \mathbf{I}_{n-\ell-1} & & \\ & \frac{\sin\epsilon\sin\hat{\psi}_{\ell,n}}{\epsilon} & \frac{\sin\epsilon\cos\hat{\psi}_{\ell,n}}{\epsilon} & \\ & & & \mathbf{I}_{N_{\mathrm{tx}}-n} \end{array}\right] \tag{7.34}$$

$$\mathbb{E}\left[\cos\psi_{\ell,n}\mathbf{G}_{\ell,n}^*\right] = \left[\begin{array}{ccc} \frac{\cos\hat{\psi}_{\ell,n}\sin\epsilon}{\epsilon}\mathbf{I}_{\ell-1} & & \\ & \mathbf{J}_{\ell,n} & \\ & & \frac{\cos\hat{\psi}_{\ell,n}\sin\epsilon}{\epsilon}\mathbf{I}_{N_{\mathrm{tx}}-n} \end{array}\right] \tag{7.35}$$

with

$$\mathbf{J}_{\ell,n} = \left[\begin{array}{ccc} \frac{\epsilon+\cos\epsilon\cos 2\hat{\psi}_{\ell,n}\sin\epsilon}{2\epsilon} & & -\frac{\cos\epsilon\cos\hat{\psi}_{\ell,n}\sin\epsilon\sin\hat{\psi}_{\ell,n}}{\epsilon} \\ & \frac{\cos\hat{\psi}_{\ell,n}\sin\epsilon}{\epsilon}\mathbf{I}_{n-\ell-1} & \\ \frac{\cos\epsilon\cos\hat{\psi}_{\ell,n}\sin\epsilon\sin\hat{\psi}_{\ell,n}}{\epsilon} & & \frac{\epsilon+\cos\epsilon\cos 2\hat{\psi}_{\ell,n}\sin\epsilon}{2\epsilon} \end{array}\right] \tag{7.36}$$

and

$$\mathbb{E}\left[\sin\psi_{\ell,n}\mathbf{G}_{\ell,n}^*\right] = \left[\begin{array}{ccc} \frac{\sin\hat{\psi}_{\ell,n}\sin\epsilon}{\epsilon}\mathbf{I}_{\ell-1} & & \\ & \mathbf{S}_{\ell,n} & \\ & & \frac{\sin\hat{\psi}_{\ell,n}\sin\epsilon}{\epsilon}\mathbf{I}_{N_{\mathrm{tx}}-n} \end{array}\right] \tag{7.37}$$

with

$$\mathbf{S}_{\ell,n} = \left[\begin{array}{ccc} \frac{\cos\epsilon\cos\hat{\psi}_{\ell,n}\sin\epsilon\sin\hat{\psi}_{\ell,n}}{\epsilon} & & -\frac{\epsilon-\cos\epsilon\cos 2\hat{\psi}_{\ell,n}\sin\epsilon}{2\epsilon} \\ & \frac{\sin\hat{\psi}_{\ell,n}\sin\epsilon}{\epsilon}\mathbf{I}_{n-\ell-1} & \\ \frac{\epsilon-\cos\epsilon\cos 2\hat{\psi}_{\ell,n}\sin\epsilon}{2\epsilon} & & \frac{\cos\epsilon\cos\hat{\psi}_{\ell,n}\sin\epsilon\sin\hat{\psi}_{\ell,n}}{\epsilon} \end{array}\right]. \tag{7.38}$$

Note that the complexity of obtaining the error covariance matrix is similar to the complexity of recovering the matrix $\hat{\mathbf{V}}$ from the quantized angles. Figure 7.2 shows the accuracy of the analytical approximation.

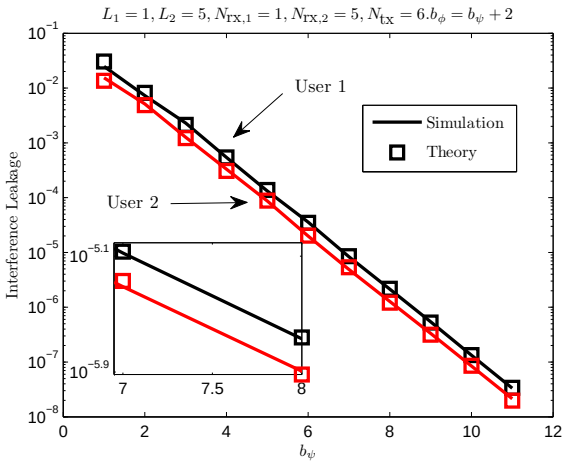


Figure 7.2: Analytical and empirical interference leakage per spatial stream $\left(\frac{1}{L_u}\,\text{tr}\;\mathbf{C}_{u,j}\right)$ for a two-user scenario. The theoretical approximation is calculated using (7.26) and the closed form expressions for the expected values (7.30) - (7.37) . The simulation interference leakage was averaged over 100 independent MIMO channels with independent complex Gaussian entries.

7.6 MCS selection

The MCS selection block consists of a function μ that takes as input the set of post-processing SNR values of user u and the number of spatial streams L_u, and computes the higher MCS that meets the FER constraint for those SNR values, i.e.

$$\mu\left(\boldsymbol{\gamma}_u, L_u\right) \;=\; \arg\max_{c\in\mathcal{C}} \eta\left(c, L_u\right) \;\text{ subject to }\; \text{FER}\left(\boldsymbol{\gamma}_u, c_u\right) \leq p_0. \tag{7.39}$$

The post-processing SNR values are calculated by applying the approximations (7.21) and (7.26) in (7.4) to incorporate the interference leakage estimate. An important observation is that in (7.39) the rate is being maximized, not the throughput. The reason is that for small values of p_0, the feasible points meet $\eta\left(c, L_u\right)\left(1 - \text{FER}\left(\boldsymbol{\gamma}_u, c_u\right)\right) \approx \eta\left(c, L_u\right)$. This approximation simplifies the problem, since estimating the actual value of $\text{FER}\left(\boldsymbol{\gamma}_u, c_u\right)$ is not required, rather it is only necessary to discriminate whether it is above the desired threshold p_0 or not.

We use a machine learning inspired approach to solve (7.39). Essentially, we classify features derived from the channel into the highest MCS that meets the target FER constraint. The

classifier is made up of individual classifiers that distinguish whether a certain MCS and number of spatial streams are supported by the current channel. Note that this is slightly different from conventional machine learning in that there is a target average error rate, whereas machine learning usually involves avoiding classification errors altogether. We will follow a supervised learning approach to solve this problem, which includes two separated tasks: feature extraction and classification.

7.6.1 Feature extraction

In machine learning, the curse of dimensionality is well known: the larger the dimension of the feature vector, the exponentially more data is required [142]. To reduce the dimensionality of $\boldsymbol{\gamma}_u$, we exploit insights made in [47] about performance in coded bit interleaved MIMO-OFDM systems. In particular, it was recognized that performance was invariant to subcarrier ordering, i.e., $\text{FER}\left(\boldsymbol{\gamma}_u, c_u\right) = \text{FER}\left(\boldsymbol{\Pi}\boldsymbol{\gamma}_u, c_u\right)$, with $\boldsymbol{\Pi}$ any permutation matrix. Therefore, the reduced dimension feature vector should be invariant to subcarrier ordering. Similar to [47], we use a subset of the ordered SNR values as our feature vector. Define the *ordered SNR vector* $\tilde{\boldsymbol{\gamma}}_u = \left[\tilde{\gamma}_{u,1}, \ldots, \tilde{\gamma}_{u,NL_u}\right]^T$ as a vector formed by ordering the elements in $\boldsymbol{\gamma}_u$ in ascending order. For example, $\tilde{\gamma}_{u,i}$ denotes the i-th smallest SNR value (among all carriers and spatial streams) of user u. We obtain our feature vector $\mathbf{f}$ by selecting a subset of the entries of $\tilde{\boldsymbol{\gamma}}_u$. Other approaches for dimensionality reduction, like principal component analysis [143], may alternatively be applied, but in our simulations we did not see a significant benefit.

7.6.2 Classification

The objective of the classification task is to estimate the highest MCS supported by the channel, as characterized by the feature vector $\mathbf{f}$. Following a similar approach as in [50], we use a set of classifiers $\delta_{c,L}(\mathbf{f})$ to discriminate whether the current channel will support transmission with MCS c and L spatial streams while meeting the FER constraint. More formally, given a set of M training samples $\{(\boldsymbol{\gamma}_i, p_{\text{tr},i})\}_{i=1}^{M}$, with $p_{\text{tr},i}$ the FER of the i-th channel, the input data for the classifier is the set $\{(\mathbf{f}_i, \nu_i)\}_{i=1}^{M}$, with

$$\nu_i = \begin{cases} 1 & \text{if } p_{\text{tr},i} \leq p_0 \\ -1 & \text{if } p_{\text{tr},i} > p_0 \end{cases} \tag{7.40}$$

and $\mathbf{f}_i$ the subset of the ordered SNR of the i-th channel. The classifier is a function of the feature input vector that maps

$$\delta_{c,L} : \ \mathbf{f} \rightarrow \{-1, 1\}\,. \tag{7.41}$$

There are several ways to construct a classifier; we choose the popular SVM. An SVM determines the class of a sample by the use of linear boundaries (hyperplanes) in high dimensional spaces. Operating in a high dimensional space is enabled by the use of a Kernel function $K(\mathbf{x}_1, \mathbf{x}_2)$ that maps $\mathbf{x}_1$ and $\mathbf{x}_2$ to vectors $\phi(\mathbf{x}_1)$ and $\phi(\mathbf{x}_2)$ lying in a Hilbert space, and performs the inner product in that space $\langle \phi(\mathbf{x}_1), \phi(\mathbf{x}_2) \rangle$. The Kernel function $K(\mathbf{x}_1, \mathbf{x}_2)$ has a very simple form for properly chosen ϕ. In many cases, perfect separation by a hyperplane is not possible (or not desirable, since it would lead to non-smooth boundaries) and a penalization term is introduced to take into account the misclassified training samples. Formally, the classifier is

$$\delta_{c,L}(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{M} \alpha_i \nu_i K\left(\mathbf{x}, \mathbf{f}_i\right) + b \right) \tag{7.42}$$

where α_i is obtained as the result of the optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \alpha_i \alpha_j K\left(\mathbf{f}_i, \mathbf{f}_j\right) - \sum_{i=1}^{M} \alpha_i \\ \text{subject to} & \sum_{i=1}^{M} \nu_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C,\, i = 1, \ldots, M \end{array} \tag{7.43}$$

and b can be obtained by solving $\delta_{c,L}(\mathbf{f}_i)\nu_i = 1$ for any training sample $\mathbf{f}_i$ such that $0 < \alpha_i < C$ [53]. The parameter C has to be adjusted to trade off smoothness and training misclassification rate. High values of C result in very irregular boundaries caused by very small training errors, and low values of C result in large training errors caused by smooth boundaries. We use the radial basis function kernel

$$K(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\rho^2} \right), \tag{7.44}$$

where parameter ρ^2 is used to tradeoff bias and variance: small values of ρ tend to take into account only the nearby training points, leading to high variance classifiers, and large values of ρ result in biased results. The parameters ρ and C are selected using a cross-validation approach [53].

For a given number of streams L_u, the overall classifier chooses the MCS with a higher rate among those predicted to meet the FER constraint. The MCS selection function μ (7.39) is implemented as

$$\mu\left(\boldsymbol{\gamma}_u, L_u\right) = \arg\max_{c \in \mathcal{C}} \left\{ \eta\left(c, L_u\right) \right\} \text{ s.t. } \delta_{c,L}(\boldsymbol{\gamma}_u) = 1. \tag{7.45}$$

7.7 User and mode Selection

Performing optimal user and mode selection requires an exhaustive search over all possible combinations of users and number of streams per user. To overcome this challenge, we propose a greedy approach, similar to [128–130], where the streams are added one by one until the utility function $\nu(\mathbf{t})$ does not increase. In each iteration, one spatial stream is added to the user whose increment in the number of spatial streams led to a higher throughput. The algorithm continues until the maximum number of spatial streams is reached, or when an increment in the number of spatial streams lead to a lower sum rate.

An important observation is that the greedy algorithm is not fair, in the sense that a user could be assigned all the spatial streams if it led to a a higher objective function, in our case the sum rate. The objective function, however, could be modified to encourage fairness. For example, metrics that are concave in the rates would give more utility to assigning spatial streams to unscheduled users. This change in the objective function does not require any other change in the user selection algorithm, or the other pieces of our link adaptation procedure. The entire proposed link adaptation algorithm is summarized in Algorithm 2.

Algorithm 2: Link Adaptation Algorithm

$L_u = 0 \ \forall u$
$\mathcal{R} \leftarrow 0$
while $\sum_{u=1}^{U} L_u < N_{\text{tx}}$ **do**
  for Each user u with $L_u < N_{\text{rx,u}}$ **do**
    Calculate matrices $\mathbf{F}_v[n]$, $\mathbf{G}_v[n]$, $\mathbf{B}_v[n]$ for all users v, for the spatial streams set $\{L_1, L_2, \ldots, L_u + 1, \ldots, L_K\}$ following the procedure in Section 7.5.
    Calculate interference leakage $\mathbf{R}_{u,j}$ as in (7.23).
    Calculate post-processing SNR values $\boldsymbol{\gamma}_v \forall v$ as in (7.4).
    $c_v \leftarrow \mu(\gamma_v, L_v) \, \forall v$. {Calculate optimum MCS for all users}
    $t_v \leftarrow \eta(c_v, L_v) \, \forall v$ {Calculate the corresponding rate}
    $\mathcal{R}_u \leftarrow \nu(\mathbf{t})$ {Utility metric if we incremented L_u by 1}
  end for
  $j \leftarrow \arg\max_u \{\mathcal{R}_u\}$ {User whose increment in L_u leads to a higher rate}
  if $\mathcal{R}_j \geq \mathcal{R}$ **then**
    $L_j \leftarrow L_j + 1$
    $\mathcal{R} \leftarrow \mathcal{R}_j$
  else
    Stop algorithm.
  end if
end while

Table 7.1: MCS in IEEE 802.11ac with the corresponding data rates in 20MHz channels [62].

MCS	Data Rate (Mb/s)	MCS	Data Rate (Mb/s)
BPSK 1/2	6.5	64-QAM 2/3	52
QPSK 1/2	13	64-QAM 3/4	58.5
QPSK 3/4	19.5	64-QAM 5/6	65
16-QAM 1/2	26	256-QAM 3/4	78
16-QAM 3/4	39		

7.8 Simulation results

To validate the performance of the proposed link adaptation method, we performed simulations using parameters from the physical layer of IEEE 802.11ac. The studied scenario comprises a 4-antenna transmitter communicating with three 2-antenna receivers over a 20MHz channel (52 OFDM carriers) with an 800ns guard interval. The frame length was set to 128 bytes. Perfect CSI was assumed at the receiver and different levels of CSI at the transmitter. The FER constraint for the link adaptation problem was set to $p_0 = 0.1$. The set of MCS for optimization with the associated data rate for one spatial stream is shown in Table 7.1.

The training of the classifiers $\delta_{c,L}$ was performed as follows. The training set was generated in a single user setting with perfect CSI by simulating different channels for all MCS and NSS values. For each of the channels, the complete transmit-receive chain was simulated, e.g. coding, interleaving, equalization, decoding. The resulting SNR $\tilde{\boldsymbol{\gamma}}_i$ values and FER $p_{\mathrm{tr},i}$ were stored for each of the channels. The channels were generated in the time domain with a 4-tap MIMO channel with iid Gaussian entries with 30 different noise levels, corresponding to SNR values between 5 and 50 dB. For each training sample, the feature vector $\mathbf{f}$ consisted on 4 equispaced SNR values (including the first and last ones) from the ordered SNR vector $\tilde{\boldsymbol{\gamma}}_i$. The class ν_i of each training sample was adjusted to -1 if the measured FER was above the desired threshold, i.e., $p_{\mathrm{tr},i} > p_0$, and 1 otherwise. We do not consider multiuser or limited feedback CSI in the training set, since we assumed that the SNR information $\tilde{\boldsymbol{\gamma}}_i$ was enough to predict the FER performance, regardless of whether the SNR values are the result of performing limited feedback BD precoding or not.

The parameters ρ and C of the SVM classifier were chosen before training the system. We followed the usual K-fold cross validation procedure [53] (with $K = 4$) to select these parameters. The SVM was implemented with the `LIBSVM` software package [144].

We compare the performance of the SVM classifier with an average SNR and an exponential effective SNR classifier. The training and the test sets were generated independently, each

one containing 6000 samples. For each sample, the FER was estimated by simulating the transmission of 10^3 frames for each of the generated channels, so a small part of the classification errors may be caused by an imperfect FER estimation. This error, however, is expected to affect all the classifiers in the same way. The average SNR classifier discriminates the class using a threshold γ_{th} on the average SNR, designed to minimize the training set error. Formally, the average SNR classifier is

$$\delta_{c,L}^{\text{Av. SNR}}(\boldsymbol{\gamma}) = \text{sign}\left(\frac{1}{NL_u}\sum_{l=1}^{L_u}\sum_{n=1}^{N}\gamma_l[n] - \gamma_{\text{th}}\right). \tag{7.46}$$

The exponential effective SNR consists on a generalized mean of the SNR values

$$\gamma_{\text{eff}} = -\frac{1}{\beta}\log\left(\frac{1}{NL_u}\sum_{l=1}^{L_u}\sum_{n=1}^{N}\exp\left(-\beta\gamma_l[n]\right)\right) \tag{7.47}$$

that is compared with a threshold $\gamma_{\text{eff, th}}$, designed to minimize the training set error. The parameter β depends on the MCS, and is also selected to minimize the training set error. We can write the effective SNR classifier as

$$\delta_{c,L}^{\text{Eff. SNR}}(\boldsymbol{\gamma}) = \text{sign}\left(\gamma_{\text{eff}} - \gamma_{\text{eff,th}}\right). \tag{7.48}$$

In Table 7.2 we show the accuracy results for the tested classifiers. We can see that the SVM classifier outperforms the average and effective SNR classifiers, and in many cases the classification error is below 1%. For example, for 16-QAM 1/2, $L = 1$, the SVM classifier misclassifies around 1 sample out of 200 (0.417%), whereas the average SNR classifier makes approximately 20 times more errors (9.73%). In that case, the error rate of the effective SNR classifier is 1.51%, approximately 4 times more than the SVM classifier. The classification gain with respect to the Av. / Eff. SNR classifier is calculated as $\left(\text{Error}_{\text{Av. / Eff. SNR}} - \text{Error}_{\text{SVM}}\right)/\text{Error}_{\text{Av. / Eff. SNR}}$, and in the 16-QAM 1/2 case for the Av. SNR classifier is approximately $\frac{20-1}{20} \approx 0.95$. This classification gain reflects the percentage of errors made by the average or effective SNR classifier that would be corrected by the SVM. For example, a classification gain of 100% means that the SVM classifier is perfect, and a classification gain of 0% means that the SVM has the same accuracy as the average SNR classifier. We can see that in some MCS the Eff. SNR classifier outperforms the SVM classifier, but in average the SVM performs 22.34% better than the Eff. SNR classifier. The average gain with respect to the Av. SNR classifier is much higher, above 67%. This result confirms that SVM-based link adaptation algorithms outperform state-of-the-art effective SNR classifiers, currently used widely in LTE, among other communication

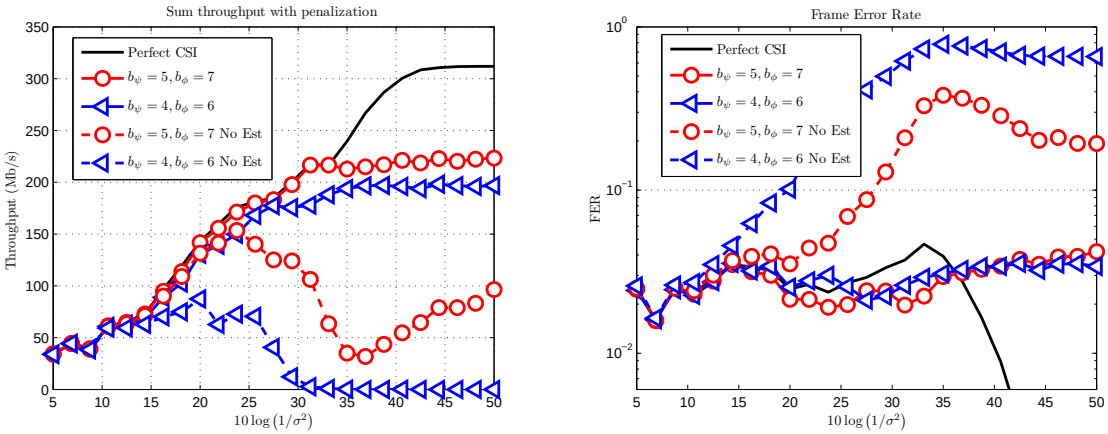
Table 7.2: Classification errors and accuracy gain of the SVM classifier with respect to an Average and Effective SNR classifier (all in %).

	SVM				Av. SNR				Gain Av. SNR				Eff. SNR				Gain Eff. SNR			
	L=1	L=2	L=3	L=4	L=1	L=2	L=3	L=4	L=1	L=2	L=3	L=4	L=1	L=2	L=3	L=4	L=1	L=2	L=3	L=4
BPSK 1/2	0.45	1.33	1.05	0.88	9.4	2.9	1.78	2.82	95.2	54.0	41.1	68.6	0.866	1.2	1.08	1.46	48.1	-11.1	3.08	39.8
QPSK 1/2	0.433	1.12	1	0.683	3.92	7.87	2.33	3.32	95.2	54.0	41.1	68.6	1.1	1.35	1.18	0.9	60.6	17.3	15.5	24.0
QPSK 3/4	1.43	3.57	4.78	2.08	9.22	6.38	6.7	3.78	84.4	44.1	28.6	44.9	2.75	1.4	3.35	2.83	47.9	-154	-42.8	26.5
16-QAM 1/2	0.417	0.6	1.22	0.8	9.73	4.11	1.97	2.45	95.7	85.4	38.1	67.3	1.51	1.58	1.68	1.8	72.5	62.0	27.7	55.6
16-QAM 3/4	1.58	3.87	3.63	1.9	7.63	7.15	4.92	3.28	79.3	45.9	26.1	42.1	3.07	3.63	3.1	2.7	48.4	-6.42	-17.2	29.6
64 QAM 2/3	0.433	1.5	1.66	1.15	4.5	5.11	3.31	2.36	90.3	70.6	49.7	51.4	1.72	2.52	2.33	2.87	74.8	40.4	28.6	59.9
64 QAM 3/4	1.03	2.68	3.05	2.05	6.53	6.1	5.47	3.12	84.2	56.0	44.2	34.2	2.57	3.28	3.08	4.13	59.7	18.3	1.08	50.4
64-QAM 5/6	0.783	1.85	1.57	2.33	7.87	6.38	2.82	5.67	90.0	71.0	44.3	58.8	2.57	3.03	1.88	3.01	69.5	39.0	16.8	22.6
Average	1.65				5.03				67.2				2.24				26.3			

systems [145].

Complete system simulations were run for three different CSI levels at the transmitter: perfect CSI and limited feedback CSI with $(b_\psi = 5, b_\phi = 7)$ and $(b_\psi = 4, b_\phi = 6)$. Simulations were also run for the $(b_\psi = 6, b_\phi = 8)$ case, the higher rate available in IEEE 802.11ac, but the results are not shown as they are indistinguishable from the perfect CSI ones. In all cases a MU feedback was assumed (i.e., including SNR information of all carriers), even for $(b_\psi = 4, b_\phi = 6)$, which is only available for SU mode in IEEE 802.11ac. The misclassified samples, i.e., the cases where the selected MCS led to a FER value over the predefined threshold, were penalized and computed as zero throughput. This penalization is set to remove the advantage of selecting an MCS with a higher throughput, but without meeting the outage constraint, and is a common procedure when evaluating link adaptation algorithms [47,146]. The channels were generated independently from the training samples, but with the same statistical distribution. The channels for the three receivers were generated independently and with the same distribution, so fairness is automatically induced in this scenario. In other scenarios with users with different average received power, fairness can be introduced by changing the objective function, as explained in Section 7.4.

Figure 7.3a illustrates the sum throughput as a function of the noise variance. As expected, in the perfect CSI case, the throughput increases with the SNR, as the interuser interference is completely avoided in this setting, and lower noise levels enable the use of higher rate MCS. For the limited feedback CSI we study two different cases. The first case, represented by solid lines, is the result of following the complete link adaptation algorithm, including the interference estimation procedure. The second case, in dashed lines, does not include the interference estimation block (i.e., assumes $\mathbf{C}_{u,j} = \mathbf{0}$), thus is overestimating the actual SNR. For the interference estimation case, we see that $(b_\psi = 5, b_\phi = 7)$ follows the trend of the perfect CSI curve with a slightly lower throughput until an SNR value of around 35dB, and then flattens with a maximum throughput around 220Mb/s. Performance loss is more significant with the lower rate feedback $(b_\psi = 4, b_\phi = 6)$ for moderate SNR values, reaching an error floor around 190Mb/s



(a) Sum throughput.

(b) Frame error rate.

Figure 7.3: Simulation results for different feedback rates, with and without interference estimation. 4-tap Gaussian channel model.

in the interference limited regime. We can see that the perfect CSI curve flattens at 312Mb/s, which is the throughput of 4 streams using 256-QAM 3/4, the maximum rate MCS in Table 7.1. If the interference estimation is not performed (the curves tagged as *No Est*) then the results are dramatically different. At low SNR values, the evolution of these curves is similar to the more sophisticated algorithm, as in this region the system is noise limited rather than interference limited. As the SNR increases, there is a huge degradation with respect to the perfect CSI case. Performance is penalized at high SNR where throughput decreases dramatically. This unexpected behavior is caused by the overestimation of the SNR due to the interuser interference that is not taken into account, which causes the classifier to choose an MCS with a higher rate than the channel can support. This leads to misclassified samples, leading to high FER values that drive the throughput towards zero. From a learning perspective, the feature set does not contain enough information to perform the adaptation. This information is implicitly included in the SNR values when performing interference estimation.

In Figure 7.3b, we plot the evolution of the FER with the average SNR. We can see that in the perfect CSI case as well as in the cases where the interference is estimated, the FER constraint of 0.1 is always met. This result shows the robustness of the proposed approach even with limited feedback. If the interference is not estimated, then the FER grows up to 1 due to the mismatch between the selected MCS and the actual SNR values, which causes the throughput to decrease, as previously explained. The classifier is able to correctly perform MCS selection in a multiuser setting despite being trained in a single user scenario, thus showing also

the robustness of the classifiers against changes on the channel distribution. This plot shows the importance of performing the interference estimation procedure.

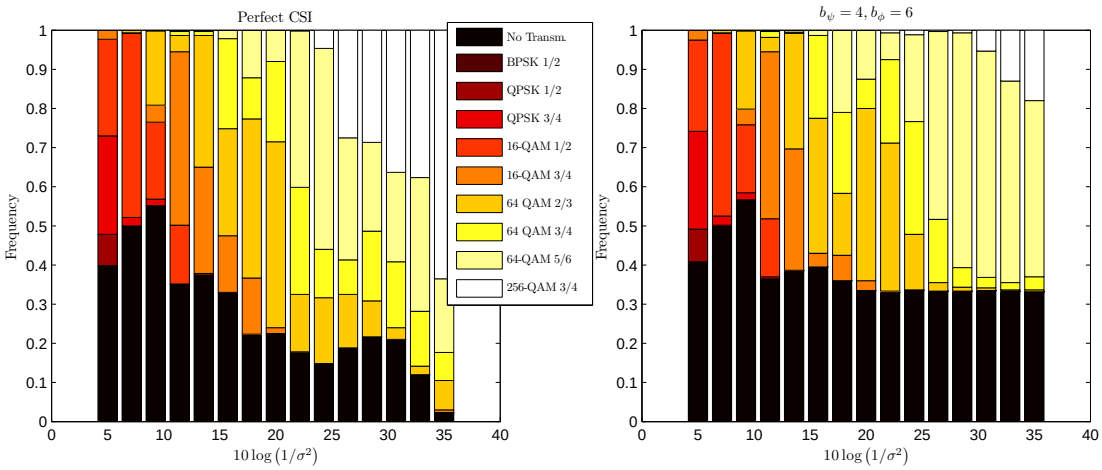


Figure 7.4: Evolution of the frequency of the different MCS with the average SNR for the perfect CSI case and limited feedback with interference estimation.

In Figure 7.4, we plot the fraction of the selected MCS for different SNR values, for both perfect CSI and limited feedback. In the perfect CSI case, the number of scheduled users is usually low for low SNR values (the *No Transmission* frequency, representing the case where a user is not scheduled for transmission, is quite high) and the MCS are robust, while for higher SNR the *No Transmission* frequency decreases and higher rate MCS are employed for transmission. In the limited feedback case the behavior is similar to perfect CSI for low SNR values, but for higher SNR the interuser interference forces to use more robust MCS and approximately one third of the time the *No Transmission* MCS is selected.

Now we study the frequency of a given number of users being selected in Figure 7.5. In the perfect CSI case the number of users grows up to three (i.e., all users are scheduled most of the time), while in the limited feedback case transmits to only two users most of the time. The reason is that in the limited feedback case the interuser interference increases with the scheduled number of users, which is the limiting factor in the high SNR regime. The user selection algorithm is able to identify the suboptimality of an aggressive multiuser transmission by estimating the residual interuser interference. The effect of scheduling only two users out of three was also observed in Figure 7.4, where for high SNR values the *No transmission* MCS is selected 33% of the time. This partial scheduling of the users can lead to unfair scenarios where the maximization of the sum rate causes one user not to be scheduled. This behavior can be corrected by changing the objective function of the LA problem, as explained in Section 7.4.

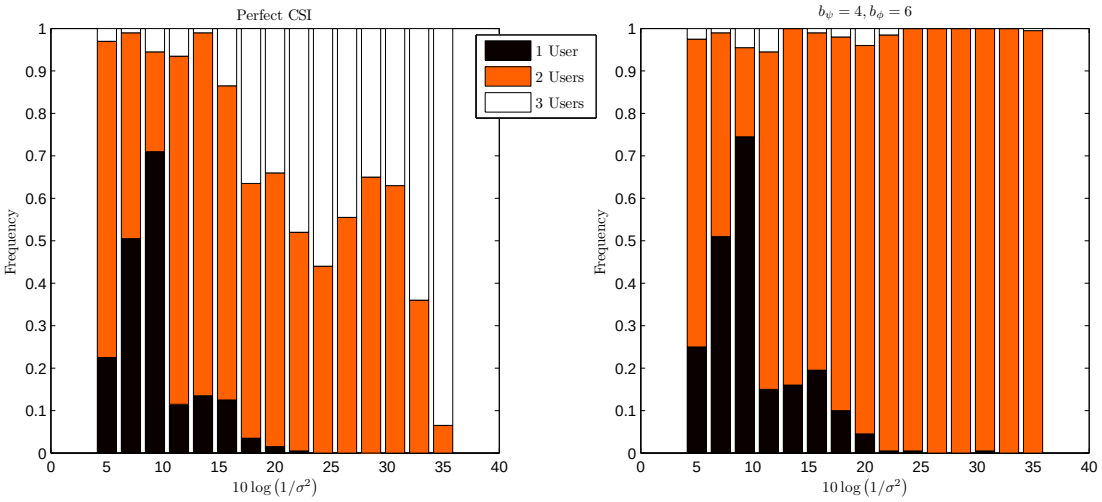
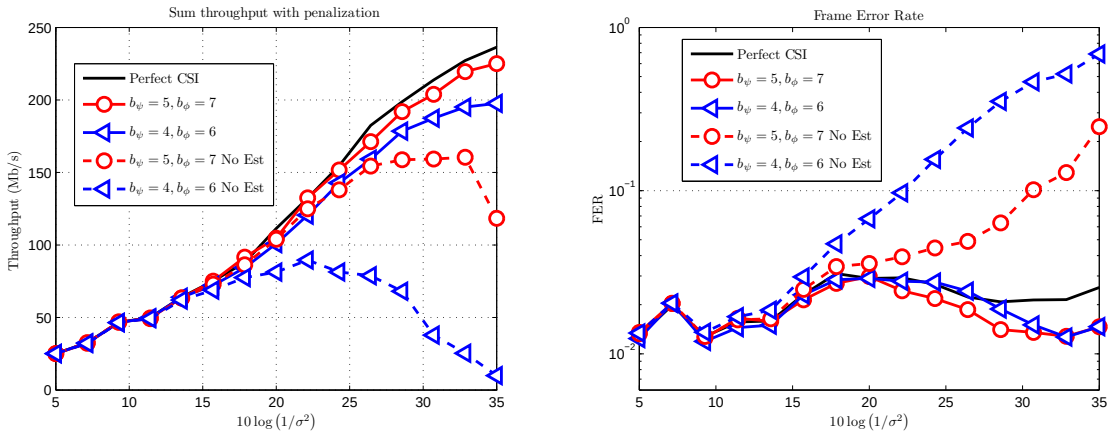


Figure 7.5: Evolution of the frequency of the number of users scheduled for MU transmission with the average SNR for the perfect CSI case and limited feedback with interference estimation.

We also run simulations using TGac channel model B [147] to show the robustness of the proposed approach against changes in the environment or channel model. The classifier is the same as in the previous section, i.e., is trained with 4-tap MIMO channels with iid Gaussian entries. The 802.11 channel model B presents realistic characteristics not present in the 4-tap channel, such as delay taps with different power, and correlation among antennas. The objective of this experiment is to show if the ordered SNR feature vector suffices to characterize the performance of the system even in the presence of a change in the environment. The throughput and frame error rate results are shown in Figure 7.6. We can see that the link adaptation algorithm is able to keep the FER under the required threshold. Once again, interference estimation is shown to be necessary to correctly select the number of users and MCS.

We compared the performance of the proposed adaptation procedure with the same algorithm, but using an effective SNR and average SNR classifier instead of SVM. We run the simulations under TGac channel model B, and assuming perfect CSI at the transmitter. The results are shown in Figure 7.7. The average SNR classifier offers a bad performance, as predicted by the results in Table 7.2. The effective SNR has a slightly worse performance than SVM, especially at low SNR values. At moderate SNR values (around 20dB), SVM offers approximately a 20% gain in throughput.



(a) Sum throughput (b) Frame Error Rate

Figure 7.6: Simulation results for different feedback rates, with and without interference estimation. TGac channel model B. Classifier trained with Gaussian channel model.

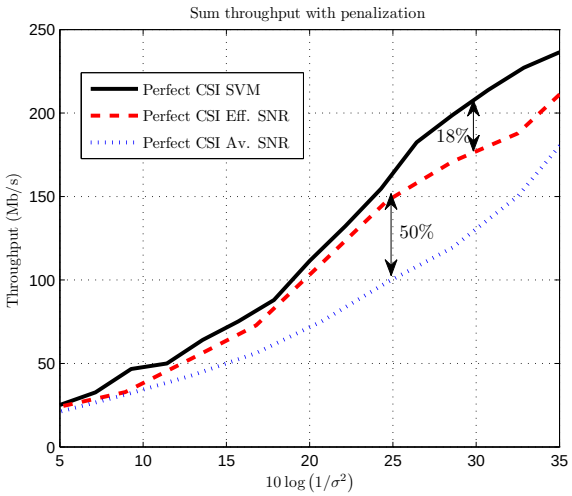


Figure 7.7: Comparison between SVM, Eff. SNR and Av. SNR classifiers

7.9 Conclusions

In this chapter we presented a framework for link adaptation in multiuser MIMO-OFDM networks with limited feedback information. The link adaptation problem is formulated as a maximization of the sum rate subject to a FER constraint. Focusing on the multiuser capabilities in IEEE 802.11ac, we developed a data-driven algorithm that performs user selection, mode selection, MCS selection and takes into account limited feedback information. We showed that estimating the interference due to imperfect CSI is crucial to achieve a good performance in the multiuser MIMO setting. Depending on the feedback rate, and by estimating the residual interference, the transmitter is able to identify the error floor caused by multiuser transmission in the high SNR regime. We conclude that machine learning classifiers can be used in a multiuser setting, even in limited feedback scenarios. Performing interference estimation, however, is crucial for the algorithm performance. The machine learning classifier is also shown to be robust to changes in the statistical distribution of the channel. For example, the information acquired in an open environment can be used to effectively perform adaptation in an office setting, where more frequency selectivity is expected.

The content of this chapter is the result of a research stay in UT Austin under the supervision of Prof. Robert W. Heath Jr., and was partially presented in EUSIPCO 2013 [148] and accepted for publication in IEEE Transactions on Wireless Communications [149]

Chapter 8

FER Prediction under Practical Impairments

Contents

8.1 Introduction

In the previous chapter we solved the problem of link adaptation in MIMO-OFDM by resorting to the use of machine learning techniques. The modulation and coding schemes (MCS) were selected based on machine learning classifiers. These classifiers used as input the signal to noise ratio (SNR) information, obtained from the received feedback information and from the analytical interference estimation.

One of the problems of the approach of Chapter 7 is the impossibility to incorporate practical impairments or different transmission parameters (like codeword length or noise distribution) into the frame error rate (FER) prediction. In IEEE 802.11ac [62], for example, the frames have variable length, which is going to affect the performance of the system. For a constant channel state, longer packets will experience higher FER values [43]. Also, different noise distributions will lead to different FER values, even with the same SNR value. Although the assumption of Gaussian noise is common in system design, a generalized Gaussian distribution can model more accurately some communication scenarios with Laplacian noise [150] or tropospheric impulse noise [151]. Learning-based approaches were developed to overcome these problems [47, 50, 52]. Within the learning framework, the FER prediction is based on observed FER samples, so the effect of impairments is already captured in these measurements. In this setting, the effect of having FER samples with different codeword length or different noise distribution has not been studied.

Previous work [40, 42, 47] focused on predicting the FER from SNR information, without taking into account possible changes in the noise distribution or frame length.

In this chapter, we design FER predictors that are able to capture the effect of these practical impairments. In Section 8.2, we design a machine learning regressor, similarly to the one developed in Chapter 7, but that includes parameters regarding the practical impairments. In Section 8.3 we develop an analytical approach to FER estimation with different codeword length. This approach is based on a maximum likelihood estimation problem, and a parametrization of the FER that depends on the codeword length.

8.2 Learning-based FER estimation

8.2.1 Introduction

In this section, we design machine learning FER predictors that include the effect of different codeword length and different noise distributions. We review effective SNR metrics (ESM) and machine learning FER predictors, and propose modifications to include additional parameters in the prediction. We design a new *semi-parametric* FER predictor by combining ideas from ESM and machine learning. The results show that ESM loses accuracy when practical impairments are present, and that the proposed methods can be used to overcome this problem.

8.2.2 System model

Consider a point to point communication multiple input multiple output (MIMO) orthogonal frequency division multiplexing (OFDM) system where a transmitter, equipped with T antennas, communicates with a receiver, equipped with R antennas. Communication takes place over an N-carrier OFDM physical layer. For every carrier $n = 1, \ldots, N$ we denote by $\mathbf{H}_n \in \mathbb{C}^{R\times T}$ the MIMO channel, by $\mathbf{n}_n \in \mathbb{C}^R$ the received noise vector, and by $\mathbf{x}_n \in \mathbb{C}^T$ the transmit signal. The received signal $\mathbf{y}_n \in \mathbb{C}^R$ is

$$\mathbf{y}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{v}_n \; n = 1, \ldots, N. \tag{8.1}$$

We restrict our analysis to transmitters using linear precoders $\mathbf{F}_n \in \mathbb{C}^{T\times M}$ to spatially conform the transmit symbols $\mathbf{s}_n \in \mathbb{C}^M$ and receive equalizers $\mathbf{G}_n \in \mathbb{C}^{M\times R}$. The number of spatial streams (NSS), also called the *mode*, is denoted by M.

We assume perfect channel state information at both transmit and receive ends. Therefore, we can apply singular value decomposition (SVD) precoding so

$$\mathbf{r}_n = \mathbf{G}_n\mathbf{H}_n\mathbf{F}_n\mathbf{s}_n + \mathbf{G}_n\mathbf{v}_n = \mathbf{\Lambda}_n\mathbf{s}_n + \mathbf{w}_n \; n = 1, \ldots, N \tag{8.2}$$

where $\mathbf{\Lambda}_n$ is a diagonal matrix including the first M singular values of matrix $\mathbf{H}_n$, and $\mathbf{w}_n \triangleq \mathbf{G}_n\mathbf{v}_n$. Design of power allocation is out of the scope of this work, and is assumed to be included in $\mathbf{\Lambda}_n$. The effect of SVD precoding is the decomposition of the MIMO channel in a set of M scalar channels, each one with an input-output relationship described by

$$r_{n,i} = \lambda_{n,i}s_{n,i} + w_{n,i}. \tag{8.3}$$

Therefore, each symbol $s_{n,i}$ passes through a flat fading channel with an SNR value of $\gamma_{n,i} = \frac{|\lambda_{n,i}|^2}{\sigma^2}$. We define the SNR vector $\boldsymbol{\gamma}$ as

$$\boldsymbol{\gamma} = \left[\gamma_{1,1}, \ldots, \gamma_{N,1}, \ldots, \gamma_{N,M}\right]^T. \tag{8.4}$$

In this section, we consider zero-mean generalized complex Gaussian noise [152]. This generalized model allows to treat Laplacian and Gaussian noise as special cases. The probability density function of the real and imaginary parts of the noise is

$$f(x) = \frac{\rho}{2\beta\Gamma(1/\rho)} \exp\left(-\left|\frac{x}{\beta}\right|^{\rho}\right) \tag{8.5}$$

with β the scale parameter, and ρ the shape parameter. Roughly speaking, the parameter ρ changes the rate of decay of the tails of the probability distribution, and β changes the variance

for a fixed ρ. For example, if $\rho = 1$ the noise is Laplacian, and if $\rho = 2$ the noise is Gaussian. The variance of the complex noise is

$$\sigma^2 = \frac{\beta^2}{2} \frac{\Gamma(3/\rho)}{\Gamma(1/\rho)}. \tag{8.6}$$

The transmitted symbols $s_{m,i}$ are the result of processing blocks of bits. Every block of bits is independently processed, and constitutes a frame. The size L of the frame is variable, and depends on the size of higher layer protocol data units. The frame is constituted after performing forward error correction (FEC) coding over blocks of bits, interleaving, and constellation mapping. The transmitter selects the MCS from a discrete set $\mathcal{C} = \{c_1, \ldots, c_C\}$.

The problem we address in this section is how to estimate the FER associated to a channel state, MCS, ρ and L. Particularly, for each MCS value, we are interested in a function

$$\eta(\boldsymbol{\gamma}, \rho, L) \tag{8.7}$$

that maps the set of SNR values $\boldsymbol{\gamma}$, the frame length L and the shape parameter ρ to a FER value. We focus first on the case of Gaussian noise and fixed codeword length L to illustrate two different approaches to FER prediction. We extend these metrics in Section 8.2.4 to deal with practical impairments.

8.2.3 FER prediction techniques

The involved structure of practical coding schemes makes the analytical study of the FER function complicated. In this section, we classify the FER prediction approaches into *parametric* and *non-parametric*. Although both approaches need to use empirical FER results, the main difference is that the parametric approaches require adjusting some parameters following a mean square error (MSE) fitting, for example, while non-parametric methods require adjusting the actual model, usually by cross-validation techniques. In this section, we review ESM and learning FER predictors.

Parametric FER prediction

Parametric techniques assume some functional relationship between the SNR values and the FER, with some parameters to be adjusted according to empirical measurements. This functional relationship is usually expressed as the composition of two different mappings, λ and

γ_{eff} [40]

$$\eta(\boldsymbol{\gamma}) = \lambda(\gamma_{\text{eff}}(\boldsymbol{\gamma})). \tag{8.8}$$

The first one is a mapping from the SNR vector $\boldsymbol{\gamma}$ to an ESM γ_{eff}, defined as the SNR value of an additive white Gaussian noise (AWGN) channel with the same FER as the fading channel under study. This mapping is as a generalized mean that maps the SNR values to a *quality* domain, averages the quality measurements, and maps the value back to the SNR domain. If we denote by $\Theta(\cdot)$ the quality mapping, γ_{eff} is defined as

$$\gamma_{\text{eff}}(\boldsymbol{\gamma}) = \Theta^{-1}\left(\frac{1}{M}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\Theta(\gamma_{i,j})\right). \tag{8.9}$$

In this section, we consider the Exponential Effective SNR metric (EESM) due to its analytical tractability and good accuracy. For example, the WiMAX forum recommended the EESM as the default method for FER prediction [45] in IEEE 802.16e. In EESM, the quality mapping is

$$\Theta(x) = e^{-\beta x} \tag{8.10}$$

with β the parameter to be adjusted with empirical information.

The second mapping is a function from the SNR domain to the FER domain. More precisely, $\lambda(x)$ is the FER of an AWGN channel with SNR x. This mapping is usually performed by the use of look-up tables (LUT) containing simulation results for the AWGN channel. To make a fair comparison with the non-parametric approach, which does not assume any prior FER information in AWGN, in this Section we consider a functional relationship between SNR and FER in AWGN. Particularly, we consider a generalized sigmoid function

$$\lambda(x) = \frac{1}{(1+\exp(b(x-m)))^{1/\nu}} \tag{8.11}$$

with ν, b and m to be fitted to empirical measurements. We verified by simulations that the performance of the FER predictor with the sigmoid function (8.11) and a LUT is similar.

Non-parametric FER prediction

We follow some ideas from [47], and propose to use learning-inspired methods to perform FER prediction. Based on some past samples (the training data), a regressor tries to estimate the function value (the FER in our case) in a different set of samples (the test data). Note that our approach differs from [47, 50], where the objective was to discriminate whether the FER is

above or below a certain threshold, instead of predicting the FER value.

If we assume ideal interleaving, the FER is going to be invariant to permutations in $\boldsymbol{\gamma}$, so the *ordered* SNR vector $\hat{\gamma} = \mathbf{P}\boldsymbol{\gamma}$ suffices to obtain the FER performance. $\mathbf{P}$ is a permutation matrix such that $[\hat{\gamma}]_i \geq [\hat{\gamma}]_{i+1} \ \forall i$.

The regression problem exploits the information in a set of *training data* $\{(\hat{\gamma}_i, y_i)\}_{i=1}^{S}$, consisting of pairs of SNR of different channel realizations $\hat{\gamma}_i$ and its associated FER value y_i. The objective of the regression function is to obtain the FER y_0 associated to a different channel realization $\hat{\gamma}_0$. This regression problem involves two steps: dimensionality reduction and regression.

We select a reduced dimension *feature vector* $\mathbf{f}_i$ from the data $\hat{\gamma}_i$ to avoid the curse of dimensionality [142]. We restrict our study to affine operations, so $\mathbf{f}_i = \mathbf{R}\left(\hat{\gamma}_i - \mathbf{r}\right)$, with $\mathbf{R}$ a dimensionality reduction matrix, and $\mathbf{f}_i \in \mathbb{R}^D$. We study the following dimensionality reduction techniques:

1. **Subset selection** This method simply selects some entries of vector $\hat{\gamma}_i$. Thus, $\mathbf{r} = \mathbf{0}$ and $\mathbf{R}$ is a sparse matrix with D rows taken from of the canonical base of $\mathbb{R}^{NM}$. Although the subset of selected entries could be optimized to gain some performance, in this Section we reduce our analysis to matrices $\mathbf{R}$ selecting equidistant SNR positions, including the first and last ones.

2. **Subset selection with feature scaling** This method selects some entries of the vector $\hat{\gamma}_i$, but performs first an affine transformation to make the different entries of $\hat{\gamma}_i$ zero mean and unit variance. We define the empirical mean and variance of the entries of $\{\hat{\gamma}_i\}_{i=1}^{S}$ as $\mu_k \triangleq \frac{1}{S}\sum_{i=1}^{S}[\hat{\gamma}_i]_k$ and $\sigma_k^2 \triangleq \frac{1}{S}\sum_{i=1}^{S}|\mu_k - [\hat{\gamma}_i]_k|^2$. The dimensionality reduction operation is
$$\mathbf{f}_i = \mathbf{R}\boldsymbol{\Sigma}\left(\hat{\gamma}_i - \boldsymbol{\mu}\right) \tag{8.12}$$
with $\boldsymbol{\Sigma}$ a diagonal matrix with entries $[\boldsymbol{\Sigma}]_{k,k} = \frac{1}{\sqrt{\sigma_k^2}}$, $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_{NM}]^T$, and $\mathbf{R}$ a selection matrix.

3. **Principal component analysis** Principal component analysis (PCA) estimates the mean and covariance matrix of the samples as $\boldsymbol{\mu} = \frac{1}{S}\sum_{i=1}^{S}\hat{\gamma}_i$, $\mathbf{C} = \frac{1}{S}\sum_{i=1}^{S}\left(\hat{\gamma}_i - \boldsymbol{\mu}\right)\left(\hat{\gamma}_i - \boldsymbol{\mu}\right)^T$. The feature set is obtained by projecting the training set onto the dominant eigenmodes: let $\mathbf{C} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ be the eigendecomposition of $\mathbf{C}$ with the eigenvalues sorted in decreasing order, and let $\tilde{\mathbf{U}}$ be the matrix containing the first D columns of $\mathbf{U}$. The dimensionality reduction operation is $\mathbf{f}_i = \tilde{\mathbf{U}}^T\left(\boldsymbol{\gamma}_i - \boldsymbol{\mu}\right)$.

After dimensionality reduction is performed, we build a regression function based on the

reduced dimension training data. The reduced dimension training data is the set $\{(\mathbf{f}_i,\, y_i)\}_{i=1}^{S}$. Although there are a wide variety of non-parametric regression methods, we choose local linear regression (LLR) for its simplicity. LLR approximates the function around a point $\mathbf{f}_0$ by a linear model.

$$\hat{y}_0 = \alpha + \boldsymbol{\beta}^T \mathbf{f}_0. \tag{8.13}$$

The parameters α and $\boldsymbol{\beta}$ are obtained from weighted least squares (WLS) fitting. The weights of the WLS problem depend on the distance between $\mathbf{f}_0$ and the different training samples $\mathbf{f}_i$. The WLS problem is

$$\min_{\alpha,\boldsymbol{\beta}} \sum_{i=1}^{S} K_\lambda\left(\mathbf{f}_0, \mathbf{f}_i\right) \left\|\alpha + \boldsymbol{\beta}^T \mathbf{f}_i - y_i\right\|^2 \tag{8.14}$$

where K_λ is a kernel function parametrized by the value λ.We use the radial basis function kernel:

$$K_\lambda\left(\mathbf{x}, \mathbf{y}\right) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\lambda}\right). \tag{8.15}$$

The value of λ determines a point in the bias vs variance tradeoff [53].

8.2.4 Inclusion of practical impairments

The methods described in the previous section perform FER prediction based only on SNR information. In real systems, however, there are other factors that impact the performance of a receiver. In this section we describe how to include practical impairments in parametric and non-parametric techniques, and present a new semi-parametric approach.

Parametric FER prediction

Parametric approaches are difficult to adapt to include practical impairments, as a functional relationship between the FER and the practical impairment has to be obtained or approximated. In the case of variable frame length, for example, previous work assumed the availability of a different FER predictor for every length [40, 43]. In this Section, we assume only one FER predictor for every MCS, so the parametric FER predictors do not take into account the practical impairments.

Non-parametric FER prediction

The non-parametric FER prediction techniques offer a flexible way to deal with practical impairments, since they do not assume any functional relationship between the FER and the channel

state. Thus, we can define an *extended channel state* that includes the practical impairment as part of the channel state. Assume that we have a training set where each sample is associated to a channel state vector and to a practical impairment vector. Denote $\mathbf{p}_i \triangleq [p_{i,1}, \ldots, p_{i,\mathcal{P}}]^T$ as the practical impairment vector of the i-th sample. The training set is a set of tuples $\{\hat{\gamma}_i, \mathbf{p}_i, y_i\}_{i=1}^S$. The extended channel state vector is defined as $\mathbf{e}_i = \left[\hat{\gamma}_i^T, \mathbf{p}_i^T\right]^T$. With this definition, we can redefine our training set as $\{\mathbf{e}_i, y_i\}_{i=1}^S$ and apply the dimensionality reduction methods in Section 8.2.3.

The inclusion of practical impairments in the channel state vector increases the dimensionality of the problem and, therefore, makes the use of dimensionality reduction even more important. In the following section, we design a semi-parametric approach that combines ideas from ESM and machine learning approaches to avoid the problem of dimensionality reduction.

Semi-parametric FER prediction

Parametric and non-parametric methods have some advantages and drawbacks. On the one hand, parametric methods use simple functional relationships between FER and SNR, work with a relatively low number of empirical samples, but are not flexible to accommodate practical impairments. On the other hand, non-parametric methods can deal with practical impairments in a straightforward manner, but need a large number of training samples to include additional features.

One key observation is that ESMs are designed to be *good* dimensionality reduction techniques, i.e., ideally the effective SNR $\gamma_{\text{eff}}(\hat{\gamma})$ is a sufficient statistic for FER estimation. Thus, we propose to use an alternative extended channel state vector, defined as

$$\mathbf{e}_i = \left[\gamma_{\text{eff}}(\hat{\gamma}_i), \mathbf{p}_i^T\right]^T. \tag{8.16}$$

The estimation process involves two steps. In the first one, the optimum ESM parameter β is obtained from the training samples without taking into account the practical impairments. In the second one, the value of β is used to build the extended channel state vector (8.16), and an LLR estimator is trained following the procedure in 8.2.3. Dimensionality reduction is not performed (the number of practical impairments is expected to be small, and we have already reduced the size of the SNR vector to one), but feature scaling might be necessary.

8.2.5 Simulation results

We evaluated the described FER prediction methods for a MIMO-OFDM system with a 4-antenna transmitter-receiver pair and 52 carriers, emulating the PHY of IEEE 802.11ac [62]. The MCS and its associated rates can be found in Table 7.1.

We performed different experiments to compare the performance of the FER prediction methods. First, we compared parametric and non-parametric methods when no practical impairments are present. We generated 6000 different realizations with SNR values between 0 and 30dB, and a 4-tap MIMO channel with Gaussian entries in the time domain. The frame length was set to $L = 1024$, and the noise distribution was Gaussian. We simulated the transmission over the channel with QPSK modulation, rate 3/4 convolutional code, and 4 spatial streams. We divided the 6000 data points into two different parts: the training data, comprising 80% of the points, and the test data, with 20% of the points. We trained our regressor with the 4800 samples, and tested it against the remaining 1200. We compared the ESM FER predictor with LLR with the 3 different types of dimensionality reduction: LLR-PCA (dimensionality reduction with PCA), LLR-SS (dimensionality reduction with subset selection), LLR-SS-SC (dimensionality reduction with subset selection and feature scaling). Also, we evaluated LLR-SS and LLR-SS-SC with the SNR values in decibels instead of natural units.

We selected the kernel parameter λ following a K-fold cross-validation approach [53], with K=4. This implies that every iteration in this cross-validation used 3600 samples to train the regressor and 1200 to test it. After selecting the value of λ, the complete training set was used to train the LLR.

In Figure 8.1 we show a plot of the FER estimation MSE as a function of the dimension of the feature vectors $\mathbf{f}_i$. ESM MSE is plot as a constant for comparison. We see that LLR-SS with SNR in dB outperforms ESM for number of features above 8. LLR-SS with SNR values in linear scale performs worse than ESM, and PCA offers a poor performance.

We performed similar experiments introducing practical impairments in the system. In all the cases, the division of the available samples into training, test and cross validation sets was the same as in the no-impairments case. In Figure 8.2 we show the results for different frame length. We generated 9000 realizations with the same channel and SNR distribution as before, Gaussian noise, but now varying the frame length between $L = 128$ and $L = 16386$. In this case, the estimation accuracy of ESM is hindered by the lack of information of the frame size. The inclusion of practical impairments reduces the accuracy of ESM in almost one order of magnitude. LLR with SNR in dB with L as a feature outperforms ESM, and the proposed semi-parametric approach (LLR-ESM(dB)) offers the best performance. The semi-parametric

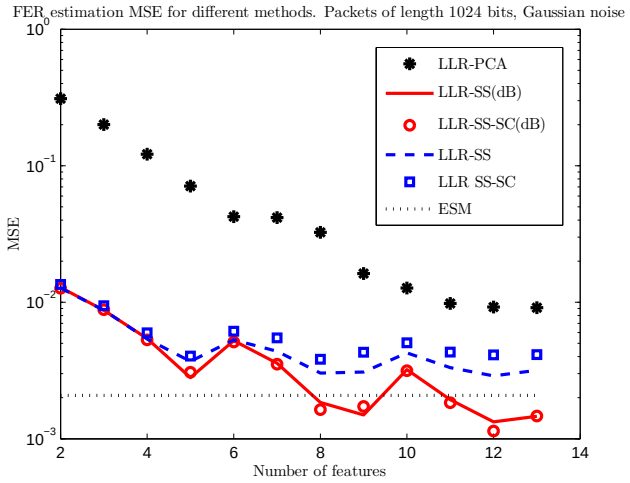


Figure 8.1: FER prediction without practical impairments.

approach is built as a 2-feature LLR, with the ESM in dB, i.e., $10\log_{10}\left(\gamma_{\text{eff}}\left(\boldsymbol{\gamma}\right)\right)$ with $\boldsymbol{\gamma}$ in natural units.

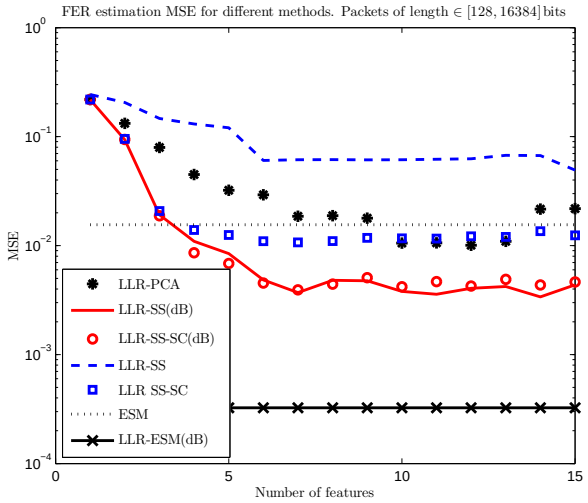


Figure 8.2: FER prediction with variable frame length.

A similar behavior is shown for the generalized Gaussian distribution in Figure 8.3. We generated 12000 channel realizations with a constant frame length of $L = 1024$ but varying the ρ parameter of the generalized Gaussian distribution between 0.1 and 4. In this case, the feature vector contains the ρ value of the corresponding channel.

In Figure 8.4 we show the result of applying parametric and semi-parametric approaches to

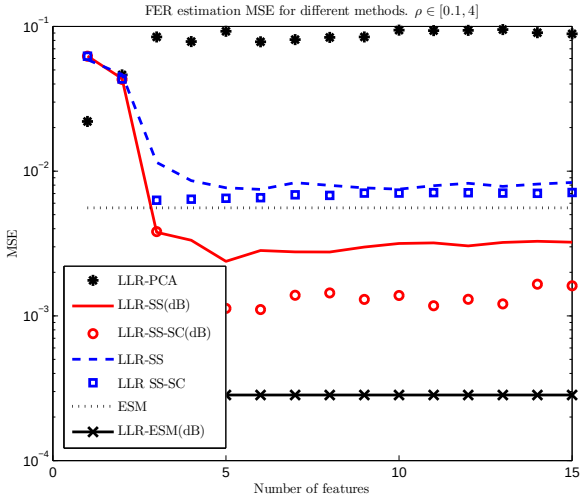


Figure 8.3: FER prediction with different noise distribution.

the problem of link adaptation with variable codeword length. A FER predictor was built for every MCS, and the MCS with a higher throughput meeting a FER constraint of $p_0 = 0.1$ was selected. The frame length was randomly selected between 128 and 13684 bits. The proposed semi-parametric approach offers up to 11% throughput gain for moderate SNR values. Also, it was observed that ESM did not meet the FER constraint in some cases due to the FER prediction inaccuracy. This results shows the advantage of taking into account practical impairments in link adaptation.

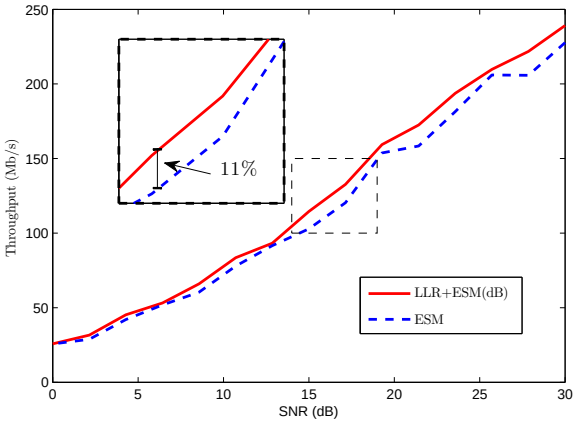


Figure 8.4: Link adaptation throughtput for variable frame length.

8.3 Maximum Likelihood FER estimation

8.3.1 Introduction

In the previous section we presented a method to perform link adaptation in the presence of practical impairments. These practical impairments are added as new features in the machine learning regression problem. Thus, we do not exploit any prior information between the FER and the different practical impairments. In this section we present an estimation theoretic method to perform FER estimation for codewords of variable length. As opposed to [40, 43], we assume the availability of a limited number of FER predictors, calibrated to estimate the FER of a small number of codeword sizes. Unlike [31, 46, 58, 59], we account for the FER prediction error of the different codeword sizes. The proposed FER estimator is derived under a binary symmetric channel (BSC) assumption, where the FER estimation is imperfect as a result of having a finite observation window. We extend the result to coded systems under frequency and space selective channels by the use of effective SNR FER predictors. The use of the proposed estimation method enables accurate FER prediction for a wide range of codeword lengths while keeping a low computational complexity.

8.3.2 FER estimation for uncoded systems

Consider a transmitter-receiver pair communicating through a noisy channel. The transmitter builds blocks of bits $[b_1, \ldots, b_L]$ of variable length L, with $b_i \in \{0, 1\}$, and the receiver observes $\left[\hat{b}_1, \ldots \hat{b}_L\right]$ at the output of the channel, with $\hat{b}_i \in \{0, 1\}$. The channel is memoryless and symmetric with error probability p, i.e., $p \triangleq \mathbb{P}\left[\hat{b}_i = 1 | b_i = 0\right] = \mathbb{P}\left[\hat{b}_i = 0 | b_i = 1\right]$. We assume that the transmitted block of bits contains an error detection code such that the receiver is able to identify the received blocks with at least one erroneous bit. We also assume that the error detection code is designed to make the missed detection probability negligible. If we denote by θ_L the probability of receiving an erroneous block of length L, we have that

$$\theta_L = 1 - \mathbb{P}\left[\bigcap_{i=1}^{L} \left(b_i = \hat{b}_i\right)\right] = 1 - (1-p)^L . \tag{8.17}$$

It can be seen from (8.17) that if the FER for a frame length L is known perfectly, we can obtain the FER for a different frame length $\tilde{L}$ as

$$\theta_{\tilde{L}} = 1 - (1 - \theta_L)^{\tilde{L}/L} . \tag{8.18}$$

In a realistic scenario, however, it is very unlikely that an exact estimate of the FER is available for any length. In general, the available FER estimate is going to be the result of observing the success and failures of frames of a certain length during a time period. Moreover, observations of frames of different length may improve our FER estimate. In the following, we formalize the problem of estimating the FER from observations of frames of different size.

Consider a communication system with ℓ different frame sizes $L_1, \ldots, L_\ell$. During an observation period, a receiver observes n_i transmissions of size L_i, out of which m_i are received with errors. m_i is binomially distributed with parameters n_i and θ_{L_i}, i.e.

$$p\left(m_i; \theta_{L_i}\right) = \binom{n_i}{m_i} \theta_{L_i}^{m_i} \left(1 - \theta_{L_i}\right)^{n_i - m_i}. \tag{8.19}$$

The maximum likelihood estimate (MLE) of θ_{L_i} given the observation m_i is the measured FER, i.e., $\hat{\theta}_{L_i} = \frac{m_i}{n_i}$. The MLE is unbiased with variance

$$\sigma_i^2 \triangleq \mathbb{E}\left[\left(\theta_{L_i} - \hat{\theta}_{L_i}\right)^2\right] = \frac{\left(1 - \theta_{L_i}\right)\theta_{L_i}}{n_i}. \tag{8.20}$$

If there are observations from only one length (i.e., $\ell = 1$), we can relate the MLE of θ_{L_1} and $\theta_{\tilde{L}}$ by using the invariance property [102] and (8.18) as $\hat{\theta}_{\tilde{L}} = 1 - \left(1 - \hat{\theta}_{L_1}\right)^{\tilde{L}/L_1}$.

If $\ell > 1$, however, the derivation of the MLE of $\theta_{\tilde{L}}$ is more involved. For a vector of observed errors $\mathbf{m} \triangleq [m_1, \ldots, m_\ell]^T$, the probability mass function of $\mathbf{m}$ parametrized by the FER $\theta_{\tilde{L}}$ can be easily obtained just by assuming independent observations and by applying (8.18) as

$$p\left(\mathbf{m}; \theta_{\tilde{L}}\right) = \prod_{i=1}^{\ell} \binom{n_i}{m_i} \left(1 - \left(1 - \theta_{\tilde{L}}\right)^{L_i/\tilde{L}}\right)^{m_i} \times \left(1 - \theta_{\tilde{L}}\right)^{(n_i - m_i)L_i/\tilde{L}}. \tag{8.21}$$

It is possible to calculate the Fisher information matrix from (8.21) and conclude that a minimum variance unbiased estimator of $\theta_{\tilde{L}}$ does not exist [102]. The MLE seems also difficult to calculate, since the likelihood function is nonconcave in $\theta_{\tilde{L}}$ and, therefore, maximizing it would require a grid search. We propose a simple and computationally efficient approach consisting on a linear combination of ℓ MLE of $\theta_{\tilde{L}}$, each one obtained from the observations from a different length. First, let us denote as $\hat{\theta}_{\tilde{L}}\left(L_i\right) \triangleq 1 - \left(1 - \hat{\theta}_{L_i}\right)^{\tilde{L}/L_i}$ the MLE of $\theta_{\tilde{L}}$ from the measurements

of length L_i. We propose to estimate $\theta_{\tilde{L}}$ by a linear combination of $\hat{\theta}_{\tilde{L}}(L_1), \ldots, \hat{\theta}_{\tilde{L}}(L_\ell)$:

$$\hat{\theta}_{\tilde{L}} = \sum_{i=1}^{\ell} \beta_i \hat{\theta}_{\tilde{L}}(L_i) \tag{8.22}$$

with $\{\beta_i\}$ a set of weights to be designed. Figure 8.5 contains a diagram of the proposed estimation scheme. Some simple direct values of β_i will serve as our baseline for comparison:

- **Average**: $\beta_i = 1/\ell \ \forall i$.
- **Closest**: $\beta_i = \begin{cases} 1 & \text{if } i = \arg\min \left| L_i - \tilde{L} \right| \\ 0 & \text{otherwise} \end{cases}$.

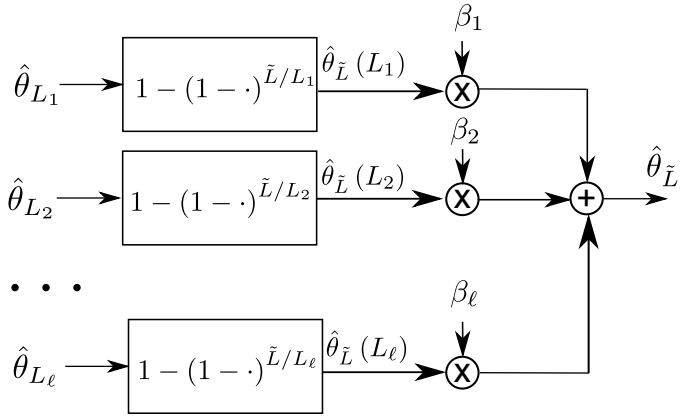


Figure 8.5: Proposed linear estimation scheme: the estimated FER is obtained as the linear combination of ℓ MLE.

Obtaining the optimum weights β_i is quite involved as the MLEs $\hat{\theta}_{\tilde{L}}(L_i)$ are biased in general (even when the MLEs $\hat{\theta}_{L_i}$ are not). Also, the variance of $\hat{\theta}_{\tilde{L}}(L_i)$ depends on the parameter to estimate. In the following, we derive a value for the β_i by assuming a sufficiently large number of observations n_i.

If we have a sufficiently large number of observations n_i, the distribution of the MLE $\hat{\theta}_{L_i}$ can be approximated as [102] $\hat{\theta}_{L_i} \sim \mathcal{N}\left(\theta_{L_i}, \sigma_i^2\right)$, so we can write

$$\hat{\theta}_{\tilde{L}}(L_i) = 1 - (1 - \theta_{L_i} + w_i)^{\alpha_i} \tag{8.23}$$

with $\alpha_i = \tilde{L}/L_i$ and $w_i \sim \mathcal{N}\left(0, \sigma_i^2\right)$. As we are using a small variance approximation, we approximate $\hat{\theta}_{\tilde{L}}(L_i)$ by the first order Taylor expansion series around $w_i = 0$ as

$$\hat{\theta}_{\tilde{L}}(L_i) \approx \theta_{\tilde{L}} + w_i \alpha_i \left(1 - \theta_{\tilde{L}}\right)^{\frac{\alpha_i - 1}{\alpha_i}}. \tag{8.24}$$

From (8.24) we can see that the MLE $\hat{\theta}_{\tilde{L}}(L_i)$ is unbiased for large n_i, as predicted by the asymptotic properties of any MLE. The variance of the estimator $\hat{\theta}_{\tilde{L}}(L_i)$ is

$$\begin{aligned} \xi_i^2 &\triangleq \mathbb{E}\left[\left(\hat{\theta}_{\tilde{L}}(L_i) - \theta_{\tilde{L}}\right)^2\right] = \sigma_i^2 \alpha_i^2 \left(1-\theta_{\tilde{L}}\right)^{2\frac{\alpha_i-1}{\alpha_i}} \\ &= \frac{\alpha_i^2\left(1-\left(1-\theta_{\tilde{L}}\right)^{1/\alpha_i}\right)\left(1-\theta_{\tilde{L}}\right)^{\frac{2\alpha_i-1}{\alpha_i}}}{n_i}. \end{aligned} \tag{8.25}$$

From the values ξ_i^2, the linear fusion weights which minimize the variance of $\hat{\theta}_{\tilde{L}}$ in (8.22) are

$$\beta_i = \frac{1/\xi_i^2}{\sum_{i=1}^{\ell} 1/\xi_i^2}. \tag{8.26}$$

Note that the MSE ξ_i depends on $\theta_{\tilde{L}}$, which is the parameter to be estimated. Thus, the parameters β_i cannot be obtained directly. It is expected, however, that the parameters β_i are not very sensitive to small changes in $\theta_{\tilde{L}}$. We propose a two-step estimation. First, obtain an initial estimate $\hat{\theta}_{\tilde{L},0}$ of $\theta_{\tilde{L}}$ by averaging the MLEs $\hat{\theta}_{\tilde{L}}(L_i)$ with weights $\beta_i = 1/\ell$. Second, use $\hat{\theta}_{\tilde{L},0}$ as the true value of $\theta_{\tilde{L}}$ to calculate the ξ_i and then obtain the optimum weights β_i according to (8.26).

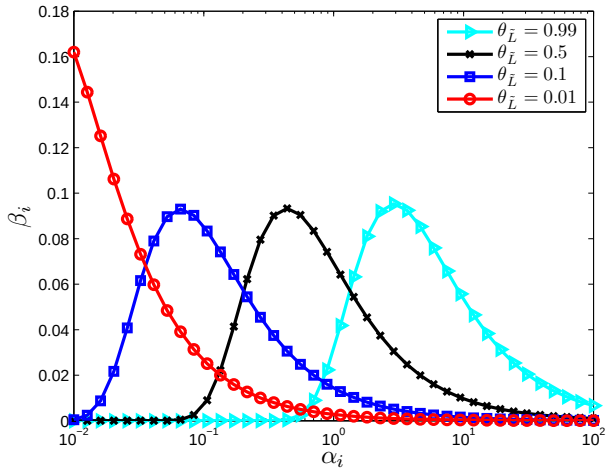


Figure 8.6: Linear combination weights β_i as a function of the error probability. We see that, in general, $\alpha_i = 1$ (same length) is not a maximum. Low α values have more weight for low FER values.

In Figure 8.6 we show the evolution of β_i with α_i for different $\theta_{\tilde{L}}$, and a constant n_i. It can be seen that the maximum weight is not given to the samples with similar L_i (i.e., $\alpha \approx 1$), and depends on the operating regime. For example, in low FER values ($\theta_{\tilde{L}} = 0.01$) more weight

is given to samples from longer packets (small α), since in those packets the error probability is going to be larger and, therefore, more error events can be observed. The opposite behavior is observed in the high FER region ($\theta_{\tilde{L}} = 0.99$), where shorter packets provide better error estimates.

We evaluated the performance of the proposed estimation approach in a BSC with error probability p. In this setting, the FER is exactly $\theta_L = 1 - (1-p)^L$, so we can compare the obtained result with the exact FER value. We set $\ell = 5$, with $\mathbf{L} = [L_1, \ldots, L_\ell] = [100, 1000, 5000, 8000, 10000]$ and $n_i = N\,\forall i$, with $N = 10, 100, 1000$. Our objective is to estimate the FER with $\tilde{L} = 2000$. We compare the results with the two simpler estimation approaches already mentioned. The results are shown in Figure 8.7, where the figure of merit is the normalized mean squared error (NMSE), defined as $\mathrm{NMSE}(dB) = 10\log_{10}\left(\frac{1}{K}\sum_{i=1}^{K}\frac{\left(\hat{\theta}_{\tilde{L},i}-\theta_{\tilde{L}}\right)^2}{\theta^2_{\tilde{L},i}}\right)$ with $\hat{\theta}_{\tilde{L},i}$ the FER estimate from the i-th realization of the observations. We averaged the results over $K = 10^4$ realizations of the observations. We can see that the proposed approach outperforms the more naive estimates for almost all values of p and N. The gain is especially noticeable when the number of observations is high ($N = 100, 1000$), and can be as large as 10dB. From the figure, we can observe that the effect in NMSE reduction when applying the proposed method is approximately the same as multiplying the number of observations by a factor of 10, especially for low p values.

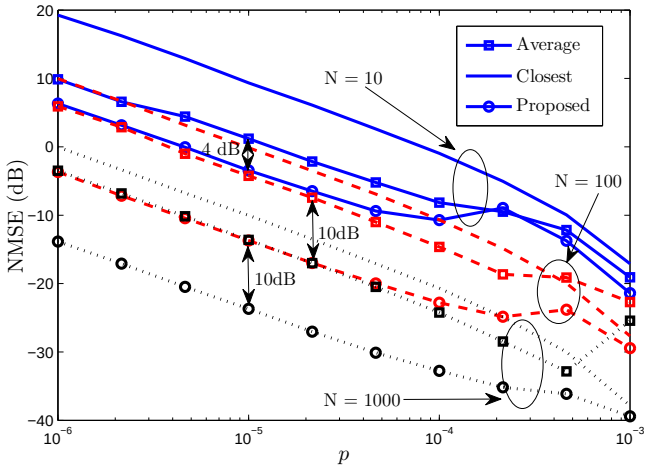


Figure 8.7: NMSE of FER prediction with different estimation methods as a function of p. Different colors are used for different codeword lengths: $N = 10$ (blue), $N = 100$ (red), $N = 1000$ (black).

We also compared the proposed method against a scenario where all the observations are of the desired length. For example, for $N = 10$, our method observes 10 frames of each length in

$\mathbf{L} = [100, 1000, 5000, 8000, 10000]$, and we compare it against the case of observing 50 frames of length 2000. In the latter case, the MLE of the FER is simply the observed FER, and its variance is given by (8.20). We show the results in Figure 8.8. The proposed method with observations of different lengths outperforms the MLE with observations of frames of the desired length, especially for low p and large N values.

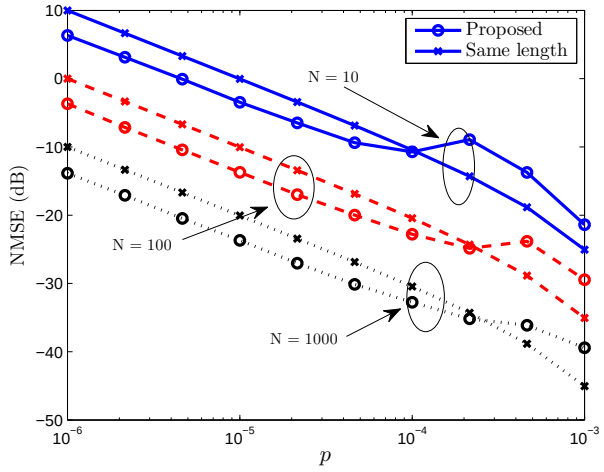


Figure 8.8: NMSE of FER prediction compared with observations of the desired length, for different p and N values.

8.3.3 FER estimation for coded systems

We exploit the insights obtained for the BSC to perform FER prediction with different codeword length in coded systems under frequency and space selective channels. Although in general it is difficult to obtain a good approximation for the FER in coded systems, the relationship between the FER and CBER p_{coded} can be approximated as $\theta_L = 1 - (1 - p_{\text{coded}})^{L/D}$, where D is a parameter that depends on the code. For example, D is the minimum distance of the code in the case of convolutional codes, and the number of coded bits in the case of LDPC codes [59]. The relationship between FER of two different lengths is given by (8.18), thus not depending on the actual value of D.

We perform FER prediction by the use of the exponential effective SNR metric (EESM). Prediction is performed in two steps. First, the set of post-processing SNR values $\gamma_1, \ldots, \gamma_K$,

are mapped to a single SNR value as

$$\gamma_{\text{eff}} = -\frac{1}{\beta} \log \left(\frac{1}{K} \sum_{i=1}^{K} e^{-\beta \gamma_k} \right) \tag{8.27}$$

where β is a calibration parameter to be fitted according to empirical results. Second, the FER estimate for a certain length L is obtained as $\hat{\theta}_L = \text{FER}_{\text{AWGN},L}(\gamma_{\text{eff}})$ with $\text{FER}_{\text{AWGN},L}$ an empirically obtained function mapping an SNR value to the FER in AWGN for a codeword length L. Note that both the calibration parameter β and the FER in AWGN are going to depend on L. Also, if FER prediction is performed for different modulation and coding schemes, a different β and FER function has to be obtained for each case.

We evaluated the proposed estimation procedure for FER prediction under the IEEE 802.11ac standard [62]. We selected MCS QPSK 3/4 with two spatial streams, a transmitter-receiver pair with 2 antennas each, and SVD precoding. We obtained FER samples for codewords with length $\mathbf{L} = \left[2^7, \ldots, 2^{14}\right]$, i.e., codeword lengths between 128 and 16384 bits. We trained 8 EESM estimators (one for each length) with 500 FER samples for different SNR values and Rayleigh channel model. The SNR values were selected between 0 and 30 dB, so FER values between 0 and 1 had to be estimated. For each of the 8 length values, we estimated the FER using the other 7 FER predictors (with the 3 proposed linear fusion rules), and compared the result with the FER prediction from the EESM estimator of codewords of the same length. Formally, the proposed FER estimator is

$$\hat{\theta}_{\tilde{L}} = \sum_{i=1, L_i \neq \tilde{L}}^{\ell} \beta_i \left(1 - \left(1 - \text{FER}_{\text{AWGN},L_i}(\gamma_{\text{eff}})\right)^{\alpha_i}\right). \tag{8.28}$$

In Figure 8.9 we show a diagram of the proposed estimation scheme.

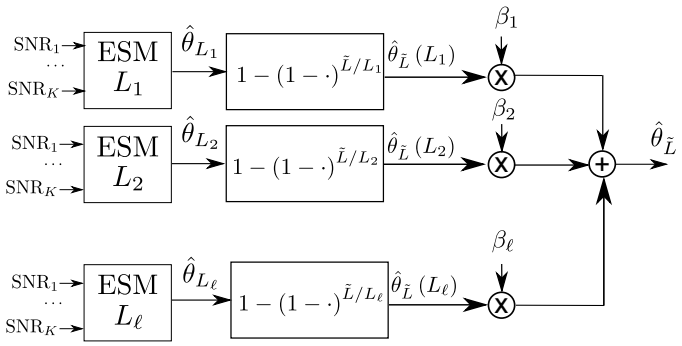


Figure 8.9: FER estimation from linear fusion of ESM estimates.

We obtained the designed weights β_i by assuming a constant n_i in (8.25). We show the

results in Figure 8.10. We can see that the proposed estimator outperforms the other linear fusion approaches for all cases, except for the higher length case, where it attains approximately the same NMSE as the *Closest* combining. Also, for 6 of the 8 lengths, the proposed estimator outperforms the EESM FER prediction trained with samples of the same length. In some cases, like $\tilde{L} = 2^{10}$, the NMSE gain with respect to the same length FER predictor is in the order of 4dB.

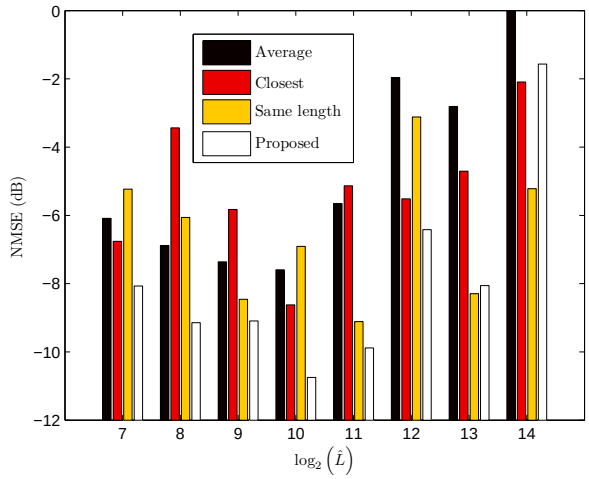


Figure 8.10: NMSE of FER prediction for different frame sizes in IEEE 802.11ac, QPSK 3/4.

8.4 Conclusions

In this chapter we presented an approach to perform FER estimation in communication systems using codewords of different length and different noise statistics. Machine learning FER estimators can incorporate the practical impairments as additional features, while traditional parametric approaches are less flexible.

Classical ESM without further modifications loses accuracy in the presence of different frame length or noise distribution. We proposed first a semi-parametric approach that combines the good properties of both parametric and non-parametric methods.

Finally, we developed an estimation theoretic approach to FER estimation with observations of codewords of different length, and a simple way to combine them. The proposed estimator with the designed linear combination is shown to outperform other simpler estimation methods. We also show that a MLE with samples of the desired length can perform worse that the proposed estimator.

The results of both sections show the importance of incorporating practical impairments into the FER predictors.

The content of this chapter was partially published in Asilomar 2013 [153] and ICASSP 2014 [154], co-authored by Prof. Robert W. Heath Jr. and Prof. Carlos Mosquera.

Chapter 9

Conclusions and Future Work

Contents

9.1 Conclusions

This thesis focuses on techniques to increase the spectral efficiency of wireless networks. Particularly, methods based on cognitive radio and link adaptation are proposed. These two lines of research cover two different approaches to the improvement of spectral efficiency: cognitive radio-based techniques provide spectral efficiency gains by the use of novel spectrum access techniques, while the use of advanced link adaptation techniques can increase the efficiency of currently deployed wireless systems with little or none modifications on the standards.

Focusing on the cognitive radio contributions, we analyzed different scenarios under the overlay paradigm. Under this paradigm, a secondary transmitter, which is trying to access the spectrum owned by a primary transmitter, knows the message to be conveyed by the latter. We focused on the problems that arise when channel state information (CSI) is not available at the transmitter, and analyzed both point to point and broadcast scenarios.

In Chapters 2, 3 and 5 we studied the effect of a cooperative secondary transmitter without CSI when the primary transmitter uses orthogonal frequency division multiplexing (OFDM). In Chapters 2 and 3 the primary receiver was assumed to be a broadcaster, so CSI acquisition was not possible. The insertion of a cooperative secondary transmitter creates an artificial multipath

that degrades the quality of service of the primary receivers, even when the received signal to noise ratio (SNR) is increased. We presented an analytical approach to quantify the effect of the secondary echoes, based on effective SNR metrics, and develop transmission techniques to alleviate the degradation caused by the artificial multipath. In this scenario, interference cancellation can be performed at the receivers if the primary message is decodable. The results of these chapters can be useful to design overlay cognitive radio networks and for broadcast network planning with single frequency operation.

In Chapter 5 the primary system was point to point, and a simple and inexpensive CSI acquisition scheme was proposed. Under this setting, we obtained the optimum signals for CSI acquisition, and the optimum power allocation to maximize the primary rate. We extended a previous CSI acquisition technique, which was designed for the single antenna time invariant channel, to the multiple input multiple output (MIMO) and time varying case. The results in this chapter offer new insights on the design of cognitive radio systems without CSI at the transmitter, and show how a simple CSI acquisition scheme can be used to increase the performance of the system.

Chapter 4 analyzes the problem of designing a secondary system that works under heavy interference of the primary system. If the transmitter does not have access to CSI, then it cannot exploit the knowledge of the primary signal to perform interference cancellation. In such a case, the optimum power allocation at the secondary transmitter can extend the coverage area of the primary system, which results in a coexistence that is beneficial for both primary and secondary users.

Link adaptation is analyzed under mobile satellite and MIMO-OFDM scenarios in Chapters 6 to 8. The satellite scenario is challenging due to the large propagation delay, which makes timely CSI difficult to acquire. In Chapter 6 we proposed adaptation techniques for the forward and return links of mobile satellite channels. In the forward link we designed two different methods, based on the use of retransmissions and multilayer coding, that rely on statistical CSI only. For the return link, we developed a novel technique that exploits both open loop and closed loop CSI. These techniques can be applied to increase the spectral efficiency of mobile satellite systems while keeping a good reliability.

The difficulty of the MIMO-OFDM system resides on the large dimensionality of CSI, and on the errors induced by the acquisition of CSI by means of a limited feedback channel. This last problem is especially challenging in systems using space division multiplexing, where CSI errors are critical. In Chapter 7 we developed an algorithm to perform multiuser transmission with limited feedback information. The algorithm includes a machine learning classifier to select the modulation and coding scheme, and a novel interference estimation technique that allows

obtaining the multiuser interference without knowing the actual values of the channel. Chapter 8 dealt with a different aspect of link adaptation, namely the effect of practical impairments such as different codeword length and noise distribution. We developed a learning-based approximation to frame error rate prediction with general practical impairments, and an estimation theoretic solution that focuses on different codeword lengths. The results in these two chapters can be applied to link adaptation in IEEE 802.11-like systems to improve performance, especially in scenarios with practical impairments and multiuser capabilities.

In a whole, the work in this thesis is based on the analysis of previously developed ideas under realistic settings. For example, ideas like the degradation of single frequency networks due to echoes and the multiuser estimation interference are novel to the best of our knowledge. These ideas may help achieving the needed spectral efficiency to accommodate the growth in data demand.

9.2 Future work

This thesis proposes some ideas that need further refinements before their real-world application. We present now some possible extensions to our work.

9.2.1 Overlay cognitive radio

- In Chapters 2 and 4 the analysis is focused on the performance of primary receivers, and no practical waveform or coding scheme is proposed for the insertion of the secondary message. Also, it is assumed that secondary users that are placed next to primary users have similar channels. This is not the case in practical scenarios, as primary users use rooftop antennas and secondary users can be mobile receivers, for example. In this thesis we restricted our analysis to the downlink of cognitive radio systems. The design of the uplink could also be a future line of work, taking into account that the terminals do not have access to the primary signal.

- Chapter 3 analyzes the performance of single frequency network (SFN) deployments under different models of fading channels. Throughout the chapter, it is assumed that the number of carriers is large enough to make the variance close to zero (or, equivalently, that the effective SNR converges to its expected value). In some cases, especially in channels with a small delay spread, this is not the case, and an analysis of variance could be able to model the effective SNR more precisely.

- The analysis in Chapters 2-4 relies on accurate information on the magnitude of the wireless channel, although not on its phase and delay. In practice, perfect magnitude information is impossible, so a sensitivity analysis could show the robustness of the proposed method against small changes in the parameters. Also, the optimization problem could be recast in a robust way to incorporate this parameter uncertainty.

- The power allocation problem in Chapter 5 is stated as an optimization problem with perfect CSI. The method is afterwards evaluated under partial CSI, acquired from the primary feedback. An alternative optimization problem, stated in a robust way, might increase the spectral efficiency of the secondary system when CSI is inaccurate.

9.2.2 Link adaptation

- In the forward link analysis of Chapter 6, the multi-layer coding scheme relies on statistical knowledge of the channel to select the modulation and coding scheme (MCS) and power weighting. In practice, however, it might not be possible to obtain prior knowledge on the distribution of the channel. Statistical knowledge of the channel can be obtained by mere channel observation, although it is not clear how this estimation scheme may affect the performance of link adaptation. A different approach could be to adapt the MCS and power weighting from the ACK/NAK interchange in an online manner, similarly to the method presented in Section 6.5.

- In Chapter 7, the interference analysis in the multiuser MIMO scenario is based on the assumption of a zero-forcing receiver. In practice, different manufacturers may implement receivers of different complexity, and the interference estimation should be different for each of them. For example, the analysis could be repeated for the widely used minimum mean squared error (MMSE) receiver. A different problem is how to acquire knowledge at the transmitter of the detection algorithm the receiver is implementing.

- The algorithm in Chapter 7 does not take into account the degradation due to channel estimation. Although the learning algorithm should be able to incorporate this degradation implicitly for the point to point case, when multiuser transmission is used the degradation is expected to be higher because of the appearance of undesired interference. One possible way to overcome this problem is to include channel estimation error in the interference estimation formulation.

- The MCS selection method of Chapter 7 makes use of pairs of observations containing frame error rate (FER) and SNR values. In a practical setting, however, an accurate

FER estimate is not available for every observed channel. If channel variation is fast, for example, every transmitted packet will undergo a different realization, and averaging different transmissions for FER estimation might not be possible. In such a case, the FER estimates will be FER = 0 or FER = 1, so a different approximation to incorporate the FER constraint is needed.

References

[1] Cisco Systems, "Cisco visual networking index: Global mobile data traffic forecast update, 2013–2018," tech. rep., 2014.

[2] J. Mitola and G. Q. Maguire Jr, "Cognitive radio: making software radios more personal," *IEEE Personal Commun. Mag.*, vol. 6, no. 4, pp. 13–18, 1999.

[3] J. Wang, M. Ghosh, and K. Challapali, "Emerging cognitive radio applications: A survey," *IEEE Commun. Mag.*, vol. 49, no. 3, pp. 74–81, 2011.

[4] OFCOM, "TV white spaces pilot."

[5] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, pp. 894–914, May 2009.

[6] D. Gurney, G. Buchwald, L. Ecklund, S. Kuffner, and J. Grosspietsch, "Geo-location database techniques for incumbent protection in the tv white space," in *Proc. IEEE DySPAN*, pp. 1–9, IEEE, 2008.

[7] S. Haykin, D. Thomson, and J. Reed, "Spectrum sensing for cognitive radio," *Proc. IEEE*, vol. 97, pp. 849–877, May 2009.

[8] L. B. Le and E. Hossain, "Resource allocation for spectrum underlay in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5306–5315, 2008.

[9] B. Maham, P. Popovski, X. Zhou, and A. Hjorungnes, "Cognitive multiple access network with outage margin in the primary system," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 3343–3353, Oct. 2011.

[10] W. Zhang and U. Mitra, "Spectrum shaping: a new perspective on cognitive radio-part i: coexistence with coded legacy transmission," *IEEE Trans. Wireless Commun.*, vol. 58, pp. 1857–1867, June 2010.

[11] A. Manolakos, Y. Noam, K. Dimou, and A. Goldsmith, "Blind null-space tracking for mimo underlay cognitive radio networks," in *Proc. IEEE GLOBECOM*, (Anaheim, CA), pp. 1223–1229, Dec. 2012.

[12] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1813–1827, May 2006.

[13] N. Devroye, P. Mitran, and V. Tarokh, "Limits on communications in a cognitive radio channel," *Communications Magazine, IEEE*, vol. 44, pp. 44–49, June 2006.

[14] M. Costa, "Writing on dirty paper," *Information Theory, IEEE Transactions on*, vol. 29, pp. 439–441, May 1983.

[15] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *Information Theory, IEEE Transactions on*, vol. 55, no. 9, pp. 3945–3958, 2009.

[16] S. K. Jayaweera, M. Bkassiny, and K. A. Avery, "Asymmetric cooperative communications based spectrum leasing via auctions in cognitive radio networks," *Wireless Communications, IEEE Transactions on*, vol. 10, pp. 2716–2724, Aug. 2011.

[17] F. Gomez-Cuba, R. Asorey-Cacheda, F. Gonzalez-Castano, and H. Huang, "Application of cooperative diversity to cognitive radio leasing: Model and analytical characterization of resource gains," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 40–49, January 2013.

[18] J. Sachs, I. Maric, and A. Goldsmith, "Cognitive cellular systems within the TV spectrum," in *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*, pp. 1–12, Apr. 2010.

[19] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, 2011.

[20] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[21] "LTE; evolved universal terrestrial radio access (E-UTRA); physical layer procedures," *ETSI TS 136 213 V8.8.0 (2009-10)*.

[22] "IEEE standard for air interface for broadband wireless access systems," *IEEE Std 802.16-2012 (Revision of IEEE Std 802.16-2009)*, pp. 1 –2542, 17 2012.

[23] "IEEE standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications," *IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007)*, pp. 1–2793, 29 2012.

[24] "Digital video broadcasting (DVB); second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications (DVB-S2)," *ETSI EN 302 307 V1.2.1 (2009-08)*.

[25] "Digital video broadcasting (DVB); second generation DVB interactive satellite system (DVB-RCS2); part 1: Overview and system level specification," *ETSI TS 101 545-1 V1.1.1 (2012-05)*.

[26] P. Wu and N. Jindal, "Coding versus ARQ in fading channels: how reliable should the PHY be?," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3363–3374, 2011.

[27] A. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218 –1230, Oct. 1997.

[28] S. T. Chung and A. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Trans. Commun.*, vol. 49, pp. 1561 –1571, Sept. 2001.

[29] A. Goldsmith and S.-G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 595 –602, May 1998.

[30] Q. Liu, S. Zhou, and G. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated arq over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1746 – 1755, Sept. 2004.

[31] C. Wang, "Dynamic ARF for throughput improvement in 802.11 WLAN via a machine-learning approach," *Journal of Network and Computer Applications*, vol. 36, no. 2, pp. 667 – 676, 2013.

[32] A. Kamerman and L. Monteban, "Wavelan®-II: a high-performance wireless LAN for the unlicensed band," *Bell Labs Technical Journal*, vol. 2, no. 3, pp. 118–133, 1997.

[33] S. Cioni, R. De Gaudenzi, and R. Rinaldo, "Adaptive coding and modulation for the forward link of broadband satellite networks," in *Proc. IEEE GLOBECOM*, vol. 6, (San Francisco, CA), Dec 2003.

[34] S. Cioni, R. De Gaudenzi, and R. Rinaldo, "Channel estimation and physical layer adaptation techniques for satellite networks exploiting adaptive coding and modulation," *International Journal of Satellite Communications and Networking*, vol. 26, no. 2, pp. 157–188, 2008.

[35] S. Cioni, R. De Gaudenzi, and R. Rinaldo, "Adaptive coding and modulation for the reverse link of broadband satellite networks," in *Proc. IEEE GLOBECOM*, vol. 2, (Dallas, TX), pp. 1101–1105 Vol.2, Nov. 2004.

[36] H. Bischl, H. Brandt, T. de Cola, R. De Gaudenzi, E. Eberlein, N. Girault, E. Alberty, S. Lipp, R. Rinaldo, B. Rislow, J. A. Skard, J. Tousch, and G. Ulbricht, "Adaptive coding and modulation for satellite broadband networks: From theory to practice," *International Journal of Satellite Communications and Networking*, vol. 28, no. 2, pp. 59–111, 2010.

[37] D. Tarchi, G. E. Corazza, and A. Vanelli-Coralli, "Adaptive coding and modulation techniques for next generation hand-held mobile satellite communications," in *Proc. IEEE ICC*, (Budapest, Hungary), pp. 4504–4508, IEEE, 2013.

[38] M. Lampe, H. Rohling, and W. Zirwas, "Misunderstandings about link adaptation for frequency selective fading channels," in *Proc. IEEE PIMRC*, vol. 2, (Lisboa, Portugal), pp. 710 – 714, Sept. 2002.

[39] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," in *Proc. IEEE PIMRC*, vol. 4, (Berlin, Germany), pp. 2306–2311 Vol. 4, Sept. 2005.

[40] Y. Blankenship, P. Sartori, B. Classon, V. Desai, and K. Baum, "Link error prediction methods for multicarrier systems," in *Proc. IEEE VTC*, vol. 6, (Los Angeles, CA), pp. 4175–4179 Vol. 6, Sept. 2004.

[41] I. Dagres, A. Polydoros, and A. Zalonis, "DR.3.3 final report: AMC design towards next generation wireless systems," *NEWCOM++*, 2010.

[42] M. Lamarca and F. Rey, "Indicators for PER prediction in wireless systems: A comparative study," in *Proc. IEEE VTC*, vol. 2, (Stockholm, Sweden), pp. 792–796, May 2005.

[43] T. Jensen, S. Kant, J. Wehinger, and B. Fleury, "Mutual information metrics for fast link adaptation in IEEE 802.11n," in *Proc. IEEE ICC*, (Beijing, China), pp. 4910 –4915, May 2008.

[44] A. Kolmogorov, *Sur la notion de la moyenne.* Atti della R. Accad. dei Lincei, 1930.

[45] R. Jain, C. So-In, and A.-k. Al Tamimi, "System-level modeling of IEEE 802.16E mobile wimax networks: Key issues," *IEEE Wireless Commun. Mag.*, vol. 15, pp. 73–79, Oct. 2008.

[46] G. Martorell, F. Riera-Palou, and G. Femenias, "Cross-layer fast link adaptation for MIMO-OFDM based WLANs," *Wireless Personal Communications*, vol. 56, no. 3, pp. 599–609, 2011.

[47] R. Daniels, C. Caramanis, and R. W. Heath Jr., "Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR ordering," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 114–126, Jan. 2010.

[48] Z. Puljiz, M. Park, and R. W. Heath Jr., "A machine learning approach to link adaptation for SC-FDE system," in *Proc. IEEE GLOBECOM*, (Houston, TX), pp. 1–5, Dec. 2011.

[49] R. Daniels and R. W. Heath Jr., "Link adaptation in MIMO-OFDM with non-uniform constellation selection over spatial streams through supervised learning," in *Proc. IEEE ICASSP*, (Dallas, TX), pp. 3314–3317, Mar. 2010.

[50] R. Daniels and R. W. Heath Jr., "Online adaptive modulation and coding with support vector machines," in *Proc. European Wireless Conference*, (Lucca, Italy), pp. 718–724, Apr. 2010.

[51] R. Daniels and R. W. Heath Jr., "An online learning framework for link adaptation in wireless networks," in *Proc. Information Theory and Applications Workshop*, (San Diego, CA), pp. 138–140, Feb. 2009.

[52] S. Wahls and H. V. Poor, "Link adaptation for bicm-ofdm through adaptive kernel regression," in *Proc. IEEE ICASSP*, (Vancouver, BC), pp. 5136–5140, IEEE, 2013.

[53] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2 ed., Feb. 2009.

[54] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural computation*, vol. 4, no. 1, pp. 1–58, 1992.

[55] R. Daniels and R. W. Heath Jr., "Modeling ordered subcarrier SNR in MIMO-OFDM wireless links," *Physical Communication*, vol. 4, no. 4, pp. 275–285, 2011.

[56] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5045–5060, Nov. 2006.

[57] J. Zhang, R. W. Heath Jr., M. Kountouris, and J. G. Andrews, "Mode switching for the multi-antenna broadcast channel based on delay and channel quantization," *EURASIP J. Adv. Signal Process*, vol. 2009, pp. 1:1–1:15, Feb. 2009.

[58] D. Qiao, S. Choi, and K. Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *IEEE Trans. on Mobile Computing*, vol. 1, no. 4, pp. 278–292, 2002.

[59] P. Tan, Y. Wu, and S. Sun, "Link adaptation based on adaptive modulation and coding for multiple-antenna OFDM system," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1599–1606, 2008.

[60] M. Lampe, T. Giebel, H. Rohling, and W. Zirwas, "PER prediction for PHY mode selection in OFDM communication systems," in *Proc. IEEE GLOBECOM*, vol. 1, (San Francisco, CA), pp. 25–29 Vol.1, 2003.

[61] ETSI TS 102 744, "Satellite Component of UMTS (S-UMTS); Family SL satellite radio interface," Oct. 2012. Draft.

[62] "IEEE draft standard for IT - telecommunications and information exchange between systems - LAN/MAN - specific requirements - part 11: Wireless LAN medium access control and physical layer specifications - amd 4: Enhancements for very high throughput for operation in bands below 6ghz," *IEEE P802.11ac/D3.0, June 2012*, pp. 1–385, jun 2012.

[63] "DekTec DTU-215 USB-2 VHF/UHF modulator Data Sheet, http://www.dektec.com/products/USB2/DTU-215/ ."

[64] "Rohde & Schwarz ETL TV Analyzer, http://www2.rohde-schwarz.com/product/ETL.html ."

[65] W. Fischer, *Digital Video and Audio Broadcasting Technology: A Practical Engineering Guide*. Springer-Verlag Berlin Heidelberg, 2008.

[66] J. Nocedal and S. Wright, *Numerical optimization*. Springer Verlag, 1999.

[67] S. Jayaweera, G. Vazquez-Vilar, and C. Mosquera, "Dynamic spectrum leasing: A new paradigm for spectrum sharing in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 2328–2339, June 2010.

[68] A. Dammann, R. Raulefs, and S. Plass, "Soft cyclic delay diversity and its performance for DVB-T in Ricean channels," in *Proc. IEEE GLOBECOM*, (Washington, D.C.), pp. 4210–4214, Nov. 2007.

[69] DVB.org, "Implementation guidelines for a second generation digital terrestrial television broadcasting system (DVB-T2) (draft TR 102 831 v0.10.4)," 2010.

[70] P. Grover and A. Sahai, "On the need for knowledge of the phase in exploiting known primary transmissions," in *Proc. IEEE DySPAN*, (Dublin, Ireland), pp. 462–471, Apr. 2007.

[71] A. Carleial, "A case where interference does not reduce capacity (corresp.)," *IEEE Trans. Inf. Theory*, vol. 21, no. 5, pp. 569–570, 1975.

[72] P. Popovski, H. Yomo, K. Nishimori, R. Di Taranto, and R. Prasad, "Opportunistic interference cancellation in cognitive radio systems," in *Proc. IEEE DySPAN*, (Dublin, Ireland), pp. 472–475, Apr. 2007.

[73] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, "IEEE 802.16m evaluation methodology document (EMD)," *IEEE 802.16 Broadband Wireless Access Working Group*, 2008.

[74] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1277–1294, June 2002.

[75] J. Lago and F. Perez-Gonzalez, "Analytical bounds on the error performance of the DVB-T system in time-invariant channels," in *Communications, 2000. ICC 2000. 2000 IEEE International Conference on*, 2000.

[76] H. Moon and D. Cox, "Efficient power allocation for coded OFDM systems," *Communications, IEEE Transactions on*, vol. 57, pp. 943–947, Apr. 2009.

[77] C. S. Park and K. B. Lee, "Transmit power allocation for BER performance improvement in multicarrier systems," in *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on*, vol. 5, pp. 2049–2053 vol.5, Sept. 2002.

[78] A. Coulson, "Narrowband interference in pilot symbol assisted OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 2277–2287, Nov. 2004.

[79] M. López and G. Still, "Semi-infinite programming," *European Journal of Operational Research*, vol. 180, no. 2, pp. 491–518, 2007.

[80] "Monte-carlo simulation methodology for the use in sharing and compatibility studies between different radio services or systems," tech. rep., European Conference of Postal and Telecommunications Administrations, 2000.

[81] A. Rico-Alvariño, C. Mosquera, and F. Pérez-González, "Overlay cognitive transmission in a multicarrier broadcast network with dominant line of sight reception," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 4128–4139, 2012.

[82] A. Rico-Alvariño, C. Mosquera, and F. Pérez-González, "Overlay spectrum reuse in a multicarrier broadcast network: Single receiver analysis," in *Proc. IEEE SPAWC*, (Cesme, Turkey), pp. 209–213, IEEE, June 2012.

[83] A. Rico-Alvariño, C. Mosquera, and F. Pérez-González, "Overlay spectrum reuse in a multicarrier broadcast network: Coverage analysis," in *Proc. IEEE SPAWC*, (Cesme, Turkey), pp. 204–208, IEEE, June 2012.

[84] A. Rico-Alvariño, C. Mosquera, and F. Pérez-González, "On the co-existence of primary and secondary transmitters in a broadcast network," in *Proc. International Conference on Cognitive Radio and Advanced Spectrum Management*, (Barcelona, Spain), Oct. 2011.

[85] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions*. Dover Publications Inc., New York, 1964.

[86] H. Jiang, P. Wilford, and S. Wilkus, "Providing local content in a hybrid single frequency network using hierarchical modulation," *IEEE Trans. Broadcast.*, vol. 56, no. 4, pp. 532–540, 2010.

[87] A. Nuttall, "Some integrals involving the Q_M function (corresp.)," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 95–96, 1975.

[88] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[89] X.-B. Liang, "Orthogonal designs with maximal rates," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2468 – 2503, oct. 2003.

[90] A. Dammann, R. Raulefs, and S. Plass, "Soft cyclic delay diversity and its performance for DVB-T in ricean channels," in *Proc. IEEE GLOBECOM*, (Washington, D.C.), pp. 4210–4214, 2007.

[91] A. Rico-Alvariño and C. Mosquera, "On the effect of echoes in hybrid terrestrial-satellite single frequency networks: Analysis and countermeasures," in *Proc. Advanced Satellite Multimedia Systems Conference and Signal Processing for Space Communications Workshop*, (Baiona, Spain), pp. 168–175, IEEE, Sept. 2012.

[92] A. Rico-Alvariño and C. Mosquera, "Analytical characterization of the single frequency network gain using effective SNR metrics," in *Proc. ISWCS*, (Ilmenau, Germany), pp. 1–5, VDE, Aug. 2013.

[93] J. Lee, J.-K. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-advanced," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, pp. 3:1–3:10, Mar. 2009.

[94] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge Univ Pr, 2004.

[95] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Requirements for evolved UTRA (E-UTRA) and evolved UTRAN (E-UTRAN). 3GPP TR 25.913 v7.3.0," 2006.

[96] OFCOM, "Technical analysis of interference from mobile network base stations in the 800 MHz band to digital terrestrial television," tech. rep., June 2011.

[97] A. Rico-Alvariño and C. Mosquera, "Overlay spectrum reuse in a broadcast network: Covering the whole grayscale of spaces," in *Proc. IEEE DySPAN 2012*, (Bellevue, WA), pp. 479–488, IEEE, Oct. 2012.

[98] C. Yu, W. Xiangming, L. Xinqi, and Z. Wei, "Research on the modulation and coding scheme in LTE TDD wireless network," in *Proc. ICIMA 2009*, pp. 468 –471, May 2009.

[99] S. Srinivasa and S. Jafar, "Cognitive radios for dynamic spectrum access-the throughput potential of cognitive radio: A theoretical perspective," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 73–79, 2007.

[100] R. Manna, R. H. Y. Louie, Y. Li, and B. Vucetic, "Cooperative spectrum sharing in cognitive radio networks with multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, pp. 5509–5522, Nov. 2011.

[101] T. Menni, E. Chaumette, P. Larzabal, and J. Barbot, "New results on deterministic Cramer-Rao bounds for real and complex parameters," *IEEE Trans. Signal Process.*, vol. 60, pp. 1032–1049, Mar. 2012.

[102] S. Kay, *Fundamentals of Statistical Signal Processing, volume 1.* Prentice Hall, 1993.

[103] A. Rico-Alvariño and C. Mosquera, "Optimum training for csi acquisition in cognitive radio channels," in *Proc. Asilomar SSC*, (Pacific Grove, CA), pp. 123–127, IEEE, Nov. 2012.

[104] A. Rico-Alvariño and C. Mosquera, "Overlay cognitive transmission in OFDM point to point systems exploiting primary feedback," in *Proc. CIP*, (Baiona, Spain), pp. 1–6, IEEE, May 2012.

[105] F. Perez Fontan, M. Vazquez Castro, C. Enjamio Cabado, J. Pita Garcia, and E. Kubista, "Statistical modeling of the LMS channel," *IEEE Trans. Vehicular Technology*, vol. 50, pp. 1549–1567, Nov. 2001.

[106] "IEEE 802.16m.Evaluation Methodology for P802.16m-Advanced Air Interface ."

[107] A. Monk and L. Milstein, "Open-loop power control error in a land mobile satellite system," *IEEE J. Sel. Areas Commun.*, vol. 13, pp. 205 –212, feb 1995.

[108] U. Erez, M. Trott, and G. Wornell, "An efficient ARQ scheme with SNR feedback," in *Proc. IZS*, (Zurich, Switzerland), pp. 88 –91, Mar. 2008.

[109] J. Arnau and C. Mosquera, "Open loop adaptive coding and modulation for mobile satellite return links," in *Proc. AIAA ICSSC*, (Firenze, Italy), Oct. 2013.

[110] C. L. C. Loo, "A statistical model for a land mobile satellite link," *IEEE Trans. Veh. Technol.*, vol. 34, pp. 122–127, Aug. 1985.

[111] D. Arndt, T. Heyn, J. Konig, A. Ihlow, A. Heuberger, R. Prieto-Cerdeira, and E. Eberlein, "Extended two-state narrowband LMS propagation model for S-Band," in *IEEE Int. Symp. Broadband Multimed. Syst. Broadcast.*, pp. 1–6, June 2012.

[112] F. Perez-Fontan, M. Vazquez-Castro, S. Buonomo, J. Poiares-Baptista, and B. Arbesser-Rastburg, "S-band LMS propagation channel behaviour for different environments, degrees of shadowing and elevation angles," *IEEE Trans. Broadcast.*, vol. 44, pp. 40–76, Mar. 1998.

[113] A. Lozano, "Interplay of Spectral Efficiency , Power and Doppler Spectrum for Reference-Signal-Assisted Wireless Communication," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5020–5029, 2008.

[114] N. Jindal and A. Lozano, "A Unified Treatment of Optimum Pilot Overhead in Multipath Fading Channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, 2010.

[115] M. R. McKay, P. J. Smith, H. a. Suraweera, and I. B. Collings, "On the Mutual Information Distribution of OFDM-Based Spatial Multiplexing: Exact Variance and Outage Approximation," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3260–3278, July 2008.

[116] S. Yun and C. Caramanis, "Reinforcement Learning for Link Adaptation in MIMO-OFDM Wireless Systems," in *Proc. IEEE GLOBECOM 2010*, (Miami, FL), pp. 1–5, Dec. 2010.

[117] J. F. Kurose and K. W. Ross, *Computer networking*, vol. 2. Pearson Education, 2003.

[118] M. Nakamura, Y. Awad, and S. Vadgama, "Adaptive control of link adaptation for high speed downlink packet access (hsdpa) in w-cdma," in *Proc. WPMC*, vol. 2, (Honolulu, HI), pp. 382–386 vol.2, Oct. 2002.

[119] T. Cui, F. Lu, V. Sethuraman, A. Goteti, S. P. Rao, and P. Subrahmanya, "Throughput optimization in high speed downlink packet access (HSDPA)," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 474–483, 2011.

[120] S. Haykin, *Adaptive filter theory.* Prentice Hall, 2002.

[121] J. Arnau, A. Rico-Alvariño, and C. Mosquera, "Adaptive transmission techniques for mobile satellite links," *Proc. AIAA ICSSC, Ottawa, Canada*, 2012.

[122] A. Rico-Alvariño, J. Arnau, and C. Mosquera, "Statistical cross layer adaptation in fast fading mobile satellite channels," in *Proc. IEEE GLOBECOM*, Dec. 2013.

[123] A. Rico-Alvariño, J. Arnau, and C. Mosquera, "Balancing closed and open loop csi in mobile satellite link adaptation," in *Proc. Advanced Satellite Multimedia Systems Conference and Signal Processing for Space Communications Workshop*, (Livorno, Italy), Sept. 2014.

[124] M. Esslaoui, F. Riera-Palou, and G. Femenias, "Fast link adaptation for opportunistic multiuser MIMO-OFDM wireless networks," in *Proc. ISWCS*, (Aachen, Germany), pp. 372 –376, Nov. 2011.

[125] M. Esslaoui, F. Riera-Palou, and G. Femenias, "A fair MU-MIMO scheme for IEEE 802.11ac," in *Proc. ISWCS*, (Paris, France), pp. 1049 –1053, Aug. 2012.

[126] Z. Chen, W. Wang, M. Peng, and F. Cao, "Limited feedback scheme based on zero-forcing precoding for multiuser MIMO-OFDM downlink systems," in *Proc. ICST WICON*, (Singapore), pp. 1 –5, Mar. 2010.

[127] S. Yun, C. Caramanis, and R. W. Heath Jr., "Distributed link adaptation for multicast traffic in MIMO-OFDM systems," *Physical Communication*, vol. 4, no. 4, pp. 286–295, 2011.

[128] G. Dimic and N. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Trans. Signal Process.*, vol. 53, pp. 3857–3868, Oct. 2005.

[129] M. Kobayashi and G. Caire, "Joint beamforming and scheduling for a MIMO downlink with random arrivals," in *Proc. IEEE ISIT*, (Seattle, WA), pp. 1442–1446, July 2006.

[130] R. Chen, Z. Shen, J. Andrews, and R. W. Heath Jr., "Multimode transmission for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 56, pp. 3294–3302, July 2008.

[131] Cisco Systems, "White paper: 802.11ac: The fifth generation of wi-fi," 2012.

[132] H. Kim, K. Kim, Y. Han, and S. Yun, "A proportional fair scheduling for multicarrier transmission systems," in *Proc. IEEE VTC-Fall*, (Los Angeles, CA), pp. 409–413 Vol. 1, Sept. 2004.

[133] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, pp. 461–471, Feb. 2004.

[134] M. Sadek, A. Tarighat, and A. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1711 –1721, may 2007.

[135] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath Jr., and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, pp. 3658–3663, Sept. 2006.

[136] G. Golub and C. Van Loan, *Matrix computations*. Johns Hopkins University Press, 1996.

[137] J. Roh and B. Rao, "Efficient feedback methods for MIMO channels based on parameterization," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 282 –292, Jan. 2007.

[138] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, pp. 84 – 95, Jan. 1980.

[139] M. A. Sadrabadi, A. K. Khandani, and F. Lahouti, "Channel feedback quantization for high data rate MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3335 –3338, Dec. 2006.

[140] H. V. Henderson and S. Searle, "The vec-permutation matrix, the vec operator and Kronecker products: A review," *Linear and multilinear algebra*, vol. 9, no. 4, pp. 271–288, 1981.

[141] R. Zamir and M. Feder, "On lattice quantization noise," in *Proc. DCC*, (Snowbird, UT), pp. 380 –389, Mar. 1994.

[142] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, pp. 55 – 63, Jan. 1968.

[143] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37 – 52, 1987. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

[144] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[145] T. Tao and A. Czylwik, "Performance analysis of link adaptation in LTE systems," in *International ITG Workshop on Smart Antennas*, pp. 1–5, 2011.

[146] Y.-S. Choi and S. Alamouti, "A pragmatic PHY abstraction technique for link adaptation and MIMO switching," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 960–971, 2008.

[147] TGac Channel Models Special Committee, "TGac channel model addendum," *IEEE 802.11-09/0308r0*, May 2009.

[148] A. Rico-Alvariño and R. W. Heath Jr., "Learning-based link adaptation in multiuser MIMO-OFDM," in *Proc. EUSIPCO*, (Marrakech, Morocco), Sept. 2013.

[149] A. Rico-Alvariño and R. W. Heath Jr, "Learning-based adaptive transmission for limited feedback multiuser MIMO-OFDM," *IEEE Trans. Wireless Commun.*, 2014.

[150] H. Shao and N. Beaulieu, "Block coding for impulsive laplacian noise," in *Proc. IEEE ICC*, (Cape Town, South Africa), pp. 1–6, May 2010.

[151] J. Miller and J. B. Thomas, "Detectors for discrete-time signals in non-gaussian noise," *IEEE Trans. Inf. Theory*, vol. 18, no. 2, 1972.

[152] S. Nadarajah, "A generalized normal distribution," *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685–694, 2005.

[153] A. Rico-Alvariño and R. W. Heath Jr., "Link adaptation in MIMO-OFDM with practical impairments," in *Proc. Asilomar SSC*, (Pacific Grove, CA), Nov. 2013.

[154] A. Rico-Alvariño, R. W. Heath Jr., and C. Mosquera, "FER prediction with variable codeword length.," in *Proc. IEEE ICASSP 2014*, (Florence, Italy), May 2014.