Cross-layer Aspects in OFDMA Systems

Feedback, Scheduling and Beamforming

PATRICK SVEDMAN

Doctoral Thesis in Telecommunications
Stockholm, Sweden 2007

© Patrick Svedman, June 2007

Tryck: Universitetsservice US AB
Abstract

This thesis mainly studies the downlink of a wireless multiuser system, where the transmitter has limited knowledge about the communication channels of the users. Key techniques to improve the performance of such systems are, for instance, multiple antennas, multiuser diversity and orthogonal frequency-division multiple access (OFDMA). Common for these techniques is that, to exploit them fully, a cross-layer approach has to be adopted. This means that the scheduling and the signal designs are done jointly and based on parameters from several communication layers.

Multiuser diversity can be used to significantly increase system throughput in wireless communication systems. The idea is to schedule users when they experience good channel conditions and let them wait when the channels are weak. In this thesis, several aspects of OFDMA systems that exploit multiuser diversity are investigated. An adaptive reduced feedback scheme for OFDMA is proposed. It significantly reduces the total feedback overhead while maintaining a multiuser diversity gain. The scheme uses clusters of sub-carriers as feedback units and feeds back information about the fading peaks only. It adapts to the number of users so that less feedback per user is required if there are many users in the system. With such a selective feedback scheme, there is a risk that the scheduler has no instantaneous channel quality information for some parts of the spectrum. Better uses for these sub-carriers are investigated. In addition, an alternative based on the channel quality feedback of some uniformly spaced sub-carriers is proposed. The scheduler estimates the channel quality on the other sub-carriers.

Channel-aware scheduling is necessary in order to exploit multiuser diversity. A modified proportional fair (PF) scheduler is proposed. It incorporates individual target bit-rates and delays and a tunable fairness level. An opportunistic beamforming scheme for clustered OFDMA is presented and evaluated. A key aspect of the opportunistic beamforming scheme is that it induces artificial frequency selectivity for users with relatively flat channels. Several aspects of the proposed system are evaluated by means of simulations. In the simulations, the clustered beamforming with the modified PF scheduler performs better than three comparison systems. The modified PF scheduler manages to divide the resources according to the user targets, while at the same time exploiting the multiuser diversity as well as the standard PF algorithm.

In many scenarios, the largest gains from having multiple antennas at the base-
station come from space-division multiple access (SDMA). In the downlink, this means that data is transmitted to several users simultaneously by using several beams. Opportunistic space-division OFDMA is proposed and evaluated. An enhancement that exploits temporal channel correlation is able to boost the throughput significantly.

SDMA based on subspace packings is proposed and evaluated. A set of beam-forming matrices (a subspace packing) is made a priori available at the base-station and at all users. In each block, one of the matrices is used for multi-beam transmission. The users pick and feed back the index of one preferred column (beam) from one of the matrices, and the corresponding SINR, which includes all potential inter-beam interference. This enables scheduling of spatially compatible users and accurate rate adaptation, with relatively little feedback. Three different subspace packings are considered and evaluated with simulations. For the i.i.d. Rayleigh fading channel, Grassmannian subspace packings were the best choice. Moreover, a method to further reduce the feedback for large packings is proposed and evaluated. It is based on the arrangement of beams in a graph and the feedback of a neighbor index. Numerical results show that the feedback can be significantly reduced with only small performance losses, even for relatively fast fading channels.
Acknowledgements

This thesis would not have been possible without all the people around me. First of all, I want to thank Professor Björn Ottersten for giving me the opportunity to work within this field. His support, advice and enthusiasm gave me the necessary boosts to overcome many of the challenges on the way. He has been able to optimally adapt with time the balance between firm guidance and confiding freedom.

Along the way, I have had the privilege to work with several excellent supervisors and mentors. I would like to thank Professor Sarah Kate Wilson for encouraging, inspiring and pushing me during my first steps in this field. I am also grateful to Dr. Mats Bengtsson for always having time to discuss and provide answers to issues regarding anything from LaTeX to communication problems. A special thanks to Professor Leonard J. Cimini, Jr., who let me visit his research group at the University of Delaware for five months. I would also like to thank Dr. Eduard Jorswieck for many inspiring discussions and guidance.

Studies without fellow students would be very tedious. I want to thank all colleagues and friends on Plan 4 for a great fellowship and interesting discussions. Also alumni and fellow GST students deserve to be mentioned, as well as all other friends. A special thanks to Karin Demin, Annika Augustsson and Anna Thöresson-Bergh for making all administrative issues running so smoothly and to Nina Unkuri and Andreas Stenhall for the best computer support. I also want to acknowledge Peter von Wycza and Simon Järmyr for proofreading parts of the thesis.

I wish to thank Professor Ana I. Pérez-Neira for taking the time to act as opponent for this thesis and Professor Lang Tong, Professor Mark T. Smith and Professor Abbas Mohammed for acting as committee.

Lastly, I want to express my greatest love and gratitude to Lotta, and also my family for supporting and believing in me.

Patrick Svedman
Stockholm, May 2007
# Contents

## 1 Introduction
- 1.1 Background .............................................. 1
- 1.2 Multiuser Diversity and Opportunism ..................... 2
- 1.3 Multiple Antennas ........................................ 4
- 1.4 OFDM and OFDMA ........................................... 6
- 1.5 Cross-layer Aspects on Multi-antenna Systems with Low Rate Feedback ........................................... 7
- 1.6 Outline and Contributions .................................. 11

## 2 Multi-cell OFDMA System Model
- 2.1 Overview ............................................. 17
- 2.2 Multi-cell Downlink OFDMA Baseband Model ............... 17
- 2.3 Baseband Model Details .................................... 22

## 3 Reduced Feedback for OFDMA
- 3.1 Overview ............................................... 27
- 3.2 Introduction ............................................. 27
- 3.3 Reduced Feedback by Clustering and Adaptive Selection ...... 28
- 3.4 Reduced Feedback - Reduced Performance? .................. 33
- 3.5 Adaptive Feedback Rate Derivation ........................ 37
- 3.6 On Sub-carrier Variations Within a Cluster ................. 39

## 4 Using Unclaimed Sub-carriers
- 4.1 Overview ............................................... 47
- 4.2 Introduction ............................................. 47
- 4.3 System Model ............................................. 48
- 4.4 Three Ways to Use the Unclaimed Sub-carriers ............... 48
- 4.5 Simulation Results ....................................... 50
- 4.6 Summary ................................................. 53
- 4.7 LMMSE Channel Estimator ................................. 57
## Reduced Feedback by SNR Estimation at the Transmitter

5.1 Overview ................................................................. 59
5.2 Introduction ............................................................ 59
5.3 Reduced Feedback Scheme .......................................... 60
5.4 Covariance Information at the Transmitter ..................... 61
5.5 SNR Estimators ......................................................... 62
5.6 System Simulation ..................................................... 65
5.7 Summary ................................................................. 67
5.8 Sub-carrier SNR Covariance ................................. 69

## Modified Proportional Fair Scheduling for OFDMA

6.1 Overview ................................................................. 71
6.2 Introduction ............................................................ 71
6.3 Proportional Fair ....................................................... 73
6.4 Modified Proportional Fair ........................................ 73
6.5 Summary ................................................................. 78
6.6 Scheduling Metric Derivation ................................. 79

## Opportunistic Beamforming for OFDMA

7.1 Overview ................................................................. 81
7.2 Introduction ............................................................ 81
7.3 Opportunistic Beamforming for Clustered OFDM .......... 83
7.4 Exploiting Temporal Channel Correlation in the Scheduling and Beamforming ......................... 84
7.5 Opportunistic Beamforming Frame Structure .............. 85
7.6 Summary ................................................................. 85

## Numerical Results for Chapters 6-7

8.1 Overview ................................................................. 89
8.2 Simulation Environment and Assumptions ................. 89
8.3 Simulation Results .................................................... 92
8.4 Summary ................................................................. 95

## Opportunistic SD-OFDMA

9.1 Overview ................................................................. 101
9.2 Introduction ............................................................ 101
9.3 Opportunistic SD-OFDMA – Basic Scheme ............... 103
9.4 Exploiting Temporal Channel Correlation ................ 106
9.5 Performance Evaluation ............................................ 107
9.6 Summary ................................................................. 111
### CONTENTS

**10 Reduced Feedback SDMA Based on Subspace Packings**  
10.1 Overview .................................................. 115  
10.2 Introduction ............................................. 115  
10.3 System Model ............................................ 116  
10.4 Reduced Feedback SDMA Based on Subspace Packings .......... 117  
10.5 Subspace Packings ....................................... 121  
10.6 Numerical Results ...................................... 123  
10.7 Conclusions .............................................. 127  
10.A The SINR Penalty When Using Inner-product as Beam-selection Criterion ........................................... 129  
10.B Proof of Theorem 10.2 .................................... 131

**11 Beam-graphs for Beamforming Codebooks**  
11.1 Overview .................................................. 133  
11.2 Introduction ............................................. 133  
11.3 Beam Graph Construction .................................. 134  
11.4 Numerical Results ...................................... 136

**12 Conclusions** ............................................. 139

**A Notation and Acronyms** .................................... 143  
A.1 Notation ................................................... 143  
A.2 Common Symbols and Functions ................................ 144  
A.3 Acronyms .................................................. 145

**Bibliography** .............................................. 149
Chapter 1

Introduction

1.1 Background

The last decade in the wireless world has meant a steady move from pure voice services to more mixed voice and data services. People are beginning to use their mobile devices more like their internet-connected stationary computers. Most mobile devices differ from stationary computers in significant ways, for example the screen size and the input interface. Still, we can expect an increasing demand for and usage of internet browsing, e-mail, file downloading, multiplayer gaming, streaming media, video telephony, instant messaging, business applications etc. with mobile devices. This vision holds both opportunities and challenges for telecom operators and vendors. There is an opportunity to increase revenue by billing as many customers as possible for new services they are willing to pay for. One major challenge is to design future communication systems that can support many users while still offering a high enough user-perceived service quality. The “high enough” user-perceived service quality level depends on the price of the service. Generally, there is a tradeoff between the offered service quality and the number of users that the system can accommodate. Hence, the business model of the operator affects the desirable behavior of the communication system. An operator that aims at guaranteeing high service quality can support few, but highly billable, customers in the system. An operator that aims at offering medium service quality levels to more users, can instead bill more customers less.

The traditional design paradigm in the telecommunications industry has been to guarantee the users a certain Quality-of-Service (QoS). Enough resources were given to each user to provide some margin for communication disturbances. The goal was an “anytime anywhere” service level. However, if the communication resources were already occupied, a new user would not be admitted to the system. The system worked in a circuit switched fashion, with very strict delay requirements. The internet world has a slightly different perspective. New users are basically always admitted, but if the network or a particular service is heavily loaded some users
CHAPTER 1. INTRODUCTION

might experience a degraded service level. The offered service level is “best effort”. This kind of system typically works in a packet switched manner, where packets of data are routed through a network, resulting in more variable packet delays and typically quite high packet loss probabilities.

The last years, packet-based “best effort” approaches have become more popular in wireless communication systems and research. This is due to the increase of wireless data traffic, but also to the emergence of the concept of multiuser diversity.

1.2 Multiuser Diversity and Opportunism

Diversity in communications is similar to the everyday life concept of diversity: “Do not put all the eggs in the same basket!”. If you drop the basket, all eggs are lost. Diversity in communications means that the same message is sent over several random channels, i.e. the same egg is put in multiple baskets. If one channel turns out to be bad, there is still a chance of successful reception through the other channels. The different channels can for instance be different frequencies (frequency diversity), different transmit and/or receive antennas (antenna diversity) or different time-instants (time diversity) [Wor98].

Diversity is needed because the wireless channel quality is random. This randomness is also called fading [TV05]. The major loss in signal power is usually due to the distance and objects between the transmitter and the receiver (path loss and shadow fading). This kind of fading changes relatively slowly. Fast fading however, arises from the constructive or destructive superposition of time-delayed copies of the signal, arriving via different propagation paths (multipaths). As the paths change, due to movement, the fading can change rapidly. Furthermore, if the time-delays of the different propagation paths differ significantly (time-dispersive channel), the fading may differ between frequencies in the received signal. This is called frequency-selective fading. A channel that is not frequency-selective is called flat fading. The channel gains as a function of time of two flat fading users are illustrated in Figure 1.1. The user that moves fast is far away from the base-station and consequently, this user has a higher path loss.

Multiuser diversity means that there are several users that want to communicate, each with different and time-varying communication channel quality [TV05]. The multiuser diversity gain can be obtained by letting the user with the best instantaneous channel quality communicate. A fundamental property of multiuser diversity is that as the number of users increases, the expected quality of the best channel increases. Multiuser diversity differs from classical diversity, e.g. time or frequency diversity, in some fundamental ways. Classical diversity is a means to increase the reliability of communication over a fading channel. The diversity combats the fading. Multiuser diversity, on the other hand, is a way to exploit the fading. The objective is not reliability, but rather to increase the system throughput. Furthermore, for classical diversity methods, there is no need to consider fairness. If an antenna or a frequency band is never used due to poor fading, there
is no real loss. When the same thing happens to a user, there may be negative consequences.

The multiuser diversity effect was first noticed in an information theoretic context. Knopp and Humblet studied the uplink of a single-cell under a sum-power constraint, with users experiencing flat fast fading channels [KH95a]. They assumed full channel state information (CSI) at the users and at the base-station. It turns out that the sum capacity (sum of the simultaneous user capacities) is maximized if, for each time-instant, the user with the best channel gain is allowed to transmit. The optimal solution also includes a power-control law which uses more transmit power for strong channels than for weak channels. This is an opposite strategy to conventional power control, which uses transmit power to compensate for weak channels. The capacity is the fundamental upper bound on the rate at which error-free communication is possible. The sum capacity achieving solution for the corresponding downlink scenario is similar. At each time-instant, the base-station should transmit to the user with the strongest channel. Related information theoretic results for fading multiuser channels can be found in [KH95b, Tse97, TH98, HT98, LG01a, LG01b].

To exploit multiuser diversity, channel access has to be given to users with instantaneously good channel conditions. The scheduling scheme that achieves capacity gives channel access to the user with the best channel conditions. This maximum throughput scheme is potentially very unfair to users with weak slow
fading condition, e.g., users that are far away from the access point. A more fair scheduler is the proportional fair (PF) scheduler [VTL02]. The PF scheduler tries to schedule a user when its instantaneous channel quality is high in relation to its average channel quality. Proportional fair scheduling is more suitable for data traffic than for voice. The reason is that users can not be guaranteed any data-rate or delay. In systems that exploit multiuser diversity, there is a general tradeoff between latency and sum throughput. If the systems allows long waiting periods for the users, the scheduler may almost always schedule users that are on their fading peaks. If, on the other hand, only small delays are tolerated, the scheduler may be forced to schedule users that have unfavorable instantaneous channels.

In multiuser diversity schemes, the scheduler must know the instantaneous channel quality of the users. In a time-division duplex (TDD) system, the same frequency band is used for both uplink and downlink communication during different time-slots. The downlink channel quality can be estimated from the uplink quality and vice versa, provided that the up and downlink channels are reciprocal and that the time variations are not too fast in relation to the duplex time. In a frequency-division duplex (FDD) system, where the uplink and downlink communication are in different frequency bands, this is more difficult. Instead, the instantaneous downlink quality, for instance, can be fed back to the scheduler via the uplink. Note that is may be possible to estimate the downlink channel from the uplink also in FDD systems [BM05].

### 1.3 Multiple Antennas

Multiple transmit and receive antennas can be used to provide diversity in space (antenna or spatial diversity). With several receive antennas, several different copies of the transmitted signal are obtained through different spatial channels. The different signal copies can be combined at the receiver to increase the communication reliability or to increase the data rate. With several transmit antennas, the spatial channels can be used to transmit different signals, so-called spatial multiplexing, which could increase the data rate significantly. Overviews of multiple-input multiple-output (MIMO) schemes are given in [GSS⁺03, VP06]. The tradeoff between spatial diversity and spatial multiplexing is studied in [ZT03]. An overview of MIMO capacity results for various fading and CSI assumptions is given in [GJJV03, TV05].

It is usually assumed that the receiver can track the MIMO channel. The primary difference between different MIMO schemes is the degree of channel state information (CSI) at the transmitter. Several spatial diversity schemes like space-time coding require no CSI at the transmitter [LS03]. With channel state information at the transmitter, it is possible to use the multiple transmit antennas to form a beam that directs the transmitted energy towards the desired user, so-called beamforming. In multi-antenna systems with only partial CSI, large performance gains can still be achieved. For instance, it is possible to combine space-time block
codes with partial CSI at the transmitter, as in [JS04]. If no instantaneous CSI can be made available at the transmitter, it is still possible to do beamforming based on the slowly changing statistical channel information [VM01, JVG01, JB04]. The full CSI is typically continuous. Hence, to allow for a finite rate feedback, the CSI needs to be quantized. An overview of vector quantization techniques for CSI feedback reduction is given in [LHSH04]. Vector quantization in the multiuser downlink context is treated in [KBS07]. A common wisdom for MIMO designs is that more and higher quality CSI at the transmitter allows higher performance in the communication link.

For the multiuser downlink with multiple transmit antennas and full CSI at the transmitter, the sum capacity achieving scheme is based on the non-linear dirty paper precoding [GH83, Cos83]. It is able to transmit information-bearing signals to several users simultaneously in such a way that a large part of the interference is suppressed. Schemes that allow simultaneous transmission to multiple users by using the spatial domain are called space-division multiple access (SDMA) schemes. Dirty paper precoding, and also sub-optimal low-complexity schemes like Tomlinson-Harashima precoding [Tom71, HM72] rely on full channel state information of all users, which is a rather limiting assumption in non-reciprocal systems.

Multiuser diversity can be exploited also in the context of multiple antenna systems. Multiuser diversity schemes rely on the fast fading of the users. If the users are moving slowly their fading will likely be slow, which can lead to a low multiuser diversity gain or long idle periods for the users. This problem can be reduced if opportunistic beamforming is applied. The traditional use of multiple antennas is to provide diversity to combat fading. Opportunistic beamforming is an opposite philosophy in that it aims at increasing the fading rate of the users [VTL02]. Opportunistic beamforming is not based on CSI at the transmitter, but randomly selects beamformers that are changed regularly. The random beamforming will give many users low signal strengths, but if there are many users in the system the beam will likely “hit” a user. In addition, compared to single antenna transmission, the interference level will fluctuate faster with opportunistic beamforming, called opportunistic nulling. Combined with a multiuser diversity scheduler, opportunistic beamforming enables multiuser diversity gain and fairness in addition to beamforming gain, even if users are moving slowly [VTL02].

The opportunistic beamforming concept was extended to SDMA in [SH05], where several random, but orthogonal beams were used. Relying on multiuser diversity, a suitable user for each beam could be found with a high probability, provided that there were enough users in the system. Then, the system throughput was shown to scale with the number of antennas as for dirty paper precoding.

It is worthwhile to note that MIMO schemes that use explicit CSI at the transmitter usually require high quality CSI to work well, which may severely degrade performance in real systems with noisy channel estimates [ZO03, YG06, Rup02]. For opportunistic beamforming, only the channel quality has to be estimated and fed back, which is allows for coarser quantization. On the other hand, it is feasible to schedule users in a round-robin fashion in CSI-at-the-transmitter schemes,
enabling feedback from only one or a few users at a time. For opportunistic schemes, all users have to feed back all the time. Opportunistic schemes are more feasible in data systems where the requirements on QoS are lower. Furthermore, multiuser diversity schemes are more suitable in systems with many users, for obvious reasons.

1.4 OFDM and OFDMA

Orthogonal frequency-division multiplexing (OFDM) is a modulation technique that is a promising candidate for future wireless broadband systems. It has been implemented in several wireless systems, such as digital audio broadcast (DAB) [DAB01] and wireless LAN [IEE99]. It is also used in wired communication systems such as ADSL [ADS02]. OFDM is currently being considered for use in the downlink of the long-term evolution of the 3G system [EFK+06].

OFDM divides the available bandwidth into orthogonal low-rate sub-carriers. Usually, the Discrete Fourier transform (DFT) is used to modulate the information onto the sub-carriers [WE71]. OFDM has several attractive features. For instance, the need for advanced equalization in multipath environments is avoided. OFDM is also attractive in a multiuser perspective. Since the bandwidth is divided into orthogonal sub-carriers, it is possible to schedule different users simultaneously on different frequency bands [WCLM99]. This enables an extension of the multiuser diversity concept to not only include temporal fading, but also fading in the frequency domain. Orthogonal frequency-division multiple access (OFDMA), i.e. several simultaneous users on different frequencies, is more suitable in the downlink. For uplink OFDMA to work, the received signals from the users need to be synchronized in time and frequency, which is practically difficult [vdBBB+99].

However, there are several practical problems with OFDM [HMCK03]. To maintain the orthogonality between the sub-carriers, carrier frequency synchronization has to be accurate and the guard period between consecutive OFDM symbols has to be longer than the channel impulse response. Furthermore, OFDM signals suffer from a high peak-to-average power ratio (PAR), which puts high requirements on the power amplifiers and the dynamic ranges of digital-to-analog and analog-to-digital converters (DAC and ADC).

The original idea of multiuser diversity, also described in Section 1.2, is based on scheduling of flat fading users in time [KH95a, VTL02]. In time-dispersive multipath channels however, the users experience frequency-selective fading. Equalization is a way to combat this kind of fading [Pro01]. The basic principle is to amplify the weaker frequencies to effectively create a frequency flat channel. In a single-carrier system with frequency-selective fading and equalization at the receivers, energy is transmitted also on frequency bands that are in deep fades.

Instead, if OFDM is used to cope with the time-dispersive channels, the frequency-selective fading can be exploited. OFDM divides the communication channel into orthogonal sub-channels, called sub-carriers, each using a small part of the available frequency band. The frequency-selective fading will result in dif-
1.5. **CROSS-LAYER ASPECTS ON MULTI-ANTENNA SYSTEMS WITH LOW RATE FEEDBACK**

Different sub-carrier gains. This enables the exploitation of multiuser diversity in the frequency domain, i.e. users can be scheduled also on their frequency fading peaks [SOAS03, WOS+03]. The effective frequency-domain communication channel will consist of orthogonal parts from different users. With a good resource allocation algorithm, the transmission of energy on badly fading parts of the spectrum can be avoided.

Multiuser diversity systems use adaptive modulation and coding instead of power control to achieve the target error rates. By scheduling users with good instantaneous channel conditions, higher order modulation can be used and high system throughput can be achieved [VTL02].

For the time-dispersive single-input-single-output (SISO) downlink channel with an average sum-power constraint, the capacity-achieving scheme is to transmit to the user with strongest channel on each frequency band [TV05], together with a power waterfilling in time and frequency. This suggests that OFDMA is a suitable technique for the multiuser downlink [CSN+03].

### 1.5 Cross-layer Aspects on Multi-antenna Systems with Low Rate Feedback

In digital communication systems, it is common to divide the functionality into layers [Hal96]. The two lowest layers are usually called the physical layer and the medium access control (MAC) layer. Simply put, on the transmitting side, the physical layer converts the bits into an analog signal, suitable for the physical medium. On the receiving side, the physical layer converts the received analog signal back to bits. Roughly, the MAC layer controls which data streams the physical layer should work with and also handles QoS to some extent. The traditional design paradigm emphasizes layer transparency, which means that a layer does not need to know the inner workings of the layer below or above. A layer only needs to know which functions the layers below and above provide. This is beneficial for many reasons. For instance, the physical layer can be changed if the communication medium (e.g. from copper-line to fiber) is changed, without having to modify the upper layers. While this philosophy is advantageous from a design perspective, it may lead to inferior performance.

In wireless multi-antenna systems, the beamforming and signal modulation traditionally belong to the physical layer, since they are closely coupled with the physical medium and the parameters therein (e.g. the CSI). The scheduling, however, traditionally belongs to the MAC layer, since it deals with handling the data streams of the users and their QoS. In most previous wireless systems, like GSM, users have been scheduled in a round-robin fashion, i.e. sequentially after each other, regardless of the instantaneous physical communication conditions.

With the information theoretic results that brought forth multiuser diversity, the necessity to co-design parts of the physical and MAC layers in order to increase the performance has become clear. Multiuser diversity schemes rely on the ability of
CHAPTER 1. INTRODUCTION

Figure 1.2: A coarse model of the multiuser multi-antenna wireless communication system that is the subject of this thesis.

the scheduler to schedule users with favorable fading conditions, which is a typical physical layer parameter. For non-orthogonal multiple access schemes, like SDMA, the ability of the beamforming design to perform well highly depends on which users that were scheduled. The more spatially separated the scheduled users are, the higher performance can be expected. Therefore, in many new MIMO schemes, the user selection and beamforming design has merged into one, usually highly complex, block. The design and operation of such a scheduling and beamforming block highly depends on how much CSI that is available at the transmitter, and in which form. In the case of finite-rate feedback links from the users, the CSI can only be partial. Therefore, the form of the partial CSI (i.e. the feedback link), the scheduling and the beamforming must be jointly designed to enable multiuser diversity and/or SDMA. In this thesis, we focus on the design of low-rate feedback mechanisms and the corresponding scheduling and choice of beamforming, in particular for OFDMA, multiuser diversity and multi-antenna systems.

In Figure 1.2, a rough block diagram of the system we are studying in this thesis is shown. Below, a description of the parts is given, together with common assumptions that we make. The structure itself is also an assumption. The parts that are drawn with solid lines constitute the data-flow, where the information for the users is processed. The parts that are drawn with dashed lines constitute the control-flow, which is necessary for the data-flow to function properly. Figure 1.2 is rough and many parts of a real wireless communication system are omitted for simplicity. Note that the figure shows the downlink of a multiuser system.

User Data Buffers

The binary data for each of the $K$ active users is provided by the upper layers. Throughout this thesis, we assume that there always is enough data in the buffers. Queue-aware scheduling and cross-layer design is treated in more detail in [SLCZ04,
Scheduling

The scheduling decides from which user buffer to take data to encode and transmit. If SDMA is employed, data to several users will be transmitted simultaneously. The number of simultaneous users for SDMA is denoted $B$ and is called the SDMA factor. If the system uses OFDMA, data to $B$ users per sub-carrier can be encoded and transmitted.

AMC - Adaptive Modulation and Coding

Adaptive modulation and coding (AMC) is an important component of any multiuser diversity system. For AMC, the scheduler decides the coding and modulation scheme to be used for each user. For an OFDMA system, each user may have different AMC levels on different sub-carriers and on different data packets. The AMC block maps a sequence of bits onto a sequence of error correction encoded and modulated data symbols. The AMC level should adapt to the varying channel conditions of the user, so that a given error rate is not exceeded. For good channel conditions, many bits per data symbol can be transmitted, whereas for poor channel conditions, few bits per data symbol can be transmitted.

BF - Beamforming

The beamforming (BF) block maps the $B$ data symbols onto the $N_t$ transmit antennas. In the OFDMA context, there is one BF block per sub-carrier. In this thesis, we only study linear precoding or beamforming. This means that the $N_t$ symbols at the output of the BF block is a linear combination of the $B$ input symbols. This linear mapping is represented by the matrix beamforming matrix $W$, which may be unique for each sub-carrier.

MOD - Modulation

The modulation (MOD) block maps the time-discrete data symbols onto time-continuous signals, which can be transmitted into the world (the channel). There is one modulation block per transmit antenna. If OFDM is used, there is one OFDM modulator within each MOD block. Furthermore, the modulation block may change the frequency range of the transmitted signal, so that government regulations are met.

Channel

The signals are transmitted through a wireless communication channel. Different models of the channel have been used throughout the thesis and are discussed when they are introduced. In general the channel distorts the transmitted signals, due to
various physical phenomena. Throughout the thesis, it is assumed that the received signals also contain an additive white Gaussian noise component.

\( K \) Users

The multiuser system contains \( K \) active users. In different parts of the thesis, they are assumed to have one or several receive antennas. The receiver structure is not discussed in any great detail here or in the thesis. We only consider coherent detection, which means that the users are time- and frequency-synchronized with the base-station and that they are able to estimate how the channel distorts the transmitted signals. This is typically done by transmitting known symbols at known instants. The users need to know when they are scheduled and with which AMC level, in order to be able to detect and decode the data.

Feedback Link

It is possible for the users to estimate the downlink channel state by observing the distortion of known transmitted pilot signals. However, in systems with non-reciprocal up- and downlink channels (e.g. in FDD systems), this information is not readily available at the base-station. In order to enable the adaptation to the instantaneous channel conditions of the users, through scheduling, AMC and BF, the users need to feed back this information to the base-station. The users need to feed it back fast, otherwise it will be outdated and rather useless. This poses a large challenge in current wireless system design, where the users possibly are moving at very high speeds. In this thesis, however, we assume that the feedback link is without errors or delay. This assumption is rather optimistic, but may be justified if channel quality prediction is considered [FSES04]. Furthermore, in the numerical throughput evaluations for the downlink, we do not subtract the uplink signalling overhead.

Scheduling and Beamforming Design

The scheduling and beamforming design block controls the operation of the other low-level blocks in the transmitter in Figure 1.2. It makes decisions based on, for instance, the current and historic feedback information, previous decisions and the QoS-requirements of the users. Users and AMC should be selected, possibly for each sub-carrier, as well as beamforming matrices. These three decisions are highly inter-connected, and should for the best performance be done jointly.

The transmit power can, in general, be adapted in all three blocks in the data path. Throughout this thesis, we assume a maximum short-term transmit power per sub-carrier. If a sub-carrier is not used, the transmit power is not transferred to other sub-carriers. This is discussed more in Section 2.2. Furthermore, the sub-carrier power is distributed over the transmit antennas with a sum-power constraint.
1.6 Outline and Contributions

In this section, an outline of the thesis is presented along with an overview of its contributions.

Chapter 2, Multiuser OFDMA System Model

A large part of the thesis deals with OFDMA systems. In this chapter, a multi-cell, multi-antenna, downlink OFDMA system model is established.

Chapter 3, Reduced Feedback for OFDMA

In this chapter, a reduced feedback scheme for OFDMA is presented. We assume that the users will be scheduled on the different sub-carriers based on instantaneous CSI. Such a channel-aware scheduler with hypothetically perfect CSI for all users would presumably not schedule users on sub-carriers where they experience bad channel quality. Therefore, detailed information about these sub-carriers is not necessary to feed back. Furthermore, the channel quality on adjacent OFDM sub-carriers is typically highly correlated. This enables the granularity of the feedback to be reduced, by grouping several sub-carriers into clusters. Hence, the feedback can be reduced by

1. using clusters of sub-carriers as feedback unit and

2. feeding back information about only the best clusters.

Additionally, we assume that if there are few active users, it is likely that each user is scheduled on several clusters. If there are many active users, it is likely that they are scheduled only on their best cluster. Consequently, we introduce an adaptive feedback scheme, which lets the users feed back more information if there are few competing users and less information if there are many competing users. This helps avoid the linear increase in total feedback with the number of users. The adaptive feedback function can be designed based on, for instance, the expected spectral usage.

The work in this chapter was published in


CHAPTER 1. INTRODUCTION

Chapter 4, Using Unclaimed Sub-carriers

In this chapter, we elaborate on OFDMA systems with selective feedback, for instance as described in Chapter 3. If all users feed back information only about their best sub-carriers, there is a risk that the scheduler has no instantaneous information about some sub-carriers. Several approaches to use these unclaimed sub-carriers are proposed and evaluated by means of single-cell simulations.

The work in this chapter was published in


Chapter 5, Reduced Feedback by SNR Estimation at the Transmitter

In this chapter, we propose a feedback scheme for OFDMA which relies on channel quality (here SNR) estimation at the transmitter. Pilot symbols are typically placed in the time-frequency grid in the OFDM downlink, to enable channel estimation. In a similar fashion, the users feed back the channel quality of some sub-carriers in the time-frequency grid. This enables the scheduler to estimate the channel quality on the other sub-carriers. Several estimation methods are compared by means of simulations.

The work in this chapter was published in


Chapter 6, Modified Proportional Fair Scheduling for OFDMA

In this chapter, first a review of multiuser OFDM resource allocation is presented. The single-carrier PF scheduler of [VTL02] is extended to incorporate user individual QoS requirements as well as a tunable fairness level. The scheduling metric is derived from a utility function. It includes as special cases the maximum throughput, the PF and the max-min scheduling metrics. Additionally, a method to couple the opportunistic beamformer of Chapter 7 with the scheduler is presented.

The work in this chapter was published in


Chapter 7, Opportunistic Beamforming for OFDMA

Opportunistic beamforming was originally designed for a flat fading single-carrier scenario [VTL02]. In this chapter, an extension to a multi-carrier scenario is proposed. The extension is compatible with the reduced feedback scheme proposed in the previous chapter.

The work in Chapters 3 and 7 was published in


Chapter 8, Numerical Results for Chapters 6-7

The impact of different beamforming schemes on the system throughput depends on what kind of scheduling that is applied. The performance of different scheduling and beamforming methods depends on what kind of CSI that is available, which is connected to the used feedback scheme. Therefore we present combined numerical simulation results on the different proposed schemes in Chapters 2-6. Comparisons to several alternatives are made. The results show that the clustered beamforming scheme performs better than the alternatives. The modified proportional fair scheduler manages to differentiate the QoS for most users according to their target QoS.

Chapter 9, Opportunistic SD-OFDMA

In the opportunistic beamforming in Chapter 7, one symbol per sub-carrier was transmitted to the scheduled user. However, with multiple transmit antennas, it is possible to transmit multiple symbols to multiple users simultaneously, without interference. One way to achieve this, without needing full CSI at the transmitter, is to extend to opportunistic beamforming principle to multiple random but orthogonal beams. We propose an opportunistic space-division OFDMA (SD-OFDMA) scheme, which is an extension to the single-carrier version in [SH05]. We propose an OFDMA-specific enhancement to the scheme, which exploits the temporal correlation between blocks. Beams on clusters which resulted in a high sum-rate are not changed, whereas beams on clusters with a low scheduled sum-rate are regenerated. The chapter is concluded with results from multi-cell simulations.

The work in this chapter was published in

Chapter 10, Reduced Feedback SDMA Based on Subspace Packings

Opportunistic beamforming and SDMA was treated in Chapter 7 and 9, respectively. The base-station formed one or several beams randomly and relied on the multiuser diversity to provide users that could communicate on the beams.

Another related approach is to let the base-station and all users share a set of eligible beams, called a codebook. The users continuously estimate the channel and feedback which of the beams in the codebook that suits them best, together with the corresponding channel quality.

In this chapter we propose a feedback and beamforming concept for SDMA, based on such codebooks. A common problem in SDMA schemes with reduced feedback is that when the base-station has designed the multiple beams, it usually does not know exactly how much inter-beam interference the different scheduled users will experience. How to then efficiently assign rates to the users becomes an issue. This inconvenience is avoided in the proposed concept, as the post-scheduling SINR is directly available at the users when they select their preferred beam. Additionally, the feedback information implicitly defines which users that are spatially compatible. The codebooks are constructed of subspace packings. Three candidate packings are proposed and evaluated by simulation.

Chapter 11, Beam-graphs for Beamforming Codebooks

In this chapter, a method to reduce the feedback in codebook-based beamforming schemes is presented. It relies on the temporal channel correlation between consecutive blocks. All beams in the codebook are arranged in a beam-graph, so that similar beams are made neighbors. Instead of having the users feed back a beam index each block, a neighbor direction index can be fed back. The can significantly reduce the feedback rates with only a small performance loss.

The work in Chapter 10 and 11 was presented in


Contributions Outside the Scope of the Thesis

In this section, a brief overview of publications outside the scope of the thesis is given.

In [SBO04], the problem of estimating the Packet Error Rate (PER) of a coded communication system from a particular channel realization is addressed. This issue is especially relevant in the simulation of large communication networks. In particular, the connection between the channel realization and the PER of a coded and interleaved OFDM system is studied.
1.6. OUTLINE AND CONTRIBUTIONS


An overview of cross-layer scheduling aspects for multiuser MIMO systems is given in [AHSA+06]. Based on an information theoretic framework, several fundamental and practical issues with resource allocation for multiuser MIMO communications are discussed. Both single-carrier and OFDMA-specific aspects are included.


In [JSO07], single- and multi-beam opportunistic beamforming is considered. The impact of SNR, spatial correlation, user distribution and number of transmit and receive antennas is analyzed. It was concluded that the single-beam performance increases with the spatial correlation and user spread, whereas the opposite is true for multi-beam transmission. Furthermore, it was concluded that single-beam transmission is more suitable at high SNR, whereas multi-beam transmission is more suitable at low SNR.


Multiuser diversity scheduling for OFDMA was combined with an on-off water-filling scheme in [WSC04]. By not admitting some weak users, transmit power can be redistributed and performance improved.


In [BSZZ05], two multi-antenna transmission schemes that use partial CSI are compared in an OFDMA context. One scheme uses the second order statistics of the channel for beamforming with spatial interference suppression. The other scheme is an opportunistic beamforming scheme, similar to the one presented in Chapter 7. The results showed a higher system throughput for the opportunistic scheme, but with a less fair rate distribution among the users.

A similar comparison was done in [BSZ06], but now with an opportunistic SD-OFDMA scheme, as described in Chapter 9. Furthermore, multiple receive antennas were used. The smart antenna scheme outperformed the opportunistic scheme, in a scenario with few active users. The paper also contains an attempt to connect the simulated system performance with a cost analysis, assuming a system with one-hop relays.


In [LSZ+04, ZJL+04, ZSLZ05], results from experiments on a real-time DSP-based MIMO test bed are described. In [LSZ+04], the impact of receiver ADC imperfections on the error-rate performance was investigated. In [ZJL+04, ZSLZ05], the implementation a two-cell MIMO system was described. In the studied indoor scenario, the importance of transmitter interference suppression was demonstrated.


Large parts of Chapters 2, 3, 6, 7 and 8 were part of the licentiate thesis [Sve04].

Chapter 2

Multi-cell OFDMA System Model

2.1 Overview

This chapter contains a multi-cell downlink OFDMA baseband model. It is used in large parts of the thesis, occasionally in a simplified form.

2.2 Multi-cell Downlink OFDMA Baseband Model

In this chapter, the multi-cell baseband communication model is presented for single-input-single-output (SISO), multiple-input-multiple-output (MIMO) and multiple-input-single-output (MISO) downlink systems. The aim is to motivate the use of (2.1), (2.4) and (2.6) in the following chapters. OFDM is treated in more detail in [HMCK03].

Consider the downlink of an FDD OFDMA cellular system with $I$ base-stations that are synchronized in time, i.e. the OFDM symbols are transmitted simultaneously from all base-stations. This synchronism is advantageous in OFDMA multiuser diversity systems, due to the synchronism of training periods, which allows the users to predict the inter-cell interference (ICI). It is assumed that the users can

- estimate the full channel from the base-station they are connected to and
- predict the channel power on each sub-carrier from every base-station.

The estimation is necessary for coherent data detection and the prediction is necessary for the feedback and scheduling. Throughout the thesis, it is assumed that the scheduling decisions are made based on the channel conditions during the first data symbol in the corresponding block.

We assume a short-term power constraint per sub-carrier. Either the maximum transmit power is used for a sub-carrier, or no transmit power. This assumption is motivated in the following bullets.
1. In multiuser diversity systems with limited feedback that are studied in this thesis, adaptive power allocation typically requires additional signalling. The feedback quantized channel quality indicators are based on training, using a fixed power allocation. If the transmit power per sub-carrier is allowed to increase between the training and data transmission, the ICI may increase in an unpredictable way during data transmission, causing outages in adjacent cells. Furthermore, in SDMA systems, it is not clear how additional transmit power translates into increased channel quality, even if ICI is disregarded. Another aspect of allowing adaptive power allocation with AMC adaptation at the transmitter is that it introduces the need for forward link signalling of the chosen AMC level, which is not necessary if the feedback channel quality remains valid.

2. When each OFDM symbol is dedicated to one user, the gain from adaptive power allocation compared to equal power for all sub-carriers can be large. The reason is that transmit power is not wasted on weak sub-carriers. In multiuser OFDMA systems with channel aware scheduling and rate adaptation, the gain from adaptive power allocation is small, especially if the number of users is large [JL03, SL05a]. This is due to the fact that transmit power is generally not wasted on poor channels, since on each sub-carrier, a user with a good channel can be picked.

3. Transmit power regulations are usually not based on total transmit power, but in the form of a spectrum mask. The spectrum mask typically specifies a uniform maximum power spectral density in the communication band around the center frequency and a much lower maximum power spectral density for the out-of-band signal component. By always using a fixed, uniformly distributed and maximum power over the sub-carriers, the regulations are easily fulfilled.

Note that if some sub-carriers occasionally are not used, this is a form of coarse power allocation.

Other assumptions are:

- The channel is time invariant during one OFDM symbol.
- The frequency offset between all transmitters and receivers is zero.
- All transmitters and receivers use the same baseband sampling frequency.
- The OFDM cyclic prefix is longer than the channel impulse responses.

**SISO Model**

Assume that the base-stations transmit with one antenna and the users receive with one antenna (single-input-single-output or SISO). Omitting the time index,
the received symbol on sub-carrier \( n \) of a user in cell 0, synchronized to base-station 0 is

\[
y_n = \sqrt{P} H^0_n c^0_n + \sqrt{P} \sum_{i=1}^{I-1} \tilde{H}_i^i c_i^i + z_n \tag{2.1}
\]

where

- \( P \) is the transmit power per sub-carrier in all cells,
- \( c_i^i \) is the complex-valued symbol transmitted on sub-carrier \( n \) by base-station \( i \),
- \( H_n^i \) is the complex-valued sampled frequency response on subcarrier \( n \) between base-station \( i \) and the user,
- \( \tilde{H}_n^i = e^{j\phi_n^i} H_n^i \) is the rotated sampled frequency response on subcarrier \( n \) between base-station \( i \) and the user where the phase-shift \( \phi_n^i \) is due to the user not being synchronized to the base-station (see Appendix 2.A), and
- \( z_n \) is additive white Gaussian noise (AWGN) with variance \( \sigma_z^2 \).

See Appendix 2.A for more details on how the base-band model is derived. Here, it is assumed that all \( I \) base-stations transmit symbols on sub-carrier \( n \). In case they do not, the corresponding base-stations can be removed from the ICI sum. In the simulations presented in Chapter 8 and 9, the sum of the interference signals will be assumed to have a Gaussian distribution. Since frequency synchronization between all transmitters and receivers is assumed, the inter-carrier interference in the inter-cell interference signals is zero.

Assuming that \( E[|c_n^i|^2] = 1 \) and that the channel realizations are given, the instantaneous received SINR on sub-carrier \( n \) can be written as

\[
\gamma_n = \frac{|H_n^0|^2}{\sum_{i=1}^{I-1} |H_n^i|^2 + \sigma_z^2/P} \tag{2.2}
\]

**MIMO Model**

Now, assume that all base-stations use \( N_t \) antennas to transmit \( B \) symbols per sub-carrier\(^1\) and that all users have \( N_r \) receive antennas (multiple-input-multiple-output or MIMO). Base-station \( i \) applies on sub-carrier \( n \) the beamforming matrix \( W_n^i \) in order to distribute the symbols among the antennas. Assuming that the SISO model holds for each transmit and receive antenna pair, a user in cell 0 receives on sub-carrier \( n \)

\[
y_n = \sqrt{\frac{P}{B}} H_n^0 y_n^0 W_n^0 c_n^0 + \sqrt{\frac{P}{B}} \sum_{i=1}^{I-1} \tilde{H}_n^i y_n^i W_n^i c_n^i + z_n \tag{2.3}
\]

\(^1\)The model is easily extended to the case with different numbers of transmitted symbols on different sub-carriers and from different base-stations \( (B_n^i) \).
where \((\cdot)^T\) denotes transpose and

- \(\mathbf{c}_n^i = [c^i_{n,1} \cdots c^i_{n,B}]^T \in \mathbb{C}^{B \times 1}\) is the vector of symbols transmitted from base-station \(i\) on sub-carrier \(n\),
- \(\mathbf{W}_n^i = [w^i_{n,1} \cdots w^i_{n,B}] \in \mathbb{C}^{N_t \times B}\) is the beamforming matrix of base-station \(i\) on sub-carrier \(n\), normalized so that \(\text{Tr}(\mathbf{W}_n^i \mathbf{W}_n^i^H) = B\),
- \(\mathbf{H}_n^i \in \mathbb{C}^{N_r \times N_t}\) is the frequency domain channel matrix on sub-carrier \(n\) from base-station \(i\) to the user,
- \(\tilde{\mathbf{H}}_n^i = e^{j\phi_i} \mathbf{H}_n^i\) where \(\phi_i\) is an unknown scalar and
- \(\mathbf{z}_n \in \mathbb{C}^{N_r \times 1}\) is additive spatially and temporally white noise with covariance \(\sigma_z^2 \mathbf{I}_{N_r}\).

Now, assume that the user applies a linear receiver \(\mathbf{v}_n \in \mathbb{C}^{N_r \times 1}\) on sub-carrier \(n\) and that \(c_{n,m}^0\) is dedicated for the user, i.e.

\[
d_n = \mathbf{v}_n^H \mathbf{y}_n = \sqrt{\frac{P}{B}} \mathbf{v}_n^H \mathbf{H}_n^0 \mathbf{w}_n^0 \mathbf{c}_n^0 + \sqrt{\frac{P}{B}} \mathbf{v}_n^H \sum_{i=1}^{l-1} \tilde{\mathbf{H}}_n^i \mathbf{W}_n^i \mathbf{c}_n^i + \mathbf{v}_n^H \mathbf{z}_n
\]

\[
= \sqrt{\frac{P}{B}} \mathbf{v}_n^H \mathbf{H}_n^0 \mathbf{w}_n,m \mathbf{c}_n,m + \sqrt{\frac{P}{B}} \mathbf{v}_n^H \sum_{k \neq m} \mathbf{H}_n^0 \mathbf{w}_n,k \mathbf{c}_n,k + \mathbf{v}_n^H \mathbf{z}_n
\]

\[
= \sqrt{\frac{P}{B}} \mathbf{v}_n^H \sum_{i=1}^{l-1} \mathbf{H}_n^i \mathbf{W}_n^i \mathbf{c}_n^i + \mathbf{v}_n^H \mathbf{z}_n
\]

\[
= \alpha_{n,m} \mathbf{c}_n,m + \sum_{k \neq m} \beta_{n,k} \mathbf{c}_n,k + \sum_{i=1}^{l-1} \zeta_i^i \mathbf{c}_n^i + \tilde{\mathbf{z}}_n,
\]

where

- \(\alpha_{n,m} = \sqrt{\frac{P}{B}} \mathbf{v}_n^H \mathbf{H}_n^0 \mathbf{w}_n,m\) is the complex scalar effective channel gain,
- \(\beta_{n,k} = \sqrt{\frac{P}{B}} \mathbf{v}_n^H \mathbf{H}_n^0 \mathbf{w}_n,k\) is the complex scalar inter-beam interference gain from beam \(k\),
- \(\zeta_i^i = \sqrt{\frac{P}{B}} \mathbf{v}_n^H \mathbf{H}_n^i \mathbf{W}_n^i \mathbf{c}_n^i \in \mathbb{C}^{1 \times B}\) is the inter-cell interference gain from the beams in cell \(i\), and
- \(\tilde{\mathbf{z}}_n = \mathbf{v}_n^H \mathbf{z}_n\) is complex-valued AWGN with variance \(\sigma_z^2 \|\mathbf{v}_n\|^2\).
The linear receiver assumption is made for simplicity. There is no reason to believe that more advanced receiver structures would change any conclusions, since the thesis mainly considers the design of the transmitter.

The beam-SINR for beam \( m \) on sub-carrier \( n \) is

\[
\gamma_{n,m} = \frac{|\alpha_{n,m}|^2}{\sum_{k \neq m} |\beta_{n,k}|^2 + \sum_{i=1}^{I-1} \|\zeta_i^n\|^2 + \|v_n\|^2 \sigma_z^2}.
\] (2.5)

In the multiple-input-single-output (MISO) setup with only one transmitted beam, i.e. \( B = N_r = 1 \), there is no inter-beam interference and the multi-antenna transmission is transparent to the user. The transmitter beamforming and MISO-channel can be combined into a complex scalar effective channel, \( G_n^i = H_n^{iT} W_n^i \). Note that \( H_n^{iT} \in \mathbb{C}^{1 \times N_r} \) is a vector. The capital letter denotes frequency domain channel in this case, which may be slightly confusing. The received signal is then

\[
y_n = \sqrt{P} G_n^0 c_n^0 + \sqrt{P} \sum_{i=1}^{I-1} G_n^i c_n^i + z_n.
\] (2.6)

Assuming \( E[|c_n^i|^2] = 1 \) and that the channel realizations are given, the received SINR on sub-carrier \( n \) is

\[
\gamma_n = \frac{|G_n^0|^2}{\sum_{i=1}^{I-1} |G_n^i|^2 + \sigma_z^2/P}
\] (2.7)

which is similar to the SISO SINR, but with the SISO channels replaced by the effective channels, \( G_n^i \).
Appendix 2.A Baseband Model Details

Here, the baseband model for multiuser multi-cell downlink OFDM communication is studied. Carrier frequency modulation methods of the baseband OFDM signal do not differ from the traditional models [Pro01]. In this thesis, perfect frequency synchronization between transmitters and receivers is assumed. An FDD system is considered, where all base-stations transmit on the same frequency band, but a different band than the users use for uplink communication. Therefore, there is no interference between users in the downlink, but only from other base-stations. Hence, it is sufficient to study the communication link to one user. The received signal at the other users is similar.

Assume that \( I \) base-stations are deployed in the area of the user. The channels from all base-stations to the user are assumed to be time-dispersive, with impulse responses \( h^i(t) \), where \( i \) is the base-station index. The time dispersion of each channel impulse response, \( \Delta^i \), is at most \( \Delta_{CP} \), which is the length of the OFDM cyclic prefix [PR80]. The time-delays between the base-stations and the user, \( \tau^i \), are included in the impulse responses, so that \( h^i(t) = 0 \) for all \( t < \tau^i \) and for all \( t \geq \Delta^i + \tau^i \).

Base-station \( i \) transmits baseband OFDM symbols

\[
x^i(t) = \sum_{l=-\infty}^{\infty} x^i_l(t - lT_o)
\]

where \( T_o = NT + \Delta_{CP} \) is the OFDM symbol period, \( N \) is the number of subcarriers, and \( x^i_l(t) \) is the \( l^{th} \) OFDM symbol:

\[
x^i_l(t) = \begin{cases} 
\sum_{k=0}^{N-1} c_{k,l}^i e^{j2\pi kt/NT} & \text{when } t \in [-\Delta_{CP}, NT) \\
0 & \text{otherwise}
\end{cases}
\]

Assume that the studied user communicates with base-station 0. The user is assumed to be perfectly synchronized to this base-station. For convenience, let \( \tau^0 = 0 \), meaning that \( \tau^i \) can be seen as the difference in physical delay to the user between base-station \( i \) and 0. The user receives the signal

\[
y(t) = x^0(t) * h^0(t) + \sum_{i=1}^{I-1} x^i(t) * h^i(t) + z(t) = y^0(t) + y^{ICI}(t) + z(t)
\]

where \( * \) denotes convolution and \( z(t) \) is AWGN. The received signal consists of the desired signal, \( y^0(t) \), inter-cell interference, \( y^{ICI}(t) \), and AWGN, \( z(t) \). The timing relation between the different terms in the sum of received signal is illustrated in Figure 2.1.

We can study the reception of one OFDM symbol without loss of generality. The received signal \( y(t) \) is sampled at \( t = mT \) for \( m = 0, \ldots, N - 1 \), which means
2.A. BASEBAND MODEL DETAILS

Figure 2.1: The received signal is the sum of the received signals from all base-stations $y^i(t)$ (plus noise). Since the user is synchronized to base-station 0, the delay $\tau^0$ can be set to zero. The delays from the other base-stations $\tau^i$ are then the additional delay compared to base-station 0.

that the cyclic-prefix samples are thrown away. The decision variables are obtained after performing the DFT of the samples.

$$y_n = \frac{1}{N} \sum_{m=0}^{N-1} y(mT)e^{-j2\pi mn/N} = y^0_n + y^{IC1}_n + z_n$$

for $n = 0 \ldots N - 1$. The desired signal part is successfully demodulated as

$$y^0_n = \frac{1}{N} \sum_{m=0}^{N-1} \int_0^{\Delta_0} h^0(\tau)x^0(mT-\tau)d\tau e^{-j2\pi mn/N}$$

$$= \frac{1}{N} \sum_{m=0}^{N-1} \int_0^{\Delta_0} h^0(\tau) \sum_{k=0}^{N-1} c^0_k e^{j2\pi k(mT-\tau)/NT} d\tau e^{-j2\pi mn/N}$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} c^0_k \int_0^{\Delta_0} h^0(\tau)e^{-j2\pi k\tau/NT} d\tau \sum_{m=0}^{N-1} e^{-j2\pi m(n-k)/N}$$

$$= \sum_{k=0}^{N-1} c^0_k H^0_k \left(\frac{k}{NT}\right) \delta(n-k)$$

$$= c^0_n H^0_n = c^0_n H^0_n$$

(2.12)
where $H^0(f)$ is the Fourier transform of $h^0(t)$. Hence, the desired signal part in the demodulated signal is the desired symbols $c_n^0$ multiplied with the frequency response of the channel, $H^0(f)$, sampled at $f = n/NT$.

The inter-cell interference term, $y_n^{\text{IC}}$, contains contributions from all interfering base-stations. Due to the different propagation times to the user, the sampling and demodulation will not be correctly synchronized. Let $I_{\text{No ISI}}$ denote the set of interfering cells for which $\Delta^i + \tau^i < \Delta_{CP}$. Sampling the signal from any cell in this set at $t = mT$ for $m = 0, \ldots, N - 1$ gives samples from only one OFDM symbol, hence no ISI. Let $I_{\text{ISI}}$ denote the other cells. Hence, $y_n^{\text{IC}} = y_n^{\text{No ISI}} + y_n^{\text{ISI}}$. Sampling and demodulation of the signals from the cells in $I_{\text{No ISI}}$ gives

\[
y_n^{\text{No ISI}} = \sum_{i \in I_{\text{No ISI}}} \frac{1}{N} \sum_{m=0}^{N-1} \int_{\tau^i}^{\tau^i + \Delta^i} h^i(\tau) x^i(mT - \tau) d\tau \ e^{-j2\pi mn/N}
\]

\[
= \sum_{i \in I_{\text{No ISI}}} \frac{1}{N} \sum_{m=0}^{N-1} \int_{\tau^i}^{\tau^i + \Delta^i} h^i(\tau) \sum_{k=0}^{N-1} c_k^i e^{j2\pi k(mT - \tau)/NT} d\tau \ e^{-j2\pi mn/N}
\]

\[
= \sum_{i \in I_{\text{No ISI}}} \frac{1}{N} \sum_{k=0}^{N-1} c_k^i \int_{\tau^i}^{\tau^i + \Delta^i} h^i(\tau) e^{-j2\pi k\tau/NT} d\tau \sum_{m=0}^{N-1} e^{-j2\pi m(n-k)/N}
\]

\[
= \sum_{i \in I_{\text{No ISI}}} \sum_{k=0}^{N-1} c_k^i \tilde{H}^i \left( \frac{k}{NT} \right) \delta(n-k) = \sum_{i \in I_{\text{No ISI}}} c_i^i \tilde{H}^i \left( \frac{n}{NT} \right)
\]

\[
= \sum_{i \in I_{\text{No ISI}}} c_i^i \tilde{H}^i_n
\]

where $\tilde{H}^i(f) = e^{-j2\pi f \tau^i} F\{h^i(t + \tau^i)\}$, which is the rotated frequency response of the non-delayed channel, $h^i(t + \tau^i)$, from the $i^{th}$ base-station.

The channels from the base-stations that belong to the set $I_{\text{ISI}}$ have a time-delay $\tau^i$ and an impulse response length $\Delta^i$ such that $\tau^i + \Delta^i \geq \Delta_{CP}$. Some of the received samples at $mT$, $m \in \{0, \ldots, N - 1\}$, contain contributions from two different OFDM symbols, making the $y_n^{\text{ISI}}$ term more complicated. Demodulation by DFT of a sum of interfering non-synchronized signals is closely related to the phenomenon of cross-talk in DSL, and has been investigated thoroughly in this context [GK02]. According to [SCM97], $y_n^{\text{ISI}}$ can be written as

\[
y_n^{\text{ISI}} = \sum_{i \in I_{\text{ISI}}} c_i^i \frac{N - \epsilon^i}{N} \tilde{H}^i \left( \frac{n}{NT} \right) + z_n^i
\]

(2.14)
where $\epsilon_i$ is the number of samples within the DFT window that contain contributions from two symbols, and $\tilde{z}_i$ is an extra noise term that is introduced since the orthogonality of the system is disturbed. It can be assumed that $\sum_{i\in I_{ISI}} \tilde{z}_i = \tilde{z}_n$ is approximately Gaussian, motivated by the general central limit theorem for sums of unequally distributed random variables [Cra61]. For simplicity, assume that $\epsilon_i \ll N$ for all $i$, resulting in

$$y_{ISI}^n = \sum_{i \in I_{ISI}} c_i^i \tilde{H}_i \left(\frac{n}{NT}\right) + \tilde{z}_n.$$  \hfill (2.15)

It is reasonable to assume that the ICI signals with most delay are also among the weakest, which helps to motivate the previous approximations. To reduce the effects of asynchronous interference signals, windowing of the OFDM symbols can be applied [WE71]. This technique removes or reduces the discontinuities between consecutive OFDM symbols.

The sampling and demodulation of the receiver noise $z(t)$ is as for standard OFDM. Assume that the receiver has a filter $f(t)$ that band-limits the signal, with frequency response

$$F(f) = \begin{cases} 1 & f \in \left[-\frac{A}{T}, \frac{A}{T}\right] \\ 0 & \text{otherwise} \end{cases}$$  \hfill (2.16)

for some large integer $A$. Then consider the filtered noise $z_f(t) = z(t) * f(t)$, where $f(t) = \frac{2A}{T} \text{sinc}(\frac{2AT}{T})$. Furthermore, assume that $A$ is large enough so that the fact that the frequencies of the OFDM signals outside the filter were cut off can be ignored. To avoid aliasing in the sampled received signal, the receiver sampling frequency has to be at least $\frac{2A}{T}$. The autocorrelation of the filtered noise becomes

$$E[z_f(t)z_f(s)^*] = \frac{4A^2}{T^2} E \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z(t-\tau) \text{sinc} \left(\frac{2\tau A}{T}\right) z(s-\sigma)^* \text{sinc} \left(\frac{2\sigma A}{T}\right) d\tau d\sigma \right]$$

$$= \frac{2N_0A^2}{T^2} \int_{-\infty}^{\infty} \delta(t-\tau-s+\sigma) \text{sinc} \left(\frac{2\tau A}{T}\right) \text{sinc} \left(\frac{2\sigma A}{T}\right) d\tau d\sigma$$

$$= \frac{2N_0A^2}{T^2} \int_{-\infty}^{\infty} \text{sinc} \left(\frac{2\tau A}{T}\right) \text{sinc} \left(\frac{(s-t+\tau)2A}{T}\right) d\tau$$

$$= \frac{N_0}{2} \int_{-\infty}^{\infty} F(f) e^{j2\pi f(s-t)} df$$

$$= \frac{N_0A}{T} \text{sinc} \left(\frac{(s-t)2A}{T}\right)$$  \hfill (2.17)

where $N_0/2$ is the power spectral density of $z(t)$. Sampling the filtered noise and
demodulating by DFT gives the frequency domain noise \( \tilde{z}_n \) with autocorrelation

\[
E[\tilde{z}_l \tilde{z}^*_m] = E \left[ \frac{1}{N} \sum_{n=0}^{N-1} z_f(nT)e^{-j2\pi nl/N} \frac{1}{N} \sum_{k=0}^{N-1} z_f(kT)^*e^{j2\pi km/N} \right]
\]

\[
= \frac{N_0A}{N^2T} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} \text{sinc}(2A(n-k))e^{j2\pi(km-ln)/N}
\]

\[
= \frac{N_0A}{N^2T} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} \delta(n-k)e^{j2\pi(km-ln)/N}
\]

\[
= \frac{N_0A}{N^2T} \sum_{n=0}^{N-1} e^{j2\pi n(m-l)/N} = \frac{N_0A}{NT} \delta(m-l).
\]

(2.18)

Hence, \( \tilde{z}_n \) is additive white Gaussian noise with variance \( \frac{N_0A}{NT} \). The total AWGN is then \( z_n = \tilde{z}_n + \tilde{z}_n \) with variance \( \sigma^2_z \).

Combining (2.11) - (2.18) gives

\[
y_n = c^0_nH^0_n + \sum_{i=1}^{I} c^i_n \tilde{H}^i_n + z_n, \tag{2.19}
\]

which is the multi-cell baseband SISO model proposed in Section 2.2.
Chapter 3

Reduced Feedback for OFDMA

3.1 Overview

This chapter contains:

- A reduced feedback scheme for OFDMA, that adapts the feedback rate per user, as a function of the number of users.

- A numerical evaluation of the scheme.

3.2 Introduction

The sum-rate capacity of the frequency-selectively fading SISO downlink channel with an average sum-power constraint is achieved by transmitting to the best user on each frequency, together with a waterfilling power allocation across time and frequency [Tse97, TV05]. For an OFDMA system, the corresponding strategy would be to transmit to the best user on each sub-carrier. In order to achieve this, the base-station requires channel quality information (CQI) about all users and sub-carriers. In systems with non-reciprocal channels, like FDD, this implies a large amount of feedback from the users. The CQI can be in the form of SINR, supportable rate or preferred adaptive coding and modulation level, etc..

Consider a system with $K$ active users using OFDM with $N$ sub-carriers. To exploit multiuser diversity, the scheduler requires CSI about the different users. The full channel state information of all users consists of $KN$ complex numbers, and even more if multiple antennas are used. For scheduling, a scalar (e.g. the SINR) per sub-carrier is sufficient, hence totally $KN$ real numbers. Assuming quantization of the channel quality to $M_{AMC}$ levels, $KN \log_2(M_{AMC})$ bits need to be fed back. Also this feedback information can create a very large overhead. In this chapter, feedback reduction techniques with small performance reduction are considered.

Two main approaches to the problem have been taken. One approach is to define a, possibly adaptive, channel quality threshold. If the sub-carrier quality of
a user is above the threshold, the user is allowed to feed back information. Another approach to reduce the feedback is to allow each user to feed back the channel quality of a, possibly adaptive, number of their best sub-carriers.

The threshold scheme is a form of coarse quantization. In [Joh03, FEM03], the effect of quantized channel quality feedback on a single-carrier multiuser diversity system was studied. Both conclude that a large portion of the full-feedback throughput can be achieved even with highly quantized feedback. The quantization thresholds were adapted to the number of active users, with higher threshold values when more users were active.

In [GA03, GA04], a threshold-based feedback scheme for single-carrier systems was proposed and evaluated that lets users feed back their detailed channel quality if it is above a threshold. The threshold was adapted to the number of users to attain a certain scheduling outage probability, i.e. the probability that no user fed back CQI. Again, the conclusion was that a large part of the multiuser diversity gain can be maintained while reducing the feedback load significantly. The problem with scheduling outages was treated in [HAOG05, HAOG06], where a second round of feedback was allowed when an outage occurred. In [HAGO05, HAGO07], scheduling outages were avoided by adapting the threshold level with time depending on how many users that actually fed back. A relative threshold, instead of an absolute, was considered in [YAG04].

The threshold based scheme was taken further in [SN05], where each user was allowed to feed back only one bit, ‘1’ if the channel quality was above the threshold and ‘0’ otherwise. If the threshold was properly adapted to the number of users, it was shown that the throughput growth rate with the number of users was the same as for a full feedback scheme.

This approach was applied to OFDMA in [SNA04], where each user fed back one bit per sub-carrier. The same asymptotic throughput growth rate as for the full feedback scheme was shown. Point-to-point OFDM with one bit per sub-carrier feedback was studied in [RVG06].

### 3.3 Reduced Feedback by Clustering and Adaptive Selection

In this section, we propose a reduced feedback scheme for OFDMA based on three components:

- grouping adjacent sub-carriers into clusters,
- only feeding back information about the strongest clusters and
- adapting the number of fed back clusters to the number of active users.
3.3. REDUCED FEEDBACK BY CLUSTERING AND ADAPTIVE SELECTION

Feedback Reduction By Clustering

To reduce the amount of overhead without sacrificing too much in performance, the following measures can be taken. Since the correlation between sub-carriers generally is high, due to the limited delay spread in the channel, the $N$ sub-carriers can be divided into $Q$ clusters of $R$ adjacent sub-carriers, which can be used as feedback units. The feedback information is a measure of the channel quality in the cluster, for instance, the minimum or average sub-carrier supportable rate within the cluster. In a well-designed system, the cluster-size, $R$, is chosen so that the sub-carriers within one cluster are highly correlated. Having small clusters offers better feedback accuracy for the sub-carriers in the cluster, but does not reduce feedback much. Having large clusters, on the other hand, reduces feedback more, but increases the risk of users feeding back supportable rates lower than necessary for some sub-carriers. Hence, finding a suitable cluster-size in practice involves many parameters, for instance channel statistics, the adaptive modulation and coding properties and uplink signalling constraints, requiring detailed system simulations. Still, analysis of the channel statistics can give some hints on suitable cluster-sizes. This is analyzed in more detail in Appendix 3.B and simulated in Section 3.4. Clustering is illustrated in Figure 3.1. The figure shows the channel gains of a user across the 64 sub-carriers. The sub-carriers are grouped in clusters of 4 adjacent sub-carriers. The cluster borders are depicted with vertical lines. In this example, there are significant channel variations within several of the 16 clusters.

The clustering of sub-carriers in an OFDM link was proposed already in [CDS96]. Clustered OFDM was also considered in [GBR01, SOAS03, WOS+03], which did not focus on the feedback aspect. Note that sub-carrier clustering could also be applied to threshold-based feedback schemes, which were discussed in Section 3.2.

Selective Cluster Feedback

Furthermore, a scheduler exploiting multiuser diversity usually does not schedule users on their weaker clusters. Hence, the amount of feedback information can
be further reduced by letting each user feed back information only about its $S$ strongest clusters. This, however, introduces the need to also feed back the indices of $S$ clusters, giving a total feedback rate of $SK \log_2(QM_{AMC})$ bits per transmission block. The feedback reduction is illustrated in Figure 3.2. The total feedback rate per scheduling block as a function of the cluster-size, $R$, and the number of feedback clusters, $S$ is illustrated in Figure 3.3.

Henceforth, the set of $N$ sub-carriers is denoted $\mathcal{N} = \{1, \ldots, N\}$. The set of clusters is denoted $\mathcal{Q} = \{1, \ldots, Q\}$. The set of sub-carriers that belong to cluster $q$ is denoted $Rc(q)$. User $k$ feeds back channel quality information about the clusters in $S_k$, and $|S_k| = S$. 

Figure 3.2: Illustration of the reduced feedback scheme. Instead of feeding back the supportable rates of all 64 individual sub-carriers in the uppermost figure, the minimum supportable rate within the clusters are fed back. In the middle figure, only the sub-carrier with minimum channel gain within each cluster remains, symbolizing the feedback reduction after clustering. Furthermore, only the rates of the strongest clusters are fed back. In the bottom figure, only the channel gains of the strongest clusters remain. In this example, the clusters-size is 4 and the rates of the 3 strongest clusters are fed back.
3.3. REDUCED FEEDBACK BY CLUSTERING AND ADAPTIVE SELECTION

Figure 3.3: This figure shows the total feedback in bits per scheduling block for a cell with 32 users, 512 sub-carriers and 8 adaptive modulation modes as a function of the cluster-size, $R$, and the number of fed back clusters, $S$. This can be compared to the feedback without reduction, which would be almost 50 kbit per scheduling block.

Adaptive Feedback

When $S < Q$, i.e. the users don’t feed back all clusters, there is a non-zero probability that the base-station receives no information about at least one sub-carrier. In [GA03], this was called scheduling outage (in the single-carrier context).

**Definition 3.1 (Scheduling outage).** A scheduling outage is the event that there is at least one cluster that no user feeds back, i.e. $\bigcup_{k=1}^{K} S_k \subset Q$.

In the case of scheduling outage, the base-station may choose between not using the sub-carrier or transmitting blindly- or semi-blindly. Various uses of these unclaimed sub-carriers are considered in Chapter 4. Here, we assume that the unclaimed sub-carriers remain unused.

**Definition 3.2 (Spectral usage).** The spectral usage, $\bar{U}$, is the fraction of the clusters that at least one user has fed back, i.e.

$$\bar{U} = \frac{1}{Q} \left| \bigcup_{k=1}^{K} S_k \right| = \frac{U_K}{Q}$$
Naturally, a low spectral usage is undesired. With no instantaneous CQI about a large part of the downlink spectrum, the scheduler is not able to fully exploit the multiuser diversity. In Appendix 3.A, the probability mass function (PMF) and expected value of $\bar{U}$ are derived. With the number of clusters, $Q$, fixed, it is possible to find an adaptive feedback scheme, that adapts $S$ to the number of users $K$. It is described in the following:

- Offline, determine the desired function $S(K)$ to fulfill a target spectral usage. The target can for instance be in either of the following forms,

1. $E[\bar{U} | S, K, Q] \geq \bar{U}_{\text{target}}$ or
2. $\Pr(\bar{U} \leq \bar{U}_{\text{threshold}} | S, K, Q) \leq P_{\text{target}}$.

The design parameters $\bar{U}_{\text{target}}$, $\bar{U}_{\text{threshold}}$ and $P_{\text{target}}$ are used to trade off spectral usage with feedback overhead. Since $S$ is an integer, the target can only be met approximately.

- Each user must feed back at least one cluster, no matter how many users are active\(^1\).

- Depending on the number of active users, the base-station broadcasts the number of clusters, $S$, each user is to feed back. Since the number of users typically is a slowly changing parameter, this induces very little feed forward overhead. To reduce the complexity of the feedback protocol, the adaptive feedback function $S(K)$ can be additionally quantized.

This scheme has several positive effects. The aggregate feedback rate as a function of the number of users can be kept fairly constant. For a very high number of active users, they each feed back only their best cluster. For few users, they each feed back many clusters, so that a large part of the bandwidth can be used. The adaptive feedback scheme balances the loss from scheduling outages with the feedback rate. Note that this adaptive scheme can also be applied on a sub-carrier level, if clustered feedback is not used.

Figure 3.4 illustrates the adaptive feedback scheme. In this example, the number of clusters $Q = 8$ and the number of modulation levels $M_{\text{mod}} = 8$. For the fixed feedback scheme, each user feeds back the modulation level and index of the 3 strongest clusters ($S = 3$). For the adaptive feedback rate, $S$ is chosen so that the probability that less than 80% of the clusters are fed back is approximately 0.2. In (a), the resulting $S$ for the adaptive scheme is displayed together with the fixed $S$ for the fixed scheme. For few users ($K < 7$), each user feeds back more information than in the fixed scheme, but less when there are many users ($K > 12$). In (b), the target probability is shown. The abrupt changes in the probability for the adaptive scheme for large $K$ are due to the integer granularity of $S$. From (c), it is clear

\(^1\)It would be possible to relax $S$ to be a rational number, and let each user feed back only a fraction of the blocks, in a round-robin fashion.
3.4. **REDUCED FEEDBACK - REDUCED PERFORMANCE?**

The cluster-size is an important design parameter. A larger cluster-size implies that fewer clusters have to be fed back to maintain a specific spectral usage. However, it is important that the cluster bandwidth is not much larger than the coherence bandwidth of the channel. Otherwise, the reported supportable rate will not properly represent all the sub-carriers within the cluster. This will lead to less accuracy.

---

**Figure 3.4:** In (a)-(d), the adaptive feedback rate scheme is compared to a fixed feedback scheme as a function of the number of users, $K$.

(a) The feedback per user (clusters)

(b) The probability that the spectral usage is below 75%

(c) The expected spectral usage

(d) Sum feedback of all users (bits per block)

that the adaptive feedback scheme results in a relatively constant expected spectral usage. The total feedback rate for all users in bits per scheduling block is depicted in (d). Since, the feedback rate per user is fixed for the fixed scheme, the total feedback rate increases linearly with the number of users. For the adaptive scheme, the total feedback rate remains fairly constant for the whole range of $K$.

---

**3.4 Reduced Feedback - Reduced Performance?**

The cluster-size is an important design parameter. A larger cluster-size implies that fewer clusters have to be fed back to maintain a specific spectral usage. However, it is important that the cluster bandwidth is not much larger than the coherence bandwidth of the channel. Otherwise, the reported supportable rate will not properly represent all the sub-carriers within the cluster. This will lead to less accuracy.
in the choice of adaptive modulation.

Here, a simulation example is presented to illustrate the effect of the clusters-size. 32 users in one cell have channels generated according to a model of time-dispersive channel [MAS+98]. In this example, no inter cell interference (ICI) is considered. The base-station has one transmit antenna and the users have one receive antenna. The impulse responses are of the form

\[ h(\tau) = \sum_{l=0}^{L-1} \beta_l \delta(\tau - lT), \]  

(3.1)

where \( T = 50 \) ns is the sample period. The channel taps \( \beta_l \) are independent, complex Gaussian variables with an exponentially decaying power delay profile

\[ E[|\beta_l|^2] = Ae^{-\frac{lT}{\nu}}, \]  

(3.2)

where \( \nu \) is the expected root mean square (RMS) delay spread and \( A \) is chosen so that \( E\left[\sum_{l=0}^{L-1} |\beta_l|^2\right] = 1 \). The channels of the users are assumed to be independent but with the same power delay profile. We have used 64 sub-carriers and a cyclic prefix of 800 ns (as in HIPERLAN/2).

The downlink rate in bits per OFDM symbol is estimated as

\[ C = \frac{R}{2} \sum_{q \in Q} \log_2(1 + \frac{\gamma_q}{\Gamma}), \]  

(3.3)

where \( R \) is the number of sub-carriers per cluster, \( Q \) is the set of clusters that have been assigned to users, \( \gamma_q \) is the fed back SNR in cluster \( q \) for the assigned user, and \( \Gamma \) is the gap corresponding to a symbol error rate of \( 10^{-4} \) for QAM [CDEF95, GC97]. Note that if a scheduling outage occurs on cluster \( q \), then \( \gamma_q = 0 \). The users feedback the SNR of the weakest sub-carrier in the cluster. It is assumed that all users estimate and feedback the SNRs perfectly. For each cluster in the downlink, the user with the highest reported SNR is scheduled. This maximizes the throughput in the SISO downlink. Also note that (3.3) assumes that the modulation can be perfectly adapted to the reported SNRs.

Figure 3.5 shows the downlink throughput in bits per OFDM symbol for different cluster-sizes, \( R \), as a function of the mean RMS delay spread of the channels, \( \nu \). Each user feeds back the minimum sub-carrier supportable rate within the 4 strongest clusters (\( S = 4 \)). The throughput of the full feedback system is also shown, in which the users feed back the SNR’s of all sub-carriers. The full feedback scheme performs best, since the scheduler can select the user with the highest SNR on each sub-carrier. For the reduced feedback scheme, the throughput depends on the channel delay spread. For relatively flat channels (low delay spread), large clusters (\( R = 8 \) and \( R = 16 \)) give the highest throughput. The reason for this is twofold. Firstly, the fed back value for each cluster is highly representative for all sub-carriers within the cluster, which gives high accuracy in the adaptive
3.4. REDUCED FEEDBACK - REDUCED PERFORMANCE?

Figure 3.5: The total cell throughput in bits per OFDM symbol is displayed as a function of the RMS delay spread of the channel. In the full feedback scheme, each user feeds back information about all sub-carriers (no clustering or reduced feedback). For the reduced feedback scheme, large cluster-sizes, \( R \), are suitable for relatively flat channels (low delay spread), whereas small cluster-sizes are more suitable for more frequency-selective channels (large delay spread). In this example, the number of users is 32. In the reduced feedback scheme, each user feeds back the information about the 4 strongest clusters.

modulation. Secondly, users with strong channels will likely experience high SNRs across all sub-carriers, due to the low delay spread. The feedback of four large clusters enables the scheduler to allocate a large portion of the OFDM symbol to the strongest user. For large clusters, the probability that some clusters are left unassigned (since no user fed them back) is very low. For low delay spread and small clusters \( (R = 1) \), the strongest users will be assigned only a small part of the OFDM symbol. Also, the expected spectral usage for \( R = 1 \) is 87% whereas it is 100% for \( R \geq 4 \). For increasing delay spread, the frequency selectivity of the channels increases. This is exploited by the schemes using smaller clusters, since the users may be scheduled only on their frequency fading peaks. Large clusters are not suitable for larger delay spreads since there can be significant channel variations within the cluster. Since the SNR of the weakest sub-carrier within the cluster is fed back in this example, the base-station will use a too low modulation level for most sub-carriers within the large clusters.
From this example, one can see that a suitable cluster-size depends on the average channel delay spread of the users. Since it is not possible to use different cluster-sizes for different users, a compromise that gives good performance for most expected channels is a suitable choice. For this example and delay spread in the range of Figure 3.5, a cluster-size of 2 or 4 sub-carriers seems to be appropriate. For the channel model used in this example, it is possible to derive the probability that a sub-carrier within a cluster deviates significantly from the average within the cluster. It is a function of the covariance matrix of the channel as well as the cluster-size. The details are given in Appendix 3.B.
Appendix 3.A  Adaptive Feedback Rate Derivation

Consider the OFDMA model, where the sub-carriers are divided into $Q$ clusters. Each of the $K$ users independently chooses $S$ clusters for feedback with uniform prior probability. In terms of sets, user $k$ chooses $S_k \subset Q$ for feedback. In this appendix, the probability mass function (PMF) of the discrete random variable $\bar{U} \in [0, 1]$, $p_{\bar{U}}(u) = \Pr(\bar{U} = u)$, is derived. In Definition 3.2, $\bar{U}$ was defined as the spectral usage, i.e. the ratio of the clusters that at least one user had fed back. Since, $\bar{U} = U_K/Q$, we can equivalently study $U_K = |\bigcup_{k=1}^K S_k| \in Q$, which is the number of clusters that have been fed back by at least one of the $K$ users. The PMF, $\Pr(U_K = u)$, is 0 for $u < S$.

Let the random variable $U_k$ denote the number of different clusters, out of the $Q$ clusters, that have been fed back when $k$ users have made their pick. The probability that less than $S$ clusters are picked is zero, as is the probability that more than $Q$ are picked. The non-zero probabilities can be stacked in a vector and computed recursively from

$$P_k = \begin{pmatrix} \Pr(U_k = S) \\ \Pr(U_k = S + 1) \\ \vdots \\ \Pr(U_k = Q) \end{pmatrix} = A \begin{pmatrix} \Pr(U_{k-1} = S) \\ \Pr(U_{k-1} = S + 1) \\ \vdots \\ \Pr(U_{k-1} = Q) \end{pmatrix} = AP_{k-1} \quad (3.4)$$

where the lower triangular matrix $A$ denotes the probability transitions when one more user picks clusters.

$$A = \begin{pmatrix} \Pr(U_k = S | U_{k-1} = S) \\ \Pr(U_k = S + 1 | U_{k-1} = S) \\ \vdots \\ \Pr(U_k = Q | U_{k-1} = S) \\ \vdots \\ \Pr(U_k = Q | U_{k-1} = Q) \end{pmatrix}. \quad (3.5)$$

The elements of $A$ are computed as

$$\Pr(U_k = S + t | U_{k-1} = S + r) =
$$

$$\begin{cases} \binom{S + r}{S - (t - r)} \binom{Q - S - r}{t - r} & \text{if } r \leq t \text{ and } t - r \leq S \\ \binom{Q}{S} & \text{otherwise} \end{cases} \quad (3.6)$$

Since the first user ($k = 1$) picks $S$ different clusters,

$$P_1 = \begin{pmatrix} \Pr(U_1 = S) \\ \Pr(U_1 = S + 1) \\ \vdots \\ \Pr(U_1 = Q) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.7)$$
CHAPTER 3. REDUCED FEEDBACK FOR OFDMA

Figure 3.6: The PMF of $\bar{U}$ is shown in (a) and (b) for 5 and 10 users and different feedback rates $S$. The number of clusters, $Q$, is 64. $S/Q = 1/32$ means that each user feeds back 2 clusters.

and

$$P_K = A^{K-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$  

Hence, the PMFs of $U_K$ and $\bar{U}$ are given by

$$p_{U_K}(u) = \begin{cases} 0, & \text{for } u \in \{0, \ldots, S - 1\} \\ P_K(u - S + 1), & \text{for } u \in \{S, \ldots, Q\} \end{cases}$$

$$p_{\bar{U}}(u) = \frac{1}{Q} p_{U_K}(u),$$

where $P_K(i)$ is the $i^{th}$ element of $P_K$. The CMF of $\bar{U}$ is simply

$$\Pr(\bar{U} \leq u) = \frac{1}{Q} \sum_{i=0}^{u} p_{\bar{U}}(u).$$  

For a fixed $S$, the expected spectral usage can be computed as

$$E[\bar{U}] = \sum_{u=S}^{Q} u p_{\bar{U}}(u).$$  

(3.10)

To find the feedback rate $S$ such that $\Pr(U_K \leq u)$ does not exceed a threshold for all $K$, a full search of all integer $S \in \{1, \ldots, Q - 1\}$ can be done. Similarly, to find the minimum $S$ as a function of $K$ such that $E[\bar{U}]$ is above a threshold,
a full search over $S$ for different $K$ can be done. Note that the adaptive feedback scheme is designed offline. The expected spectral usage can also be found directly as [CB06]

\[
E[U] = 1 - \left(1 - \frac{S}{Q}\right)^K.
\tag{3.11}
\]

In Figure 3.6, the PMF of $\hat{U}$ is shown for a system with 64 clusters ($Q = 64$). In the left Figure, $p_G(u)$ for 5 users is shown for various feedback rates $S$. In order to enable a high spectral usage (above 80%), each user needs to feedback $3Q/8 = 24$ clusters. In the left figure, 10 users are active. To get approximately the same spectral usage, each user needs to feedback $2Q/8 = 16$ clusters.

**Appendix 3.B  On Sub-carrier Variations Within a Cluster**

**Correlation Between Sub-carriers**

The correlation between the $k^{th}$ and $n^{th}$ sub-carriers is

\[
\rho_{kn} = \frac{E[H_k H_n^*]}{\sqrt{\text{var}(H_k) \text{var}(H_n)}}. \tag{3.12}
\]

Here, the channel model (3.1)-(3.2) with a sample spaced exponentially decaying power delay profile is used. It is stated below for reference. The taps in the channel impulse response are spaced $T$ apart,

\[
h(\tau) = \sum_{l=0}^{L-1} \beta_l \delta(\tau - lT),
\]

where the taps are zero-mean, independent and with variances

\[
E[|\beta_l|^2] = \sigma_l^2 = Ae^{-lT/2\nu}, \tag{3.13}
\]

with $A$ being a normalization constant such that $\sum_{l=0}^{L-1} \sigma_l^2 = \sigma_h^2$, i.e.

\[
A = \frac{\sigma_h^2}{\sum_{l=0}^{L-1} e^{-lT/2\nu}} = \sigma_h^2 \frac{1 - e^{-T/2\nu}}{1 - e^{-LT/2\nu}}. \tag{3.14}
\]
The numerator in (3.12) can be found as

\[
E[H_k H_n^*] = E \left[ \frac{1}{N^2} \sum_{m=0}^{L-1} \beta_m e^{-j2\pi mk/N} \sum_{l=0}^{L-1} \beta_l^* e^{j2\pi ln/N} \right]
\]

\[
= \frac{1}{N^2} \sum_{l=0}^{L-1} \sigma_l^2 e^{-j2\pi l(k-n)/N} \tag{3.15}
\]

\[
= \frac{A}{N^2} \sum_{l=0}^{L-1} \left( e^{-T/2\nu - j2\pi(k-n)/N} \right) l
\]

\[
= \frac{A}{N^2} \frac{1 - e^{-LT/2\nu - j2\pi L(k-n)/N}}{1 - e^{-T/2\nu - j2\pi(k-n)/N}}. \tag{3.16}
\]

Since \( H_k \sim \mathcal{CN}(0, \sigma_k^2/N^2) \), the correlation between sub-carrier \( k \) and \( n \) is in general

\[
\rho_{kn} = \frac{\sum_{l=0}^{L-1} \sigma_l^2 e^{-j2\pi l(k-n)/N}}{\sigma_k^2}, \tag{3.17}
\]

and in particular if the channel power delay profile of the taps is exponentially decaying, as in (3.13), the correlation can be written as

\[
\rho_{kn} = \frac{1 - e^{-T/2\nu}}{1 - e^{-LT/2\nu - j2\pi L(k-n)/N}} \frac{1 - e^{-LT/2\nu - j2\pi L(k-n)/N}}{1 - e^{-T/2\nu - j2\pi(k-n)/N}}. \tag{3.18}
\]

As an illustration, the magnitude of the correlation, \( |\rho_{kn}| \) according to (3.18), between sub-carriers as a function of the distance between the sub-carriers, \( k - n \) is shown in Figure 3.7 for three different delay spreads \( \nu \). The parameters are as in Section 3.4, i.e. \( N = 64 \), \( L = 16 \) and \( T = 50 \) ns.

**Probability of a Weak Sub-carrier Within a Cluster**

In the reduced feedback scheme presented in Section 3.3, the users feed back the channel quality of a cluster. The channel quality can be chosen as the channel quality of the worst sub-carrier in the cluster, the average channel quality within the cluster or some other function of the channel quality of the sub-carriers. In all cases, it is beneficial if the channel quality of the sub-carriers within the cluster does not vary too much. To help evaluate the appropriate cluster-size \( R \), the probability, \( P_i \), that the gain of the \( i \)th sub-carrier in a cluster is less than \( \alpha \) times the average gain of the cluster is derived. For convenience, consider the cluster with lowest
3.B. ON SUB-CARRIER VARIATIONS WITHIN A CLUSTER

Figure 3.7: The correlation magnitude $|\rho_{kn}|$ between the sub-carriers $k$ and $n$ as a function of $k - n$ according to (3.18).

index, i.e. $i \in \{1, \ldots, R\}$.

\[
P_i = \Pr \left( |H_i|^2 < \frac{\alpha}{R} \sum_{r=1}^{R} |H_r|^2 \right) \\
= \Pr \left( |H_i|^2 - \frac{\alpha}{R} \sum_{r=1}^{R} |H_r|^2 < 0 \right) \\
= \Pr \left( \begin{bmatrix} h^H (A_i - \frac{\alpha}{R} I_R) h < 0 \end{bmatrix} \right),
\]

where $h = [H_1 \cdots H_R]^T$ and $A_i$ is an all zero $R \times R$ matrix except for the $(i, i)^{th}$ element, which equals one. It is assumed that $h \sim \mathcal{CN}(0, R_h)$. Therefore, $h = R_h^{1/2} u$, where $u \sim \mathcal{CN}(0, I_R)$ and $R_h^{1/2} R_h^{1/2} = R_h$ is Hermitian.

Thus, (3.19) can be rewritten as

\[
\Pr (h^H D h < 0) = \Pr \left( \begin{bmatrix} u^H R_h^{1/2} DR_h^{1/2} F u < 0 \end{bmatrix} \right).
\]
Since \( \mathbf{D} \) is diagonal, \( \mathbf{F} \) is also Hermitian. Therefore, it can be decomposed as \([HJ85]\)

\[
\mathbf{F} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H,
\]

where \( \mathbf{U} \) is unitary and \( \mathbf{\Lambda} \) is a diagonal matrix with the eigenvalues of \( \mathbf{F} \), \( \lambda_1, \ldots, \lambda_R \), as diagonal elements. Since \( \tilde{\mathbf{u}} = [\tilde{u}_1 \cdots \tilde{u}_R]^T = \mathbf{U}^H \mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_R) \)

\[
P_i = \Pr(\tilde{\mathbf{u}}^H \mathbf{\Lambda} \tilde{\mathbf{u}} < 0)
= \Pr \left( \sum_{r=1}^{R} \lambda_r |\tilde{u}_r|^2 < 0 \right)
= \Pr \left( \sum_{r \in \mathcal{L}} \lambda_r |\tilde{u}_r|^2 < 0 \right)
= \Pr \left( \sum_{r \in \mathcal{L}} X_r < 0 \right)
= \Pr (X < 0),
\]

where \( \mathcal{L} \subseteq \{1, \ldots, R\} \) is the set of indices in the sum for which the eigenvalues \( \lambda_r \) are non-zero. The random variables \( X_r = \lambda_r |\tilde{u}_r|^2 \) are exponentially distributed with PDFs \( p_{X_r}(x) \)

\[
p_{X_r}(x) = \frac{1}{\lambda_r} e^{-x/\lambda_r} \Theta(x), \quad r \in \mathcal{L}^+ \\
p_{X_r}(x) = -\frac{1}{\lambda_r} e^{-x/\lambda_r} \Theta(-x), \quad r \in \mathcal{L}^-
\]

where \( \Theta(x) \) is the Heaviside step function, \( \mathcal{L}^+ = \{r : \lambda_r > 0\} \) and \( \mathcal{L}^- = \{r : \lambda_r < 0\} \). The characteristic functions, \( \Phi_{X_r}(f) \), for \( X_r \) are given by

\[
\Phi_{X_r}(f) = \frac{1}{1 + j2\pi f \lambda_r},
\]

for all \( r \in \mathcal{L} \). Hence, the characteristic function and PDF of \( X = \sum_{r \in \mathcal{L}} X_r \) are

\[
\Phi_X(f) = \prod_{r \in \mathcal{L}} \frac{1}{1 + j2\pi f \lambda_r}, \quad p_X(x) = \int_{-\infty}^{\infty} \Phi_X(f) e^{j2\pi fx} df.
\]

Assuming no duplicate singular values, partial fraction expansion gives \([Wum94]\)

\[
\Phi_X(f) = \sum_{r \in \mathcal{L}} \frac{B_r}{1 + j2\pi f \lambda_r}.
\]
Now,

\[ P_i = \Pr (X < 0) \]

\[ = \int_{-\infty}^{0} p_X(x)dx \]

\[ = \int_{-\infty}^{0} \int_{-\infty}^{\infty} \Phi_X(f)e^{j2\pi fx}dfdx \]

\[ = \int_{-\infty}^{0} \int_{-\infty}^{\infty} \sum_{r \in L} B_r e^{j2\pi fx} dfdx \]

\[ = \sum_{r \in L^+} B_r \int_{-\infty}^{0} \frac{1}{\lambda_r} e^{-x/\lambda_r} \Theta(x)dx \]

\[ - \sum_{r \in L^-} B_r \int_{-\infty}^{0} \frac{1}{\lambda_r} e^{-x/\lambda_r} \Theta(-x)dx. \]

The integrals over the positive \( \lambda_r \) are all zero, which finally gives us the desired probability,

\[ P_i = \sum_{r \in L^-} B_r \int_{-\infty}^{0} \frac{1}{\lambda_r} e^{-x/\lambda_r} \Theta(x)dx = \sum_{r \in L^-} B_r. \] (3.20)

Hence, the probability that the \( i^{th} \) sub-carrier in a cluster is below \( \alpha \) times the mean can be found exactly by the procedure above. To find the probability that any sub-carrier channel quality is below the mean is more cumbersome. However, by observing that the sub-carrier on the cluster edge is typically the least correlated with the average, designing the cluster size for the outage probability of the edge sub-carrier seems reasonable. The probability that any sub-carrier is in outage is left for future work. A loose upper bound is given by,

\[ \Pr \left( \bigcup_{i=1}^{R} E_i \right) < \sum_{i=1}^{R} P_i < RP_{\text{max}}, \]

where \( E_i \) is the event that sub-carrier \( i \) is in outage and \( P_{\text{max}} \) is the sub-carrier with highest outage probability, i.e. the edge sub-carrier.

In Figure 3.8, the outage probability for the edge sub-carrier is shown. The sub-carrier correlation in the previous section has been used, with the parameters described in Section 3.4 (64 sub-carriers). The upper two curves represent the probability that the edge sub-carrier quality, \( |H_i|^2 \), is below 50% of the mean, where the two lower represent the probability that it is below 10% of the mean. Two different delay spreads, \( \nu \) are shown. Based on these probabilities, it is possible to trade off

- The feedback rate (through \( R \))
Figure 3.8: The probability that the edge sub-carrier quality is below $\alpha$ times the mean, $P_1$, is shown as a function of the cluster size $R$. In the two upper curves, $\alpha = 0.5$ and for the two lower, $\alpha = 0.1$. The two curves represent two values of the channel delay spread $\nu$.

- The fading margin in the adaptive modulation
- The packet outage probability (through the sub-carrier outage probability).

Comment on Feed Forward Overhead

As discussed in the previous sections, the overhead of measurement information from the users back to the base-station can be significant. However, the feedforward overhead can also be significant in an adaptive resource allocation system, especially if the reallocation rate is high. The scheduler has to inform the users about the scheduling decisions. If there are $K$ users in the cell, competing for the $Q$ clusters, the feedforward overhead is $Q \log_2(K)$ bits per scheduling block. For each cluster, the base-station can start the block by sending an identification number of the user ($\log_2(K)$ bits) that was allocated to the cluster. Unfortunately, the base-station has to use the lowest modulation order to send the user identification, in order to guarantee that all users competing for the particular cluster can receive the user identification correctly. The modulation order that will be used in the cluster during the rest of the scheduling block does not have to be sent, since the selected user knows which supportable rate it fed back. If the resource allocation algorithm
is allowed to change the power allocation and adaptive modulation level, also the selected modulation level has to be fed forward.
Chapter 4

Using Unclaimed Sub-carriers

4.1 Overview

This chapter contains:

- Three approaches to use the unclaimed sub-carriers, which occur in OFDMA systems, where the users claim sub-carriers by selective feedback.

- Single-cell numerical evaluations of the proposed methods.

4.2 Introduction

In Chapter 3, an adaptive feedback scheme for the FDD OFDMA downlink was proposed. The total feedback load was reduced by letting each user feed back information only about the sub-carriers with relatively good channel quality. There is however a risk that all users experience good channels on the same sub-carriers. The consequence could be that the scheduler, for many sub-carriers, receives no channel quality information. What to do with these unclaimed sub-carriers that no user claims is the topic of this chapter. One option is to move transmit power from unclaimed sub-carriers to sub-carriers where users can be scheduled. In this thesis, however, we assume a power constraint per sub-carrier (c.f. Section 2.2). Several approaches are considered. The first approach uses the unclaimed sub-carriers for pilot symbols in addition to the grid of fixed pilot sub-carriers. Another alternative is to assign each unclaimed sub-carrier to a user that fed back sub-carrier channel quality information close to the unclaimed sub-carrier. In this chapter, clustered OFDMA is not considered, but the problem with unclaimed clusters is equivalent.

In order to compensate for the effect of the channel, each user has to estimate the channel frequency response for each sub-carrier that they will use for data reception. The channel estimation is based on a-priori known pilot symbols that are transmitted on some sub-carriers [TSD04]. We assume that the users estimate
the channel by Linear MMSE channel estimation, which is described in more detail in Appendix 4.A.

On each sub-carrier, a different user can be scheduled. The scheduling is updated regularly with an interval of several OFDM symbols, here called a block. As in Chapter 3, the scheduling is based on limited feedback from each active user in the form of the supportable number of bits per symbol for a number of the strongest sub-carriers. The sub-carriers that no user fed back are here called unclaimed sub-carriers.

For each block, the set of sub-carriers can be divided into three disjoint sets, \( N_p \cup N_c \cup N_u = N = \{1, \ldots, N\} \), where \( N_p \) is the predefined set of fixed pilot sub-carriers, \( N_c = \bigcup_{k=1}^{K} S_k \) is the set of claimed sub-carriers (where \( S_k \) are the sub-carriers the \( k^{th} \) user fed back), and \( N_u \) is the set of unclaimed sub-carriers. For the claimed sub-carriers, at least one user fed back the instantaneous channel quality, and for the unclaimed sub-carriers, the scheduler has received no instantaneous channel information. Note that the users do not consider the fixed pilot sub-carriers in \( N_p \) for feedback. This pilot placement is not optimal [ATV02], but it is possible to apply the ideas presented here to any a priori pilot placement. Also note that the sets \( N_c \) and \( N_u \) are random, since they depend on the instantaneous channels of the users. The spectral usage, as in Definition 3.2, is \( |N_c|/|N \setminus N_p| \). It is assumed that transmit power cannot be transferred from \( N_u \) to \( N_c \). How to use the bandwidth in \( N_u \) is the topic of this chapter.

### 4.3 System Model

In this chapter, the SISO OFDMA downlink of an FDD cellular wireless communication system is considered, as in Section 2.2. However, inter-cell interference is not considered here. Disregarding the base-station index and inter-cell interference term in (2.1), the received signal of a user on sub-carrier \( n \in N = \{1, \ldots, N\} \) can be written as

\[
y_n = \sqrt{P} H_n c_n + z_n,
\]

where \( \sqrt{P} \) is transmit power per sub-carrier, \( c_n \) is the transmitted symbol on sub-carrier \( n \), \( H_n \) is the sub-carrier frequency response of the user and \( z_n \) is AWGN with variance \( \sigma_z^2 \). We assume that \( P = 1 \) and \( E[|c_n|^2] = 1 \). The sub-carrier SNR, \( \gamma_n \), is then

\[
\gamma_n = \frac{|H_n|^2}{\sigma_z^2}.
\]

### 4.4 Three Ways to Use the Unclaimed Sub-carriers

In this section, we propose three alternatives which are evaluated in Section 4.5. In the **Pilot** method, pilots are transmitted on the unclaimed sub-carriers, in addition to the fixed sub-carriers. In the **Random** method, no fixed pilots are transmitted.
4.4. THREE WAYS TO USE THE UNCLAIMED SUB-CARRIERS

Instead, the users need to rely only on the pilots on the unclaimed sub-carriers. In the Data method, users are scheduled on the unclaimed sub-carriers.

**Pilot**

In the Pilot method, the fixed pilots in $\mathcal{N}_p$ are accompanied by pilot symbols in the set of unclaimed sub-carriers, $\mathcal{N}_u$. The users can detect which sub-carriers are in the set $\mathcal{N}_u$ if the base-station transmits a user index (e.g. 0) that is reserved for the dynamically assigned pilot sub-carriers. For the sub-carriers in $\mathcal{N}_c$, the base-station transmits the user index of the scheduled user, either on a control channel or directly on the sub-carriers. The Pilot approach could increase the accuracy of the channel estimation for the sub-carriers in $\mathcal{N}_c$, leading to lower error rates. However, the adaptive channel estimation would lead to higher receiver complexity.

**Random**

It is possible that the pilot sub-carriers in $\mathcal{N}_p$ would be suitable for high-order modulation for some users. If the opportunistic feedback and scheduling principle is extended to all sub-carriers, the set of fixed pilot sub-carriers, $\mathcal{N}_p$, is made empty. Instead, pilots are only transmitted in the random set $\mathcal{N}_u$. The disadvantage of this method is that the channel estimation quality is more random. The advantage is that the users can choose sub-carriers to feed back from $\mathcal{N}_u$ instead of only $\mathcal{N} \setminus \mathcal{N}_p$, thereby increasing the effect of the multiuser diversity.

**Data**

It is difficult for the scheduler to assign a user and modulation level to an unclaimed sub-carrier without jeopardizing the transmitted packet. No specific CQI is available to the scheduler, other than that the unclaimed sub-carrier is weaker than all the claimed sub-carriers. However, if the unclaimed sub-carrier is almost adjacent to a claimed sub-carrier, it is not likely that the quality of the unclaimed sub-carrier deviates much from the quality of the claimed sub-carrier (see Appendix 3.B for a derivation of this probability, which is a function of the channel delay-spread and the distance to the adjacent sub-carrier). Then, it seems to be a good choice to schedule the user from the adjacent claimed sub-carrier, but with a one-step lower modulation level. We call this the Data method. It requires no extra forward signaling of modulation levels for the sub-carriers in $\mathcal{N}_u$, since the user knows that it did not feed back that sub-carrier, and will therefore try to detect symbols from a constellation-size one level below the constellation-size of the adjacent claimed sub-carrier. Only the scheduling decision has to be signaled.
4.5 Simulation Results

The methods described above, as well as two reference methods, have been evaluated. They are shortly summarized here.

- **Fixed**: Only $N_p$ is used for pilots. $N_u$ is unused.
- **Perfect**: Similar to the **Fixed** method, but the channel is perfectly estimated at the receivers.
- **Pilot**: Both $N_p$ and $N_u$ are used for pilots.
- **Random**: Only $N_u$ is used for pilots, whereas $N_p$ is empty.
- **Data**: Only $N_p$ is used for pilots and $N_u$ is used for data symbols, assigned to the scheduled user closest in frequency, but at a lower rate.

A cell with $K$ users has been simulated, using the channel model with exponentially decaying power delay profile, as in Section 3.4. The time-evolution of the channel taps is generated through Jake’s model [DBC93]. The number of taps $L = 60$ and the sampling time is $T = 260$ ns. The number of sub-carriers is 256, the OFDM symbol time is 82.16 µs, including the 15.6 µs cyclic prefix. The carrier-frequency is 2 GHz. We assume that the receivers use a separate power prediction algorithm for the sub-carrier SNRs. Hence, in this numerical evaluation, the fed back supportable rates available at the scheduler are assumed to be without errors or delay.

On each sub-carrier, the user with the highest reported supportable rate is scheduled and assigned a modulation level from BPSK, QPSK, 16-QAM, 64-QAM or 256-QAM. The adaptive modulation thresholds are based on a target bit-error rate of $10^{-4}$. The throughput is computed as the number of bits in the successfully received packets, which are 128 bits long. A packet is considered erroneous if at least one bit is erroneous. Error correcting codes are not considered here.

The users estimate the channel on each sub-carrier using the LMMSE estimator described in Appendix 4.A using $N_{\text{pilot}} = 6$ pilot symbols. The pilots are taken from the same OFDM symbol as the sub-carrier to be estimated and chosen so that the sub-carrier to be estimated is located in the middle of the 6 pilots. The fixed pilots in $N_p$ are spaced 8 sub-carriers apart in each OFDM symbol. The average SNR for all users is set to 10 dB.

Each of the $K$ users feeds back the supportable number of bits per symbol for their $S < N$ strongest sub-carriers in each block. The allowed number of bits per symbol is 1, 2, 4, 6 and 8. For each sub-carrier, the base-station schedules the user with the highest supportable number of bits per symbol. The scenarios $K = 10$ and $K = 50$ users in the cell have been studied. For 10 users, each user feeds back $S = 28$ sub-carriers, which gives $E[|N_u|] \approx 59$ sub-carriers (c.f. Chapter 3). For 50 users, $S = 6$, which gives $E[|N_u|] \approx 58$. An important difference in the structure of $N_u$ between few and many users is that for few users, the unclaimed sub-carriers tend to be adjacent. The reason is that the fed back sub-carriers of a user often
4.5. SIMULATION RESULTS

Figure 4.1: The throughput for the proposed methods to use the unclaimed sub-carriers. The frequency selectivity increases along the x-axis. The number of users in the cell is 10.

are adjacent, especially for lower delay-spreads. For many users, the sub-carriers in $\mathcal{N}_u$ tend to be more uniformly spread in the OFDM symbol, since more users feedback fewer sub-carriers. The pilot-spacing of 8 gives $|\mathcal{N}_p| = 32$ sub-carriers.

Figures 4.1-4.6 show the system throughput, bit-error rate (BER) and packet-error rate (PER) for the different methods as a function of the channel RMS delay spread $\nu$. The performance loss due to channel estimation errors can be seen as the difference between the Perfect and Fixed methods, both of which do not utilize the unclaimed sub-carriers.

In Figure 4.1, where there are 10 users in the cell, the throughput of Perfect increases with the increased channel frequency variability due to the opportunistic scheduling. For Fixed, the throughput initially increases, but as the delay spread reaches a certain point, the effect of the reduced channel estimation quality dominates. By using the unclaimed sub-carriers for additional pilots, Pilot manages to reduce some of the loss due to channel estimation errors at high delay spreads. For the Data method, which schedules users on unclaimed sub-carriers, the throughput is even higher than for Perfect for low delay-spreads. This is due to the fact that Perfect does not use the unclaimed sub-carriers. However, as the frequency variability increases with the delay spread, the reduced channel estimation quality as well as the loss in adaptive modulation accuracy on the unclaimed sub-carriers makes the throughput of Data reach that of Fixed. The overall performance for
the Random scheme is poor, except when the channels are almost flat, i.e. for low delay spread, where the penalty of the random pilot placement is very small.

The underlying bit-error rates for the 10-user scenario are shown in Figure 4.2. The BER increases with increased channel variability, except for Perfect, which decreases slightly. This is due to the slightly higher channel gains for the scheduled users, which for this particular average SNR (10 dB) and 10 users implies more margin in the adaptive modulation on average. The higher modulation level thresholds are separated further apart (in dB) than the lower. Of the methods that do not enjoy perfect channel estimation, Pilot has the lowest error rates, due to the additional transmitted pilots on the sub-carriers that no user fed back. Even though the BER of Data is higher than for Fixed, the throughput is higher, since more bits are transmitted.

The packet-error rates are displayed in Figure 4.3. The packet-error rates connect the BER with the throughput. Due to the higher BER, the scheme with random pilot placement performs catastrophically for higher delay-spread. However, the additional pilot symbols on the unclaimed sub-carriers for the Pilot scheme helps to reduce the packet errors, compared to Fixed.

In Figures 4.4 and 4.6, the similar results are given, but for 50 users. The throughput for Perfect is higher than for 10 users, due to the increased multiuser diversity. Note that the BER for Perfect is slightly higher than for 10 users. The
4.6. SUMMARY

In the simulations, it could be observed that the approach to transmit low-rate data on the sub-carriers that no user claimed in the feedback showed superior performance for low to moderate channel RMS delay spread. For the higher delay spreads, additional pilots on the unclaimed sub-carriers gave the highest system throughput due to the improved channel estimation. The scheme with completely random pilot placement proved to be the worst in terms of throughput, especially for channels with high frequency variability.

In the simulations, the fixed pilot spacing has not been varied. The effect of varying the fixed pilot spacing is equivalent to varying the channel RMS delay spread, which has been done here. Furthermore, the selected scheduling method

Figure 4.3: The packet-error rate for the proposed methods to use the unclaimed sub-carriers. The frequency selectivity increases along the x-axis. The number of users in the cell is 10.

reason is that, for the particular average SNR is these simulations and 50 users, the highest modulation level is selected more often, but with a smaller margin. An important difference to the 10-user case is that Random performs better. This is due to the fact that for 50 users, the sub-carriers in $|\mathcal{N}_u|$ are typically more uniformly spaced in the OFDM symbol. For 10 users, the unclaimed sub-carriers are often located on adjacent sub-carriers, which leads to worse channel estimation on the other sub-carriers for Random.
Figure 4.4: The throughput for the proposed methods to use the unclaimed sub-carriers. The frequency selectivity increases along the x-axis. The number of users in the cell is 50.

has an impact on performance. In the simulations, the max-throughput scheduling has been used. Since the additional pilots in the Pilot scheme are beneficial for all users, it is reasonable to assume that this method would perform as well with other schedulers. The Data scheme should also perform similarly with other scheduling methods. However, the user-scheduling on the unclaimed sub-carriers should not be changed. For the Random scheme, the situation would not change much with another scheduling method, due to the poor channel estimation performance.

In the simulations in this section, inter-cell interference was not considered. The Fixed scheme could be expected to perform relatively slightly better if inter-cell interference would have been included, since not using the unclaimed sub-carriers at all would give a lower average inter-cell interference.
Figure 4.5: The bit-error rate for the proposed methods to use the unclaimed sub-carriers. The frequency selectivity increases along the x-axis. The number of users in the cell is 50.
Figure 4.6: The packet-error rate for the proposed methods to use the unclaimed sub-carriers. The frequency selectivity increases along the x-axis. The number of users in the cell is 50.
Appendix 4.A LMMSE Channel Estimator

Here, we describe the linear minimum mean square error (LMMSE) channel estimator that the receivers use in order to be able to coherently detect the transmitted symbols.

The sub-carrier and OFDM symbol indices are denoted \((n, t)\). For the channel model used in the simulations, the correlation between two sub-carriers can be computed as (c.f. Section 3.B)

\[
r_f(\Delta n) = E[H(n + \Delta n, t)H(n, t)^*] = \frac{1 - e^{-T/2\nu} 1 - e^{-LT/2\nu} - j2\pi L\Delta n / N}{1 - e^{-T/2\nu} 1 - e^{-LT/2\nu} - j2\pi \Delta n / N},
\]

where \(\Delta n\) is the spacing in sub-carriers. The approximation is valid when \(L\) is large. The temporal correlation is given by

\[
r_t(\Delta t) = J_0(2\pi f_d \Delta t),
\]

where \(J_0(x)\) is the zero-order Bessel function of the first kind, \(\Delta t\) is the time-difference and \(f_d\) is the maximum Doppler frequency.

If the channel impulse response taps \(\beta_l(t)\) are zero-mean complex Gaussian distributed and independent of the noise, so will the sub-carrier frequency response \(H(n, t)\). Then, the LMMSE estimate of \(H(n, t)\) is [Kay93]

\[
\hat{H}(n, t) = r_{Hp}(n, t)R_{pp}^{-1}(n, t)p(n, t),
\]

where \(p(n, t) = [p(n_1, t_1) \cdots p(n_P, t_P)]^T \in \mathbb{C}^{N_{\text{pilot}} \times 1}\) is the vector of least-squares estimated frequency responses at the \(N_{\text{pilot}}\) pilot positions \((n_1, t_1) \cdots (n_P, t_P)\). These \(N_{\text{pilot}}\) pilots are used to estimate the channel on sub-carrier \(n\) and OFDM symbol \(t\). The estimate of the channel at pilot sub-carrier \(n_i\) and OFDM symbol \(t_i\) is given by

\[
p(n_i, t_i) = \frac{y(n_i, t_i)}{c(n_i, t_i)} = H(n_i, t_i) + \frac{z(n_i, t_i)}{c(n_i, t_i)},
\]

where \(c(n_i, t_i)\) is a known unit-energy pilot symbol. The cross covariance vector between the non-pilot sub-carrier that is to be estimated and the pilot sub-carriers, \(r_{Hp}(n, t) \in \mathbb{C}^{1 \times N_{\text{pilot}}}\) is

\[
r_{Hp}(n, t) = E[H(n, t)p(n, t)^H] = [r_H(n - n_1, t - t_1) \cdots r_H(n - n_P, t - t_P)],
\]

and the covariance matrix of the pilot sub-carriers, \(R_{pp}(n, t) \in \mathbb{C}^{N_{\text{pilot}} \times N_{\text{pilot}}}\) is
given by

$$
\mathbf{R}_{pp}(n, t) = \mathbb{E}[\mathbf{p}_k(n, t)\mathbf{p}_k(n, t)^H] = \\
\begin{bmatrix}
    r_H(0, 0) & \cdots & r_H(n_1 - n_P, t_1 - t_P) \\
    \vdots & \ddots & \vdots \\
    r_H(n_P - n_1, t_P - t_1) & \cdots & r_H(0, 0)
\end{bmatrix} + \sigma_z^2 \mathbf{I}_{N_{\text{pilot}}}. \quad (4.5)
$$

The frequency response correlation function

$$
\begin{align*}
    r_H(\Delta n, \Delta t) &= \mathbb{E}[H(n + \Delta n, t + \Delta t)H(n, t)^*] \\
    &= \sigma_H^2 r_f(\Delta n)r_t(\Delta t), \quad (4.7)
\end{align*}
$$

where $\sigma_H^2$ is the total average power of the channel impulse response. The frequency correlation depends on the impulse response RMS delay spread and the temporal correlation depends on the maximum Doppler frequency of the user. The $N_{\text{pilot}}$ pilot sub-carriers used for estimation of the channel on $(n, t)$ should be chosen wisely, for instance to maximize the $l_1$ norm of the cross covariance (4.4). The filter order $N_{\text{pilot}}$ is a trade-off between estimation accuracy and complexity.

The complexity of the LMMSE estimator is high, since it requires a matrix inversion of a $N_{\text{pilot}} \times N_{\text{pilot}}$ matrix, but several low-complexity alternatives exist, for instance [ESvdB+98, YCL01]. In this chapter we considered the LMMSE as described above, even though low-complexity approximations could be applied. Here, we assume that the users employ a low-complexity algorithm to predict the channel qualities on all sub-carriers (for the feedback) and the more complex LMMSE algorithm only on the sub-carriers where the users are scheduled (for data detection). It can also be noted that the LMMSE estimator requires the channel correlation and the noise power to be known at the receiver.
Chapter 5

Reduced Feedback by SNR Estimation at the Transmitter

5.1 Overview

This chapter contains:

- A reduced feedback design for OFDMA, which relies on channel quality estimation at the base-station.

- Single-cell numerical evaluations of the proposed methods.

5.2 Introduction

In this chapter, we propose and evaluate a different feedback strategy from what was considered in Chapter 3. In classical pilot-symbol aided OFDM, the transmitter sends known pilots on predefined positions on the time-frequency grid. The receiver can then estimate the frequency response of the channel and receive data coherently. An overview of channel estimation methods for OFDM is given in [CEPB02, KHJ03]. In this chapter, we reverse this process, by letting the users feed back SNRs from a grid of sub-carriers to the base-station, which then estimates the channel quality on the sub-carriers that were not fed back. In fact, since we consider an opportunistic system, the users only need to feed back SNRs for the parts of the spectrum where the channel gain is high. Consequently, the base-station considers the users for scheduling only in those parts. This effectively reduces the feedback load. We propose and evaluate MMSE and LMMSE estimators as well as interpolation.
5.3 Reduced Feedback Scheme

The considered system model is as in Section 4.3. On each sub-carrier, a different user can be scheduled. The scheduling is updated regularly with an interval of several OFDM symbols, here called a block. The set of sub-carriers is denoted $\mathcal{N}$ and the set of sub-carriers eligible for feedback is denoted $\mathcal{N}_{fb} \subset \mathcal{N}$, called feedback sub-carriers. The feedback sub-carriers are typically uniformly spaced in the OFDM symbol. The sub-carrier sets here are assumed to be ordered. Below, we will use the notion of adjacency of two elements in a set, by which is meant that one of the elements is directly followed by the other element in the set. In this chapter, the CQI of user $k$ on sub-carrier $n$ is in the form of the SNR, $\gamma_{n,k}$. The reduced feedback scheme is described in the following bullets.

1. All users estimate or predict the channel quality, $\gamma_{n,k}$ on all sub-carriers.

2. A user can be scheduled on a particular sub-carrier only if it fed back information about the two adjacent sub-carriers in $\mathcal{N}_{fb}$ that are closest to the particular sub-carrier. Each user feeds back the SNRs of the sub-carriers in $\mathcal{N}_{fb}$ that enable the highest possible sum-rate. Note that adjacency in $\mathcal{N}_{fb}$ does not mean adjacency in $\mathcal{N}$. The set of sub-carriers that user $k$ feeds back is denoted $\mathcal{N}_k \subseteq \mathcal{N}_{fb}$ with cardinality $|\mathcal{N}_k| = S$. In other words, user $k$ feeds back the indices in $\mathcal{N}_k$ and $\gamma_{n,k} \ \forall n \in \mathcal{N}_k$.

3. Based on the feedback from the users, the base-station estimates the SNR on the other sub-carriers. In the bullets below, the estimation for the $k^{th}$ user is considered.

   - The sub-carrier SNRs in $\mathcal{N}_k$ do not need to be estimated, since they were already fed back.
   - The sub-carrier SNRs between two adjacent sub-carriers in $\mathcal{N}_{fb}$ are estimated only if the two adjacent sub-carriers are in $\mathcal{N}_k$.
   - All other sub-carrier SNRs are estimated to 0.

4. Based on the estimated channels in the previous step, the resources are allocated among the users. Any scheduler may be used.

The base-station only estimates the SNR of a sub-carrier if it is located in frequency between two sub-carriers that are adjacent in both $\mathcal{N}_k$ and $\mathcal{N}_{fb}$. The reason is that the estimation quality may not be reliably estimated on the other sub-carriers. Hence, all other sub-carrier SNRs are estimated to 0, resulting in that the user is not scheduled on those sub-carriers. This is not a problem if a channel-aware scheduler is used, which anyway does not schedule users on their weakest sub-carriers.

As an example, consider a system with 128 sub-carriers, i.e. $\mathcal{N} = \{0, 1, \ldots, 127\}$, as in Figure 5.1. The sub-carriers eligible for feedback is every eighth, i.e. $\mathcal{N}_{fb} = \{0, 8, \ldots, 122\}$. From $\mathcal{N}_{fb}$, user $k$ selects the $S = 8$ sub-carriers which yield the
5.4 Covariance Information at the Transmitter

The MMSE estimators presented in the next section assume that the transmitter knows the covariance between sub-carrier SNRs. For the channel model in Section 3.4, the covariance between sub-carriers is (see Section 3.B)

$$E[H_{n_1}H^*_{n_2}] = \frac{(1 - e^{-T/2\nu})(1 - e^{-LT/2\nu-j2\pi L(n_1-n_2)/N})}{(1 - e^{-LT/2\nu})(1 - e^{-T/2\nu-j2\pi(n_1-n_2)/N})}. \quad (5.1)$$

For simplicity, we here assume that $E[|H_n|^2] = 1$. The sub-carrier SNR is exponentially distributed with mean

$$\gamma_n = \frac{1}{\sigma_z^2} E[|H_n|^2] = \frac{1}{\sigma_z^2}. \quad (5.2)$$
CHAPTER 5. REDUCED FEEDBACK BY SNR ESTIMATION AT THE
TRANSMITTER

and covariance

\[ E[\gamma_{n_1} \gamma_{n_2}] - \bar{\gamma}_{n_1} \bar{\gamma}_{n_2} = \frac{1}{\sigma_z^4} \left( \frac{1 - e^{-T/\nu} - 2e^{-T/2\nu}}{1 - e^{-LT/\nu} - 2e^{-LT/2\nu}} \right) \times \left( \frac{1 + e^{-LT/\nu} - 2e^{-LT/2\nu} \cos(2\pi L(n_1 - n_2)/N)}{1 + e^{-T/\nu} - 2e^{-T/2\nu} \cos(2\pi (n_1 - n_2)/N)} \right) \]  

(5.3)

\[ = \frac{|E[H_{n_1} H^*_{n_2}]|^2}{\sigma_z^4}. \]  

(5.4)

The expression in (5.3) is derived in Appendix 5.A for the channel model in Section 4.3 with exponentially decaying power delay profile. We also show that (5.4) is valid in general when \( H_{n_1} \) and \( H_{n_2} \) are zero-mean complex Gaussian variables. The notation \( \bar{X} \) here means the expected value of the random variable \( X \). The sub-carrier covariance in (5.1) is needed for the MMSE estimator presented in Section 5.5 and the SNR mean and covariance in (5.2)-(5.3) is needed for the LMMSE estimator in Section 5.5.

The covariance information can be obtained in several ways. Since the channel covariance typically is a slowly changing parameter, compared to the instantaneous SNR, the additional overhead of feeding back the covariance from each user would be rather small. Additionally, the covariance can often be parameterized, for instance by \( L \) and \( T/\tau \) in (5.1), in order to reduce the feedback even more. An alternative, that implies no extra feedback, is that the base-station estimates the sub-carrier covariance from the fed back SNRs of the sub-carriers in \( \mathcal{N}_k \). A third option for obtaining downlink covariance information at the base-station without feedback is to estimate it from the uplink [Zet97].

5.5 SNR Estimators

In this section, three estimators are presented, that estimate the sub-carrier SNRs at the transmitter based on the few fed back SNRs of the sub-carriers in \( \mathcal{N}_k \). Only a sub-set of the sub-carriers are estimated, according to the scheme described in Section 5.3. The other sub-carrier SNR estimates are set to zero. In this section, estimators with two observations (SNRs) are considered. The LMMSE estimator and interpolation are easily extended to more observations. The extension of the MMSE estimator does not seem to be straightforward, however.

**MMSE**

The MMSE estimator of the SNR takes advantage of the underlying PDF of the SNRs. To derive the estimator, it is assumed that the channel of each sub-carrier is a zero-mean complex Gaussian random variable.
5.5. SNR ESTIMATORS

The MMSE estimator of the SNR of subcarrier $n_3$, $\gamma_{n_3}$ from the known SNRs of two other sub-carriers, $\gamma_{n_1}$ and $\gamma_{n_2}$, is given by [Kay93]

$$\hat{\gamma}_{n_3}^{\text{MMSE}} = E[\gamma_{n_3}|\gamma_{n_1}, \gamma_{n_2}].$$

(5.5)

Even though the joint PDF $f(\gamma_{n_1}, \gamma_{n_2}, \gamma_{n_3})$ cannot, in general, be expressed in closed form, the closed form expression for the MMSE estimator is [SHO06]

$$E[\gamma_{n_3}|\gamma_{n_1}, \gamma_{n_2}] = \frac{1 + \alpha + |\beta| \cos(\angle \beta - \angle R_{12})}{R_{33}^{-1}} I_0(2\sqrt{|\gamma_{n_1}\gamma_{n_2}|}) I_1(2\sqrt{|\gamma_{n_1}\gamma_{n_2}|}) \frac{R_{13}^{-1}(R_{23}^{-1})^*}{R_{33}^{-1}},$$

(5.6)

where $\angle \{\cdot\}$ is the argument (phase) operator, $R$ is the positive definite covariance matrix of $h$, $h = \frac{1}{\sigma_z} [H_{n_1} H_{n_2} H_{n_3}]^T$, and the notation $R_{kl}^{-1}$ is to be interpreted as $[R^{-1}]_{kl}$. The functions $I_0(\cdot)$ and $I_1(\cdot)$ denote the modified Bessel functions of the first kind (zeroth and first order, respectively). The parameters $\alpha$, $\beta$ and $\gamma$ are defined as

$$\alpha \triangleq \gamma_{n_1} \left| R_{13}^{-1}\right|^2 + \gamma_{n_2} \left| R_{23}^{-1}\right|^2, \quad \beta \triangleq 2\sqrt{\gamma_{n_1}\gamma_{n_2}} \frac{R_{13}^{-1}(R_{23}^{-1})^*}{R_{33}^{-1}},$$

(5.8)

and

$$\zeta \triangleq \frac{|R_{12}|}{R_{11}R_{22} - |R_{12}|^2} = \frac{\sqrt{\text{Cov}[\gamma_{n_1}, \gamma_{n_2}]} \gamma_{n_1} \gamma_{n_2}}{\gamma_{n_1}\gamma_{n_2} - \text{Cov}[\gamma_{n_1}, \gamma_{n_2}]},$$

(5.9)

where $\text{Cov}[\cdot, \cdot]$ denotes covariance. It should be noted that $\zeta$ is only a function of the mutual statistics of $\gamma_1$ and $\gamma_2$, and independent of their relation to $\gamma_3$.

Note that the computational complexity of evaluating (5.6) is low, since the fraction of the modified Bessel functions can be efficiently computed from a few terms of the series expansions.

LMMSE

The linear minimum mean square error estimator of size 2 for the sub-carrier SNR, $\gamma_{n_3}$, based on the fed back observations $\gamma_{n_1}$ and $\gamma_{n_2}$ is given by [Kay93]

$$\hat{\gamma}_{n_3}^{\text{LMMSE}} = \tau_{n_3} + R_{n_3x} R_{xx}^{-1} (x - \bar{x})$$

(5.10)

where

$$x = [\gamma_{n_1}, \gamma_{n_2}]^T$$

$$R_{n_3x} = E[\gamma_{n_3}x^H] - \tau_{n_3} \bar{x}^H$$

$$R_{xx} = E[xx^H] - \bar{x}\bar{x}^H.$$
The covariances are assumed to be known at the transmitter. For the channel model presented in Section 4.3, the covariances are given by (5.3). The LMMSE estimator with more observations is of the same form as above.

Interpolation

The estimation problem in this chapter can be seen as an interpolation based on the fed back SNRs. Therefore, the statistical estimators that use two observations are compared with a linear interpolator that requires only the fed back SNR values of the sub-carriers in \( N_k \). The interpolation-based estimation extends naturally to more observations, for instance by using piecewise cubic spline interpolation.

Performance Evaluation

The performance of the estimators is evaluated by considering the SNR estimate of sub-carrier \( n \) from the SNRs of the sub-carriers \( n_2 = n - 8 \) and \( n_3 = n + 8 \). In total, there are \( N = 128 \) sub-carriers, and \( T/\nu = 2 \). In Figure 5.2 the probability of under- and over-estimating the SNR by \( \alpha \) dB is displayed. As can be seen, the LMMSE estimate tends to under-estimate, whereas the interpolation tends to over-estimate the SNR.
5.6 System Simulation

Simulation Assumptions

In order to evaluate the impact of the estimators in Section 5.5, a multiuser OFDMA single-cell system has been simulated. The cell is populated by $K = 8$ users with i.i.d. channels according to Section 4.3 and temporal block-fading according to Jake’s model [DBC93], with a carrier frequency assumption of 2 GHz and user speeds of 20 m/s. The channel RMS delay spread is 200 ns. The number of sub-carriers $N = 128$, the OFDM symbol duration is 38.33 $\mu$s and the block length equals 16 OFDM symbols. The average SNR is 10 dB for all sub-carriers.

In the simulations, the users are assumed to estimate the downlink channel perfectly. Furthermore, the users are assumed to predict the sub-carrier SNRs perfectly (e.g. by [SA03]), so that the fed back information is not outdated. The feedback and SNR estimation scheme is according to Section 5.3, with $S = 4$ sub-carrier SNRs per user being fed back each block. In the figures, the spacing between the sub-carriers in $N_{fb}$ varies between 8 and 32. On each sub-carrier, the user with highest estimated SNR is scheduled and assigned a modulation order from BPSK, QPSK, 16-QAM, 64-QAM or 256-QAM. The adaptive modulation thresholds are based on a target bit error rate of $10^{-3}$ and the transmit power is allocated equally over the sub-carriers. The throughput is computed as the number of bits in the successfully received packets, which are 128 bits long. A packet is considered erroneous if at least one bit is erroneous. The users estimate the channel perfectly. The bit and packet errors here are due to the AWGN.

The three estimators in Section 5.5 are compared to two alternatives, which both estimate the sub-carrier SNRs perfectly at the transmitter. The scheme denoted Perfect in the figures works just as the other estimators, in that it only computes estimates on some sub-carriers. On the sub-carriers that are not between two adjacent sub-carriers in both $N_k$ and $N_{fb}$, it estimates zero SNR, as in Figure 5.1. Hence, the difference between the performance of Perfect and the other estimators is only due to SNR estimation inaccuracy. For the scheme denoted Full, the base-station has full CQI of all users and sub-carriers. Hence, it always schedules the strongest user on all sub-carriers. The MMSE, LMMSE and interpolation-based estimators use two observations to estimate one SNR.

Simulation Results

In this section, the performance as a function of the feedback spacing is studied. The feedback spacing is the distance (in sub-carriers) between adjacent sub-carriers in $N_{fb}$. For a low feedback spacing, the estimation quality of the estimated sub-carriers is relatively good, resulting in few erroneous decisions. A large spacing gives poorer SNR estimates at the base-station since the input observations to the estimators are less correlated with estimated parameter. On the other hand, a large feedback spacing enables more sub-carriers per user to be estimated, assuming a
CHAPTER 5. REDUCED FEEDBACK BY SNR ESTIMATION AT THE TRANSMITTER

Throughput [Mbps] vs. Feedback spacing [sub-carriers]

Figure 5.3: The system throughput for different estimators as a function of the spacing in the feedback grid.

fixed $S$. This can give a higher system throughput since an overall strong user can be scheduled on more of its sub-carriers.

Figures 5.3-5.5 show the sum throughput, BER and PER, respectively, as a function of the feedback spacing. $Full$ has the overall highest throughput, since it can always accurately schedule the strongest user on each sub-carrier. Also $Perfect$ estimates the sub-carrier SNRs perfectly. However, for low feedback spacings, the sum throughput of $Perfect$ is much lower than for $Full$. This is due to relatively low number of sub-carrier SNRs per user $Perfect$ estimates. If one user has a stronger channel than the other users on most sub-carriers, $Perfect$ schedules it only on its strongest sub-carriers, resulting in lower sum throughput. For the feedback spacing 32, almost all sub-carriers of each user are estimated at the transmitter, resulting in a similar throughput of $Perfect$ and $Full$.

For the non-ideal SNR estimators using MMSE, LMMSE and interpolation-based estimation, a similar effect as for $Perfect$ occurs at low feedback spacings. Only a relatively small number of sub-carrier SNRs per user are estimated. However, the resulting BER and PER for low feedback spacings is low, similar to $Perfect$ and $Full$, which estimate the SNRs perfectly at the transmitter. Note that for all schemes, the users estimate the channel perfectly. The reason the BER is lower than the target BER is that a finite modulation set has been used, resulting in a significant SNR margin on many sub-carriers. The throughput for the non-ideal es-
Figure 5.4: The BER for different estimators as a function of the spacing in the feedback grid.

Estimators increases with the feedback spacing to a certain point and then decreases. The decrease is due to the inaccurate estimation at the transmitter for high feedback spacings. Both the BER and the PER increases significantly. This indicates that there is an optimal feedback spacing, given the channel delay spread, the feedback rate $S$ and the number of users $K$. Of the three non-ideal estimation schemes, MMSE performs slightly better than the other two.

### 5.7 Summary

A reduced feedback scheme for OFDMA was proposed and evaluated. It is based on the feedback of the channel quality of a few sub-carriers per user. Based on the feedback sub-carriers, the base-station estimates the channel quality of the other sub-carriers, if it can be done reliably. Otherwise, the sub-carrier channel quality is estimated to zero. MMSE, LMMSE and interpolation-based channel-quality estimation was proposed.

In the system simulation, the MMSE estimator performed slightly better than the other estimators. We concluded that the feedback spacing trades off estimation accuracy with the number of sub-carriers that are estimated at the transmitter. In the simulations, the maximum throughput scheduler was used. For a scheduler with fairness, we expect the performance gap between full and the incomplete CQI
to be smaller. Since a fair scheduler typically only schedules a user on its best sub-carriers, the loss from not estimating the worst sub-carriers of a strong user becomes negligible. The estimators in the simulation study used two input observations. The LMMSE estimator and interpolator can easily be extended to more observations. The numerical study in [SHO06] indicates that no significant performance gain can be expected, since the two observations used here are the most correlated with the parameter. The additional gain from using an MMSE estimator compared to a linear interpolator can be motivated if the covariance information is already available at the transmitter and if the additional small complexity is not an issue at the base-station.

This chapter only considered instantaneous channel quality estimation. For users with significant temporal channel correlation between blocks, the channel quality could be estimated based on instantaneously fed back and previously fed back information. This is left for future research.
Appendix 5.A  Sub-carrier SNR Covariance

In this appendix, the covariance between sub-carrier SNRs is derived for the channel model used in the previous chapters. The time-dispersive sample-spaced channel impulse response is given by

$$h(m) = \sum_{l=0}^{L-1} \beta_l(t) \delta(m-l) \quad (5.11)$$

where $\beta_l \sim \mathcal{CN}(0, \sigma_l^2)$ are independent taps with sum power $\sum_{l=0}^{L-1} \sigma_l^2 = \sigma_h^2$ and $L < N$ is the number of taps. Note that the power delay profile of $h(m)$ is not specified yet.

To find the SNR covariance between sub-carrier $k$ and $n$, we compute $E[|H_k|^2|H_n|^2]$,

$$E[|H_k|^2|H_n|^2] = \frac{1}{N^4} \left( E\left[ \sum_{l=0}^{L-1} \beta_l e^{-j2\pi lk/N} \sum_{m=0}^{L-1} \beta_m^* e^{j2\pi mk/N} \right] - \sum_{p=0}^{L-1} \beta_p e^{-j2\pi pn/N} \sum_{q=0}^{L-1} \beta_q^* e^{j2\pi qn/N} \right)$$

$$= \frac{1}{N^4} \sum_{l,m,p,q} \left( E[\beta_l \beta_m^*] E[\beta_p \beta_q^*] + E[\beta_l^* \beta_q] E[\beta_m \beta_p^*] \right)$$

$$\times e^{-j2\pi(lk-mk+pn-qn)/N}. \quad (5.12)$$

To arrive at (5.13), we used that the expectation of the product of four scalar complex zero-mean jointly Gaussian variables can be written as [JS88]

$$E[\beta_l \beta_m^* \beta_p \beta_q^*] = E[\beta_l \beta_m^*] E[\beta_p \beta_q^*] + E[\beta_l^* \beta_q] E[\beta_m \beta_p^*],$$

and $E[\beta_l \beta_p] = 0$ for all $l, p$. Furthermore, since $E[\beta_l \beta_m^*] = \sigma_l^2 \delta(l-p)$, (5.13) can be simplified to

$$E[|H_k|^2|H_n|^2] = \frac{1}{N^4} \sum_{l=0}^{L-1} \sigma_l^2 \sum_{p=0}^{L-1} \sigma_p^2$$

$$+ \frac{1}{N^4} \sum_{l=0}^{L-1} \sigma_l^2 e^{-j2\pi l(k-n)/N} \sum_{p=0}^{L-1} \sigma_p^2 e^{-j2\pi p(n-k)/N} \quad (5.14)$$

$$= \frac{\sigma_h^4}{N^4} + \frac{1}{N^4} \left( \sum_{l=0}^{L-1} \sigma_l^2 e^{-j2\pi l(k-n)/N} \right)^2 \quad (5.15)$$

$$= \frac{1}{N^4} \left( \sigma_h^2 + |E[H_k H_n^*]|^2 \right) \quad (5.16)$$
From (3.15), we know that
\[ E[|H_n|^2] = \frac{\sigma_h^2}{N^2}. \]

Thus, the SNR covariance is obtained as
\[ E[\gamma_k \gamma_n] - \overline{\gamma_k} \overline{\gamma_k} = E\left[ \frac{|H_k|^2 |H_n|^2}{\sigma_z^4} \right] - E\left[ \frac{|H_k|^2}{\sigma_z^2} \right] E\left[ \frac{|H_n|^2}{\sigma_z^2} \right] = \frac{1}{N^4 \sigma_z^4} \left( \sigma_h^2 + |E[H_k H_n^*]|^2 \right) - \frac{\sigma_h^4}{N^4 \sigma_z^4} = \frac{|E[H_k H_n^*]|^2}{N^4 \sigma_z^4}, \quad (5.17) \]

where \( \sigma_z^2 \) is the AWGN variance on each sub-carrier. Note that (5.17) holds for any power delay profile of the channel impulse response \( h(m) \).

Now, we apply the previously used exponentially decaying power delay profile
\[ E[|\beta_l|^2] = \sigma_l^2 = Ae^{-LT/2\nu}, \quad (5.18) \]
where \( \nu \) is the RMS delay spread and \( T \) is the tap spacing, which is equal to the sampling time. The constant \( A \) is used to normalize the average channel power to \( \sigma_h^2 \), i.e. (see (3.14))
\[ A = \sigma_h^2 \frac{1 - e^{-T/2\nu}}{1 - e^{-LT/2\nu}}. \]

Plugging (5.18) into the second term of (5.14) (the first term vanishes, as in (5.17)) gives
\[ E[\gamma_k \gamma_n] - \overline{\gamma_k} \overline{\gamma_k} = \frac{A^2}{N^4 \sigma_z^4} \left( \sum_{l=0}^{L-1} \left( e^{-lT/2\nu-j2\pi(k-n)/N} \right)^i \right)^2 = \frac{A^2}{N^4 \sigma_z^4} \left( 1 - e^{-LT/2\nu-j2\pi L(k-n)/N}/1 - e^{-T/2\nu-j2\pi L(k-n)/N} \right)^2 = \frac{A^2}{N^4 \sigma_z^4} \left( \frac{1 + e^{-LT/\nu} - 2e^{-LT/2\nu} \cos(2\pi L(k-n)/N)}{1 + e^{-T/\nu} - 2e^{-T/2\nu} \cos(2\pi(k-n)/N)} \right) = \frac{\sigma_h^2}{N^4 \sigma_z^4} \left( \frac{1 + e^{-T/\nu} - 2e^{-T/2\nu} \cos(2\pi(k-n)/N)}{1 + e^{-LT/\nu} - 2e^{-LT/2\nu} \cos(2\pi L(k-n)/N)} \right) \times \frac{(1 + e^{-LT/\nu} - 2e^{-LT/2\nu} \cos(2\pi L(k-n)/N))}{(1 + e^{-T/\nu} - 2e^{-T/2\nu} \cos(2\pi(k-n)/N))}. \quad (5.19) \]

Normalizing so that \( E[|H_n|^2] = 1 \) gives (5.3) and (5.4).
Chapter 6

Modified Proportional Fair Scheduling for OFDMA

6.1 Overview

This chapter contains:

- A short overview of multiuser diversity scheduling, especially for OFDMA.
- A modified Proportional Fair scheduler (M-PF). The scheduler can handle individual bit-rate and delay requirements and incorporates a variable fairness level.

6.2 Introduction

To exploit multiuser diversity, a proper scheduler has to be used. To maximize the cell throughput, the user with the highest supportable rate should always be selected [TV05, RPNM05]. If the user channels are not equally distributed, this scheduler may give unfairly distributed rates. The proportional fair (PF) scheduler attempts to combine fairness and the exploitation of multiuser diversity [Kel97, Qua01, VTL02]. The PF scheduler is briefly presented in Section 6.3. For channel-aware schedulers that try to schedule users on their fading peaks, there is a connection between sum throughput and individual packet delays, especially if the channels fade slowly. A stationary user may experience a deep fade for a long time, resulting in a long idle period and a growing packet queue [And04]. There is also a tradeoff between sum throughput and fairness, in the sense that a system that achieves little variation between the individual user rates has lower average sum throughput than a system that allows radically different user rates [BPN06].

Recently, numerous variations of PF have been proposed, of which only a few are mentioned here. In [BH02], delay sensitive users are given priority in the single-carrier PF algorithm when their delays reach a certain threshold. A similar ap-
CHAPTER 6. MODIFIED PROPORTIONAL FAIR SCHEDULING FOR OFDMA

proach was considered in [KKH02] where unfairly treated users are successively given higher priority. In [ALS+03], the single-carrier PF scheduler is extended to OFDM. We will use the recommended scheme in [ALS+03] as a comparison, calling it PF.

Many results on sub-carrier, bit and power allocation schemes for multiuser OFDMA are available, both for the single- and multi-antenna scenarios, see for instance [WCLM99, RC00, ZL04, ZL03, ECV03, BPNI04]. Optimization problems of different forms are solved to satisfy the minimum rate requirements of the different users. However, these schemes only consider resource allocation for one transmission block. Hence, resources may be wasted on a demanding user that is in a deep fade on most sub-carriers. The philosophy of the PF scheduler is different in that it can let a user wait several transmission blocks if it is fading unfavorably. Hence, a particular QoS can not be guaranteed, but rather a best effort QoS. The QoS parameters are for instance user throughput and delay.

Recently, resource allocation schemes that directly aim at user satisfaction instead of QoS have emerged, e.g. [SL03]. Also, the business model of the operator influences the desirable behavior of the scheduler [BLZZ03]. A good scheduler in a system with few high-paying users is not necessarily good in a system with many moderately-paying users. Recent studies on user satisfaction indicate that the satisfaction of an already "well-served" user increases only marginally by increasing the service level even further [JMK+01, EL03]. However, if the service level is decreased below some level, the satisfaction level drops significantly. This is well captured by the logarithmic utility function which connects user satisfaction with the QoS. In fact, PF maximizes the sum utility when the utility function of each user is the logarithm of the average rate [Kel97]. Proportional fairness for multi-carrier systems was considered in [KH05]. A theoretical framework for utility-maximization in OFDMA systems with infinitesimal sub-carrier bandwidth is given in [SL05a] and optimal sub-carrier and power allocation was characterized. In [SL05b], practical sub-carrier and power allocation algorithms were presented.

In this chapter, we propose a modified PF (M-PF) scheduler for OFDM. It differs from previous PF schedulers in that it can accommodate different quality-of-service (QoS) classes and that it has a tunable fairness level. Utility functions with a system-wide parameter controlling fairness were also mentioned in [Kel97, SL05a]. A generalized utility function, similar to the one in this chapter, was proposed in [MW00]. A fairness parameter was applied to the PF scheduling metric in [YSSL06]. We show that the M-PF scheduler corresponds to a log-like utility function, to which the PF utility function is a special case. The M-PF scheduler allows users to have different utility functions, defined through their target bit-rates and maximum scheduling delays. Simulation results in Chapter 8 show that most users within a QoS class are equally distributed in terms of bit-rates and delay. If the QoS class is well defined for a particular application and most users are near their target QoS, the average user's satisfaction level will be high.
6.3 Proportional Fair

The proportional fair (PF) scheduling algorithm was designed for the downlink of the single carrier CDMA/TDMA IS-856 system, also known as 1xEV or HDR [Qua01, BGP00]. In this system, voice and data traffic is employed on different carriers. The data traffic part of the system exploits multiuser diversity by letting all active users feed back their estimated instantaneous supportable rates and giving channel access according to the PF algorithm. The system has 12 different adaptive coding and modulation levels and supports 60 active users per sector. Multiuser diversity and the proportional fair algorithm are also being considered for the high-speed downlink packet access of UMTS [PELT06].

For each transmission block $m$, the scheduler receives the instantaneous supportable bit-rate from all active users, $C_k(m)$. Furthermore, the scheduler keeps track of the average bit-rate of each user, $\bar{R}_k(m)$, in a historical time-window. For each transmission block $m$, the user $k^*(m)$ is scheduled according to

$$k^*(m) = \arg \max_k \frac{C_k(m)}{\bar{R}_k(m)}.$$  \hspace{1cm} (6.1)

This is basically the maximum throughput scheduling, but weighted by the inverse of the historical bit-rate. If a user has not been scheduled for a long time, it will be favored. If a user has been given channel access recently and therefore has a high historical throughput, it will need a very high channel quality, $C_k(m)$, to be scheduled. This gives a compromise between multiuser diversity and fairness. The time-update of the historical bit-rate, $\bar{R}_k(m)$, can be done in several different ways. A low complexity update equation that also has low memory requirements is

$$\bar{R}_k(m+1) = \begin{cases} (1 - \frac{1}{t_c})\bar{R}_k(m) + \frac{1}{t_c}C_k(m) & \text{if } k^*(m) = k \\ (1 - \frac{1}{t_c})\bar{R}_k(m) & \text{if } k^*(m) \neq k \end{cases}$$  \hspace{1cm} (6.2)

where $t_c$ is the parameter that determines the historical time-window [VTL02]. The update equation (6.2) is an exponentially weighted filter that includes all historical rates in the average rate.

6.4 Modified Proportional Fair

In future communication systems, users will likely have different requirements, for instance real-time data, voice or background downloading of data. This is not taken into account by the PF algorithm which a priori considers all users equal. In this section, we modify the PF algorithm to take user requirements in terms of bit-rates and scheduling delays into account. Note that queue-lengths and queuing delay are not considered here. The presentation of the modified proportional fair (M-PF) scheduling algorithm in this section assumes the clustered OFDMA downlink with selective feedback, as proposed in Section 3. Note that the M-PF scheduler could be applied in a single-carrier system as well with small modifications.
CHAPTER 6. MODIFIED PROPORTIONAL FAIR SCHEDULING FOR OFDMA

Preliminaries

Assume that each user, $k$, belongs to a QoS class with parameters $R_k$ and $T_k$, the target rate in bits/s and averaging window time in seconds, respectively. The averaging window time $T_k$ reflects the maximum scheduling delay of the QoS-class. Long delay windows allow longer periods of not being scheduled, whereas a short delay window will help schedule the user more often. The delay parameter $T_k$ can equivalently be expressed in blocks, $N_k = T_k / N_{tb} T_o$, where $N_k$ is the averaging window time in blocks, $N_{tb}$ is number of OFDM symbols per block and $T_o$ is the OFDM symbol time.

The CQI at the transmitter at block $m$ is composed of $S$ instantaneous rate values for each of the $K$ users corresponding to the minimum supportable rate in a cluster. This information is gathered in the set of supportable rates $\{C_{q,k}(m)\}$, where for each $k$ only $S$ rates are non-zero (the ones that were fed back). The variable $S$ denotes the number of fed back clusters and $Q$ is the number of clusters.

In the following,

- $\bar{X}$ denotes the average measure over the historical time window,
- $\tilde{X}$ denotes the measure over only the current block that is to be scheduled, i.e. what the scheduler can influence, and
- $\hat{X}$ denotes the new average measure, including both the historical time window and the current block.

The average rate, $\bar{R}_k(m)$, reflects the average bit-rate of user $k$ during the historical averaging window time $T_k$. The scheduled rate of user $k$, $\tilde{R}_k(m)$ is to be decided by the scheduler,

$$\tilde{R}_k(m) = R \sum_{q=1}^{Q} \tilde{R}_{q,k}(m) = R \sum_{q=1}^{Q} \xi_{q,k}(m) C_{q,k}(m)$$

(6.3)

where $R$ is the number of sub-carriers per cluster, $\tilde{R}_{q,k}(m)$ is the instantaneous scheduled rate of user $k$ on cluster $q$ and $\xi_{q,k}(m)$ is the indicator function with the property that $\xi_{q,k}(m) = 1$ if user $k$ is scheduled on cluster $q$ during block $m$ and otherwise zero. At most one user can be scheduled per cluster, i.e. $\sum_k \xi_{q,k}(m) \in \{0,1\}$.

The average rate after scheduling is a linear combination of the scheduled rate the current block and previous average rate, as

$$\hat{R}_k(m) = \alpha_k \tilde{R}_k(m) + \beta_k \bar{R}_k(m)$$

(6.4)

where $\alpha_k$ and $\beta_k$ are positive weighting factors with the property that $\alpha_k + \beta_k = 1$. Since $\beta_k = 1 - \alpha_k$, there is only one degree of freedom in the choice of the weighting factors. The weighting factors are user-specific since they depend on the length of the averaging time-window, which is user-specific.
6.4. **MODIFIED PROPORTIONAL FAIR**

For the single-carrier proportional fair scheduler in Section 6.3, \(\alpha_k = \frac{1}{t_c}\) and \(\beta_k = (1 - \frac{1}{t_c})\). Since all previous rates are included in the average rate, the averaging window parameter \(t_c\) does not exactly specify the number of blocks that are averaged.

In this work, we compute \(\bar{R}_k(m)\) as the true average rate within the time window \(T_k\) seconds or \(N_k\) blocks,

\[
\bar{R}_k(m) = \frac{1}{N_k} \sum_{p=1}^{N_k} \tilde{R}_k(m - p).
\]  

(6.5)

Hence, the parameters in (6.4) are given by

\[
\alpha_k = \frac{1}{N_k + 1}, \quad \beta_k = \frac{N_k}{N_k + 1}.
\]

Note that the exponentially decaying averaging in (6.2) could also be used here. However, using the true average bit-rate emphasizes scheduling of users that have not been scheduled during the time window.

For each user, the scheduler keeps track of the average relative rate of each user:

\[
\tilde{B}_k(m) = \frac{\bar{R}_k(m)}{R_k}.
\]

\(\tilde{B}_k(m)\) is a measure of how well the user has met its rate requirements in the historical averaging window time, \(T_k\):

- \(\tilde{B}_k(m) > 1\): User \(k\) has exceeded its target rate.
- \(\tilde{B}_k(m) = 1\): User \(k\) has just met its target rate.
- \(\tilde{B}_k(m) < 1\): User \(k\) has undershot its target rate.

As for the average relative rate, the instantaneous and new average relative rates are defined, respectively, as

\[
\tilde{B}_k(m) = \frac{\tilde{R}_k(m)}{R_k} \\
\hat{B}_k(m) = \frac{\hat{R}_k(m)}{R_k} = \alpha_k \tilde{B}_k(m) + \beta_k \bar{B}_k(m).
\]

**Utility Function**

Utility functions are used to describe the connection between user utility, or satisfaction, and QoS-parameters, e.g. average rate. The classical proportional fair scheduler maximizes the sum of the users’ utilities if the utility function is \(U(\tilde{R}_k) = \ln(\tilde{R}_k)\) [Kel97]. The logarithmic utility function indicates that users with
low average rates benefit more in utility from being scheduled than users with high average rates. In this section, we propose a more general utility function, that in addition is user-specific, to capture the heterogeneous user requirements.

Consider the following class of concave and differentiable utility functions,

\[
U_k(\hat{B}_k) = \frac{R_k\beta_k}{\alpha_k} \frac{1}{1-\kappa}(\hat{B}_k^{1-\kappa} - 1),
\]

(6.6)

where \(\kappa \in [0, 1) \cup (1, \infty)\) is a fairness parameter and the time index \(m\) is left out for brevity. In previous work, utility functions are usually functions of the average user bit-rate. Here, the utility function depends on the average relative rate \(\hat{B}_k\), which is natural if the users use different applications with different target bit-rates corresponding to an acceptable user satisfaction level. Furthermore, the user-specific weighting constant in the utility function, \(\frac{R_k\beta_k}{\alpha_k}\), scales the utility function according to the QoS-parameters of the user.

In Appendix 6.A, we show that (6.6) simplifies to several known utility functions when the QoS-parameters \(R_k, \alpha_k\) and \(\beta_k\) are equal for all users. When \(\kappa \to 1\), (6.6) simplifies to the utility function of a proportional fair scheduler. When \(\kappa = 0\), (6.6) simplifies to the utility function of the maximum sum-rate scheduler. When \(\kappa \to \infty\), (6.6) corresponds to the utility function of the max-min scheduler [BG91].

Note that the appropriate utility function of a particular user depends on the type of application used. Also, note that the fairness parameter \(\kappa\) is common for all users and can only be tuned on a system level.

In Figure 6.1, the effect of the fairness parameter on the utility function is visualized. The QoS-parameters are set to \(\frac{R_k}{\alpha_k} = 1\) and \(\beta_k = 0.99\). The utility function for \(\kappa = 0\) is linear, corresponding to the linear increase of utility with the average rate for all rates. Increasing \(\kappa\) gives a steeper function for small average rates. This corresponds to a scheduler that prefers to allocate users with low average rates, since a great increase in utility can be achieved. The choice of \(\kappa\) for a practical system depends on the desired resource allocation behavior, which in turn depends on the type of traffic that is expected. For more elastic traffic [She95], i.e. traffic that tolerates packet delays gracefully, for instance file downloads, the fairness parameter can be set to a low value. For traffic that relies on a steady data flow, a higher \(\kappa\) should be selected. Note that the delay behavior is also controlled by the user-specific QoS-parameters in the utility function.

Modified Proportional Fair Scheduling Metric

Assume that \(\beta_k \hat{R}_k(m) \gg \alpha_k \hat{R}_k(m)\). Then, the scheduler that maximizes the sum-

utility

\[
\sum_{k=1}^{K} U_k(\hat{B}_k(m)),
\]

(6.7)
Figure 6.1: The effect of the fairness parameter $\kappa$ on the utility function.

where $U_k(\hat{B}_k(m))$ is defined in (6.6), is the following; For each cluster, $q$, the scheduler chooses a user according to the criterion

$$k_q^*(m) = \arg \max_k \left\{ \frac{C_{q,k}(m)}{\hat{B}_k(m)^\kappa} \right\}.$$  \hspace{1cm} (6.8)

This claim is proven in Appendix 6.A. In practice, a small positive regularization term can be added to the denominator to stabilize the system for the cases when $\hat{B}_k(m) = 0$. The assumption can be motivated since $\beta_k \gg \alpha_k$ when the historical averaging window time is much larger than one transmission block. It is reasonable to assume that the instantaneous scheduled rate $\tilde{R}_k(m)$ is not much greater than the historical average rate $\bar{R}_k(m)$ since an OFDMA system offers many more scheduling opportunities than the "all-or-nothing" allocation of single-carrier systems, resulting in a steadier data flow for the users [RKPN05].

**Scheduling Metric Simplifications**

The scheduling metric (6.8) can be simplified by setting the target bit-rate, $R_k$, equal for all users, i.e. $R_k = R_0$. Then, $\hat{B}_k(m) = \bar{R}_k(m)/R_0$, and (6.8) simplifies to

$$k_q^*(m) = \arg \max_k \left\{ \frac{C_{q,k}(m)}{R_k(m)^\kappa} \right\}.$$
CHAPTER 6. MODIFIED PROPORTIONAL FAIR SCHEDULING FOR OFDMA

Setting $\kappa = 0$ gives the maximum throughput scheduler,

$$k_q^*(m) = \arg \max_k \{C_{q,k}(m)\}.$$  

Setting $\kappa = 1$ gives the proportional fair scheduling metric,

$$k_q^*(m) = \arg \max_k \left\{ \frac{C_{q,k}(m)}{R_k(m)} \right\}.$$  

Letting $\kappa \to \infty$, the user $k$ with lowest $\bar{R}_k(m)$ is scheduled

$$k_q^*(m) = \lim_{\kappa \to \infty} \arg \max_k \left\{ \frac{C_{q,k}(m)}{R_k(m)^\kappa} \right\} = \arg \min_k \{ \bar{R}_k(m) \},$$

since $C_{q,k}(m)$ is bounded. This corresponds to the max-min scheduler.

6.5 Summary

The single-carrier proportional fair scheduler was modified

- to OFDMA,
- to let each user have a unique target rate and target scheduling delay and
- to incorporate a global fairness level.

The system does not guarantee a fulfilment of the target rate or delay, but tries to combine channel-aware scheduling and relative fairness (relative to the parameters of the users). The corresponding utility function was shown to be a log-like function, provided that the averaging window is long enough. The scheduler can be tuned with the fairness parameter between the two extremes max-min scheduling and maximum throughput scheduling. In Chapter 8, the algorithm is numerically evaluated.
Appendix 6.A  Scheduling Metric Derivation

In this appendix, we elaborate on the utility function (6.6) and derive the scheduler in Section 6.4.

Utility Function Simplifications

The proposed general utility function in (6.6) is restated here,

\[ U_k(\hat{B}_k) = \frac{R_k \beta_k^\kappa}{\alpha_k} \frac{1}{1 - \kappa} (\hat{B}_k^{1-\kappa} - 1). \]  (6.9)

When the QoS-parameters of all users are equal and the fairness parameter \( \kappa \) equals zero, the utility function becomes linear and equal for all users,

\[ U_k(\hat{B}_k) = C_0(\hat{R}_k - D_0), \]

which corresponds to the classical maximum rate scheduler, with \( C_0 \) and \( D_0 \) being constants.

When the QoS-parameters of all users are equal and the fairness parameter approaches 1,

\[ \lim_{\kappa \to 1} U_k(\hat{B}_k) = \lim_{\kappa \to 1} C_1 \ln (\hat{B}_k^{1-\kappa}) = C_1 \ln \hat{B}_k = C_1 \ln \hat{R}_k - D_1 \]

the utility function corresponds to that of the classical proportional fair scheduler, with \( C_1 \) and \( D_1 \) being constants. In the third step, the power series expansion of \( \hat{B}_k^{1-\kappa} \) was used. When \( \kappa \to \infty \) and the QoS-parameters of the users are equal, the utility function corresponds to that of the max-min scheduler, which was also shown in Section 6.4.

Scheduling Metric Derivation

In the following the claim in Section 6.4 is established. The problem of maximizing (6.7) is highly complex, since it involves a joint user allocation on all clusters. Here, it is relaxed by assuming that \( \beta_k \hat{R}_k(m) \gg \alpha_k \hat{R}_k(m) \). This assumption is valid if the averaging window is long (\( \beta_k \gg \alpha_k \)) and the scheduled rate is in the same order as the average rate (\( \hat{R}_k(m) \approx \hat{R}_k(m) \)). This assumption enables the decoupling of the joint scheduling.

Denote the optimal instantaneous relative rates \( \{\hat{B}_{q,k}^*(m)\} \). The optimal allocation has the following property: given the allocation of all other clusters, the \( q^{th} \)
cluster is allocated to the user that achieves the highest utility-increase. In other words,

\[ k_q^* (m) = \arg \max_{k'} \sum_k U_k \left( \alpha_k \sum_{i \neq q} \tilde{B}_{i,k}^*(m) + \delta(\hat{k} - k') \frac{\alpha_k R}{R_k} C_{q,k}(m) + \beta_k \tilde{B}_k(m) \right) \]

\[ = \arg \max_{k'} \sum_k U_k \left( \alpha_k \sum_{i \neq q} \tilde{R}_{i,k}(m) + \delta(\hat{k} - k') \frac{\alpha_k R}{R_k} C_{q,k}(m) + \beta_k \bar{R}_k(m) \right), \]

where \( \delta(\cdot) \) is the Kronecker delta and \( R \) is the number of sub-carriers per cluster. Denote the potential relative rate in cluster \( q \) for user \( k \),

\[ \tilde{B}_{q,k}^*(m) = \frac{\alpha_k \hat{R}_k(m) + \beta_k \bar{R}_k(m)}{R_k}. \]

The relaxing assumption that \( \beta_k \bar{R}_k(m) \gg \alpha_k \hat{R}_k(m) \geq \alpha_k \sum_{i \neq q} \tilde{R}_{i,k}^*(m) \) means that the user allocations during the current block have little impact on the new average rates. Hence,

\[ k_q^* (m) = \arg \max_{k'} \sum_k U_k \left( \delta(\hat{k} - k') \alpha_k \hat{B}_{q,k}(m) + \beta_k \bar{B}_k(m) \right). \] (6.10)

This implies that the scheduling problem is separable and can be solved for each cluster separately. Note that \( \delta(\hat{k} - k_q^*) \tilde{B}_{q,k}(m) = \tilde{B}_{q,k}(m) \).

Now, consider the problem of scheduling a user for the \( q^{th} \) cluster, based on (6.10). For all but the scheduled user, \( \tilde{B}_{q,k}(m) = 0 \). Therefore, the user allocation that maximizes the sum-utility in (6.10) is to choose the user with highest marginal utility increase if scheduled,

\[ k_q^* (m) = \arg \max_k U_k \left( \alpha_k \hat{B}_{q,k}(m) + \beta_k \bar{B}_k(m) \right) - U_k \left( \beta_k \bar{B}_k(m) \right) \approx \alpha_k \hat{B}_{q,k}(m) \frac{d}{d\hat{B}_k} U_k(\beta_k \bar{B}_k(m)), \] (6.11)

where the first order linear approximation is used with the assumption again that the rates of the current block perturbs the average rates marginally. From (6.9), the derivative is easily found as

\[ \frac{d}{d\hat{B}_k} U_k(\hat{B}_k) = \frac{R_k \beta_k^\kappa}{\alpha_k \hat{B}_k^{-\kappa}}. \] (6.12)

Combining (6.11) and (6.12) gives (6.8), which concludes the proof.
Chapter 7

Opportunistic Beamforming for OFDMA

7.1 Overview

This chapter contains:

- An introduction to opportunistic beamforming.
- An opportunistic beamforming design for clustered OFDM. By applying different beamforming weights in different clusters, but the same within a cluster, frequency variability is induced in a favorable way.

7.2 Introduction

The concept of multiuser diversity relies on the fading of the users. If the channel quality of each user varies with time, it is possible to schedule users that are near their fading peaks. In practice, however, some users may be stationary or moving very slowly. For such users, the time between fading peaks may be long, which can result in long delays.

Opportunistic beamforming was introduced in [VTL02] to cope with this problem. A flat fading single-carrier scenario was considered where only one user at a time could be scheduled. Figure 7.1 roughly describes the three phases of opportunistic beamforming.

1. First, a random beam is formed and a training signal is transmitted. The SINR is estimated by each user in the cell using this training signal. This scheme assumes that all nearby base-stations change their beams and transmit their training signals simultaneously. Also, the training signals of the different cells have to be distinguishable by the users, so that the inter-cell interference can be predicted.
2. The estimated SINR, or equivalently the supportable rate, is fed back by each user to the base-station.

3. Based on the fed back information, the scheduler selects one user for communication and sends the user identification followed by the data using the fed back supportable modulation level and the same beam as in 1.

4. After the block of data has been transmitted, the process starts over again with the base-station forming a new random beam and transmitting a training signal.

It is not necessarily the user with the strongest received signal power that will have the highest received SINR. The opportunistic beamforming in the adjacent cells can give high fluctuations in the interference level, giving a low interference level for some users and a high level for others. This opportunistic nulling can act as an interference suppression mechanism. It is clear that an opportunistic system has a higher total throughput if there are many active users in the system to select from. Several modifications to opportunistic beamforming have been proposed, for instance [KHGT05], where multiple beams are tried before one is selected.

The opportunistic beamforming in [VTL02] assumes single-antenna receivers and uses a beamforming vector of the form

\[ \mathbf{w} = \left( \begin{array}{c} \alpha_1 e^{j\theta_1} \\ \vdots \\ \alpha_{N_t} e^{j\theta_{N_t}} \end{array} \right) \]

where \( \alpha_i \) are random variables normalized so that \( \sum \alpha_i^2 = 1 \) and the random \( \theta_i \) can vary between 0 and 2\( \pi \). The parameters in \( \mathbf{w} \) are regenerated each transmission block, inducing additional temporal variability to the effective channel \( \mathbf{h}_k^T \mathbf{w} \).

The highest received signal power at user \( k \) is obtained if the beamforming weights happen to be in the beamforming configuration, \( \mathbf{w} = \mathbf{h}_k^* / \| \mathbf{h}_k \| \). Then, disregarding inter-cell interference, the received SNR is

\[ \gamma_{BF} = \frac{\| \mathbf{h}_k \|^2}{\sigma_z^2} \] (7.1)

where \( \sigma_z^2 \) is the receiver AWGN power. It is shown in [VTL02] that as the number of users grow, the likelihood that one of the users will be in the beamforming configuration goes to 1. The SNR in (7.1) can be compared to the SNR if only one transmit antenna is used,

\[ \gamma_{SISO} = \frac{|h_k|^2}{\sigma_z^2} \] (7.2)

where \( h_k \) is the scalar complex channel gain of the \( k^{th} \) user. Instead, if for example Alamouti space-time block-coding [Ala98] is used, the SNR is

\[ \gamma_{Alamouti} = \frac{\| \mathbf{h}_k \|^2}{2\sigma_z^2} \] (7.3)
The increased received SNR for the beamforming configuration compared to the single transmit antenna SNR is called the beamforming gain. The gain comes at the cost of additional feedback. In addition to the gain from selecting a user that matches the beam well, opportunistic beamforming can exploit the multiuser diversity, i.e. schedule a user with high $\|h\|^2$. Note that, for opportunistic beamforming to achieve these gains, there has to be many users to choose from.

Opportunistic beamforming can be extended directly to OFDM, as outlined in [VTL02]. Since such a system consists of parallel flat fading channels, opportunistic beamforming as described above can be applied to each sub-carrier.

### 7.3 Opportunistic Beamforming for Clustered OFDM

In an OFDMA multiuser diversity system, it is favorable if the users fade, both in time and frequency. There is not much multiuser diversity gain to be exploited in a system with only stationary and flat fading users. The opportunistic beamforming scheme proposed in this section is a way to induce additional effective channel variability in both time and frequency. The scheme is based on the clustered OFDMA described in Chapter 3.

We assume a base-station with $N_t$ transmit antennas and $K$ users with one receive antenna each. To reduce the amount of feedback from the users to the base-station scheduler, the feedback scheme proposed in Chapter 3 is used.

The opportunistic beamforming used in this chapter is similar to that used in [VTL02, LLRS03]. Assuming $I$ synchronous base-stations and omitting the time index (c.f. Section 2.2 and (2.6)), the $k^{th}$ user that belongs to base-station 0 receives, on the $n^{th}$ sub-carrier,

$$y_{n,k} = \mathbf{H}^T_{n,k} \mathbf{w}^0_n \sqrt{P} c^0_n + \sqrt{P} \sum_{i=1}^{I-1} \tilde{\mathbf{H}}^i_{n,k} \mathbf{w}^i_n c^i_n + z_{n,k}$$

(7.4)

where $c^i_n$ is the unit-energy transmitted symbol from base-station $i$, $z_{n,k}$ is additive white Gaussian noise, $\mathbf{H}^i_{n,k} \in \mathbb{C}^{N_t \times 1}$ is the baseband frequency response vector from the $N_t$ antennas of the $i^{th}$ base-station to the $k^{th}$ user on the $n^{th}$ sub-carrier, $\mathbf{w}^i_n \in \mathbb{C}^{N_t \times 1}$ is the transmitter beamforming vector for the $n^{th}$ sub-carrier and $i^{th}$ base-station, and $P$ is the transmit power per sub-carrier. For the interfering channels, $\tilde{\mathbf{H}}$ means that the frequency response is rotated, which does not change the received SINR. The random beamforming vectors $\mathbf{w}^i_n$ are uniformly distributed on the unit sphere. Let $G^i_{n,k} = \mathbf{H}^T_{n,k} \mathbf{w}^i_n$ denote the complex-valued effective baseband frequency response from base-station $i$ to user $k$ on sub-carrier $n$, so that

$$y_{n,k} = G^0_{n,k} c^0_n + \sqrt{P} \sum_{i=1}^{I-1} G^i_{n,k} c^i_n + z_{n,k}.$$ 

(7.5)
Hence, the users effectively experience SISO channels on all sub-carriers. We propose a clustered beamforming design for $w_n^i$. For clustered beamforming (CL-BF), the $w_n^i$ are identical within one cluster of sub-carriers, but independent between the clusters. Hence, the sub-carrier correlation within the clusters is maintained and the sub-carrier correlation between the clusters is reduced. This enables the feedback of one channel quality value per cluster since the effective channel within a cluster remains fairly constant. The channel quality of two adjacent clusters may, however, be radically different due to the different beamformers. By having different beamforming weights in different clusters (CL-BF), additional variability in the frequency dimension is induced. This is an extension of the idea in [VTL02] where different beamforming weights were used in different transmission blocks to increase the temporal variability. Figure 7.2 illustrates the effect of clustered opportunistic beamforming on two channels.

In the simulations in Chapter 8, the clustered beamforming is compared to a scheme where the same opportunistic beamforming is used across all sub-carriers, here called equal beamforming (EQ-BF). For equal beamforming, $w_n^i = w^i$ for all $n$. Hence, only one IDFT has to be performed at the transmitter, compared to $N_t$ for CL-BF, but the frequency variability is not changed particularly. The beamforming weights are constant during one transmission block, which consists of a number of OFDM symbols, but change from block to block.

A second alternative to CL-BF is delay diversity (DD). In a delay diversity system, the same signal is transmitted on all antennas but time-delayed on a subset of the antennas. This results in an artificially longer channel impulse response and more diversity also in the frequency domain. The advantage of such an approach is its low complexity. The first drawback of a delay diversity solution is that the length of the cyclic prefix has to be increased to incorporate the longest length of the artificial impulse response of any user. The second drawback is that it increases the frequency variability not only between the clusters, but also within the clusters. This leads to less accuracy in the feedback rates.

To illustrate the effect of clustered opportunistic beamforming and delay diversity, the effective gains across the sub-carriers of a user are shown in Figure 7.3. The used simulation parameters are described in Section 8.2. The performance of the described schemes is evaluated in Chapter 8.

### 7.4 Exploiting Temporal Channel Correlation in the Scheduling and Beamforming

For slowly moving users, the temporal channel correlation between transmission blocks is high. If clustered opportunistic beamforming is used, this can be exploited by the base-station to help users that are not achieving their desired QoS. By keeping the beamforming weights for those users’ strongest cluster during the next transmission block, the probability that the cluster is strong also during the next
transmission block is increased. Now, assume that the modified PF scheduler of Section 6.4 is used. Then, the modification can be expressed as follows.

1. A user, $k$, is considered to be well below its target rate if $\bar{B}_k < \eta$, where $\eta < 1$ is a threshold value.

2. For each user, $k$, that is well below its target rate, keep the beamforming weights on the strongest cluster of the user during the next transmission block.

Note that this method does not require knowledge of the users’ speeds.

### 7.5 Opportunistic Beamforming Frame Structure

In a multiuser diversity system, all users have to track their SINR. SINR tracking and prediction is an easier problem than full channel estimation [FSES04]. Before each scheduling decision all users feed back their supportable rates based on the estimated or predicted SINR. However, in a system with opportunistic beamforming, tracking and prediction of the SINR needs to be coordinated with the change of beamforming weights. This brings the need for a training period with the new beamforming weights before the feedback and scheduling. Since there is a processing and feedback delay between the training and the scheduling decision, a frame structure as in Figure 7.4 is feasible. The training with the new beamforming weights, $w(m + 1)$, can be done within the previous transmission block, which uses the previous beamforming weights, $w(m)$.

Since the users estimate their SINR during the training period, it is important that both the signal and the interference power do not change between the training and the data transmission. This can be guaranteed by synchronizing the base-stations and not allowing shifting of transmit power from unused to used sub-carriers.

### 7.6 Summary

In this chapter, an opportunistic beamforming structure for clustered OFDMA was proposed. In order to induce frequency variability, in addition to the temporal variability offered by the single-carrier opportunistic beamforming, different beamformers are employed on different clusters. The same beamformer is used on all sub-carriers within a cluster to maintain the high correlation and enable clustered feedback and scheduling. A simple modification that could fix a beam over a longer time period was proposed to improve the rates of the most unfairly treated users. The performance of the schemes in this chapter is numerically evaluated in Chapter 8.
Figure 7.1: This figure illustrates the phases of single-carrier opportunistic beamforming. In phase 1., the base-station forms a random beam and transmits a training signal using this beam. Each user estimates the SINR from this training signal. The SINR or the corresponding supportable rate is fed back to the base-station in 2., which will be the basis for the scheduling decision. In 3., the base-station transmits a data block to the scheduled user using the fed back supportable rate and the corresponding beam. After the finished transmission, the base-station forms a new beam and transmits the training signal again in 4. and so on.
Figure 7.2: In these three plots, the effect of clustered beamforming on a slowly moving user is illustrated. The two uppermost plots show the magnitudes of the channels from the two transmit antennas to a user. The lower plot shows the effective channel after clustered opportunistic beamforming is applied to the both channels. It can be seen that the channel variability in both time and frequency is increased, which is advantageous in a multiuser diversity system. The number of sub-carriers is 128, the cluster-size is 4 sub-carriers and the beamforming pattern is changed every $16^{th}$ OFDM symbol.
Figure 7.3: Example of effective gains of a user for different transmitter schemes. The lowest plot shows the gains for a SISO channel. By adding one transmit antenna and delay diversity in the middle plot, more frequency variability is induced. In the uppermost plot, the effective gains for clustered beamforming is shown. The cluster-size, $R$, is 4 sub-carriers.

Figure 7.4: Frame structure for opportunistic beamforming. During the $m^{th}$ transmission block, the beamforming weights, $w(m)$, are used. The scheduling and adaptive modulation in the next transmission block $(m + 1)$ is based on measurements of training symbols (Tr) with the new beamforming weights, $w(m + 1)$. Due to the feedback delay, the training (Tr) with the new beamforming weights, $w(m + 1)$, is done inside the previous transmission block $m$. 
Chapter 8

Numerical Results for Chapters 6-7

8.1 Overview

This chapter contains:

- A description of the parameters and assumptions in the multi-cell simulations.
- Simulation results for the schemes proposed in Chapters 6-7.

8.2 Simulation Environment and Assumptions

To evaluate the performance of the modified proportional fair (M-PF) scheduler and the clustered opportunistic beamforming (CL-BF) presented in the previous chapters, the downlink of an FDD system with seven cells has been simulated. The performance is evaluated in terms of sum-rate and distribution of rates and scheduling delays. Results are collected from one cell which is surrounded by six interfering cells. All base-stations are assumed to be synchronous and use the same frequency band for the downlink. All base-stations use the same beamforming and scheduling schemes. The users are assumed to perfectly estimate the SINR on all sub-carriers and the feedback delay is assumed to be zero. The adaptive feedback rate is such that the probability that less than 80% of the sub-carriers can be assigned is 0.2 (see Section 3.3).

If we assume that the sum of the interference is Gaussian, the downlink channel consists of parallel approximately Gaussian sub-channels, with contributions from different users. In the simulations, the supportable number of bits per symbol for the $n^{th}$ sub-carrier and the $k^{th}$ user in cell 0 is estimated as

$$T_{n,k} = \frac{1}{2} \log_2 \left( 1 + \frac{\gamma_{n,k}}{\Gamma} \right)$$  \hspace{1cm} (8.1)

where the notation from (2.6)-(2.7) is used and $\Gamma$ is the gap corresponding to a symbol error rate of $10^{-4}$ for QAM [CDEF95, GC97]. Thus, it is assumed that the
### Table 8.1: Simulation parameters. In some Figures, $\kappa$ and $N_t$ are varied.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling frequency</td>
<td>4 MHz</td>
</tr>
<tr>
<td>Number of sub-carriers $N$</td>
<td>128</td>
</tr>
<tr>
<td>Cyclic prefix length</td>
<td>12 $\mu$s</td>
</tr>
<tr>
<td>Total OFDM symbol period $T_o$</td>
<td>$32+12=44$ $\mu$s</td>
</tr>
<tr>
<td>Sub-carrier spacing</td>
<td>31.25 kHz</td>
</tr>
<tr>
<td>Total bandwidth</td>
<td>4 MHz</td>
</tr>
<tr>
<td>Total transmit power $P_{tot}$</td>
<td>500 W</td>
</tr>
<tr>
<td>Average SNR at cell boundary</td>
<td>10 dB</td>
</tr>
<tr>
<td>Carrier frequency</td>
<td>1900 MHz</td>
</tr>
<tr>
<td>Number of transmit antennas $N_t$</td>
<td>2</td>
</tr>
<tr>
<td>Antenna separation</td>
<td>$4\lambda$</td>
</tr>
<tr>
<td>Maximum user speed</td>
<td>100 km/h</td>
</tr>
<tr>
<td>Maximum relative Doppler</td>
<td>0.77%</td>
</tr>
<tr>
<td>Cell radius</td>
<td>1 km</td>
</tr>
<tr>
<td>Distance between base-stations</td>
<td>2 km</td>
</tr>
<tr>
<td>Number of interfering base-stations</td>
<td>6</td>
</tr>
<tr>
<td>Cluster-size $R$</td>
<td>2 sub-carriers</td>
</tr>
<tr>
<td>Fairness parameter $\kappa$</td>
<td>0.5</td>
</tr>
<tr>
<td>Weight-keeping threshold $\eta$</td>
<td>0.8</td>
</tr>
<tr>
<td>Transmission block $N_{tb}$</td>
<td>16 OFDM symbols</td>
</tr>
<tr>
<td>Simulation time</td>
<td>$1536T_o \approx 68$ ms</td>
</tr>
</tbody>
</table>

Adaptive modulation can handle also non-integer number of bits per symbol. The supportable rate in bits per second on sub-carrier is $n$ for user $k$ is $C_{n,k} = T_{n,k}/T_o$, where $T_o$ is OFDM symbol duration. We assume that the scheduled users achieve the rates they estimate from (8.1).

The system parameters are described in Table 8.1. The 3GPP spatial channel model for MIMO simulations for urban environments is used [3GP03]. It is modified for omni-directional antennas by changing the antenna gain pattern to be uniform. The users are uniformly distributed over the cells and their speeds are uniformly distributed between 0 and 100 km/h. Temporal correlation over the simulated OFDM symbols is added using Jakes model [DBC93]. In one simulation, flat channels are used. The same channel model is used for this case, but the taps of each impulse response are summed to one tap.

Three user-classes are considered: one class with high bit-rate requirements and short delay-window, corresponding to a user using a real-time application, one class with medium bit-rate and delay requirements, and one class with a low bit-rate and long delay-window, corresponding to a user downloading data in the background. Half of the users belong to Class 3 and one quarter of the users belong to Class 1 and 2 each. In practice, each user can have an individual rate and delay requirement.

- **Class 1:** $R = 256$ kbps and $T = 20N_{tb}T_o = 14$ ms
• Class 2: $R = 128$ kbps and $T = 40N_{tb}T_o = 28$ ms

• Class 3: $R = 64$ kbps and $T = 60N_{tb}T_o = 42$ ms

The opportunistic beamforming schemes with PF schedulers are compared to a conventional beamforming scheme with round-robin scheduling (Conv. BF RR). In the conventional beamforming scheme, the beamforming is computed independently for each OFDM sub-carrier. This scheme requires full CSI for the scheduled user at the transmitter, which is also assumed. The base-stations have no channel knowledge of the users in the adjacent cells, so interference nulling is not possible. The transmitter selects the beamforming weights for each sub-carrier to be in the beamforming configuration, which maximizes the received SNR (c.f. (7.1)). Water-filling of the transmit power across the sub-carriers is done, giving the sub-carriers with higher effective gain more power [CT91]. This may result in some sub-carriers not being used. For the conventional beamforming scheme, round-robin scheduling is used, with fixed allocated time-slots for each user. Users with high-rate requirements will be assigned correspondingly more time-slots. One time-slot here is equal to one transmission block. One problem with the conventional beamforming scheme is that the inter-cell interference may change between channel estimation and data transmission, due to transmit power waterfilling in adjacent cells. This would degrade the performance of this scheme. In the simulations, however, we neglect this effect and assume that the interference level during the data transmission is known when the modulation level is selected. However, the waterfilling of power does not take the frequency-selective ICI power into account.

The M-PF scheduler proposed in Chapter 6.4 is also compared to the OFDM proportional fair extension recommended in [ALS+03], which will be called standard PF. In this scheme as in single-carrier PF, the users have no particular QoS requirements, but compete for the channel with equal priority. The user-specific historic throughput is updated after the allocation of all sub-carriers using (6.2). The sub-carrier assignment order is as for M-PF.

To evaluate the performance of CL-BF, it is compared to equal beamforming weights across the sub-carriers (EQ-BF), delay diversity with delay half the length of the cyclic prefix (see Section 7.3) and SISO.

Delay is an important QoS measure. In the simulations, we have defined the maximum scheduling delay a user experiences as the maximum time between the reception of two consecutive packets. We have defined a packet to be 512 bits. A user which is rarely scheduled will have a high maximum delay. Maximum delay is also partly a function of the rate of the user. Users with very low bit-rates will experience longer delays, even if they are scheduled each OFDM symbol. Packet outages are not considered here.
Figure 8.1: Cell throughput as a function of the number of users in each cell is displayed for two transmit antennas. Conventional beamforming with round-robin scheduling, clustered beamforming with M-PF and PF scheduling, equal beamforming on all clusters, single transmit antenna and delay diversity with M-PF scheduling are compared.

8.3 Simulation Results

The cell throughput is the sum of the average bit-rates of all users in one cell. In Figure 8.1, the cell throughput for different beamforming and scheduling schemes is compared as a function of the number of users in each cell. The schemes based on clustered beamforming perform best. The low cell throughput of the conventional beamforming scheme is due to the inability of the round-robin scheduler to avoid inter-cell interference. For the opportunistic schemes, the sub-carriers with high ICI are probably not even fed back. In an environment free of ICI, the conventional beamforming scheme outperforms the opportunistic systems except when many users are active [SWO04]. The cell throughput of the delay diversity scheme decreases due to the increased cyclic prefix overhead. There is only a small difference in total throughput between single antenna transmission (SISO) and equal opportunistic beamforming on all clusters (EQ). The reason is that for Rayleigh fading channels (which the channel model closely resembles) with relatively fast fading, opportunistic beamforming does not change the fading statistics [VTL02]. The small gain comes from the slowly fading users that can be scheduled more often for EQ, thereby decreasing the risk of having to transmit on a bad channel. For
8.3. Simulation Results

Figure 8.2: The CDF of the bit-rates of the 32 users for clustered beamforming with M-PF and PF scheduling. Note that the CDF for PF is for all users. The sum throughputs of the two schemes are nearly the same for 32 users.

CL-BF however, the fading rate in both time and frequency is increased, resulting in more efficient scheduling. Multiuser diversity schemes with low rate feedback usually perform worse for few users. By adapting the feedback rate, as described in Chapter 3, we keep the spectral usage high even for few users.

The M-PF scheduler is designed to give users the rates they require, but not much more. This is not possible in the standard PF scheduler. To evaluate how well this discrimination works, user rate CDFs are displayed in Figure 8.2 when there are 32 users per cell. The cell throughputs for CL-BF M-PF and CL-BF PF are nearly the same for 32 users (see Figure 8.1). However, the M-PF scheduler manages to differentiate the rates of the different QoS-classes as can be seen in Figure 8.2. For Class 1, roughly 80% of the users exceed their target rate, for Class 2 more than 90% of the users exceed their target rate and for Class 3 more than 95% of the users achieve their target rate. Since users are not treated differently by the PF algorithm, the CDF for PF All is for all users. The rates of the users within the QoS classes are fairly concentrated. This indicates that the system is fair. If the rate requirements of the classes are properly set and there are not too many users in the system, most users will also be satisfied.

In Figure 8.3, CDFs of maximum delay for CL-BF, M-PF, and CL-BF PF are displayed. For all three classes and also for PF, most users experience a maximum delay below 10 transmission blocks, which equals 7 ms. The low delay can be
explained by the many time-frequency slots that are available to the scheduler (at most $Q$ users can be scheduled in the same OFDM symbol). The maximum delay of users in QoS Class 3 (background download) is larger, which is acceptable due to the lower delay requirements.

Adding more transmit antennas in a beamforming system enables narrower beams. In the clustered opportunistic beamforming system studied here, more transmit antennas is advantageous as long as there are enough active users, as can be seen in Figure 8.4.

To evaluate the effect of the fairness parameter, $\kappa$, in the M-PF scheduler, the cell throughput is displayed as a function of $\kappa$ in Figure 8.5. The cell throughput of the PF scheduler, which has $\kappa = 1$ and equal rate and delay parameters for all users, is displayed as a comparison. As expected, the highest cell throughput is achieved with $\kappa = 0$, when the strongest user is always scheduled, without taking fairness into account. User rate and delay CDFs are shown in Figures 8.6 and 8.7. A higher fairness level gives more equally distributed users within the QoS-classes.

To further help the weakest users, the M-PF scheduler interacts with the beamformer, as described in Section 7.4. The beamforming weights of the strongest cluster of the weakest users are kept to the next transmission block. The effect of this method can be seen in Figure 8.8, which is a zoom-in on the CDF of the weakest users of QoS Class 1 with a target rate of 256 kbps. The gain of this method is limited because of the changing inter-cell interference between transmission blocks.
Even if the base-station keeps the beamforming weights on the strongest cluster, the inter-cell interference might change radically. Still a higher throughput for the weakest users can be observed if this extra information is used.

8.4 Summary

The modified proportional fair (M-PF) scheduler and clustered opportunistic beamforming (CL-BF) was numerically evaluated in a multi-cell simulator. The CL-BF performed better than single-antenna transmission, a delay diversity scheme and an opportunistic beamforming with equal beamforming weights on a sub-carriers. The M-PF scheduler performed similarly to a standard PF scheduler, but managed to differentiate the user rates according to their target rates. The user-specific delay parameters influenced the scheduling delay. The fairness parameter was shown to significantly impact both sum throughput and the user bit-rate distributions. The extension that exploits the temporal channel correlation slightly improved the rates of the weakest users.
Figure 8.5: Cell throughput for 32 users as a function of the fairness parameter $\kappa$ for clustered beamforming with M-PF scheduling. As a comparison, the throughput of clustered beamforming with standard PF is displayed. The standard PF has no fairness parameter.
Figure 8.6: User rate CDFs for QoS class 1 for the clustered beamforming with M-PF scheme are displayed for various fairness levels $\kappa$. There are 32 users in the cell.
Figure 8.7: Maximum delay CDFs for QoS Class 3 for the clustered beamforming with M-PF scheme are displayed for various fairness levels $\kappa$. A delay of 10 transmission blocks is about 7 ms. There are 32 users in the cell.
Figure 8.8: The CDFs for the 60% of the users in Class 1 with the lowest rates. Clustered beamforming with M-PF with beamforming weight keeping is compared to clustered beamforming with M-PF without beamforming weight keeping. Beamforming weight keeping means keeping of the beamforming weights for the reported strongest cluster of the weakest users during consecutive transmission blocks. The threshold for weight keeping is here $0.8 \times \text{target rate}$, which is 205 kbps.
Chapter 9

Opportunistic SD-OFDMA

9.1 Overview

This chapter contains:

- A description of opportunistic space-division OFDMA (SD-OFDMA).
- A method to exploit temporal channel correlation for opportunistic SD-OFDMA, in order to increase throughput.
- Numerical evaluation of the proposed scheme.

9.2 Introduction

The opportunistic beamforming principle was applied to OFDMA in Chapter 7. For each flat fading sub-carrier, a single random beam was generated at the base-station. However, the potentially largest increase in data-rate when using multiple antennas on a single link comes from spatial multiplexing [RC98], in which several data streams can be transmitted simultaneously over a multiple-input-multiple-output (MIMO) channel, thereby increasing the data-rate. In Chapter 7, only one data stream per sub-carrier was transmitted from the base-station antenna array. The step to opportunistic single-carrier spatial multiplexing was taken in [CHKK03], where the base-station communicated with one user at a time over one or several streams.

Space-division multiple access (SDMA) is similar to spatial multiplexing, but on a multiuser level [PP97]. In SDMA, the base-station, equipped with multiple antennas, can simultaneously transmit different data streams to several different users. The base-station multiplexes users by transmitting different signals in different spatial directions. This spatial separation typically requires detailed channel state information of the different users, so that the user-specific beams can be computed. Fortunately, the opportunistic principle can also be applied in the SDMA scenario.
In [SH05], the theoretical aspects of opportunistic SDMA for a flat fading broadcast channel were investigated. One important result was that the sum rate in a cell can grow linearly with the number of transmit antennas at the base-station if the number of users is sufficiently large. This is the same capacity growth rate as when perfect CSI is available at the transmitter and dirty paper precoding is used. In [SH05], a number of random but orthogonal beams were generated at the base-station. Using fast feedback from the users, the base-station scheduled, on each beam, the user with the highest SINR. Note that the beams were orthogonal at the transmitter but typically not at the terminals. This resulted in users receiving significant signal power from all of the transmitted beams, here called inter-beam interference. However, with enough users to choose from, the inter-beam interference of the scheduled users was low.

In this chapter, we apply the random multi-beam communication scheme in [SH05] to an OFDMA downlink. We propose a method to exploit the temporal channel correlation in order to increase the throughput, without any additional overhead. We use the fact that several beamformers can be maintained in parallel in the OFDMA downlink to find and keep successful beams over several blocks. Since multiple access is opportunistically granted through both space and frequency, we have chosen to call this scheme opportunistic space-division OFDMA (SD-OFDMA). Note that although the multiple access in frequency is orthogonal the space-division multiple access is not, resulting in additional interference.

As for the opportunistic beamforming in Chapter 7, we divide the OFDM symbol into clusters, which define the frequency granularity for feedback, scheduling and beamforming. The scheduling and beamforming is common for all sub-carriers in a cluster, but may differ between clusters. Furthermore, we assume that the users feedback information only about their best clusters, as in Chapter 3. An alternative or additional feedback reduction mechanism was presented in [WGK05], where users were allowed to feedback only if the normalized cross-correlation between their channel and the beam is above a threshold.

The main contribution of this chapter is a method to also exploit the temporal channel correlation in opportunistic SD-OFDMA. For the beams that are applied in parallel to the clusters, due to the fast feedback, the base-station knows which ones best match the instantaneous channels of the users. If the channels for some users fade slowly, it would be advantageous to keep their best beams also for the next block. In this work beams on clusters that result in the highest scheduled sum-rates are kept for the next block. Hence, user speed information is not required at the base-station. This means that the beams on some clusters will be kept and the beams on the other clusters will be regenerated. Because we have an OFDMA system with many parallel channels, the same users are likely to be scheduled again on the kept beams. In a single-carrier system with fairness restrictions this approach is less feasible since all users compete for one channel. In an OFDMA system, the randomization and keeping of beams can work in parallel, which can enable the finding of better beams for stationary users and new good beams for moving users. A single-carrier system is less flexible in this respect.
In Section 7.4, the temporal correlation was used with opportunistic beamforming to help users that are currently below their QoS targets. Previous approaches to exploit the temporal channel correlation in single-carrier random beamforming include [ALP04, KG05]. In [ALP04], the temporal correlation was used to help users with long delays. By averaging historical feedback information about preferred beams, beams can be generated for users with long waiting times. The goal of [KG05] was to improve the sum-rate performance of the random multi-beam scheme of [SH05]. Several historical successful beamforming matrices are kept in the memory. By a two-step training procedure, with one old and one new random matrix, the set of matrices in the memory is updated, and the most successful of the two is used for data transmission. Note that the training and feedback overhead is doubled.

The approach to keep the successful beams on some clusters and to randomize the beams on other clusters can be seen as a random search of beamforming configurations in the space of orthogonal beamformers. When a good beam configuration is found in a cluster, it is kept until the scheduled users have faded enough to make the beam configuration in some other cluster better. Remember that the advantage of the described methods over the completely random case is due to the correlation in time and frequency of OFDM symbols.

We also show the impact of several receiving antennas at the terminals. The ability to suppress inter-beam interference using multiple-antenna receivers can significantly improve the performance.

In Section 9.3, opportunistic SDMA for the OFDMA broadcast channel is presented. The method to exploit temporal channel correlation in the beamforming process is described in more detail in Section 9.4, and thoroughly evaluated by means of simulation in Section 9.5.

9.3  Opportunistic SD-OFDMA – Basic Scheme

The principle of the proposed SD-OFDMA scheme is described below (called Rand). A modification of this scheme is presented in Section 9.4.

1. For each sub-carrier, each base-station independently forms a number of random but orthogonal beams and transmits orthogonal training sequences on the beams.

2. Each mobile terminal estimates the SINR on each sub-carrier and each beam (the beam-SINR) and feeds back the supportable rate for a number of them to the designated base-station.

3. Each base-station schedules its users on the sub-carriers and beams according to the fed back rates. Any scheduling policy can be used.

4. The scheduling decisions are transmitted to the users, for instance by transmitting the index of the scheduled user on each beam.
5. During the rest of the block, data is transmitted to the users according to the scheduling in step 3, with the rates they estimated and fed back in step 2.

6. The same procedure, steps 1-5, is repeated for the next block for a new set of beams.

A block is a number of consecutive OFDM symbols. The $I$ base stations are equipped with $N_t$ antennas each and the users receive with $N_r$ antennas. The transmit power per sub-carrier and base-station is constrained to $P$. In [SH05], the transmit power was constrained per scheduled user. We also note that more beams generally implies more training.

Following the OFDMA MIMO model of Section 2.2, the received signal on sub-carrier $n \in \{0, \ldots, N - 1\}$, $y_n \in \mathbb{C}^{N_r \times 1}$, of a user connected to base-station 0 can be written as

$$y_n = H^0_n^T x^0_n + \sum_{i=1}^{I-1} \tilde{H}^i_n^T x^i_n + z_n$$

(9.1)

where $P$ is the transmit power per sub-carrier, $H^T_n \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix from base-station $i$ to the user on sub-carrier $n$ where $\tilde{H}^i_n$ denotes channel matrix times a scalar phase-shift due to the wrong time-synchronization, $x^i_n \in \mathbb{C}^{N_t \times 1}$ is the transmitted signal from the $i^{th}$ base-station and $z_n$ is additive white Gaussian noise with covariance matrix $\sigma^2 z^i I \in \mathbb{R}^{N_r \times N_r}$. Omitting the base station index, the transmit beamforming on each sub-carrier is similar to the single-carrier beamforming in [SH05],

$$x_n = \sqrt{\frac{P}{B}} \sum_{m=1}^{B} w^m_n c^m_n = \sqrt{\frac{P}{B}} W_n c_n$$

(9.2)

where $W_n = (w^1_n \ldots w^n_B) \in \mathbb{C}^{N_t \times B}$ is the random unitary beamforming matrix with $B$ orthonormal columns and $c_n = (c^1_n \ldots c^n_B)^T \in \mathbb{C}^{B \times 1}$ is the vector of unit average energy transmitted symbols. The transmitted signal on the $n^{th}$ sub-carrier in (9.2) is the sum of $B$ orthogonal beams, with one symbol transmitted on each beam. Note that the transmitted signal is normalized so that the average transmitted power per sub-carrier is the same regardless of the number of beams, which is not the case in [SH05].

The aggregate channel, $G_n \in \mathbb{C}^{N_r \times B}$, on the $n^{th}$ sub-carrier is

$$G_n = \sqrt{\frac{P}{B}} H^T_n W_n.$$ 

(9.3)

Hence, (9.1) can be written as

$$y_n = G^0_n c^0_n + \sum_{i=1}^{I-1} \tilde{G}^i_n c^i_n + z_n,$$

(9.4)
where $\tilde{G}_i^n = e^{j\phi^n_i}G_i^n$ and $\phi^n_i$ is unknown, as in (2.1). Assume that the aggregate channel from base-station 0 is known at the receiver and that the noise plus inter-cell interference sum on each antenna is Gaussian with known covariance (c.f. Section 2.A), $Z_n \in \mathbb{R}^{N_r \times N_r}$, that is,

$$Z_n = \sigma^2_z I_{N_r} + \sum_{i=1}^{l-1} \text{diag} \left( G_i^n G_i^n H \right)$$

where the $\text{diag}(\cdot)$ extracts the diagonal matrix (i.e. nulls the off-diagonal elements) and $Z_n$ is a diagonal matrix with the noise-plus-interference powers of the receive antennas on sub-carrier $n$ as diagonal elements. This means that the receiver assumes that the inter-cell interference is spatially white. Then, the linear MMSE receiver on sub-carrier $n$, $V_n = (v_{n,1} \ldots v_{n,B}) \in \mathbb{C}^{N_r \times B}$, is

$$V_n = \left( G_0^n G_0^n H + Z_n \right)^{-1} G_0^n.$$  

Note that the $B$ columns of $V_n$ are the beam-specific receiver beamformers. Also note that other receivers than LMMSE could be used here. The received symbol on beam $m$ is

$$d_n = v_{n,m}^H v_n.$$ 

As in Section 2.2, we henceforth assume that each user is allocated at most one beam, i.e. only one column of $V_n$ is used for data reception. Hence, the data symbol after the linear combining with $v_{n,m}$ contains the desired data symbol, interference from all other beams in the system and AWGN, as in (2.4). The beam-SINR is given by (2.5),

$$\gamma_{n,m} = \frac{|\alpha_{n,m}|^2}{\sum_{k \neq m} |\beta_{n,k}|^2 + \sum_{i=1}^{l-1} \left\| \zeta^n_i \right\|^2 + \left\| v_{n,m} \right\|^2 \sigma^2_z},$$

where

- $\alpha_{n,m} = \sqrt{\frac{P}{B}} v_{n,m}^H H_{n,m}^0 T w_{n,m}$ is the complex scalar effective channel gain,
- $\beta_{n,k} = \sqrt{\frac{P}{B}} v_{n,m}^H H_{n,m}^0 T w_{n,k}$ is the complex scalar inter-beam interference gain from beam $k$,
- $\zeta^n_i = \sqrt{\frac{P}{B}} v_{n,m}^H \tilde{H}^n_i T W^n_i \in \mathbb{C}^{1 \times B}$ is the inter-cell interference gain from the beams in cell $i$, and
- $\tilde{z}_n = v_{n,m}^H z_n$ is complex-valued AWGN with variance $\sigma^2_z \left\| v_{n,m} \right\|^2$.

The supportable rate on beam $m$ and sub-carrier $n$, $C_{n,m}$, is a function of the SINR. The supportable rate is typically highly quantized and is therefore suitable
for feedback. Based on the fed back supportable rates of all users in the cell, the base-station schedules (step 3 in Section 9.3) a user on the $n^{th}$ sub-carrier and $m^{th}$ beam with rate $r_{n,m}$ which is equal to $C_{n,m}$ of the scheduled user. If no user is scheduled on a beam and sub-carrier, $r_{n,m} = 0$. To clarify, each user estimates the supportable rate on beam $m$ and sub-carrier $n$, $C_{n,m}$, whereas the base-station assigns the rate $r_{n,m}$. Note that in step 5), a particular user typically decodes only a small subset of all the transmitted symbols, i.e., the symbols that are dedicated to the particular user.

In a typical OFDM system, there is significant correlation between adjacent sub-carriers. This was exploited in Chapter 3, where the $N$ sub-carriers were divided into $Q$ clusters of $R$ adjacent sub-carriers each ($N = QR$), so that the sub-carriers in one cluster typically were highly correlated. Then, the users only had to feedback one supportable rate per beam and cluster, which reduced the feedback rate. By only feeding back the supportable rates of the strongest clusters, the required feedback rate was further decreased, without significantly affecting performance. In this chapter, we apply the clustering principle to opportunistic SD-OFDMA. In order not to destroy the high correlation within a cluster, the same beamforming matrix is used on all sub-carriers in a cluster, during all OFDM symbols of a block. In other words, $W_k = W_l$ if sub-carriers $k$ and $l$ are within the same cluster and block and $W_k$ and $W_l$ are independent random matrices if sub-carriers $k$ and $l$ are not in the same cluster and block.

Also note that the basic opportunistic SD-OFDMA described above (henceforth called Rand) is the direct application of the single-carrier transmission scheme in [SH05] to each cluster. The differences are that, here, the transmit power is not allowed to change with the number of transmitted beams and the receivers can have multiple antennas. The same beamforming matrix is used on all sub-carriers in a cluster and during all OFDM symbols in a block. The random beamforming matrices are independent between clusters and blocks. In the next section, we propose a method (Keep) that partly removes this independence in time by exploiting the temporal correlation between blocks.

### 9.4 Exploiting Temporal Channel Correlation

In an opportunistic communication system, the channels of the users should not fade too much during a block, since the scheduling also includes channel-dependent rate-adaption. The block duration should typically be designed for the worst case, i.e., for the highest velocity. For a worst-case user, the correlation of the channels for two consecutive blocks is probably moderate. However, for the average user, the temporal channel correlation between two consecutive blocks is probably significant. For near-stationary users, the correlation will be close to 1. The method presented here (called Keep) tries to exploit the temporal correlation of the slowly moving users while also keeping track of fast-fading users.

The idea is that the beams in the frequency clusters that have the highest
scheduled sum-rates are also used for the next block. Thus, step 1) in Section 9.3 is changed, based on the outcome of step 3) in the previous block. In other words, the beamformers, \( W_n(k) \), in \( Q_{\text{keep}} \) of the \( Q \) clusters are kept for the next block \( k+1 \), where \( k \) is the block index. The notation \( W_n(k) \) means the beamforming matrix on sub-carrier \( n \) used during block \( k \). Let \( R(c) \) denote the set of sub-carriers that belong to cluster \( c \). Then, the sum-rate of cluster \( c \) is

\[
R_c(k) = \sum_{n \in R(c)} \sum_{m=1}^{B} r_{n,m}(k)
\]  

(9.8)

which is simply the sum of the scheduled rates over the beams and sub-carriers in the cluster. Now, let \( Q_{\text{max}}(k) \) be the set of clusters with highest sum-rate \( R_i(k) \) during block \( k \), i.e. \( R_i(k) \geq R_j(k) \) for all \( i \in Q_{\text{max}}(k) \) and \( j \notin Q_{\text{max}}(k) \). The size of \( Q_{\text{max}}(k) \) is \( Q_{\text{keep}} \). Let

\[
W_n(k+1) = W_n(k)
\]  

(9.9)

for all \( n \in R(c) \) and for all \( c \in Q_{\text{max}}(k) \). This means that the most successful beamformers \( W_n(k) \) in terms of sum-rate during block \( k \) in the cluster are not changed for block \( k+1 \). All other beamformers are randomized independently.

Note that the set of clusters for which the beams are kept generally changes from block to block. However, if the beams on a cluster fit some slowly moving users particularly well, the beams can successfully be kept for many consecutive blocks. If all users fade rapidly, \textit{Keep} will result in a behavior similar to the all-random method, since the aggregate channel will change significantly between blocks.

The proposed method is compatible with any scheduler that uses the instantaneous CSI provided by the fast feedback from the users, for example, proportional fair and maximum rate scheduling. It is also compatible with any feedback scheme, since it is based on the scheduled rates, not the feedback itself. Note that no extra feedback and little extra complexity is required for the \textit{Keep} scheme.

### 9.5 Performance Evaluation

#### Simulation Assumptions

A system with 9 cells has been simulated using the 3GPP spatial channel model for suburban environments [SDS+05]. Each cell has three sectors, i.e., the system has 27 sectors. The simulated area is wrapped around on a torus, such that each sector is surrounded by other sectors with similar traffic. We have assumed block fading channels. Only the OFDMA downlink is studied. The feedback channels are assumed to be free of errors and delay.

In the simulations, we assume that the base-stations are synchronized in the sense that the OFDM symbols are transmitted simultaneously and that the block timing is synchronized. Each user is connected and synchronized to one base-station. The additional propagation delay from the interfering cells is assumed to...
be less than the length of the cyclic prefix, as in the the multi-cell baseband model in Section 2.2. The base station synchronization enables the inter-cell interference to be predicted during the training period in step 2) in Section 9.3. It also prevents the inter-cell interference from changing in the middle of a block, in step 5). The scheme could also be used in a non-synchronized system, with an extra margin in the rate selection in step 2), due to the unpredictability of the inter-cell interference. Note that the base-stations work completely independently, except for the block synchronization.

In the simulations, the supportable number of bits per symbol, \( T_{n,m} \), is computed using the formula for the AWGN channel capacity and a gap, \( \Gamma \), for M-QAM corresponding to a bit error rate of \( 10^{-4} \) [CDEF95, GC97],

\[
T_{n,m} = \frac{1}{2} \log_2 \left( 1 + \frac{\gamma_{n,m}}{\Gamma} \right),
\]

where the SINR \( \gamma_{n,m} \) is computed as in (2.5). The supportable rate in bits per second on sub-carrier is \( n \) and beam \( m \) is \( C_{n,m} = T_{n,m}/T_o \), where \( T_o \) is OFDM symbol duration. The base-station schedules users according to a version of the proportional fair algorithm that updates the user rates after the allocation of all clusters and beams (the "third" scheme in [ALS+03]).

In the simulations, the adaptive feedback scheme described in Chapter 3 is used. When there are few active users in the sector, the base-station commands the users to each feed back the supportable rates of the best beams in many clusters. If there are many active users in the sector, they each feed back the supportable rates of the best beams in fewer beams. For instance, for 2, 5 and 8 users in a sector, they feed back 26, 15 and 11 beams each, respectively. For each cluster, each user feeds back at most one beam. Note that the feedback rate is independent of the number of transmitted beams. Since the number of users is a relatively slowly changing parameter, this adaptive scheme introduces very little feed forward overhead.

In Figs. 9.1-9.3, the ratio of kept beams, \( Q_{\text{keep}}/Q \), is set to 50% and the user speeds are uniformly distributed between 0 and 50 km/h. The number of transmitted beams, \( B_n \), on each sub-carrier is fixed to the number of transmit antennas, \( N_t \). We concentrate on a medium load region with between 3 and 10 active users per sector. Additional parameters are summarized in Table 9.1.

Two beamforming methods are compared:

- **Rand**: In this scheme, as described in Section 9.3, the temporal correlation is not exploited, but \( W_n(k) \) and \( W_n(k+1) \) are independent for all \( n \).

- **Keep**: This is the proposed scheme described in Section 9.4.

Note that when \( N_t = 1 \), no beamforming can be done.

**Simulation Results**

The supportable rate of a beam in a cluster is chosen as the lowest supportable rate of the beam across the cluster sub-carriers. Hence, the supportable rate of a
9.5. PERFORMANCE EVALUATION

Table 9.1: Simulation parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total bandwidth</td>
<td>3.84 MHz</td>
</tr>
<tr>
<td>Number of sub-carriers ( N )</td>
<td>128</td>
</tr>
<tr>
<td>Cluster-size ( R )</td>
<td>4 sub-carriers</td>
</tr>
<tr>
<td>Sub-carrier spacing</td>
<td>30 kHz</td>
</tr>
<tr>
<td>Cyclic prefix length</td>
<td>5 ( \mu s )</td>
</tr>
<tr>
<td>Total OFDM symbol period</td>
<td>33.33+5=38.33 ( \mu s )</td>
</tr>
<tr>
<td>Carrier frequency</td>
<td>2 GHz</td>
</tr>
<tr>
<td>Block length</td>
<td>32 OFDM symbols</td>
</tr>
<tr>
<td>Transmit power per sub-carrier and sector ( P )</td>
<td>312.5 mW</td>
</tr>
<tr>
<td>Noise power per sub-carrier and antenna ( \sigma_z^2 )</td>
<td>( 2.38 \cdot 10^{-11} ) W</td>
</tr>
</tbody>
</table>

beam is a function of the minimum beam-SINR within a cluster, which is a random variable. Its empirical cumulative density function (CDF) is shown in Figure 9.1 for 1, 2, and 4 transmit antennas. The receivers have only one antenna. The average number of users per sector is 6. The SINR when only one beam is transmitted, \( N_t = 1 \), is generally higher than when two and four beams are transmitted. The reason is twofold: firstly, when more beams are transmitted, the transmit power per beam, \( \frac{P}{N_t} \), is lower. Secondly, when more beams are transmitted, the inter-beam interference is generally higher. When only one beam per cluster is transmitted, it can use all available transmit power and there is no inter-beam interference. On the other hand, there is no spatial multiplexing gain when \( N_t = 1 \). A tradeoff is therefore between beam-SINR and number of spatial channels (beams) that can be used.

The average (over time) user bit-rate is also a random variable. Its empirical CDF is displayed in Figure 9.2. The settings are the same as for Figure 9.1. It can be seen that two beams per cluster \( (N_t = 2) \) gives slightly higher average rates for the lower-rate users than one beam, even though the beam-SINRs are lower (Figure 9.1). This is due to the spatial multiplexing gain from using more lower-rate beams simultaneously. The most suitable number of beams also depends on the noise level, as pointed out in [SH05]. With a high noise level, the significance of the inter-beam interference is lower, thereby promoting the use of more beams. With a low noise level, it is more feasible to use only one beam, since no inter-beam interference could yield very high SINRs.

Figure 9.3 shows the average sum-rate per sector as a function of the number of users. The upper set of curves is for four receive antennas \( (N_r = 4) \) and the lower set is for one receive antenna \( (N_r = 1) \). Note from Figure 9.3 that it is not advantageous to use more than one beam when the average number of users is low. The reason is that, for few users, the penalty in increased interference due to multiple beams is not compensated for by the multiplexing gain. With more users,
the likelihood of finding a user with a large beam-SINR for each beam is higher. With more receive antennas, the receivers are able to separate the transmitted beams better, thereby enabling higher spatial multiplexing gain. For all cases, the \textit{Keep} method outperforms the all-random \textit{Rand} method.

In \textit{Keep}, the beams in a ratio ($Q_{\text{keep}}/Q$) of the clusters are kept from block to block. Figure 9.4 shows the average sum-rate per sector as a function of this ratio. When the ratio is 0, \textit{Keep} coincides with the \textit{Rand} method, since no beams are kept. When it is 1, the beams on all clusters are generated only once and then kept forever. It is noteworthy that the sum-rate performance for the ratio 1 equals the performance for the ratio 0. This can be explained by noting that the system has 32 parallel channels (clusters) with different beamformers. Even if the beamformers in the different clusters remain constant all the time, there is a low risk that the base-station has to schedule a user on a weak channel, due to the diverse channels in the frequency domain. The best performance is achieved when 20-60\% of the beams are randomized for each block. For the three curves, the user speeds are uniformly distributed between 0 and 10 km/h, 50 km/h and 100 km/h respectively. Keeping beams is more advantageous when the users fade slowly. If some of the users move as fast as 100 km/h, a significant gain in sum-rate can still be expected.
Figure 9.2: CDF of user bit-rate for one receive antenna and an average of 6 users per sector. The penalty in beam-SINR in Figure 9.1 when $N_t > 1$ is compensated by the spatial multiplexing gain.

since the rates of slower users increase by keeping the best beams.

The simulation results presented here are highly influenced by the choice of scheduler. If maximum-rate scheduling would have been used as in [SH05], the sum-rate would have been higher. If additionally the users would have been stationary, the Keep algorithm could be expected to converge to a local sum-rate optimum for the beams on $Q_{\text{keep}}$ of the clusters.

9.6 Summary

For the scenario with low number of users that has been studied here, multi-beam opportunistic beamforming performs worse than single-antenna transmission. This is due to the strong inter-beam interference of the scheduled users. We observe that keeping successful beams over several blocks can give a significant performance boost. The Keep scheme outperformed single-antenna transmission for above 5 users, with larger gain if multiple receive antennas were used. However, the results show that the feasibility of multi-beam transmission is limited to the cases where the system load is high. The feasibility regions for single- and multi-beam transmission is investigated in more detail in [JSO07].
Figure 9.3: The average sum-rate per sector as a function of the number of users. The two lower curves are for one receive antenna ($N_r = 1$), whereas the two upper curves are for four receive antennas ($N_r = 4$).
Figure 9.4: The average sum-rate per sector for the *Keep* scheme with $N_t = 2$ and $N_r = 1$, as a function of the ratio of beams that are kept from block to block. For $Q_{\text{Keep}} = 0$, the scheme is identical to *Rand*. There are in average 6 users per sector. The three curves represent users having uniformly distributed speeds between 0 km/h and 10, 50 and 100 km/h, respectively.
Chapter 10

Reduced Feedback SDMA Based on Subspace Packings

10.1 Overview

This chapter contains:

- A description of a feedback and SDMA scheme based on subspace packings. The scheme is presented in a single-carrier context, but it is well suited for clustered OFDMA due to its low feedback requirements and low complexity. Two benefits of the scheme are:
  1. The feedback implicitly contains information about the spatial compatibility of the users.
  2. The feedback supportable rates are based on the post-scheduling SINR.
- A discussion on the important concept of allocation outage and the derivation of the allocation outage probability.
- The introduction of non-orthogonal Grassmannian subspace packings.
- Numerical system simulation results comparing various flavors of the proposed method with opportunistic SDMA and optimal linear precoding.

10.2 Introduction

In the previous chapters, multi-antenna transmission using random beams was considered. One advantage was that the explicit CSI was not needed at the transmitter, but only the resulting supportable rate, and, in the case of multiple beams, the beam index.

A different approach to provide partial CSI at the transmitter with low feedback load is to let the users vector quantize their channel states, or equivalently optimal
beamformers, using e.g. the generalized Lloyd’s algorithm [XG06] or random vector quantization [AYL07]. The user would feed back an index in the quantization codebook together with the supportable rate. Another way to generate the quantization codebook is to use Grassmannian line packings [MSEA03]. A Grassmannian packing is a collection of subspaces in a higher dimensional space. An optimal packing is such that it maximizes some distance measure between the two closest subspaces in the packing. For the intuitive case where the channel vector is real and three-dimensional, the optimal line packing is a collection of one-dimensional subspaces, i.e. lines, in $\mathbb{R}^3$ that pass through the origin and maximize the angle between the two closest lines. Clearly, Grassmannian line packings are promising candidates for the quantization of the i.i.d. Rayleigh fading vector channel [MSEA03, LHS03]. In the case of spatially correlated Rayleigh fading channels, the packing can be prewhitened as in [LH06]. A major disadvantage with Grassmannian packings is the lack of analytical results for most cases, but numerical methods to generate good packings exist [CHS96, Tro04].

The step from Grassmannian single-beam beamforming, i.e. line packings, to Grassmannian spatial multiplexing was taken in [LH05]. In MIMO channels, several data streams can be simultaneously transmitted to a user by precoding with a unitary matrix (instead of a vector in the single-stream case) at the transmitter. The optimal unitary matrix is quantized using a Grassmannian subspace packing.

Herein, the Grassmannian subspace packing approach is applied to the multiuser scenario. As in [LH05], unitary matrices from Grassmannian subspace packings are used for transmitter precoding. The major difference is that we consider SDMA, where one user per beam (or column in the unitary beamforming matrix) can be allocated. There is no cooperation between the receive antennas (users), and only a subset of the receiver antennas (users) are used at each time. We show that, contrary to single-user Grassmannian beamforming or spatial multiplexing, a denser packing often does not yield better downlink performance due to allocation outages. Additionally, DFT-based and non-orthogonal Grassmannian subspace packings are considered.

### 10.3 System Model

We consider the downlink of a cellular multiuser FDD system. The base-station with $N_t$ antennas communicates with $K$ single-antenna users. The wireless channels are assumed to be flat i.i.d. Rayleigh fading and quasi-stationary, i.e. constant during each block. One block is also the resource allocation period. We do not consider OFDMA explicitly here. However, the proposed scheme may be applied to each OFDMA sub-carrier of cluster. The application of the scheme to OFDMA makes low feedback overhead and complexity even more crucial.

The received signal for user $k$ at symbol $t$, $y_k(t) \in \mathbb{C}$, is

$$y_k(t) = h_k^T(t)x(t) + z_k(t),$$
where $h_k(t) \in \mathbb{C}^{N_t \times 1}$ is the MISO channel of the $k^{th}$ user, $x(t) \in \mathbb{C}^{N_t \times 1}$ is the transmitted signal from the base-station, $z_k(t) \in \mathbb{C}$ is AWGN with variance $\sigma_z^2$ and $(\cdot)^T$ denotes transpose. The quasi-stationary assumption means that $h_k(t)$ is constant for all $t$ within a block. In the following, the time-index $t$ is omitted for brevity.

The transmitted signal, $x$, is assumed to be

$$x = \frac{1}{\sqrt{B}} W s = \frac{1}{\sqrt{B}} \sum_{i=1}^{B} w_i s_i,$$

where $W = [w_1 \cdots w_B] \in \mathbb{C}^{N_t \times B}$ is the beamforming matrix and $s = [s_1 \cdots s_B]^T \in \mathbb{C}^{B \times 1}$ is the vector of i.i.d. symbols. The variable $B$ denotes the number of simultaneously transmitted symbols as well as the number of used beams. The $i^{th}$ data symbol $s_i \in \mathbb{C}$ is transmitted using the $i^{th}$ beamforming vector $w_i \in \mathbb{C}^{N_t \times 1}$. The number of simultaneous streams $B$ is henceforth called the SDMA factor. The case $B = 1$ corresponds to single-user beamforming (TDMA). The average transmit power is constrained by setting $\mathbb{E}[|s_i|^2] = 1$ and $\|w_i\| = 1$. This means that $\mathbb{E}[\|x\|^2] = 1$. We assume that each user $k$ accurately estimates the channel vector $h_k$. The transmission of training symbols has to be done without applying any beamforming matrix, to enable the estimation of $h_k$.

Now, assume that symbol $m$ is dedicated to user $k$. The received SINR is given by

$$\gamma_k = \frac{|h_k^T w_m|^2}{B \sigma_z^2 + \sum_{n \neq m} |h_k^T w_n|^2}.$$\vspace{0.5cm}

The user SINR is maximized if $h_k$ is co-linear with $w_m$ and if $h_k$ is orthogonal to $w_n$ for all $n \neq m$. Then, the SINR is $\tilde{\gamma}_k = \frac{|h_k|^2}{B \sigma_z^2}$. This means that users are perfectly spatially compatible if their channel vectors are orthogonal. In practice, this is rarely the case. Instead, the beams are typically designed to balance the inter-beam interference to fulfill QoS-constraints of the scheduled users, see e.g. [SB05].

Furthermore, it is assumed that the system supports a finite set of adaptive modulation and coding (AMC) levels. Given the error rate constraint of user $k$, the supportable AMC level is uniquely determined by the mapping function $\Gamma_k(\gamma_k)$, which is assumed to be known.

### 10.4 Reduced Feedback SDMA Based on Subspace Packings

Assume that a set of orthogonal beamforming matrices, $\mathcal{W} = \{W_1, \ldots, W_N\}$, is a priori known at both the base-station and at the users. The columns of each matrix in $\mathcal{W}$ span a $B$-dimensional subspace of $\mathbb{C}^N_t$. $\mathcal{W}$ is called a subspace packing. Each beamforming matrix consists of $B$ orthogonal columns (beams/vectors), $W_i = [w_i^1 \cdots w_i^B]$.\vspace{0.5cm}

**Theorem:** Assume that a set of orthogonal beamforming matrices, $\mathcal{W} = \{W_1, \ldots, W_N\}$, is a priori known at both the base-station and at the users. The columns of each matrix in $\mathcal{W}$ span a $B$-dimensional subspace of $\mathbb{C}^N_t$. $\mathcal{W}$ is called a subspace packing. Each beamforming matrix consists of $B$ orthogonal columns (beams/vectors), $W_i = [w_i^1 \cdots w_i^B]$.\vspace{0.5cm}

**Proof:**

1. **Step 1:** Consider a set of orthogonal beamforming matrices, $\mathcal{W} = \{W_1, \ldots, W_N\}$, which are known to both the base-station and the users.

2. **Step 2:** Each beamforming matrix $W_i$ consists of $B$ orthogonal columns (beams/vectors), $W_i = [w_i^1 \cdots w_i^B]$.

3. **Step 3:** Let $\mathcal{W}$ span a $B$-dimensional subspace of $\mathbb{C}^N_t$. This means that $\mathcal{W}$ supports $B$ simultaneous streams.

4. **Step 4:** Assume that the system supports a finite set of adaptive modulation and coding (AMC) levels. Given the error rate constraint of user $k$, the supportable AMC level is uniquely determined by the mapping function $\Gamma_k(\gamma_k)$, which is assumed to be known.

5. **Step 5:** The user SINR is maximized if $h_k$ is co-linear with $w_m$ and if $h_k$ is orthogonal to $w_n$ for all $n \neq m$. Then, the SINR is $\tilde{\gamma}_k = \frac{|h_k|^2}{B \sigma_z^2}$. This means that users are perfectly spatially compatible if their channel vectors are orthogonal. In practice, this is rarely the case. Instead, the beams are typically designed to balance the inter-beam interference to fulfill QoS-constraints of the scheduled users, see e.g. [SB05].

6. **Step 6:** Furthermore, it is assumed that the system supports a finite set of adaptive modulation and coding (AMC) levels. Given the error rate constraint of user $k$, the supportable AMC level is uniquely determined by the mapping function $\Gamma_k(\gamma_k)$, which is assumed to be known.

**Conclusion:**

The reduced feedback SDMA based on subspace packings allows for efficient resource allocation and interference mitigation in multi-user MISO systems, enabling the support of multiple users with varying QoS requirements.
The proposed beamforming and scheduling schemes work on a block basis. Before each block:

1. Each user $k$ finds the overall best beam index,

$$
(i, j)_k = \arg \max _{(i,j)} \gamma _k ((i,j)) \quad \text{(10.1)}
$$

This means that, when selecting the best beam, the user assumes that all other beams in $W_i$ (other than $w^j_i$) will be used for transmission to other users.

2. Each user feeds back the beam index $(i, j)_k$ and the corresponding supportable AMC level $\Gamma_k (\gamma _k ((i,j)_k))$.

3. According to a scheduling metric, the base-station chooses a subspace $i^*$ to be used during the next block, i.e. the base-station will use $W_{i^*}$ as beamforming matrix. For each beam in $W_{i^*}$, a user is scheduled with the fed back AMC level. For example, if the base-station uses maximum throughput scheduling, the subspace is chosen according to

$$
i^* = \arg \max _i \sum _{j=1} ^B \max _k \Gamma_k (\gamma _k ((i,j)_k)). \quad \text{(10.2)}
$$

To summarize, each user finds its best overall beam from all subspaces. Then, the user computes the supportable rate assuming it is scheduled on its best beam and other users are scheduled on the other beams in the same subspace. Finally, the base-station selects a subspace according to some scheduling metric. Only users that fed back beam-indices for that subspace are considered for scheduling.

It is important to note that the fed back AMC level, $\Gamma_k$, includes the post-scheduling inter-beam interference, regardless of which users are scheduled on the other beams. This is notable since $\Gamma_k$ is computed before the feedback and the scheduling take place.

The scheduling is simplified since the feedback implicitly shows which users are spatially compatible. It is likely that two users that fed back two different beam-indices within the same subspace have nearly orthogonal channels, since all beams within one subspace are orthogonal.

The reduction in feedback is due to the quantization of the optimal beamforming vector. It is indexed with the $(i, j)$-pair using $\log _2 (BN)$ bits. A couple of additional

---

1For $B = N_t$, (10.1) is equivalent to $\arg \max _{(i,j)} |h^T_k w^j_i|$. In Theorem 10.3 in Appendix 10.A, we upper bound the relative loss in SINR from choosing the $(i, j)$ that maximizes $|h^T_k w^j_i|$ instead of (10.1).
10.4. REDUCED FEEDBACK SDMA BASED ON SUBSPACE PACKINGS

bits are needed for the AMC level. To reduce the feedback further, Section 11 presents a method to order the $BN$ beams in a graph.

There is an important trade-off in the selection of the number of subspaces $N$. The more subspaces and beams to choose among (high $N$), the more likely that the beam a user chooses is close to optimal. This relation is illustrated in Figure 10.2 in Section 10.6. However, the feedback increases with increasing $N$. More importantly, for a limited number of users, the probability that the base-station can not allocate users to all $B$ beams in any subspace increases with $N$. If the base-station can not allocate $B$ beams, there is a significant performance loss. This is due to the fact that the allocated users assume in their AMC level computation that the other beams are used for transmission to other users, creating inter-beam interference. Theorem 10.2 presents the probability of this event, which is called allocation outage.

**Definition 10.1** (Allocation outage). Consider a system where $K$ users feed back one beam index each out of $NB$ possible. The $NB$ beams are divided into $N$ disjoint sets of $B$ beams each. An allocation outage is the event that there exists no set where all $B$ beam indices have been fed back by at least one user.

**Theorem 10.2.** Consider the allocation outage event, as in Definition 10.1. Additionally assume that the users choose beam indices independently and with equal probability. Then,

$$P_{ao} \triangleq \Pr (\text{allocation outage}) = 1 + \sum_{n=1}^{N} (-1)^{n} \binom{N}{n}$$

$$+ \frac{1}{(NB)^{K}} \sum_{n=1}^{N} \sum_{m=1}^{nB} (-1)^{n+m} \binom{N}{n} \binom{nB}{m} (NB - m)^{K}$$

(10.3)

**Proof.** See Appendix 10.B.

The result in Theorem 10.3 is useful in understanding the tradeoff between having many beams in the packing ($NB$) and a high SDMA factor ($B$) on one hand and the probability of undesired allocation outages on the other. The outage probability (10.3) for a number of packing dimensions is displayed in Figure 10.1. For few users (low $K$), higher SDMA factor ($B$) than 2 is not suitable due to the high allocation outage probability. The scheduler would rarely find a subspace where all beams could be used. This would result in the number of spatial streams less than $B$ as well as lower fed back supportable rate for the scheduled users. For 10 users, SDMA factor 2 and 6 subspaces ($K = 10$, $B = 2$ and $N = 6$), the scheduler can with more than 90% probability find a subspace where 2 users can be scheduled, which seems to be a suitable operating point. For the scenario with many active users, it is clearly suitable to use a higher SDMA factor, enabling many
Figure 10.1: Allocation outage probability $P_{ao}$ as a function of the number of users $K$. The curves represent packings with in total of 12 and 24 beams ($NB = 12$ and $NB = 24$). The beams are grouped in subspaces (beamforming matrices) of dimensions $B = 2$, $B = 3$ and $B = 4$.

spatial streams with high probability. Note that $B > N_t$ is not possible without inter-beam interference.

For a given $K$, the optimal $N$ and $B$ in terms of sum rate is an open problem. A packing with high $NB$ may enable the users to find a beam close to the optimal one, but on the other hand, increases the allocation outage probability. However, suitable packing dimensions may be found from simulations.

Figure 10.1 suggests a scheme that would adapt to the number of active users, which can be assumed to be a slowly varying parameter. Both the base-station and the users would have a set of packings with various dimensions $N$ and $B$. Depending on the number of users, the base-station would broadcast the $N$ and $B$ of the currently used packing.

In the presentation above, single antenna users have been assumed. However, the proposed scheme is compatible with most multi-antenna receiver structures, since the MIMO channel, the transmit beamformer, and the interfering beamformers are known at all receiving users. The structure proposed here does not, however, allow for the transmission of several beams to the same user.
10.5 Subspace Packings

In this section, the design of $W$ is discussed. In Section 10.4, the proposed subspace packing based SDMA scheme was described. It relies on users finding a beamforming vector from a finite predefined set that in some sense is *close to* the optimal beamforming vector. At the same time, the vectors spanning each subspace, or the columns of each $W_i$, should be orthogonal, enabling inter-beam interference free SDMA.

When constructing packings for SDMA, there is a fundamental tradeoff between 1) good vector quantization properties and 2) the orthogonality of vectors that are to be used simultaneously. For the i.i.d. Rayleigh fading vector channel, the Grassmannian line packing is optimal in terms of quantization. However, the line packing in general does not contain any orthogonal vectors, thereby penalizing SDMA. On the other hand, the optimal subspace packing gives orthogonal matrices that are maximally separated in chordal distance (see Section 10.5). For the optimal subspace packing, the ensemble of basis vectors from all subspaces, i.e. the columns of all unitary matrices, do not in general quantize the vector channel as well as the optimal line packing. This tradeoff is numerically studied in Section 10.6.

Grassmannian and DFT-based packings, which are known from literature, are presented below. However, they have not been used in the context of SDMA previously. Additionally, we propose non-orthogonal Grassmannian subspace packings that quantize the vector channel optimally, but introduce some extra inter-beam interference.

Grassmannian Subspace Packings

The distance measure between subspaces in $\mathbb{C}_t^N$ used here is the *chordal distance*. Other measures are also feasible, but the chordal distance is chosen due to its simplicity and that it yields the most symmetrical packings [CHS96, Tro04]. The chordal distance between the subspaces spanned by the $B$ orthonormal columns of $W_i$ and $W_j$ is given by

$$
\text{dist} (W_i, W_j) = \sqrt{B - \|W_i^H W_j\|^2_F},
$$

(10.4)

where $(\cdot)^H$ denotes conjugate transpose. The complex Grassmannian manifold $\mathcal{G}(B, \mathbb{C}_t^N)$ is the collection of all $B$-dimensional subspaces in $\mathbb{C}_t^N$. A Grassmannian packing is a finite set of subspaces from the manifold, spanned by the unitary matrices in $W$. An optimal packing is such that the minimum distance between any two subspaces is maximized, i.e. for $i \neq j$

$$
\hat{W} = \arg \max_{W} \min_{W_i, W_j \in W} \text{dist} (W_i, W_j).
$$

(10.5)

For $B = 1$, the problem reduces to the packing of lines in $\mathbb{C}_t^N$ that pass through the origin, or equivalently, to the packing of points on the unit sphere in $\mathbb{C}_t^N$. For
the i.i.d. Rayleigh fading channel, the optimal Grassmannian line packing is optimal in terms of, for instance, outage probability [MSEA03]. In general, there is no analytic solution to (10.5). However, for moderate problem dimensions, there are numerical methods which can provide near-optimal packings. Grassmannian packing construction is discussed in more detail in Section 10.6. Numerically constructed near-optimal Grassmannian packings are henceforth called Grassmannian packings.

**DFT-based Subspace Packings**

In [MSEA03], it was shown that the columns of the unitary space-time constellations in [HMR+00] work well as beamformers for the i.i.d. Rayleigh fading channel. The first elements of the columns of an $NB$-dimensional DFT matrix are taken as beamforming vectors.

$$w_l = \frac{1}{\sqrt{N_t}} \begin{pmatrix} e^{j2\pi(l-1)/(NB)} \\ \vdots \\ e^{j2\pi(N_t-1)(l-1)/(NB)} \end{pmatrix} \quad \text{for } l = 1, \ldots, NB.$$ 

Fortunately, there are orthogonal vectors in the DFT packing, which naturally span the $B$-dimensional subspaces in the same fashion as Grassmannian subspace packings. A rearrangement of the DFT vectors can create the necessary unitary matrices for subspace packings of higher dimension than one.

**Non-orthogonal Grassmannian Subspace Packings**

The trade-off between channel vector quantization distortion and line orthogonality was discussed in the beginning of this section. On one side, we have the Grassmannian subspace packings, with orthogonal vectors spanning the subspaces, but which lack in quantization performance. On the other side, we have the Grassmannian line packings with optimal quantization performance. A natural alternative to the Grassmannian subspace packings is to construct packings with optimal quantization performance, but with vectors spanning the subspaces that are non-orthogonal. We have constructed such packings, based on Grassmannian line packings, where the vectors have been greedily grouped to yield near-orthogonality within the subspaces. Such a packing will on average outperform the Grassmannian subspace packing in terms of channel vector quantization, but will not enable SDMA without inter-beam interference. The greedy non-orthogonal packing construction is briefly described in Table 10.1. From a set of $NB$ vectors, it creates $N$ matrices with $B$ columns each.

In [ZPN06], a multiuser transmission scheme based on a non-orthogonal Grassmannian line packing was proposed. It differs in several aspects. Firstly, it is applied in the framework of opportunistic SDMA, as in [SH05], i.e. only one beamforming
10.6. **NUMERICAL RESULTS**

1. Put all vectors from a Grassmannian line packing in the set $W_{vec}$.
2. Find the two vectors in $W_{vec}$ with lowest inner product. Remove them from $W_{vec}$ and let them be the first columns in a new matrix.
3. If the number of columns in the new matrix equals $B$, save the matrix and go to step 5. Otherwise go to step 4.
4. Find the vector in $W_{vec}$ with the lowest average inner product with the columns in the new matrix. Remove it from $W_{vec}$ and append it as a column in the new matrix. Go to step 3.
5. If $W_{vec}$ is not empty, go to step 2.
6. Now, $N$ matrices with $B$ columns each have been created.

**Table 10.1:** The greedy algorithm to construct non-orthogonal packings.

matrix is considered. Here, we generate a packing with $N$ matrices. Secondly, no construction algorithm is given.

---

**10.6 Numerical Results**

**Packing Study**

The construction of optimal Grassmannian packings is still an open problem. Several numerical methods to construct packings have been proposed, e.g. [CHS96, Tro04]. In this chapter, we have used the alternating projections technique for complex-valued subspace packings in [Tro04]. For the cases tabulated in [Tro04], very similar minimum distances were obtained. In the following results, the number of transmit antennas is 4 ($N_t = 4$).

A measure for the vector quantization performance of a packing $W$ is the average distortion [LH06],

$$d(W) = E_h \left[ \|h\|^2 - \max_{1 \leq i \leq N} \max_{1 \leq j \leq B} |h^T w_i^j|^2 \right],$$

(10.6)

where $h = \frac{g}{\|g\|}$ is an isotropically distributed unit vector and $g \in \mathbb{C}^{N_t \times 1}$ is i.i.d. zero-mean complex Gaussian.

The average distortion for a number of packings has been numerically evaluated. The results are displayed in Figure 10.2. It is intuitive that the distortion decreases with the number of beams in the packing. For Grassmannian packings with 2-dimensional subspaces ($B = 2$), the distortion is slightly higher than for line packings ($B = 1$). This is due to the fact that for $B = 2$, the subspaces (planes)
are isotropically spread out, but not the ensemble of lines (i.e. the vectors) that span the subspaces. For the DFT-based packings for a given $NB$, the quantization performance is identical for $B = 1$ and $B = 2$. This is due to the fact that the packings for $B = 1$ and $B = 2$ contain the same vectors, but for $B = 2$ arranged in orthogonal pairs (c.f. Section 10.5). To summarize, the Grassmannian packings have a superior quantization performance for i.i.d. Rayleigh fading channel vectors, especially for packings with many beams.

**System Study**

A system level simulation study has been conducted in order to evaluate the performance of packings previously discussed. One cell with one base-station and several single-antenna users has been simulated according to the system model in Section 10.3. The base-station is equipped with 4 transmit antennas ($N_t = 4$). The maximum throughput scheduler is used. For each block, it selects the beam/user combination that yields the highest sum-rate (c.f. (10.2)). The average SNR is 5 dB. The simulated channels are generated using Jake’s model with block fading.

The Grassmannian, DFT and non-orthogonal packings, as described in Section 10.5, are used in reduced feedback SDMA, as described in Section 10.4. In this study, the fed back supportable rate is not quantized.
As an additional comparison, opportunistic SDMA as in [SH05] is used (c.f. Chapter 9). This technique randomly generates one set of orthogonal beams in each block. It relies on the multiuser diversity to provide the scheduler with users that are compatible with the generated beams. This technique can be seen as a packing with only one subspace, that is randomly re-generated for each block. The Grassmannian, DFT and non-orthogonal packings with multiple subspaces exploit both the multiuser diversity and the diversity from having many subspaces. Another advantage with having packings with non-random predefined beamforming matrices is that the users may track only the underlying channel, whereas with random beams they need to estimate the discontinuous aggregate channel (i.e. random beam times channel).

Finally, the optimal linear downlink multiuser beamformer with uniform power allocation is used as a comparison. The transmit beamformers are given by the dual uplink MMSE receiver beamformers [TV05],

$$w_i = (\sigma_n^2 I + \frac{1}{\sqrt{B}} \sum_{j=1 \atop j \neq i}^B h_j h_j^H)^{-1} h_i.$$

In each block, the same users are scheduled for the LMMSE scheme as for the Grassmannian subspace packing based scheme. This guarantees that they are spatially well separable. Hence, the gap between the LMMSE and the Grassmannian
performance is mainly due to the feedback quantization distortion. Additionally, the allocation outage events (less than $B$ users scheduled) do not penalize the LMMSE scheme.

In Figure 10.3, the average sum rate as a function of the number of active users is displayed. The SDMA factor, $B$, is 2. The packing-based methods use packings with 32 subspaces ($N = 32$). The packing-based methods outperform opportunistic SDMA at the cost of 5 extra feedback bits per user and block. The DFT-based packing results in slightly worse performance than the Grassmannian. The non-orthogonal SDMA performs as well as orthogonal Grassmannian SDMA, since the loss of orthogonality from grouping 64 vectors in groups of two is low.

In Figure 10.4, the SDMA factor, $B$, is 3, but the number of subspaces, $N$, is only 4. These parameters are less beneficial for the non-orthogonal packing, since it is not possible to find 4 groups of 3 nearly orthogonal beams from a Grassmannian line packing of these dimensions. Even the opportunistic SDMA outperforms the non-orthogonal scheme. Again, the Grassmannian subspace packing outperforms the other packings. However, the performance gap to the LMMSE scheme based on perfect CSI is significant.

Figure 10.5 illustrates the relation between rate, SDMA factor and number of users. The figure shows the average sum rate for Grassmannian subspace packings with subspace dimensions 1, 2 and 3. The total number of beams, $NB$, is 12. For SDMA factors greater than 1 ($B = 2$ and $B = 3$) to perform well, many users in
the system are required. Otherwise, allocation outages will occur frequently. Even though an allocation outage does not occur, the SINR penalty from scheduling a user with a channel vector poorly aligned to the beam is greater for higher SDMA factors than for single-beam transmission ($B = 1$). This is due to the resulting non-orthogonality to the interfering beams. The optimal switching points for the SDMA factor depends on the SNR. For higher SNR, the switching points are shifted to higher number of users (to the right in Figure 10.5). A framework for finding the optimal $B$, given an SNR and the number of users, for opportunistic TDMA and SDMA is given in [JSO07].

### 10.7 Conclusions

We have proposed a scheme for reduced feedback SDMA based on subspace packings. The scheme is based on a codebook of beams (a subspace packing), common to the base-station and the users. The scheme allows all users to compute their post-scheduling SINR already before the CSI feedback. Only a few bits feedback per user and coherence time is necessary. The feedback also provides information about the spatial compatibility of the users. We highlighted a fundamental property of the scheme through the allocation outage probability. The reduced feedback SDMA scheme based on subspace packings shows potential to cover some of the
middle ground between opportunistic SDMA and SDMA based on perfect CSI. Several parameters of the method should be tuned according to the specific scenario. For instance, a scenario with many users would benefit from a packing with many subspaces and with a high SDMA factor. A scenario with few user on the other hand would benefit from a packing with fewer subspaces of lower dimension, i.e. fewer simultaneous spatial streams. The Grassmannian subspace packings showed best performance of the low-rate feedback schemes in the numerical comparison for the studied i.i.d. Rayleigh fading scenario.

Grassmannian subspace packings are well suited for the quantization of i.i.d. Rayleigh fading channels. However, in many practical systems, there is correlation between the transmitting antennas. For Grassmannian beamforming (using line packings), this can be handled by prewhitening the packing, as in [LH06]. Grassmannian subspace packings can not directly be prewhitened in the same manner, since the prewhitening destroys the orthogonality between the vectors spanning each subspace. A possible approach left for future research is to prewhiten a Grassmannian line packing, and based on that, construct a non-orthogonal subspace packing, as in Section 10.5.

The scheme proposed in this chapter is presented in a single-carrier context. However, it could be applied to each sub-carrier or cluster in an OFDMA system. The low feedback and complexity of the scheme makes it suitable for an extension to OFDMA. The selective and adaptive reduced feedback scheme in Section 3.3 could be applied by merging the concepts of scheduling outage and allocation outage.
Appendix 10.A  The SINR Penalty When Using Inner-product as Beam-selection Criterion

In many cases, it is sufficient for the users to find the \((i, j)_k\) that maximizes \(|h_k^T w_j^i|^2\) instead of the \((i, j)_k\) that maximizes \(\gamma_k((i, j))\), as in (10.1). Unfortunately, this may give a sub-optimal choice. The processing delay induced by the computation of (10.1) adds to the delay between channel estimation and data transmission. This delay should be kept to a minimum, otherwise the feedback information will be outdated. For large packings, a large number of divisions can be saved by considering the inner product instead of the SINR when selecting a beam index.

The maximum loss in SINR by using the maximum inner product beam selection criterion instead of the maximum SINR criterion is quantified in the following result.

**Theorem 10.3.** Let

\[
(a, b)_k = \arg \max_{(i, j)} \gamma_k((i, j)) \tag{10.7}
\]

\[
(c, d)_k = \arg \max_{(i, j)} |h_k^T w_j^i|^2. \tag{10.8}
\]

denote the optimal and sub-optimal beam choice for the \(k\)th user, respectively. Then, for \(1 < B < N_t\), the relative loss in SINR is bounded by

\[
\frac{\gamma_k((a, b)_k)}{\gamma_k((c, d)_k)} \leq 1 + \frac{\|h_k\|^2 (1 - \alpha)}{\sigma_z^2 B}, \tag{10.9}
\]

where \(\frac{\|h_k\|^2}{\sigma_z^2}\) is the instantaneous SNR of the user and \(\alpha\) represents the worst-case inner product of the packing \(W\),

\[
\alpha = \min_{u^T \in C_1 \times N_t} \max_{(i, j)} |u w_j^i|^2. \tag{10.10}
\]

Equality in (10.9) is obtained when \(\frac{h_k}{\|h_k\|}\) equals the minimizing \(u\) in (10.10). For \(B = 1\) and \(B = N_t\), \(\gamma_k((a, b)_k) = \gamma_k((c, d)_k)\).

**Proof.** Consider user \(k\) with channel \(h_k\). The optimal (maximum SINR from (10.7)) beam from the set \(W\) is denoted \(w_a^b\). The beam with maximum inner product (as in (10.8)) is denoted \(w_d^b\). The case \(B = 1\) is trivial, since there is no inter-beam interference. Hence, \(w_a^b = w_c^d\).

Now, consider the case \(B > 1\). If \(B = N_t\), the channel vector can always be expressed in the coordinate system spanned by the orthonormal columns of \(W_i\) for all \(i\), i.e. \(h_k = \sum_{j=1}^{B} \beta_j^i w_j^i\), where \(\beta_j^i = h_k^T w_j^i\). Thus,

\[
\gamma_k((i, j)_k) = \frac{|\beta_j^i|^2}{B\sigma_z^2 + \sum_{n \neq j} |\beta_j^n|^2} = \frac{|\beta_j^i|^2}{B\sigma_z^2 + \|h_k\|^2 - |\beta_j^i|^2},
\]
CHAPTER 10. REDUCED FEEDBACK SDMA BASED ON SUBSPACE PACKINGS

since \( \| h \|_2^2 = \sum_{j=1}^{B} |\beta_j|^2 \) due to the orthonormality of the basis vectors. It is clear that

\[
(c, d)_k = \arg \max_{(i, j)} |\beta_j|^2 = \arg \max_{(i, j)} \gamma_k ((i, j)) = (a, b)_k,
\]

when \( B = N_t \).

If \( B < N_t \), one additional unit norm vector \( v_i \), orthogonal to \( W_i \), is in general needed to express \( h_k \), with

\[
v_i = \frac{h_k - \sum_{j=1}^{B} \beta_j w_i^j}{\| h_k - \sum_{j=1}^{B} \beta_j w_i^j \|}.
\]

The channel vector can then be written as a linear combination of the basis vectors as

\[
h_k = \sum_{j=1}^{B} \beta_j w_i^j + \beta_{B+1} v_i,
\]

where \( \beta_{B+1} = h_k^T v_i \). Due to the orthonormality of the basis vectors for all \( i \),

\[
\| h_k \|_2^2 = \sum_{j=1}^{B} |\beta_j|^2 + |\beta_{B+1}|^2 = |\beta_j|^2 + \sum_{n \neq j} |\beta_n|^2 + |\beta_{B+1}|^2.
\]

Note that

\[
0 \leq \sum_{n \neq j} |\beta_n|^2 = \| h_k \|_2^2 - |\beta_j|^2 - |\beta_{B+1}|^2 \leq \| h_k \|_2^2 (1 - \alpha), \tag{10.11}
\]

where \( \alpha \) is defined in (10.10). The minimizing \( u \) in (10.10) is the worst-case unit vector, i.e. the inner product with the closest vector in the packing is minimal. Since \( \alpha \) only depends on \( W \), it can be numerically found offline. Equality in the upper bound in (10.11) is obtained when \( h_k \) is co-linear with the minimizing \( u \) in (10.10) and also in the subspace spanned by \( W_i \), i.e. \( \beta_{B+1} = 0 \). Equality in the lower bound in (10.11) is obtained when \( h_k \) is orthogonal to the interference vectors \( \{ w_n^a \}_{n \neq j} \), i.e. \( h_k = \beta_j^j w_i^j + \beta_{B+1} v_i \). The relative SINR loss is

\[
\frac{\gamma_k ((a, b)_k)}{\gamma_k ((c, d)_k)} = \frac{|\beta_a|^2}{B \sigma^2 + \sum_{n \neq a} |\beta_n|^2} \cdot \frac{B \sigma^2 + \sum_{n \neq d} |\beta_n|^2}{|\beta_c|^2}. \tag{10.12}
\]

The worst case is characterized by

- \( |\beta_a|^2 = |\beta_c|^2 \), since \( |\beta_d|^2 = \max_{(i, j)} |\beta_j|^2 \geq |\beta_a|^2 \), which means that two vectors in the packing had the same inner product and the wrong one was chosen,

- \( \sum_{n \neq b} |\beta_n|^2 = 0 \), i.e. \( h_k \) is orthogonal to all interference vectors \( \{ w_n^a \}_{n \neq b} \) and...
10.B. PROOF OF THEOREM 10.2

- \( \sum_{n \neq d} |\beta_n|^2 = \|h_k\|^2 (1-\alpha) \), i.e. \( h_k \) is spanned by \( W_c \) and \( |h_k^T w_c|^2 \) is minimal.

Hence,

\[
\gamma_k ((a, b)_k) \leq \frac{B \sigma_2^2}{\gamma_k ((c, d)_k)} = 1 + \frac{\|h_k\|^2 (1-\alpha)}{B}. \tag{10.15}
\]

\( \Box \)

Appendix 10.B Proof of Theorem 10.2

In the following proof, we will use Poincaré’s formula\(^2\) twice [Fel62, page 89].

**Theorem 10.4** (Poincaré’s formula). Let \( \{E_m\}_{m=1}^{b} \) be \( b \) events. Then,

\[
\Pr \left( \bigcup_{m=1}^{b} E_m \right) = \sum_{m=1}^{b} (-1)^{m-1} S_m, \tag{10.13}
\]

where \( \bigcup \) denotes the union of events and

\[
S_m = \sum_{1 \leq j_1 < \cdots < j_m \leq b} \Pr (E_{j_1} \cap \cdots \cap E_{j_m}), \tag{10.14}
\]

where \( \cap \) denotes AND.

In order to derive the allocation outage probability, we first prove the following Lemma.

**Lemma 10.5.** Consider a set \( \mathcal{A} \) with \( a \) elements and a subset \( \mathcal{B} \subseteq \mathcal{A} \) with \( b \leq a \) elements. Independenty, with uniform probability and with replacement, \( k \) elements are taken from \( \mathcal{A} \). The taken elements, with duplicates removed, constitute the set \( \mathcal{K} \subseteq \mathcal{A} \), with cardinality \( |\mathcal{K}| \leq \min(k, a) \). Then,

\[
\bar{P}_k(a, b) \triangleq \Pr (\mathcal{B} \subseteq \mathcal{K}) = 1 + \frac{1}{a^k} \sum_{m=1}^{b} (-1)^m \binom{b}{m} (a-m)^k. \tag{10.15}
\]

**Proof.** First, define the \( b \) dependent events \( \{E_m\}_{m=1}^{b} \), where \( E_m \) is the event that element \( m \) in \( \mathcal{B} \) is not in \( \mathcal{K} \). Then,

\[
\bar{P}_k(a, b) = \Pr (\mathcal{B} \subseteq \mathcal{K}) = 1 - \Pr \left( \bigcup_{m=1}^{b} E_m \right). \tag{10.16}
\]

First note that

\[
\Pr (E_{j_1} \cap \cdots \cap E_{j_m}) = \frac{(a-m)^k}{a^k}
\]

\(^2\)It is sometimes attributed to de Moivre or Sylvester and is also known as the inclusion-exclusion principle and the sieve principle.
for all distinct $j_1, \ldots, j_m$. Therefore, $S_m$ in Theorem 10.4 can be simplified to

$$S_m = \frac{(a - m)^k}{a^k} \sum_{1 \leq j_1 < \cdots < j_m \leq b} \frac{1}{(m)} = \frac{(a - m)^k}{a^k} \binom{b}{m},$$

since the number of ways the $m$ integer indices can be chosen so that $1 \leq j_1 < \cdots < j_m \leq b$ equals $\binom{b}{m}$. Finally, applying Theorem 10.4 to the union of events in (10.16) gives

$$\bar{P}_k(a, b) = 1 + \frac{1}{a^k} \sum_{m=1}^{b} (-1)^m \binom{b}{m} (a - m)^k.$$

Equipped with Lemma 10.5, the allocation outage probability can be determined. First, we define the $N$ dependent events $\{D_n\}_{n=1, \ldots, N}$, where $D_n$ is the event that all beam indices in the $n^{th}$ set have been fed back by at least one user. Then,

$$P_{ao} = 1 - \Pr \left( \bigcup_{n=1}^{N} D_i \right),$$

where $\bigcup$ denotes the union of events. As in the proof of Lemma 10.5, using Theorem 10.4 and the fact that

$$\Pr (D_{j_1} \cap \cdots \cap D_{j_n}) = \bar{P}_K(NB, nB)$$

for all distinct $j_1, \ldots, j_n$ and where $\bar{P}_k(a, b)$ is defined in Lemma 10.5 gives

$$P_{ao} = 1 + \sum_{n=1}^{N} (-1)^n \binom{N}{n} \bar{P}_K(NB, nB)$$

$$= 1 + \sum_{n=1}^{N} (-1)^n \binom{N}{n} + \frac{1}{(NB)^K} \sum_{n=1}^{N} \sum_{m=1}^{nB} (-1)^{n+m} \binom{N}{n} \binom{nB}{m} (NB - m)^K.$$
Chapter 11

Beam-graphs for Beamforming Codebooks

11.1 Overview

This chapter contains:

- A beam-graph method to further reduce the feedback load in multi-antenna systems that use codebook-based CSI feedback.
- A numerical evaluation using a Grassmannian line packing for feedback and beamforming.

11.2 Introduction

In this chapter, we consider a multiuser downlink with a multi-antenna base-station and single-antenna users, as in Chapter 10. One way for the users to provide the base-station with partial CSI is to use a codebook, \( W \), common to the base-station and all users. Each codeword is a vector of the same dimension as the channel vector and the beamforming vector (c.f. 10.3). The feedback from each user is an index in the codebook, and the corresponding quantized supportable rate.

For single-user beamforming, the more codewords in \( W \), the higher the downlink performance. For SDMA, as in Chapter 10, this does not necessarily hold. Still, there are many cases where feedback of the codeword index from all users would yield a significant feedback load.

When the channel changes moderately from block to block, the feedback can be reduced further by exploiting structure in the codebook. Here, a method to arrange the beams in \( W \) in a graph is proposed. The purpose is to reduce the overhead caused by the feedback of the beam-index. Instead the branch index in the graph is fed back.
For codebook based feedback and beamforming, each user needs to feed back one beam (codeword) index and one AMC level each block. This is already a significant reduction from a full channel state feedback. However, in systems with many users, high mobility and many beams in $\mathcal{W}$, a further reduction may be necessary. In mobile wireless systems, it is reasonable to assume a significant temporal channel correlation between consecutive blocks. The channel should not change too much during one block to enable accurate rate adaptation and beamforming.

For feedback in such temporally correlated systems, it is natural to apply delta modulation [Pro01]. The AMC level is straightforward to delta modulate by only feeding back ‘up one level’, ‘down one level’ or ‘remain on the current level’ commands. How to do this for the beam index is not as trivial since the unit-norm beams are vectors that lie on an $N_t$-dimensional complex sphere. As a simple example, consider a set of unit-norm vectors in $\mathbb{R}^2$ that all lie on the unit circle. For this set, the neighbor relation is easily defined in the angle of the vectors; the two vectors with closest higher and lower angle, respectively, are considered the neighbors of a vector. In higher dimensions however, the neighbor concept is not as simple.

### 11.3 Beam Graph Construction

A unit-norm vector $\mathbf{w} \in \mathbb{C}^{N_t \times 1}$ can be represented in spherical coordinates by $2N_t - 1$ angles. First, the complex-valued $\mathbf{w}$ is transformed into a real vector, $\mathbf{v} = [v_1 \cdots v_{2N_t}]^T = [\text{Re}(\mathbf{w}^T) \ \text{Im}(\mathbf{w}^T)]^T \in \mathbb{R}^{2N_t \times 1}$. Since unit-norm vectors are considered here, only the angular coordinates are of interest. They can be computed as [Has99]

$$
\theta_{2N_t-1} = \arctan \frac{v_{2N_t}}{v_{2N_t-1}} \\
\vdots \\
\theta_1 = \arctan \sqrt{\frac{\sum_{i=2}^{2N_t} v_i^2}{v_1}}.
$$

Having determined the spherical coordinates of all vectors in a packing, the nearest neighbors to each vector have to be found. To illustrate the proposed method, consider unit-norm vectors in $\mathbb{R}^3$. The vectors can be represented by two angles, $\theta_1$ and $\theta_2$. The nearest neighbors in four directions are considered. One neighbor is found in the segment of the sphere with angles greater than $\theta_1$ and $\theta_2$. Another neighbor is found in the segment with angles greater than $\theta_1$ and less than $\theta_2$ and one is found in the segment with angles smaller than $\theta_1$ and greater than $\theta_2$. The final nearest neighbor is found in the segment where the angles are less than both $\theta_1$ and $\theta_2$. The four neighbor segments of $\mathbf{v} = [-1 \ 0 \ 0]$ on the $\mathbb{R}^3$-sphere are illustrated in Fig. 11.1. Note that this is not the only way to find neighbors on a unit sphere.
11.3. BEAM GRAPH CONSTRUCTION

For vectors on a sphere in $\mathbb{R}^{2N_t}$, nearest neighbors in $2^{2N_t-1}$ directions can be found, using the same approach. Clearly, even for vectors with relatively low dimensions, one can find an overwhelming number of nearest neighbors. For instance, for four transmit antennas ($N_t = 4$), 128 nearest neighbors can be found with this approach. For packings with less than 128 beams, this means that no feedback reduction is achieved, compared with beam index feedback. However, it is not necessary to use all neighbor directions. A method to remove neighbor directions is proposed in Table 11.1.

Each neighbor direction signifies a particular change pattern in the angles of a vector. These directions can be enumerated and used as feedback entities, instead of the beam index. Now, a beam-graph can be constructed, connecting nearest neighbor vectors in all chosen directions. The graph can be constructed offline, and has to be known at both the base-station and at the users. Note that the proposed graph construction algorithm is heuristic. It can not be guaranteed to find a connected graph for a given number of neighbor directions, even if it exists.

As the channels of the users change, the preferred beams change. The users notify the base-station by feeding back the index of the neighbor direction instead of the beam index, thereby traversing through the graph. The beam-graph restricts the users to change their preferred beams to neighbors in the beam-graph. At the
1. Choose a chordal distance $d_n$. Vectors within this distance are considered to be neighbors.

2. For each vector in the packing, find the number of neighbors in each direction, within a distance $d_n$.

3. Compute the average number of neighbors for each direction.

4. Successively remove directions, starting from the direction with lowest average number of neighbors. If too many directions have 0 neighbors, increase $d_n$ and go to step 2.

5. After each removal, test the graph for connectedness. A graph is connected if all nodes can be reached from any node [May92]. If the graph is disconnected undo the removal and move on to the next direction.

6. Repeat steps 4-5 until the desired number of neighbor directions has been obtained.

Table 11.1: The algorithm for neighbor direction removal.

setup of a link or on a regular basis, the users have to feedback the whole beam index.

Although this is an ad hoc strategy, significant feedback reduction can be obtained, in particular for packings with many beams and if the number of neighbors is reduced. The penalty of using beam-graph based feedback instead of feeding back the beam-index each block is that for rapidly fading users, the channel may change faster than the graph is traversed. This leads to inaccurate beamforming and a low SINR for the user. However, there is no additional risk of outage, since the feedback channel quality is based on the inaccurate beam in the graph. Furthermore, by significantly reducing the number of neighbors in the graph, there is a risk that the user gets stuck with locally optimum beam. This can be overcome by feeding back the globally best beam-index on a regular basis, temporarily disregarding the graph. In Section 11.4, a numerical performance comparison between full beam index feedback and feedback based on a beam-graph is presented.

11.4 Numerical Results

In the results above, the whole beam index is fed back each block. In Fig. 11.2, the impact of the reduced feedback scheme based on beam-graphs, as described in Section 11.3, is evaluated. A system with one user has been simulated using a Grassmannian packing of 128 lines, i.e. $B = 1$. The beamforming vectors have been arranged in different graphs depending on the number of feedback bits per
11.4. NUMERICAL RESULTS

Figure 11.2: Average sum rate for a system with 1 user, using a Grassmannian line packing with 128 beams. For the Full curve, the whole beam index is fed back each block (7 bits per block). For the curves with relative Doppler ($r_d$) 5-20%, beam-graph feedback with different rates is used.

block. For instance, for 3 feedback bits per block, each beam (vector) in the graph is allowed to have $2^3 = 8$ neighbors. The average sum rate for the full feedback scheme, i.e. the whole beam index is fed back each block, serves as an upper bound. The performance using beam-graph feedback is shown for three values of relative Doppler, $r_d$. The relative Doppler is defined as $r_d = \frac{\text{block duration} \times f_d}{f_d}$, where $f_d$ is the maximum Doppler frequency. For example, $r_d = 20\%$ could correspond to a system with 2 GHz carrier frequency, a 0.15 ms block duration and user speed of 20 m/s. From Fig. 11.2 it is clear that a significant feedback reduction is possible without sacrificing performance to any great extent. Since the packing contains 128 beams, the full beam index feedback also means 7 feedback bits per block. The relative performance loss from reducing the feedback from 7 to 3 bits per user and block for $r_d = 20\%$ is less than 8\%. There is a gap between the full feedback performance and the beam-graph performance with 7 feedback bits. This is due to the fact that, using the beam construction method in Section 11.3, some beams will have no neighbors in some directions. Finally, note that the performance gap between full feedback and beam-graph feedback will be reduced when there are more users in the system. This is due to the fact that the rate loss due to beam quantization is accounted for in the supportable rate computation by the users. Therefore, a user with temporarily bad beam quantization will less likely be scheduled than a
user with perfect beam quantization.
Chapter 12

Conclusions

In this thesis, several aspects of multiuser OFDMA systems were discussed. A solution to the practical problem of feedback overhead was proposed and analyzed. It involves clustering of sub-carriers and feedback of CQI from only a small part of the spectrum. To overcome the low spectral usage when few users are active and the reduced feedback scheme is used, an adaptive reduced feedback scheme was derived. The idea is to let each user feed back more when there are few users in the system and less when the system is loaded.

In selective feedback schemes for OFDMA, as in this thesis, there is a risk that the scheduler has no information about some sub-carriers. Not using them at all would be wasted spectrum. That the sub-carriers were not claimed by any user means that they do not belong to the best sub-carriers of any user. Therefore, a good approach could be to transmit additional, dynamic, pilots on these sub-carriers, improving the channel estimation quality. An alternative approach would be to still allocate users for data transmission, based on other feedback information. Numerical results show that the extra pilots are beneficial for highly frequency selective channels, whereas allocating users for data transmission gives better performance for less frequency selective channels.

The user scheduling on sub-carriers where the base-station has no explicit channel quality information was more thoroughly investigated in an alternative feedback and scheduling framework. Based on the feedback of the channel quality on some uniformly spaced sub-carriers, the scheduler would be able to estimate the channel quality of the other sub-carriers. Several estimators were studied, showing similar performance.

Based on the clustering, an opportunistic beamforming scheme was proposed. It uses random beamforming in each cluster, which induces additional frequency selectivity, thereby increasing the multiuser diversity effect. Important to note is that the opportunistic beamforming scheme does not increase the fading between sub-carriers within the clusters, which enables high accuracy in the reduced feedback scheme. By changing all beamforming weights each scheduling block, the
temporal fading rate of slowly moving users is increased.

The scheduling algorithm is central in multiuser diversity schemes. In this thesis, a modification of the proportional fair scheduling method was proposed and evaluated. The proportional fair algorithm tries to combine multiuser diversity and fairness. The proposed modified proportional fair scheduler incorporates two important aspects into the standard proportional fair scheduler. Firstly, the modified scheduler allows users to have individual target bit-rates and delays. This is necessary in order to meet the diverse user requirements on future wireless voice and data traffic. Secondly, a fairness parameter was introduced into the scheduler, allowing the operator to tune the behavior of the scheduling algorithm in the wide range from a pure multiuser diversity scheduler to a “completely fair” scheduler. The proposed scheduling metric can also be used in single-carrier multiuser diversity systems.

Numerical results for the clustered opportunistic beamforming and modified proportional fair were presented jointly, since the performance of each these components is connected to the assumptions on the other. The clustered opportunistic beamforming performed well compared to several other alternatives. It also outperformed coherent beamforming with round-robin scheduling, even with few users. The modified proportional fair scheduler managed to differentiate the rates to the users and their delays while still being able to exploit the multiuser diversity.

The clustered opportunistic beamforming for OFDMA was extended to opportunistic multi-beam beamforming (SD-OFDMA). Instead of transmitting to one user per sub-carrier at a time, multiple orthogonal beams are formed. With enough active users, the scheduler can allocate suitable users on all beams, thereby enabling space-division multiple access. Numerical results show that, for a small number of users, only a small number of beams is feasible. An OFDMA-specific enhancement to the scheme was proposed, which exploits the temporal channel correlation. If a certain set of beams is particularly successful during one block, chances are high that the beams will perform well also during the next block, in particular if the scheduled users are moving slowly. By keeping the beams on the clusters with highest scheduled rate and re-randomizing the other beams, the performance can be greatly improved, even in a system with high mobility.

The last part of the thesis deals with codebook-based beamforming and SDMA. A codebook containing a large set of grouped beams is available at the base-station and at all users. Based on estimation of the channel, the users find and feed back the index of the preferred beam in the codebook. Each beam is a priori connected with a set of other orthogonal or near-orthogonal beams that will be used simultaneously. Hence, the user is able to compute and feedback the SINR including all inter-beam interference even before the scheduling. This type of structure also directly enables the scheduler to group the users into spatially compatible groups. Three designs of the codebook are considered: Grassmannian, DFT-based and non-orthogonal Grassmannian subspace packings. Simulations show superior performance for the Grassmannian packing in i.i.d. Rayleigh fading. For the design of codebooks, the number of beams and subspaces are important parameters. The allocation outage
probability was derived, which gives good guidelines on these parameter-choices.

In codebooks with many beams, the frequent feedback of the beam-index from all users may constitute significant feedback. By arranging the codebook in a graph, it is possible to reduce the feedback greatly in scenarios with temporal channel correlation. Instead of feeding back the index of the preferred beam, the users feed back the neighbor index in the graph. The graph is arranged so that neighbor beams are similar. If the number of neighbors is much greater than the total number of beams, the feedback reduction is significant.

**Future Work**

This thesis has touched on some aspects of cross-layer design for future wireless cellular systems. The theoretical limits of multiuser multi-antenna systems are relatively well understood. However, the gap between the current technology and practical schemes and the theoretical limits is still significant. Below we list a few research problems that could continue the work in this thesis.

- For OFDMA, we have considered the feedback of frequency domain CQI parameters. However, there are no more degrees of freedom in the frequency domain channel than in the time-domain channel impulse response. User feedback of the most significant time-domain channel parameters could be investigated further and compared to for instance the selective feedback in Chapter 3.

- Feedback reduction by thresholding gives excellent asymptotic multiuser diversity gains, assuming a maximum throughput scheduling. The comparison of thresholding with selective feedback using a fair scheduler would be of some interest.

- The SNR estimation at the transmitter presented in Chapter 5 could be greatly refined, for instance by also considering temporal channel correlation in the estimators and including previously fed back CQIs in the estimation. A Kalman filter approach at the transmitter to track and predict the CQI state in time and frequency would be an interesting development. The feedback design for such a transmitter CQI tracker would an interesting research topic, since the users know the CQI and can choose which observations to provide. Furthermore, an extension of the transmitter CQI estimation to more robust AMC and power allocation could be investigated.

- This thesis has not considered traffic flows and queues, but assumed user buffers to contain data at all times. The validity of this assumption is somewhat dubious, but it simplifies the design of scheduling and beamforming methods. Traffic flow and queue-length issues have been addressed to a great extent in the MAC scheduling context, but less in the domain of spatial-temporal scheduling.
• The subspace packing based SDMA approach in Chapter 10 was presented in a single-carrier context. Its application to a multi-carrier scenario could be an interesting evolution, perhaps by incorporating ideas from Chapter 3.
Appendix A

Notation and Acronyms

A.1 Notation

Plain letters, e.g. $a$ and $A$, are used for scalars. Boldface letters, e.g. $a$ and $A$, are used for vectors and matrices, respectively. Uppercase letters are also used for “frequency domain” variables, so in some cases boldface uppercase letters are used for frequency domain vectors. Calligraphic uppercase letters, e.g. $\mathcal{A}$ are used to denote sets.

\[
\begin{align*}
\bigcup_{i=1}^{2} \text{event}_i & \quad \text{The union of event}_1 \text{ and event}_2 \,(\text{event}_1 \ OR \ \text{event}_2) \\
\text{event}_1 \cap \text{event}_2 & \quad \text{event}_1 \ AND \ \text{event}_2 \\
\Pr(\text{event}) & \quad \text{The probability of event} \\
\Pr(\text{event}_1 | \ \text{event}_2) & \quad \text{The probability of event}_1, \ \text{given event}_2 \\
E[X] & \quad \text{The expected value of the random variable } X \\
E[X|Y] & \quad \text{The expected value of the random variable,} \\
& \quad \text{given the outcome of the random variable } Y \\
\var(X) & \quad \text{The variance of the random variable } X \\
p_X(x) & \quad \text{The PDF of the continuous random variable } X \text{ or} \\
& \quad \text{the PMF of the discrete random variable } X. \\
\Phi_X(f) & \quad \text{The characteristic function of the random variable } X \\
\mathcal{A} \subseteq \mathcal{B} & \quad \mathcal{A} \text{ is a subset of } \mathcal{B} \\
\mathcal{A} \subset \mathcal{B} & \quad \mathcal{A} \text{ is a strict subset of } \mathcal{B} \\
\mathcal{A} \cup \mathcal{B} & \quad \text{The union of the sets } \mathcal{A} \text{ and } \mathcal{B} \\
\mathcal{A} \setminus \mathcal{B} & \quad \text{The set of elements in } \mathcal{A} \text{ that are not in } \mathcal{B} \\
\mathbf{A} \in \mathbb{C}^{M \times N} & \quad \text{The elements of the matrix } \mathbf{A}, \text{ with } M \text{ rows and} \\
& \quad N \text{ columns, are complex.} \\
\mathbf{A} \in \mathbb{R}^{M \times N} & \quad \text{The elements of the matrix } \mathbf{A}, \text{ with } M \text{ rows and} \\
& \quad N \text{ columns, are real.} \\
\mathbf{a} \sim \mathcal{CN} (\mathbf{m}, \mathbf{R}) & \quad \text{The vector } \mathbf{a} \text{ is complex Gaussian distributed with} \\
& \quad \text{mean } \mathbf{m} \text{ and covariance } \mathbf{R} \\
\mathbf{a}^T \mathbf{A}^T & \quad \text{The transpose of the vector } \mathbf{a} \text{ and matrix } \mathbf{A}
\end{align*}
\]
\(a^*\) \(a^*\) \(A^*\) The complex conjugate of the scalar \(a\), the vector \(a\) and matrix \(A\)

\(a^H\) \(A^H\) The transpose conjugate of the vector \(a\) and matrix \(A\)

\(A^K\) The matrix \(A\) multiplied \(K\) times with itself

\(A^{1/2}\) The matrix square-root of \(A\)

\(\|a\|\) The Euclidean vector norm of \(a\)

\(\text{Tr}(A)\) The trace of the matrix \(A\)

\(x(t) * g(t)\) The convolution of \(x(t)\) and \(g(t)\)

\(\mathcal{F}\{x(t)\}\) The Fourier transform of \(x(t)\)

\(\text{arg}(a)\) The phase of the complex number \(a\)

\(\text{arg max}_x f(x)\) The \(x\) that maximizes \(f(x)\)

\(\text{max} f(x)\) The maximum value of \(f(x)\)

\(a \in [b, c)\) \(b \leq a < c\)

\(k \in \{0, \ldots, N - 1\}\) \(k\) can take on the integer values between 0 and \(N - 1\)

\(\{a : f(a) = b\}\) The set of \(a\) such that \(f(a) = b\)

### A.2 Common Symbols and Functions

Here follows the description of most of the used symbols and functions.

\(B\) The SDMA factor, i.e. the number of simultaneously transmitted symbols

\(c\) \(c\) The transmitted symbol

\(C\) The supportable rate [bps]

\(f_d\) The maximum Doppler frequency [Hz]

\(h\) \(h\) The time-domain channel

\(H\) \(H\) The frequency-domain channel

\(I_m\) The \(m \times m\) identity matrix

\(j\) The imaginary unit, \(j^2 = -1\)

\(K\) The number of active users

\(L\) The number of channel taps

\(M_{\text{AMC}}\) The number of AMC levels

\(N\) The set of sub-carriers

\(N\) The number of sub-carriers

\(N_r\) The number of receive antennas

\(N_t\) The number of transmit antennas

\(\mathcal{O}(\cdot)\) The big ordo operator

\(P\) The transmit power [W]

\(Q\) The set of clusters

\(Q\) The number of clusters

\(r_d\) The relative Doppler

\(\mathcal{R}(c)\) The set of sub-carriers that belong to cluster \(c\)

\(R\) The number of sub-carriers per cluster
\( S_k \) The set of clusters that user \( k \) fed back
\( S \) The number of fed back cluster indices and AMC levels per user
\( T \) The symbol duration [s]
\( T_o \) The OFDM symbol duration [s]
\( w \) The beamforming vector (\( w \in \mathbb{C}^{N_t \times 1} \))
\( W \) The beamforming matrix (\( W \in \mathbb{C}^{N_t \times B} \))
\( x, x \) The transmitted signal
\( y, y \) The received signal
\( z, z \) The AWGN
\( \gamma \) The SINR or SNR
\( \delta(t) \) The time-continuous Dirac delta-function
\( \delta(k) \) The time-discrete Kronecker delta-function
\( \eta \) The threshold for keeping beams
\( \kappa \) The fairness parameter
\( \nu \) The RMS delay spread [s]
\( \rho \) The correlation coefficient
\( \sigma_z^2 \) The AWGN variance
\( \xi \) The indicator function (\( \xi \in \{0, 1\} \))
\( \Gamma \) The SNR or SINR gap
\( \Upsilon \) The AMC level
\( \Theta(x) \) The Heaviside step function
\( \text{sinc}(t) \) The sinc-function \( \frac{\sin(\pi t)}{\pi t} \)
\( J_0(x) \) The zero-order Bessel function of the first kind
\( I_0(x), I_1(x) \) The modified Bessel functions of the first kind
\( (\text{zeroth and first order, respectively}) \)
\( \mathbb{G}(B, \mathbb{C}^m) \) The Grassmannian manifold of all \( B \)-dimensional subspaces in \( \mathbb{C}^m \)

A.3 Acronyms

1xEV-DO CDMA2000 Evolution - Data Optimized
3GPP 3\textsuperscript{rd} Generation Partnership Project
ADC Analog-to-Digital Converter
AWGN Additive White Gaussian Noise
BER Bit Error Rate
BF Beamforming
bps Bits per second
BPSK Binary Phase Shift Keying
CDF Cumulative Density Function
CDMA Code Division Multiple Access
CL-BF Clustered Beamforming
CMF Cumulative Mass Function
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>C/N</td>
<td>Sub-carrier-to-Noise power ratio</td>
</tr>
<tr>
<td>Conv-BF</td>
<td>Conventional Beamforming</td>
</tr>
<tr>
<td>CP</td>
<td>Cyclic Prefix</td>
</tr>
<tr>
<td>CQI</td>
<td>Channel Quality Information</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel State Information</td>
</tr>
<tr>
<td>DAC</td>
<td>Digital-to-Analog Converter</td>
</tr>
<tr>
<td>dB</td>
<td>Decibel</td>
</tr>
<tr>
<td>DD</td>
<td>Delay Diversity</td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier Transform</td>
</tr>
<tr>
<td>DSL</td>
<td>Digital Subscriber Line</td>
</tr>
<tr>
<td>EQ-BF</td>
<td>Equal Beamforming</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency-Division Duplex</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communications</td>
</tr>
<tr>
<td>H/2</td>
<td>HIPERLAN/2</td>
</tr>
<tr>
<td>HDR</td>
<td>High Data Rate</td>
</tr>
<tr>
<td>HSDPA</td>
<td>High-Speed Downlink Packet Access</td>
</tr>
<tr>
<td>IDFT</td>
<td>Inverse Discrete Fourier Transform</td>
</tr>
<tr>
<td>IFFT</td>
<td>Inverse Fast Fourier Transform</td>
</tr>
<tr>
<td>ICI</td>
<td>Inter Cell Interference</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter Symbol Interference</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>LMMSE</td>
<td>Linear MMSE</td>
</tr>
<tr>
<td>M-PF</td>
<td>Modified Proportional Fair</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-Input Multiple-Output</td>
</tr>
<tr>
<td>MISO</td>
<td>Multiple-Input Single-Output</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum Mean Square Error</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency-Division Multiplexing</td>
</tr>
<tr>
<td>OFDMA</td>
<td>Orthogonal Frequency-Division Multiple Access</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PER</td>
<td>Packet Error Rate</td>
</tr>
<tr>
<td>PF</td>
<td>Proportional Fair</td>
</tr>
<tr>
<td>PMF</td>
<td>Probability Mass Function</td>
</tr>
<tr>
<td>P/S</td>
<td>Parallel to Serial Conversion</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>QPSK</td>
<td>Quadrature Phase Shift Keying</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>RR</td>
<td>Round Robin</td>
</tr>
<tr>
<td>SDMA</td>
<td>Space-Division Multiple Access</td>
</tr>
<tr>
<td>SD-OFDMA</td>
<td>Space-Division OFDMA</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal to Interference plus Noise power Ratio</td>
</tr>
<tr>
<td>SISO</td>
<td>Single-Input Single-Output</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise power Ratio</td>
</tr>
<tr>
<td>STBC</td>
<td>Space Time Block Coding</td>
</tr>
<tr>
<td>TDD</td>
<td>Time-Division Duplex</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time-Division Multiple Access</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunication System</td>
</tr>
</tbody>
</table>
Bibliography


[ADS02] ETSI TS 101 388 Transmission and Multiplexing (TM); Access transmission systems on metallic access cables; Asymmetric Digital Subscriber Line (ADSL) - European specific requirements. ETSI, 2002.


149


BIBLIOGRAPHY


[IEE99] IEEE. Std 802.11a: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, 1999.


[Qua01] Qualcomm. 1xEV: 1xEVolution IS-856 TIA/EIA standard, airlink overview, November 2001.


