PH.D. THESIS

Resource allocation issues in broadband wireless networks with OFDM signaling

by Iordanis Koutsopoulos
Advisor: Leandros Tassiulas

CSHCN PhD 2002-6
(ISR PhD 2002-11)

The Center for Satellite and Hybrid Communication Networks is a NASA-sponsored Commercial Space Center also supported by the Department of Defense (DOD), industry, the State of Maryland, the University of Maryland and the Institute for Systems Research. This document is a technical report in the CSHCN series originating at the University of Maryland.

Web site http://www.isr.umd.edu/CSHCN/

ABSTRACT

Title of Dissertation: RESOURCE ALLOCATION ISSUES
IN BROADBAND WIRELESS NETWORKS
WITH OFDM SIGNALING

Iordanis Koutsopoulos, Doctor of Philosophy, 2002

Dissertation directed by: Professor Leandros Tassiulas
Department of Electrical and Computer Engineering

Wireless broadband technologies are anticipated to flourish in the next few years, due to the increasing demand for wireless connectivity and the need to support enhanced services and applications in local- or wide-area environments. The primary goal in a communications system is Quality of service (QoS) provisioning to users, which depends on procedures that span several communication layers. Although independent consideration of different layers simplifies system design, it often turns out to be insufficient for wireless networks. Cochannel interference between users that reuse the limited spectrum and the resulting impact of local adaptation actions on overall network performance impose layer interactions in wireless systems. The purpose of this work is to identify and study some of the

issues that arise from the synergy between the physical and the MAC layer in the context of multiple access schemes with orthogonal channels.

Using the essential feature of channel orthogonality as a baseline, our approach places emphasis on Orthogonal Frequency Division Multiplexing (OFDM), which is an emerging multiple access and signaling method for future wireless broadband networks. In OFDM, the broadband spectrum is divided into orthogonal, narrow-band subcarriers and user symbols are split into subsymbols, which are transmitted in parallel over those variable-quality subcarriers. OFDM transmission reduces the effective symbol transmission rate, simplifies equalization at the receiver and provides high immunity to inter-symbol interference and delay spread. Furthermore, it defines a framework for flexible adaptation to varying channel conditions, by allowing transmission parameter control for each subcarrier.

We first address the joint problem of channel allocation with simultaneous adaptation of modulation level and transmission power in a multi-cell OFDM network. We study the impact of those parameters on cochannel interference and channel reuse and present two classes of centralized heuristic algorithms to perform the allocation. Next, we focus on a single-cell multi-user system with modulation control and study the problem of subcarrier assignment to users subject to time resource constraints. We study and compare integral and fractional user assignment, whereby a user is assigned to one subcarrier or can be partially assigned to multiple subcarriers. In addition, we consider the synergy between link-layer ARQ protocols and physical layer parameter adaptation. We consider a simple channel monitoring method which is based on counting received ACKs and NACKs. For a single subcarrier, we show that the adaptation policy which maximizes long-term average throughput per unit time is of threshold type. We also expand our policy

to the multiple-subcarrier case with similar or different channel qualities.

In the sequel, we study the impact of smart antennas and Space Division Multiple Access (SDMA) on MAC layer channel allocation for a single-cell multi-user system. Our approach encompasses multiple access schemes with orthogonal channels, such as OFDM. We first consider the case of unlimited transceiver resources, where a separate beam can be formed for each user of a spatially separable cochannel user set in a subcarrier. We present heuristic algorithms to allocate subcarriers to users and adjust down-link beam patterns, transmission powers and rates with the objective to increase total achievable system rate and provide QoS to users in the form of minimum rate guarantees. Then, we consider the allocation problem for limited transceiver resources, which arises whenever certain reasons impose limitations on the number of beams that can be formed. We propose meaningful heuristic algorithms to jointly form beams from corresponding transceivers and assign subcarriers and transceivers to users, such that the total achievable system rate is increased.

RESOURCE ALLOCATION ISSUES
IN BROADBAND WIRELESS NETWORKS
WITH OFDM SIGNALING

by

Iordanis Koutsopoulos

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2002

Advisory Committee:

Professor Leandros Tassiulas, Chairman/Advisor
Professor Anthony Ephremides
Professor K.J. Ray Liu
Professor A. Udaya Shankar
Assistant Professor Sennur Ulukus

DEDICATION

To my parents Georgios and Maria

and my sister Argiroula

for their invaluable love and support

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my advisor, Professor Leandros Tassiulas, for his continuous support, guidance and encouragement during the entire course of my graduate studies in Maryland. His deep insight in the field and his invaluable suggestions helped me in making progress through my PhD, but also through my Masters studies. I would not exaggerate if I said that every single uttered sentence in our discussions was an invaluable learning experience for me.

I am also thankful to the members of the dissertation committee, Professors Anthony Ephremides, Ray Liu, Udaya Shankar and Sennur Ulukus for kindly consenting to serve on the defense committee and for their constructive comments on my dissertation. In particular, I am thankful to Sennur Ulukus for the enlightening discussions on several research issues. I would like to specially thank Professors Anthony Ephremides and Prakash Narayan for their advice and suggestions on my research proposal. Unfortunately, Professor Narayan was on sabbatical at the time of the defense and could not participate in the committee.

During the five years and three months that I spent in College Park, I had a chance to meet many people and make some really good new friends, who made these years much more enjoyable on and off campus. It would be the eighth chapter of the dissertation, if I had to mention their names and the experiences that I have shared with them. However, I would like to express my thanks to all my roommates

that shared with me the house in 5012 Stewart Court in all these years, for their support and help and even for bearing with me sometimes.

Furthermore, I would like to thank Dr. Subir Varma for hosting me as a summer intern in the summer of 2000 in Aperto Networks, Inc. During these three months I had the opportunity to study and understand several issues related to broadband wireless networks that eventually turned out to be connected to this thesis. From my fellow graduate students, I would especially like to thank my officemate and friend Ulas C. Kozat. I learned a lot from our discussions and from our cooperation on different research problems. Also, my thanks to Tianmin Ren, for the collaboration in various smart antenna issues and in several research papers and to Koushik Kar and Kaushik Chakraborty for useful discussions on related research topics.

Above all, I would like to thank my wonderful family in Greece. It is extremely difficult to express how much love I feel for my parents, $\Gamma\epsilon\acute{\omega}\rho\gamma\iota o\varsigma$ (Georgios) and $M\alpha\rho\acute{\iota}\alpha$ (Maria) and my sister $A\rho\gamma\upsilon\rho o\acute{\upsilon}\lambda\alpha$ (Argiroula). They supported me in every possible way and in every single moment during these years and they were always beside me although they were so many thousand miles away. They gave me strength and courage to continue my effort in difficult times and they always embraced me with love, support and encouragement. The least I can do in recognition to their love is to dedicate this thesis to them.

TABLE OF CONTENTS

List of Figures **xi**

1 Introduction **1**

1.1 Broadband wireless communications 1

1.2 Wireless networks: Layered architecture and mechanisms 3

1.2.1 Wireless channel 3

1.2.2 Quality of Service 4

1.2.3 Multiple access 6

1.2.4 Channel allocation 9

1.2.5 Physical layer adaptation 11

1.3 Orthogonal Frequency Division Multiplexing (OFDM) 15

1.3.1 OFDM transmission and reception 16

1.4 Smart antennas and Space Division Multiple Access (SDMA) 21

1.4.1 OFDM/SDMA transmission 24

1.5 Outline of dissertation 28

1.5.1 Published work 33

2 Joint channel allocation and transmission adaptation in multi-cell multi-user OFDM systems **34**

2.1 Introduction 34
2.1.1 Related work 36
2.1.2 Outline of chapter 39
2.2 System model 39
2.3 Characterization of achievable rate set 43
2.4 Problem statement 45
2.4.1 Problem complexity 47
2.5 Proposed heuristic algorithms 48
2.5.1 Algorithm A 49
2.5.2 Algorithm B 53
2.5.3 Description of Algorithms 54
2.5.4 Practical Considerations 55
2.6 Optimal solution for special cases 57
2.6.1 Problem version I 57
2.6.2 Problem version II 59
2.7 Performance results 60
2.7.1 Simulation setup 60
2.7.2 Numerical results 62
2.8 Conclusion 69

3 Carrier assignment algorithms for OFDM-based networks with channel adaptation 71

3.1 Introduction 71
3.1.1 Related work and motivation 73
3.1.2 Outline of chapter 75
3.2 System model 75

3.3 Problem statement 78
3.3.1 Problem formulation 79
3.4 Integral user assignment 81
3.4.1 Complexity of finding a feasible solution 81
3.4.2 Complexity of finding an optimal solution 82
3.4.3 Proposed heuristic algorithm 83
3.5 Fractional user assignment 86
3.5.1 The case of $N = 2$ carriers 86
3.5.2 Example 89
3.6 Performance bounds 91
3.6.1 Lagrangian relaxation 92
3.7 Further considerations and extensions 96
3.7.1 Time-varying channel quality 96
3.7.2 Infeasible problem instance 99
3.8 Performance results 100
3.8.1 Simulation setup 100
3.8.2 Numerical results 100
3.9 Conclusion 105

4 Link adaptation policies for wireless OFDM-based networks 107

4.1 Introduction 107
4.1.1 Related work and motivation 109
4.1.2 Outline of chapter 113
4.2 System model 113
4.3 Rate adaptation in a single link 115
4.3.1 Problem statement 115

4.3.2 Markov Decision Process (MDP) approach 117
4.4 Rate adaptation for multiple links 125
4.4.1 Problem statement 125
4.4.2 Special case: Multiple links of same quality 126
4.4.3 Extension to multiple links of different quality 128
4.5 Heuristic determination of thresholds for the single-link case 130
4.6 Simulation results 135
4.6.1 Simulation settings 135
4.6.2 Numerical results 137
4.7 Conclusion 143

5 Adaptive resource allocation in OFDM-based wireless networks with smart antennas 145

5.1 Introduction 145
5.1.1 Related work and motivation 146
5.1.2 Notational remarks 150
5.1.3 Outline of chapter 150
5.2 System model 151
5.3 Problem statement 154
5.4 Single-rate transmission: Proposed heuristic algorithms 157
5.4.1 Algorithm A 157
5.4.2 Algorithm B 161
5.4.3 Description of algorithms A and B 161
5.4.4 Algorithm C 163
5.4.5 Solution for a special case 166
5.5 Extensions to multi-rate transmission 167

5.5.1 Spatial separability conditions . . . 167
5.5.2 Multi-rate transmission . . . 168
5.6 Simulation results . . . 170
5.6.1 Simulation setup . . . 170
5.6.2 Comparative results . . . 171
5.7 Conclusion . . . 174

6 Adaptive channel allocation in OFDM-based smart antenna systems with limited transceiver resources 175

6.1 Introduction . . . 175
6.1.1 Related work and motivation . . . 175
6.1.2 Outline of chapter . . . 178
6.2 System model . . . 178
6.3 Channel allocation in OFDM/SDMA systems with limited transceiver resources . . . 181
6.3.1 Problem statement . . . 181
6.3.2 Proposed approach . . . 183
6.3.3 Description of the algorithm . . . 188
6.3.4 Further considerations and extensions . . . 189
6.3.5 Optimal solution for a special case . . . 192
6.4 Simulation results . . . 193
6.4.1 Simulation setup . . . 193
6.4.2 Comparative results . . . 193
6.5 Conclusion . . . 199

7 Conclusion and future work 200

7.1 Summary of contributions 200
7.2 Further extensions 204
7.2.1 Extensions to other multiple access schemes 209
7.2.2 Extensions to higher layers 210

Bibliography **212**

LIST OF FIGURES

1.1 Spectra of OFDM sub-carriers. 16
1.2 Single-user OFDM transmitter and receiver. 17
1.3 Single-user OFDM/SDMA transmitter. 24
1.4 A transceiver module that forms one beam. 25

2.1 Multi-user OFDM transmitter diagram. 40
2.2 Schematic diagram for implementation of algorithms A and B. . . . 56
2.3 The simulated wireless network. 61
2.4 Cumulative distribution function of total rate per subcarrier for algorithm A and different adaptation schemes. 64
2.5 Average rate per subcarrier for different number of available modulation levels. 65
2.6 Average rate per subcarrier for different initial SIR values. 66
2.7 Cumulative distribution function of total rate per subcarrier for algorithm B and different adaptation schemes. 68

3.1 Illustrative example of user assignment to subcarriers. 77
3.2 Average ratio of unsatisfied user rate requirements for different values of SIR per subcarrier. 103

3.3 Average efficiency of feasible solutions for different values of SIR per subcarrier. 104

4.1 Schematic diagram of an OFDM transmission system with FEC encoding. 126
4.2 Markov chain model for state transitions. 131
4.3 Illustrative example for throughput curves for rates r_0 and r_1. . . . 134
4.4 FEC code rate adaptation: Throughput efficiency for different values of the ACK threshold. 137
4.5 FEC code rate adaptation: Throughput efficiency for different values of the NACK threshold. 138
4.6 Modulation level adaptation: Throughput for different values of the ACK threshold. 139
4.7 Modulation level adaptation: Throughput for different values of the NACK threshold. 140

5.1 Block diagram of a multi-user OFDM/SDMA transmitter. 151
5.2 The structure of M transceiver modules for one subcarrier, n. . . . 152
5.3 Total achievable system rate vs. SIR threshold for $M = 4$ antennas. 171
5.4 Total residual rate vs. SIR threshold for $M = 4$ antennas. 172
5.5 Total rate vs. SIR threshold for $M = 8$ antennas. 173

6.1 Block diagram of a multi-user OFDM/SDMA transmitter with limited transceiver resources. 179
6.2 The structure of C transceiver modules. 180
6.3 The beams for assigned users in each subcarrier after the first stage of the algorithm. 184

6.4 Average throughput vs. number of transceivers for approaches A and B, for multi-path with $L = 1$ and $L = 2$ paths and $M = 4$ antennas. 195
6.5 Average throughput vs. number of transceivers for approaches A and B, for multi-path with $L = 1$ and $L = 2$ paths and $M = 8$ antennas. 196
6.6 Residual throughput vs. number of transceivers for approaches A and B, for $M = 4$ antennas. 197
6.7 Average throughput vs. SIR threshold, for unlimited number of transceivers and $M = 4$ antennas. 198

Chapter 1

Introduction

1.1 Broadband wireless communications

The field of wireless communications has seen unprecedented growth during the last two decades, while the advances in relevant enabling technologies and the increasing research interest suggest an even more prosperous future. The need for ubiquitous coverage and connectivity in all kinds of environments and the increasing user demand for mobility, flexibility and easiness of system deployment have necessitated wireless access. It is anticipated that wireless networks will establish themselves as the dominant telecommunication method in the next few years.

Inspired by the successful application of the cellular concept [1], the wireless evolution has so far gone through two generations. First generation (1G) wireless systems (e.g., AMPS, TACS) use analog transmission and support voice services. Second generation (2G) systems (e.g., GSM, IS-95, PDC) employ digital technology and provide circuit-switched, low-speed data communication services in addition to voice. On the other hand, the so-called 2.5G systems (e.g., EDGE/GPRS, HDR),

which currently operate in most countries, support more advanced services such as moderate-rate (up to 100 kbps) packet-switched data.

In 1G and 2G technologies, the main focus was on increasing system capacity in terms of established connections carrying constant-bit-rate data streams. However, recent evolutions in the telecommunications arena indicate a clear trend towards enhanced, rate-demanding services which are expected to flourish in the next years. The advent of services such as telecommuting, home-networking, video-conferencing, fast wireless/mobile Internet access and multimedia constitutes only the first manifestation of the projected demand for wide-band access to information sources of every kind. The idea of third generation (3G) systems became evident by the need to support high and diverse data rates for such heterogeneous applications. Already proposed 3G systems such as UMTS and cdma2000 are envisioned to support rates of the order of $1 - 2$ Mbps [2].

In parallel to the aforementioned wide-area cellular systems, other technologies evolve as a complement to 3G systems with the objective to provide wireless services in different environments. The wireless Metropolitan Area Network (WMAN) standard IEEE 802.16 specifies fixed Broadband Wireless Access (BWA) at the $10 - 66$GHz band for buildings communicating through exterior antennas with central base stations (BSs) which are wired to the backbone network [3]. Fixed BWA provides an alternative to cabled access networks such as fiber optic links, cable modems and digital subscriber line (DSL) links. Multichannel Multipoint Distribution Systems (MMDS) operate at the 2.5GHz band and offer broadband packet services to residential users at rates of 10 Mbps. Wireless Local Area Network (WLAN) standards are primarily concerned with wireless connectivity in a short-range environment with localized mobility, with or without the presence of

a central access point (AP), which plays the role of a BS. The WLAN standards IEEE 802.11a and 802.11b (the latter also known as WiFi) discuss localized transmission at the unlicensed bands of 5GHz and 2.4GHz and can achieve nominal rates of 54 Mbps and 11 Mbps respectively [4, 5]. The ETSI HiperLAN/2 WLAN system will also operate at the 5 GHz band and will offer rates of about 50 Mbps [6]. On the other hand, the wireless Personal Area Network (WPAN) standard IEEE 802.15 focuses on short-range interconnectivity between different equipment (printers, PDAs, home appliances, etc.). Bluetooth and HomeRF technologies provide such services at the 2.4GHz band and support rates up to 1 Mbps.

1.2 Wireless networks: Layered architecture and mechanisms

1.2.1 Wireless channel

The inherent volatility of the wireless medium constitutes the major difficulty in the design of wireless networks. The quality of a wireless link between a transmitter and a receiver depends on radio propagation parameters (path loss, shadow fading, multi-path fading) and cochannel interference. Path loss stems from wave propagation attenuation in free space. Shadow fading is caused by large obstacles such as buildings and the incurred loss is modeled as a log-normally distributed random variable. Multi-path fading arises due to additive and subtractive effect of delays and amplitudes from multiple paths.

The time-varying nature of these factors due to transmitter or receiver mobility and movement of the surrounding objects causes the quality of a narrowband

wireless link to fluctuate with time. On the other hand, a broadband wireless link is characterized both by time-varying behavior due to the aforementioned factors and by frequency selectivity caused by multi-path propagation and delay spread. The frequency-selectivity can lead to inter-symbol interference (ISI) and thus significantly degrade link quality.

The time-varying wireless channel can be completely characterized by its base-band impulse response $h(t, \tau)$, which is given by

$$h(t, \tau) = \sqrt{G\,\sigma(t)} \sum_{\ell=1}^{L} \zeta_\ell(t)\delta(\tau - \tau_\ell), \tag{1.1}$$

where G is the path loss, $\sigma(t)$ denotes time-varying shadow fading, L is the number of paths in the multi-path and $\zeta_\ell(t)$, τ_ℓ are the time-varying gain and time delay for the ℓth path. The transmitted signal is

$$s(t) = x(t)e^{j2\pi f_c t}, \tag{1.2}$$

where f_c is the carrier frequency and $x(t)$ is the complex base-band signal. This is expressed as,

$$x(t) = \sum_{i=-\infty}^{+\infty} d(i)\, g(t - iT), \tag{1.3}$$

where $\{d(i)\}_{-\infty}^{\infty}$ is the symbol sequence, T is the symbol duration and $g(\cdot)$ is the pulse shaping waveform. The signal at the receiver input is

$$r(t) = \int s(t - \tau)h(t, \tau)\, d\tau + \tilde{z}(t), \tag{1.4}$$

where $\tilde{z}(t)$ is the receiver noise process.

1.2.2 Quality of Service

The primary goal of a wireless communications system is the fulfillment of quality of service (QoS) requirements of users. Different interpretations of QoS are available,

depending on the network structure and the communication layer at which QoS is considered. Thus, in single-hop systems with transmission from a single sender to multiple users, QoS at the physical layer is synonymous to an acceptable signal-to-interference-and-noise ratio (SINR) or bit error rate (BER) at the receiver of each user. At the data link control (DLC) and medium access control (MAC) layers, QoS is usually expressed by packet error rate (PER), as well as by minimum achievable rate and maximum tolerable delay guarantees for users. At higher layers, QoS can be perceived as certain throughput, delay, or delay jitter guarantees on a session basis, or even as a form of fairness in rate allocation at the flow level. In multi-hop networks, QoS in the physical and DLC/MAC layers is defined in the same manner as in single-hop systems. However, in multi-hop networks, QoS is also meaningful at the network layer, in the form of end-to-end bandwidth or delay guarantees.

The ability of the network infrastructure to satisfy such QoS requirements and ultimately enhance system capacity depends drastically on procedures and mechanisms which span several communication layers. First, methods for efficient multiple access of users to the network need to be employed. In addition, the quality of each communication link needs to be reliably estimated. At the MAC layer, QoS guarantees can be provided by appropriate scheduling strategies, as well as sophisticated resource management and reuse methods. At the physical layer, adaptive transmission techniques provide the potential to adjust parameters such as transmission power, modulation level, symbol rate or coding rate in order to mitigate link quality fluctuations and maintain acceptable link quality. Moreover, the employment of multiple antennas at the transmitter and/or the receiver constitutes perhaps the most promising means of increasing system capacity. In the sequel, we describe some of these procedures that are considered in this study.

Such adaptive techniques are required for the down-link (link from the BS to the user) and the up-link (link from the user to the BS). The need for supporting higher rates in the down-link direction provides the stimulus for considering down-link in this work.

Two short notes about terminology before we proceed. Although the term "multiple access" is often used in the literature to refer to up-link, we use the same term to refer to user coordination at the base station before down-link transmission. Second, we use the term "base station" to refer to the central unit that coordinates users within a cell and this is the prevalent term for outdoor cellular networks. However, our approach also encompasses indoor cellular networks and WLANs, in which an access point plays the role of base station.

1.2.3 Multiple access

Multiple access schemes are employed to coordinate several users which need to access a common channel, so that the channel is shared and reused efficiently by them and user signals are distinguished at corresponding receivers. Three approaches can be identified in multiple access: connection-oriented (or fixed-assignment), connectionless (or random access) and demand-assignment methods. These methods are characterized by a tradeoff between coordination information overhead and risk of unsuccessful transmission.

Connection-oriented multiple access methods

In connection-oriented multiple access, a separate connection is created for each user session and it is maintained for the entire duration of the session, regardless if the user transmits data or not. Connection-oriented access methods can be

categorized as follows:

- Frequency Division Multiple Access (FDMA). The spectrum is divided into orthogonal, non-overlapping frequency bins and each user is assigned to one frequency. A special case of FDMA is Orthogonal Frequency Division Multiple Access (OFDMA), which is the focus of this dissertation and is explained in detail in a subsequent section.

- Time Division Multiple Access (TDMA). The spectrum is divided into orthogonal time slots and each user is assigned one (or more) slots.

- Code Division Multiple Access (CDMA). All users transmit in the entire frequency spectrum at the same time, but each user is uniquely identified through its assigned signature sequence (code), with which it modulates the transmitted bits. Signature sequences can be deterministically computed or randomly generated. Furthermore, they can be orthogonal or non-orthogonal. CDMA falls within the category of spread-spectrum multiple access (SSMA) methods. Frequency-Hopped Multiple Access (FHMA), where carrier frequencies of users are varied in a random fashion is another SSMA method.

- Space Division Multiple Access (SDMA). The separation of users is performed in space by directing the emitted energy towards each intended user through directional beams which are formed with an adaptive antenna array.

Connectionless and demand-assignment multiple access methods

Connectionless multiple access methods involve less coordination overhead than connection-oriented ones and are more suitable for networks with low traffic, where

the information stream to be transmitted is intermittent or bursty in nature. However, they are associated with higher risk of transmission failure due to potential simultaneous transmissions from other users. ALOHA, Carrier Sense Multiple Access (CSMA) and their derivatives fall within this category. In ALOHA, a user transmits with certain probability whenever it has data to transmit, while in CSMA the user listens to the channel before transmitting and transmits when the channel is free. In CSMA with collision detection (CSMA/CD), a user can also detect collision while it is transmitting and can interrupt transmission if a collision occurs.

Demand-assignment techniques use random access methods in low traffic and fixed-assignment access in high traffic. Dynamic Assignment Multiple Access (DAMA) and Packet Reservation Multiple Access (PRMA) protocols belong to this class of methods. In DAMA, users can reserve traffic channels for packet transmission through contention-free assignment of request channels. When the number of users increases, users content for the request channels. In PRMA, a user transmits with the ALOHA protocol, but if transmission occurs periodically, the user may also acquire contention-free transmission slots by reservation.

DLC layer mechanisms

The DLC layer is above the MAC layer and its purpose is to guarantee reliable communication over the link. Automatic Repeat Request (ARQ) schemes guarantee packet delivery to the intended destination. The receiver receives the transmitted packet and checks its integrity before forwarding it to higher layers. Depending on the outcome of the receiver decoder, a positive acknowledgment (ACK) or negative acknowledgment (NACK) is sent to the transmitter. Upon reception of a NACK,

the transmitter retransmits the packet, whereas when an ACK is received, the transmitter sends a new packet.

ARQ protocols are variants of three basic schemes: stop-and-wait (SW), go-back-N (GBN) and selective-repeat (SR) [7]. In SW, the transmitter must receive the ACK of a packet before proceeding to transmission of the next packet. In GBN, the transmitter sends packets continuously without waiting for ACKs. When a NACK is received for a packet, the transmitter retransmits this packet together with all subsequently sent packets, regardless of their being correctly received or not. Finally, in SR packets are transmitted continuously as in GBN, but only negatively acknowledged packets are retransmitted. SR ARQ yields the highest throughput of all three ARQ schemes.

1.2.4 Channel allocation

Channel allocation can be viewed as an integral part of multiple access that is performed at the MAC layer. Depending on the multiple access scheme, channels can be time slots, carrier frequencies or codes. If the set of users is given, an efficient channel allocation algorithm should try to minimize the number of channels needed to accommodate users and guarantee acceptable link quality for them. By minimizing the number of required channels at any time instant, the system can respond better to a potential sudden load increase or link quality deterioration. Hence, the likelihood of blocking a user is minimized. When the number of available channels is provided, the objective of channel allocation is to maximize system capacity, i.e., the number of accommodated users with acceptable link quality. If users have different rate requirements and need additional channels, the objective becomes to maximize the total achievable rate of users in the system. A

plethora of algorithms for channel allocation has been studied in literature and a comprehensive survey on the topic can be found in [8].

Since wireless spectrum is limited, efficient channel allocation is inherently connected with maximal channel reuse. If SDMA is not employed, channels can be reused to serve users in different cells that are sufficiently well separated by distances large enough, so that transmissions do not interfere with each other. If SDMA is used, channels can be reused even for users in the same cell, if the directed beams that are pointed towards users do not interfere. The channel allocation problem in a cellular network with no SDMA and with pre-assigned base stations to users is equivalent to a generalized graph-coloring problem [9]. When only cochannel interference is considered, the graph is constructed by representing each cell by a vertex, with an edge connecting two vertices if the corresponding cells must not use common channels. The problem is to assign colors (channels) to the vertices, such that no common colors are assigned to adjacent vertices and the minimum number of colors are used. Since this problem is known to be NP-hard and exhaustive search over all possible allocations is impractical for large-scale systems, most efforts in literature focus on developing efficient heuristic algorithms which may provide optimal solutions for simple networks or special cases but are suboptimal in general [10, 11].

Fixed and dynamic channel allocation

Proposed channel assignment methods fall within the categories of fixed channel allocation (FCA) or dynamic channel allocation (DCA). In FCA, a set of channels is permanently allocated to each cell and a user in a cell can utilize only channels that are assigned to that cell. A new user is admitted only if the cell where it

resides has a free channel, otherwise it is blocked. In DCA, all channels are kept in a central pool and any channel can be used by any user in any cell. However, a channel can be reused simultaneously in different cells only if the separation distance between the two cells is greater than a pre-specified minimum distance to avoid excessive cochannel interference. A new user is admitted whenever there exists a channel to be assigned to it subject to minimum reuse distance constraints.

Maximum packing (MP)

A different class of channel allocation methods encompasses the so-called packing algorithms, which perform reassignments of existing users in order to accommodate a new one. Their performance in terms of the number of accommodated users provides upper bounds for general channel allocation methods that do not have the channel reassignment option. The extreme upper bound in this class is provided by the Maximum Packing (MP) policy, which accepts a new user if there exists a global channel reassignment for existing users and the new user, so that all users can be supported and reuse constraints are satisfied [12]. For a linear cellular network with non-overlapping cells, it was shown in [13] that MP can be implemented by doing at most two user rearrangements upon arrival of a new user, which translates into a polynomial-complexity algorithm that accommodates the maximum number of users. However, this approach does not hold for two-dimensional cellular networks.

1.2.5 Physical layer adaptation

Physical layer-based adaptation techniques are employed on a link basis in order to achieve high data rate (in bits/sec) while maintaining an acceptable BER at the receiver irrespective of link quality. The controllable parameters in this work

are coding rate, modulation level and transmission power.

Coding rate

Data from higher layers arrive at the input of the block encoder in the form of a bit stream. The block encoder encodes each k-bit data block into a n-bit code word, by appending $n-k$ redundant bits, which are used by the receiver decoder for error detection and/or correction. The block code is then referred to as a (n, k) code and the code rate is k/n. A particular class of block codes which are considered in this work is Reed-Solomon (RS) forward error correction (FEC) codes. An (n, k) RS FEC code can correct up to $(n - k)/2$ errors.

Depending on the quality of the wireless link, adaptive error protection can be applied to transmitted data by varying the code rate [14, 15]. The encoder has a set of c available code rates, $\{\frac{k_i}{n}\}_{i=1}^{c}$, which can be generated for example with the aid of a punctured convolutional code [16]. In good channel conditions, few redundant bits are appended to the data block in order to provide the desired level of protection, since transmission errors are not very likely to occur. Hence, a high-rate code can be used. On the other hand, when channel conditions deteriorate, lower-rate codes with more redundant bits are required, since errors occur more often.

Modulation level

The encoded bit stream from the output of the encoder enters the modulator, which maps digital bits into analog waveforms. Each block of $b = \log_2 M$ bits from the coded bit stream constitutes a symbol and each symbol is mapped to one of M waveforms. This waveform modulates the carrier and is transmitted over

the channel. We fix our attention to quadrature amplitude modulation (QAM) schemes, for which the amplitude or phase of the carrier changes, but the frequency does not. Each waveform is associated with a signal point in the two-dimensional plane and the ensemble of signal points is the modulation constellation. We are not concerned with the mapping of the b bits to signal points, which is assumed to be accomplished with Gray encoding.

The number of transmitted bits per symbol can be adjusted with adaptive modulation techniques [17]. The modulator has a set $\mathcal{M}$ of L_0 available modulation levels in terms of number of bits per symbol, $\{b_i\}_{i=1}^{L_0}$. Thus, 2-QAM, 4-QAM, 8-QAM and 16-QAM have modulation levels of 1,2,3 and 4 bits/symbol respectively. In the presence of time-varying link quality, the objective of modulation adaptation is to increase transmission rate and maintain an acceptable BER at the user receiver. High modulation levels provide high transmission rates, but they are more susceptible to interference and noise, since signal points are densely packed in the constellation and hence the probability of error at the receiver is high. Such modulation levels should be used only in good quality channels. On other hand, low modulation levels provide lower transmission rates but can sustain more interference and noise.

It should be noted that even in the absence of cochannel interference, the use of high modulation levels is restricted by time-varying background noise in the channel. Although noise is not explicitly taken into consideration in the formulation and analysis in subsequent chapters, the existence of a minimum amount of noise is implicitly assumed by the use of a maximum possible rate which is achieved by modulation level of b_{L_0} bits/symbol.

The BER at the output of the detector when a M_i-QAM modulation level is

used (where $M_i = 2^{b_i}$ for some $b_i \in \mathcal{M}$) is expressed as a function of SINR by the approximation [18]

$$BER_i \approx 0.2e^{-1.5\frac{SINR}{M_i-1}}\,. \tag{1.5}$$

For a maximum allowable BER of ϵ, the SINR at the output of the detector should satisfy

$$SINR \geq \frac{-\ln(5\epsilon)}{1.5}(M_i - 1)\,. \tag{1.6}$$

Hence, we can map each modulation level $b_i \in \mathcal{M}$ to a minimum SINR value (SINR threshold) γ_i (in dB) through a one-to-one increasing function f, such that $\gamma_i = f(b_i)$ equals the right-hand side of (1.6). Clearly, higher modulation levels should be used only in cases of high SINR in order to guarantee an acceptable BER, while lower modulation levels can achieve the same BER at lower SINRs but with lower transmission rate.

Symbol rate

In addition to modulation level, the transmitter can adjust the symbol rate by varying the duration of transmitted symbols as a means of combating ISI [19]. In a link with time-varying multi-path characteristics, the objective of symbol rate control is to increase transmission rate subject to the requirement that delay spread should not exceed a certain fraction of the symbol duration. A high symbol rate with associated small symbol duration yields high transmission rate, but it is more vulnerable to ISI and delay spread. Hence, it should be used when delay spread is small enough and does not constitute a significant fraction of the symbol duration. On the other hand, a low symbol rate with large symbol duration is less vulnerable to delay spread and can be employed even in cases of larger delay spread.

Transmission power

Transmission power control is another technique to control cochannel interference and ensure acceptable link quality. The basic idea is to adjust the transmission power at each transmitter, such that SINRs at receivers are acceptable. The problem of achieving acceptable SINR for a set of cochannel transmitter-receiver pairs through power control has been studied and solved by Zander in [20]. The maximum achievable common SIR, γ^* is

$$\gamma^* = \frac{1}{\lambda^* - 1}\,, \tag{1.7}$$

where λ^* is the maximum positive real eigenvalue of a matrix that contains the link gains from all transmitters to all receivers.

1.3 Orthogonal Frequency Division Multiplexing (OFDM)

Orthogonal Frequency Division Multiplexing (OFDM) is one of the proposed modulation and multiple access technique for wireless broadband access [21]. OFDM is included in the IEEE 802.11a and ETSI HiperLAN/2 standards for WLANs, as well as in the digital audio/video broadcasting (DAB/DVB) standards in Europe. It has also been proposed by IEEE 802.15 and IEEE 802.16 working groups for WPANs and fixed BWA respectively. OFDM is based on the principle of multicarrier transmission, also known as Discrete Multi-Tone (DMT), which was applied earlier in high bit-rate DSLs [22].

In OFDM, the wide-band spectrum is divided into orthogonal narrow-band subcarriers as in frequency division multiplexing (figure 1.1). The bit stream is

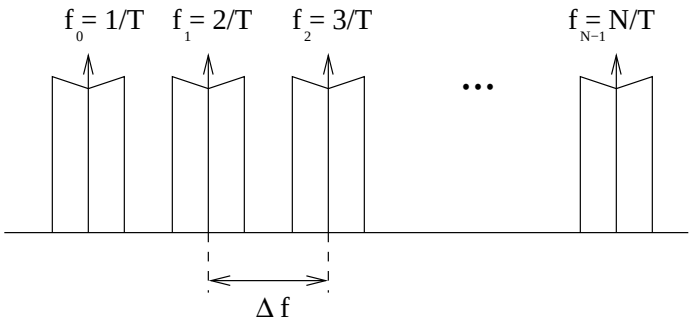


Figure 1.1: Spectra of OFDM sub-carriers.

split into subsets, each of which constitutes a subsymbol. Each subsymbol modulates a different subcarrier and several subsymbols of a user are transmitted in parallel over these low-rate subcarriers. Modulation and demodulation of subcarriers during transmission and reception are implemented with inverse discrete Fourier transform (IDFT) and DFT respectively. The orthogonality of signals in different subcarriers is preserved by appropriate selection of the frequency spacing between the subcarriers. Due to this orthogonality, the signals are separated at the receiver.

1.3.1 OFDM transmission and reception

The schematic diagram of a single-user OFDM transmitter and receiver with N subcarriers is depicted in figure 1.2. The bit stream is divided into bit groups and each bit group constitutes one OFDM symbol. Assuming that OFDM symbols do not interfere with each other, it suffices to concentrate on one OFDM symbol. The OFDM symbol is further divided into N bit subgroups. The bits in the nth subgroup are fed into the nth modulator and modulate the nth subcarrier, $n = 0, \ldots, N-1$. The complex subsymbol d_n at the output of the nth modula-

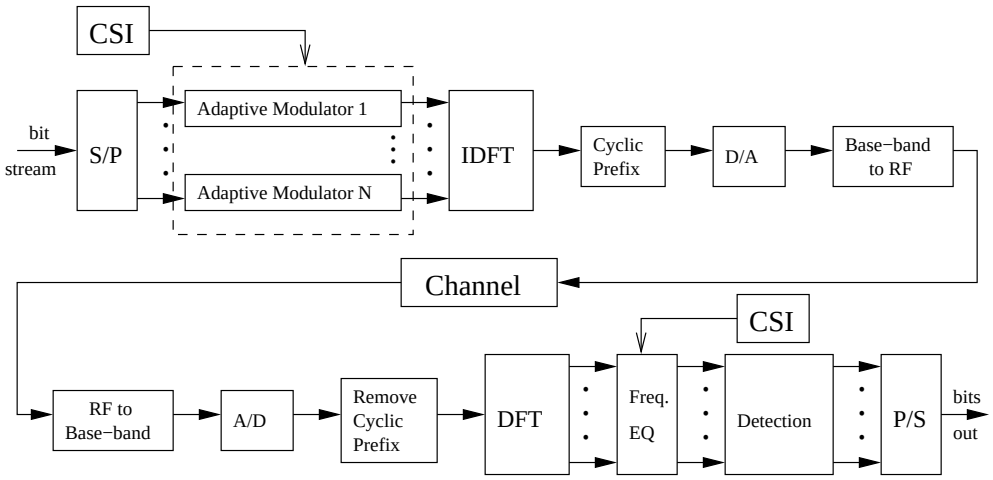


Figure 1.2: Single-user OFDM transmitter and receiver.

tor is selected from a QAM or QPSK constellation and the modulation level of d_n depends on the number of allocated bits in the nth subcarrier. The number of allocated bits per subcarrier depends on subcarrier quality. Better quality subcarriers can carry more bits and maintain acceptable BER at the receiver. All subsymbols are then fed into an IDFT module and are transformed into time samples $\{x_i\}_{i=0}^{N-1}$, where x_i is

$$x_i = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} d_n e^{j2\pi in/N}, \tag{1.8}$$

where $1/\sqrt{N}$ is a scale factor. A cyclic prefix of ν time samples with total duration larger than the maximum delay spread is appended to the N time samples, as a means of eliminating ISI. The sequence $\{x_i\}_{i=0}^{N-1}$ is then passed to a D/A converter, whose output is the continuous signal,

$$x(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} d_n e^{j2\pi nt/T}, \ 0 \le t \le T, \tag{1.9}$$

where T is the symbol duration. The pulse-shaping filter $g(t)$ is taken to be normalized to unit. Note that the signal in the frequency domain consists of N

$sinc(\pi fT)$ functions, each shifted in frequency by $1/T$, where each such function corresponds to the Fourier transform of the unit pulse. Due to the property of the $sinc(\pi fT)$ function that is zero at integer multiples of $1/T$, the subsymbols at different subcarriers can be distinguished at the receiver.

The base-band signal $x(t)$ is up-converted and transmitted through the channel. At the receiver, the signal is translated to base-band and its cyclic prefix is removed. If the channel is invariant for the duration of one OFDM symbol, (1.1) simplifies to

$$h(t) = \sum_{\ell=1}^{L} \beta_\ell \, \delta(t - \tau_\ell), \tag{1.10}$$

where all propagation effects are captured by the parameter β_ℓ. The signal after down-conversion is

$$r(t) = \sum_{\ell=1}^{L} \beta_\ell e^{-j2\pi f_c \tau_\ell} x(t - \tau_\ell) + z(t), \tag{1.11}$$

where $z(t)$ is the base-band noise process. Then, the signal is digitized by being sampled at time points kT/N, for $k = 0, \ldots, N-1$. The kth sample is given as

$$r_k = \frac{1}{\sqrt{N}} \sum_{\ell=1}^{L} \sum_{n=0}^{N-1} d_n \xi_\ell(n) e^{j2\pi nk/N} + z_k, \tag{1.12}$$

where

$$\xi_\ell(n) = \beta_\ell e^{-j2\pi(f_c + n/T)\tau_\ell} \tag{1.13}$$

captures the different impact of the ℓth path delay on different subcarriers and z_k are noise samples. The time samples $\{r_k\}_{k=0}^{N-1}$ enter the DFT module and the subsymbol at subcarrier n is given as

$$R_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} r_k e^{-j2\pi nk/N}. \tag{1.14}$$

After some algebraic manipulations and by using the orthogonality property we have

$$R_n = d_n \sum_{\ell=1}^{L} \xi_\ell(n) + z_n = g_n d_n + z_n,\ n = 0, \ldots, N-1. \tag{1.15}$$

where z_n is the noise level at subcarrier n. The received subsymbols are scaled versions of the transmitted ones and the complex parameter g_n captures the effects of the multi-path channel at subcarrier n.

In order to retrieve the transmitted symbol, the receiver needs channel state information (CSI) in terms of frequency-domain channel transfer function values at subcarrier frequencies. Channel estimation can be performed with pilot symbols that are interspersed with transmitted data symbols. A pilot symbol e consists of known subsymbols $\{e_n\}_{n=0}^{N-1}$. The received pilot subsymbol at subcarrier n after DFT is $y_n = e_n g_n + z_n$. Then, the minimum-mean-squared-error (MMSE) estimate of the complex gain g_n is obtained as

$$\tilde{g}_n = \frac{y_n}{e_n} = g_n + \frac{z_n}{e_n},\ n = 0, \ldots, N-1. \tag{1.16}$$

The estimates $\tilde{g}_n$ are used for frequency-domain equalization (FEQ), namely compensation for the phase and amplitude of received subsymbols prior to detection. Given that the transmitter communicates the utilized modulation level of each subcarrier at the receiver, the Maximum Likelihood (ML) detector decides about the transmitted subsymbol based on $R_n/\tilde{g}_n$. In this study, we assume that perfect CSI is available at the transmitter and the receiver. For slowly time-varying channels, the transmitter can obtain reliable CSI with feedback from the receiver. Assuming that all transmitted subsymbols are normalized to unit power, the signal-to-noise ratio (SNR) at the receiver at the nth subcarrier is,

$$SNR_n = \frac{G_n}{\sigma^2}, \tag{1.17}$$

where σ^2 is the noise variance and $G_n = |g_n|^2$ is the link gain of subcarrier n. When the transmitter uses power level P_n for subcarrier n, a term $\sqrt{P_n}$ multiplies subcarrier n in (1.8). Then, $SNR_n = G_n P_n/\sigma^2$.

Remark: The IDFT of $\{d_n\}_{n=0}^{N-1}$ in (1.8) gives complex-valued time samples $\{x_n\}_{n=0}^{N-1}$. In order to ensure a real-valued transmitted signal, we construct $N' = 2N$ subsymbols by defining $a_{2N-n} = a_n^*$, for $n = 1, \ldots, N-1$, with a new subsymbol $a_0' = \Re(a_0)$ and $a_N = \Im(a_0)$, where $*$ denotes complex conjugate and $\Re$, $\Im$ denote real and imaginary parts. We will assume that such a technique is applied and focus on the N subcarriers.

Advantages of OFDM

The subcarrier spacing of $1/T$ in OFDM results in much higher spectral efficiency than that of simple frequency division multiplex. OFDM transmission increases the effective symbol duration and reduces the effective symbol transmission rate, since information is essentially transmitted over narrow-band subcarriers. Thus, it provides high immunity to ISI and delay spread. In addition, since the frequency-selective broadband channel is divided into a set of frequency non-selective subcarriers, the equalization procedure at the receiver simplifies to a scalar multiplication for each subcarrier. Furthermore, OFDM provides additional flexibility in adapting transmission to varying link conditions, by allowing adaptation for each subsymbol in a subcarrier [23].

1.4 Smart antennas and Space Division Multiple Access (SDMA)

Smart antennas for transmission or reception are recognized as the prominent means of overcoming wireless channel impairments and providing high data rates [24]. Several companies (e.g., Iospanwireless, Metawave, Navini, Arraycomm) aim at commercial products based on smart antennas. Furthermore, smart antennas have been considered for inclusion in several existing wireless standards (e.g., the smart wireless LAN (SWL) [25] system, for IEEE 802.11). Although multiple-input-multiple-output (MIMO) systems have recently received considerable attention [26], the use of multiple receive antennas at the user side is still difficult to implement in certain cases due to size limitations and high cost of multiple down-conversion RF circuits. In this work, we study the down-link of systems with a smart antenna array at the base station and a single omni-directional antenna at the receiver of each user.

A smart antenna array can dynamically adapt its radiation pattern by changing the amplitudes and phases of the excitation currents of each antenna element. Several beam patterns can be formed simultaneously and each beam corresponds to a specific user. Depending on the amount of correlation between paths from different transmit antennas to each receiver and the amount of available CSI at the transmitter and receiver, smart antennas can provide significant benefits in different perspectives.

First, multiple transmit antennas provide transmit diversity, which helps in mitigating fading. The advantage of transmit diversity is based on the fact that if multiple replicas of the same signal are sent over independently fading channels,

the probability of all of them being faded is much less than the probability of one being faded. Diversity benefits are realized if antenna elements are spaced several wavelengths far from each other, so that fading processes of corresponding paths to receiver are uncorrelated.

With respect to the amount of available CSI at the transmitter and receiver, two broad categories of schemes can be identified. The first category comprises schemes where feed-forward or training information is provided to the receiver but no feedback to the transmitter exists, so that CSI is available only at the receiver. At the transmitter, some kind of processing is required to spread data across multiple antennas, while at the receiver CSI information is exploited by ML decoding techniques in order to retrieve the transmitted data. Space-time coding techniques, which combine channel code design with symbol mapping onto multiple antennas have recently been proposed for such cases [27].

When the transmitter has perfect CSI through feedback from the receiver, transmit beamforming can be used in order to efficiently suppress interference and achieve high SINR at the receiver. The available CSI at the transmitter pertains to knowledge of spatial signatures of the user. For a M-element antenna array, the spatial signature is a M-dimensional complex vector, whose entries denote signals received at each antenna element when the user is transmitting alone. In an environment with a single line-of-sight (LOS) path, the spatial signature is a vector pointing to the physical location of the user. In an environment with multi-path, each entry of the spatial signature vector is a superposition of multi-path components coming from different directions. When the transmitter knows the spatial location and multi-path channel characteristics of the user perfectly, it can steer the main lobe of the beam pattern to the direction of the intended user

and can place nulls in the beam pattern in the directions of interfering users. A thorough treatment of beamforming and associated signal processing algorithms is included in [28, 29]. Beamforming has been shown to achieve channel capacity in the information-theoretic sense, if perfect CSI is available at the transmitter [30].

Transmit beamforming divides the space into several spatial channels and can thus implement Space Division Multiple Access (SDMA). SDMA can be combined with any other multiple access scheme with or without orthogonal channels. SDMA can provide substantial capacity benefits, since it enables intra-cell channel reuse by several spatially separable users. For example, if SDMA is combined with TDMA, the same time slot can be utilized for transmission to several intra-cell users, provided that they can be appropriately separated during transmission, so that SINRs at corresponding receivers are acceptable. An adaptive antenna array with M elements can provide M degrees of freedom and can thus separate at most M users in the same channel, depending on the relative locations of the users and the level of noise and fading in the communication channel.

An important issue in SDMA is the determination of beams that guarantee spatial separability. In the up-link, spatial separation of cochannel users is performed at the base station with the use of appropriate filtering algorithms. The user separation problem is decomposed into independent problems, one for each user and the beams (filtering vectors) can be easily computed [31]. However, user separation in the down-link is more cumbersome, since the beam that corresponds to one user affects interference level at all receivers. In addition, since user receivers are distributed and are not usually equipped with multiple antennas, they cannot cooperate to perform joint signal detection, as in the up-link.

In this work, we concentrate on the class of problems associated with smart

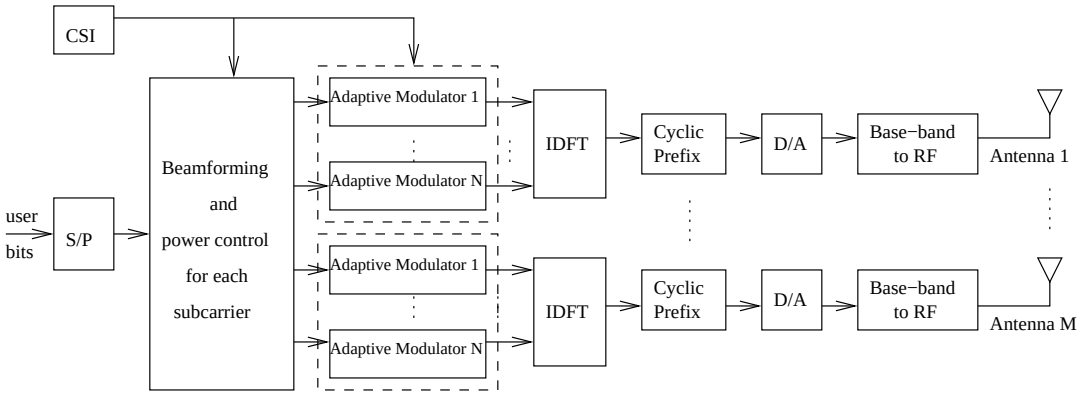


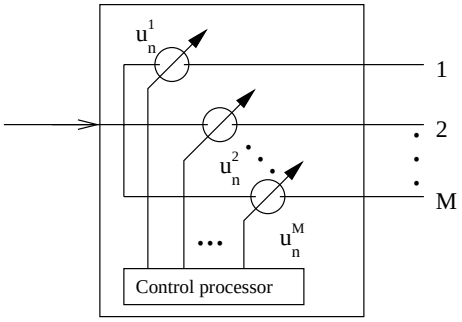
Figure 1.3: Single-user OFDM/SDMA transmitter.

antennas, where CSI is available at both the transmitter and receiver. We focus on the issue of down-link beamforming under the assumption of a multiple access scheme with orthogonal channels, such as OFDM. Our approach encompasses the cases when CSI is deterministic or when it involves a statistical characterization of the channel.

1.4.1 OFDM/SDMA transmission

We describe OFDM/SDMA transmission for a single-user system with N subcarriers and M antennas at the transmitter. The schematic diagram for the OFDM/SDMA transmitter is illustrated in figure 1.3.

As in the case of OFDM single-antenna transmission, we can study separately each OFDM symbol. The bit stream is segmented into bit groups, the OFDM symbols, and the bits of each symbol are further segmented into N parallel streams, each of which is a subsymbol. Subsymbols enter the beamforming and power allocation module and beamforming with M antenna weights is performed, so that M parallel sets of N streams are formed. A separate beam $\mathbf{u}_n = (u_n^1, u_n^2, \ldots u_n^M)^T$ is



Transceiver module for beam formation
in one subcarrier.

Figure 1.4: A transceiver module that forms one beam.

formed by a dedicated transceiver (figure 1.4) and a power $\sqrt{p_n}$ is assigned to each subcarrier n. Beams are normalized, i.e., $\|\mathbf{u}_n\| = 1$, where $\|\cdot\|$ denotes ℓ_2-norm of a complex vector. Then, user bits are forwarded into M parallel modules of N modulators. Within each such module, subsymbols modulate different subcarriers and the complex subsymbol at the output of the nth modulator is d_n. Next, subsymbols are transformed into N time-domain samples as in (1.8). After the cyclic prefix addition and the D/A conversion, continuous signals are transmitted in parallel from the M antennas. The signal transmitted by the mth antenna is

$$x_m(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \sqrt{p_n} u_n^m d_n e^{j2\pi nt/T}, \ 0 \le t \le T. \tag{1.18}$$

In order to simplify the analysis, we assume that multi-path channel characteristics are similar across antennas. This model corresponds to the situation where antenna elements are placed relatively close to each other and there do not exist scatterers close to the BS that decorrelate fading channels. The multi-path channel between the mth antenna and the user is

$$h_m(t) = \sum_{\ell=1}^{L} \beta_\ell \, \delta\left(t - \tau_\ell + \tau_\ell^m\right), \tag{1.19}$$

where L is the number of paths, β_ℓ is the complex gain of the ℓth path and τ_ℓ is its time delay with respect to a reference antenna element. The term

$$\tau_\ell^m = \frac{\delta}{c}(m-1)\cos\theta_\ell \tag{1.20}$$

captures the delay difference between the mth antenna element and the reference element, where δ is the distance between two antennas, θ_ℓ is the angle of the ℓth path with respect to the array and c is the electro-magnetic wave propagation speed. The received signal after down-conversion is

$$r(t) = \sum_{m=1}^{M}\sum_{\ell=1}^{L}\beta_\ell x_m(t-\tau_\ell+\tau_\ell^m) + n(t). \tag{1.21}$$

The signal is digitized by being sampled at time points kT/N, for $k = 0, \ldots, N-1$ and the kth sample is,

$$r_k = \frac{1}{\sqrt{N}}\sum_{m=1}^{M}\sum_{\ell=1}^{L}\sum_{n=0}^{N-1}\sqrt{p_n}u_n^m\xi_\ell(n)d_n \mathrm{e}^{j2\pi nk/N}e^{j2\pi(f_c+n/T)\tau_\ell^m} + z_k, \tag{1.22}$$

where $\xi_\ell(n)$ is given by (1.13). The subsymbol at subcarrier n after the DFT is

$$R_n = \sum_{m=1}^{M}u_n^m\sum_{\ell=1}^{L}d_n\sqrt{p_n}\xi_\ell(n)e^{j2\pi(f_c+n/T)\tau_\ell^m} + z_n. \tag{1.23}$$

Define the mth element of the $M\times 1$ antenna steering vector $\mathbf{v}_n(\theta_\ell)$ at direction θ_ℓ and subcarrier n as

$$v_n^m(\theta_\ell) = e^{-j2\pi(f_c+n/T)\tau_\ell^m}. \tag{1.24}$$

Then, the vector

$$\mathbf{a}_n = \sum_{\ell=1}^{L}\xi_\ell^*(n)\mathbf{v}_n(\theta_\ell) \tag{1.25}$$

is called spatial signature of the user at subcarrier n and captures angular and multi-path properties of the user at that subcarrier. The received signal at subcarrier n is

$$R_n = \sqrt{p_n}\left(\mathbf{a}_n^H\mathbf{u}_n\right)d_n + z_n. \tag{1.26}$$

If all L paths in the multi-path are resolvable, their angles, gains and delays are deterministically known to the transmitter. Then, the SNR at the nth subcarrier is

$$SNR_n = \frac{p_n \left(\mathbf{u}_n^H \mathcal{H}_n \mathbf{u}_n\right)}{\sigma^2} \tag{1.27}$$

where matrix $\mathcal{H}_n$ is defined as,

$$\mathcal{H}_n = \sum_{\ell_1=1}^{L} \sum_{\ell_2=1}^{L} (\xi_{\ell_1}(n)\xi_{\ell_2}^*(n)) \mathbf{v}_n(\theta_{\ell_1})\, \mathbf{v}_n^H(\theta_{\ell_2}). \tag{1.28}$$

If CSI is provided in the form of a statistical characterization of the parameters above, the gain β_ℓ can be modeled as a complex Gaussian random variable with zero mean and variance A_ℓ and the delay τ_ℓ is uniformly distributed in $[0, T]$. The expected useful received signal power is $\mathbb{E}\left\{\left|\sqrt{p_n}(\mathbf{a}_n^H \mathbf{u}_n) d_n\right|^2\right\} = p_n \left(\mathbf{u}_n^H \mathcal{H}_n \mathbf{u}_n\right)$, with

$$\mathcal{H}_n = \sum_{\ell_1=1}^{L} \sum_{\ell_2=1}^{L} \mathbf{v}_n(\theta_{\ell_1})\, \mathbf{v}_n^H(\theta_{\ell_2})\, \mathbb{E}\left\{\xi_{\ell_1}(n)\xi_{\ell_2}^*(n)\right\}, \tag{1.29}$$

where user symbols are normalized to unit power. If paths are uncorrelated,

$$\mathbb{E}\left\{\xi_{\ell_1}(n)\xi_{\ell_2}^*(n)\right\} = \begin{cases} 0, & \text{if } \ell_1 \neq \ell_2 \\ A_\ell, & \text{if } \ell_1 = \ell_2 = \ell, \end{cases} \tag{1.30}$$

and

$$\mathcal{H}_n = \sum_{\ell=1}^{L} A_\ell\, \mathbf{v}_n(\theta_\ell) \mathbf{v}_n^H(\theta_\ell)\,. \tag{1.31}$$

The matrix $\mathcal{H}_n$ is called spatial covariance matrix of the user at subcarrier n and in general it has $\text{rank}(\mathcal{H}_n) > 1$, unless all paths have the same variance, i.e., they are identically distributed. The expected SNR at subcarrier n at the receiver is again given by (1.27).

Deterministic CSI at the transmitter is difficult to obtain in practice, since it requires exact knowledge of spatial signature of the user, which means that

angular and multi-path characteristics for each path are known. CSI in terms of spatial covariance matrix $\mathcal{H}_n$ is more common. The spatial covariance matrix can be estimated in the up-link by sampling the received vector signal $\mathbf{x}$ for each subcarrier n in N_s snapshots and obtaining measurements $\{\mathbf{x}_n(q)\}_{q=1}^{N_s}$. Known pilot symbols can be used for this purpose. Then, the estimate of $\mathcal{H}_n$ is obtained by a simple averaging, i.e.,

$$\hat{\mathcal{H}}_n = \frac{1}{N_s} \sum_{q=1}^{N_s} \mathbf{x}_n(q) \mathbf{x}_n^H(q) \,. \tag{1.32}$$

With time division duplexing (TDD) and the assumption of reasonably slow time variation of the channel, the BS can use this estimate to adapt the down-link beamforming vector.

1.5 Outline of dissertation

The underlying philosophy in all problems that are considered in this dissertation is a synergy between the physical and the MAC layer. Wireless networks have inherited the traditional layered architecture from wire-line networks, where each communication layer is treated as a separate entity with its own adaptable parameters and constraints. While independent consideration of different layers leads to simplified protocol design, it often proves to be insufficient and suboptimal when dealing with wireless systems. First of all, the wireless medium is shared between several users. This property also holds for wire-line access methods such as CSMA/CD, where users use the channel for transmission one at a time, otherwise there is collision. The difference in wireless medium is that transmissions destined for different users can take place simultaneously on the same channel, provided that they do not cause excessive interference to each other.

Cochannel interference between users that reuse the limited spectrum and the resulting impact of local adaptation actions on overall network performance impose layer interactions in wireless systems. Physical layer parameters, such as transmission power or modulation level have considerable impact on multiple access of users in a common channel, since they affect interference levels as well as the amount of interference that can be sustained in the channel. Adaptation of such parameters affects not only QoS at the physical layer (e.g., BER, SINR), but also the perceived QoS at higher layers (e.g., achievable transmission rate). Furthermore, decisions at the MAC layer of one cell affect interference at neighboring cells and hence trigger appropriate physical-layer adaptation actions. The existence of smart antennas at the physical layer raises significant issues at the MAC layer, since intra-cell channel reuse by users depends jointly on beamforming and channel allocation. Therefore, the MAC layer protocols need to exploit the additional flexibility provided by physical layer adaptation, while the physical layer actions need to obtain a more "network-wide" view and consider the impact of parameter adaptation of one user on other links and users. A cross-layer design would provide the required adaptivity in all layers, so that the best possible QoS is obtained.

In this dissertation, we address such cross-layer issues in the context of multiple access schemes with orthogonal channels. Using the essential feature of channel orthogonality as a baseline, our approach places emphasis on OFDM, which presents some novel challenges in resource allocation and provides additional flexibility in adapting transmission to varying channel conditions. The problems that are considered in this dissertation are organized as follows.

In chapter 2, we consider cooperation of the MAC and the physical layer in the context of OFDM for multi-cell multi-user networks. We address the joint

problem of channel allocation with simultaneous adaptation of modulation level and transmission power for each orthogonal channel in OFDM. Our objective is to study the impact of these parameters on cochannel interference and channel reuse, which essentially affect capacity. We start by characterizing the complexity of the problem and the set of achievable rates in this setting. Then, we present two classes of centralized heuristic algorithms, which sequentially construct cochannel sets of users. The first class of algorithms uses greedy criteria to create preferences for user assignment. These criteria are induced interference to other cochannel users, received interference from them and amount of rate increment. The second class of algorithms is based on providing a high minimum SIR in the subcarrier. Some simple special cases of the problem are also identified and solved optimally. Numerical results illustrate the performance benefits of this unified approach.

In chapter 3, we focus on a single-cell multi-user system and study the problem of carrier assignment to users under time resource constraints. In the previous chapter, emphasis was placed on subcarrier reuse in different cells with a network-wide perspective. In this chapter, we focus on the use of adaptive modulation to create preferences for subcarrier allocation to users within a cell, such that a user is allocated to the subcarrier in which it uses fewer time slots to satisfy rate requirements. We first consider the case where subcarrier quality for a user remains fixed within a time frame. We study integral and fractional user assignment, whereby a user is assigned exclusively to one subcarrier or can be partially assigned to more than one subcarriers. For integral user assignment, we identify the complexity of finding a feasible or an optimal solution and provide a heuristic algorithm for subcarrier assignment. For fractional user assignment, we formulate the problem as a linear programming one and present an algorithm that finds

the optimal assignment for a special case. Our algorithms are categorized in the class of heuristics that emanate from Lagrangian relaxation, which is used to obtain performance bounds for our algorithms. Our approach is also extended to time-varying subcarrier quality. The performance of our heuristic algorithms is evaluated by comparison to known lower bounds.

In chapter 4, we consider the synergy between link-layer ARQ protocols and physical layer parameter adaptation in the context of OFDM. Transmission rate is controlled by FEC coding rate and modulation level adaptation. We consider a simple link monitoring method, which is based on counting received ACKs and NACKs and we investigate the class of adaptation policies that correspond to this method. We start by addressing the problem for one user and one subcarrier. We formulate the problem as a Markov Decision Process (MDP) one and we show that the rate adaptation policy that maximizes long-term average throughput per unit time is of threshold type. The optimal policy suggests that transmission rate should be increased whenever the number of successive ACKs exceeds a threshold and it should be decreased whenever the number of successive NACKs exceeds a threshold. We identify the difficulty in realizing this policy and present a suboptimal heuristic method to estimate the thresholds and perform the adaptation. Next, we expand our policy to the case of one user and multiple subcarriers and investigate the impact of several system parameters on the optimal policy for the cases of equal- and different-quality subcarriers. Numerical results validate our policy and denote a considerable improvement in throughput under such adaptation techniques.

In chapter 5, we investigate the impact of smart antennas on MAC layer channel allocation in a single-cell multi-user OFDM system. We consider the case of

unlimited transceiver resources, where a separate beam can be formed for each user in a spatially separable cochannel user set in a subcarrier. We start with the case of single-rate transmission and present heuristic algorithms to allocate subcarriers to users and adjust down-link beamforming vectors and transmission powers, with the objective to increase total achievable user rate and provide QoS to users in the form of minimum rate guarantees. Our algorithms fall within two classes. The first class encompasses greedy algorithms with criteria such as induced and received interference or minimum SIR in a subcarrier to perform the allocation. In these algorithms, power control is activated whenever it is necessary. The second class uses the principle of SIR balancing per subcarrier and employs joint adaptation of beamforming vectors and powers. Next, we extend these principles to the case of multi-rate transmission and state conditions under which a user rate vector is achievable. Numerical results illustrate the comparative performance of algorithms and the relative impact of power control and beamforming on performance.

In chapter 6, we study the problem of channel allocation for OFDM-based smart antenna systems with limited transceiver resources. This issue arises whenever implementation complexity and cost, space inadequacy or other specifications impose limitations on the number of beams that can be formed at the base station. Since each beam serves different users only if they are assigned to different subcarriers and users that reuse the same subcarrier need to be served by beams of different transceivers, the problems of subcarrier and transceiver assignment are coupled. The problem becomes even more challenging since users experience interference from other transceivers that use the same subcarrier. We propose meaningful heuristic algorithms to jointly form beams from corresponding transceivers and assign subcarriers to users, such that the total achievable system rate is increased.

Our algorithms consist of two stages. First, the assignment is performed under no transceiver limitations. Then, the allocation is adjusted for limited transceivers by beam unification based on spatial properties of users, beam cross-correlations and induced interference. Numerical results quantify the performance of these techniques and provide design guidelines for realistic systems.

In chapter 7, we summarize the contributions of this dissertation and present some directions for future study. We discuss the need for a distributed version of the centralized algorithm of chapter 2. We mention the issue of fairness in rate allocation and argue that our policies can also be viewed in the context of admission control. Next, we propose the incorporation of power control in the subcarrier allocation of chapter 3. We also discuss the arising issues from extending the approach of chapter 4 to the multi-user case. As a future step from chapter 5, we suggest to study beamforming for a linear cellular system. Furthermore, we discuss the scheduling issues that arise when the problems that we consider in chapters 2 and 6 are addressed at the packet level. It is emphasized that our approaches can be appropriately modified so as to encompass other multiple access schemes with or without orthogonal channels, such as TDMA and CDMA. Finally, we make some statements about applying such cross-layer approaches to higher layers.

1.5.1 Published work

Most of the results in this dissertation have been published previously or have been accepted for publication. The materials of chapters 2, 3 and 4 have been presented in part in [32, 33, 34]. The problem of chapter 5 has been presented in part in [35]. A paper towards the same route of thought but with a unified approach for TDMA, CDMA and OFDMA is accepted for publication [36].

Chapter 2

Joint channel allocation and transmission adaptation in multi-cell multi-user OFDM systems

2.1 Introduction

In order to adhere to the volatility of the wireless medium, combat the existing interference and ultimately increase achievable data rates, the adoption of sophisticated adaptation techniques is required. When each BS in a cellular network acts independently from other BSs, it is responsible only for users within the coverage area of its cell. The BS is aware of the channel quality of all users and allocates channels to users for down-link transmission [37]. Each channel is allocated to the user that experiences the least interference in it. In that context, transmission parameter adaptation for each user is performed in a straight-forward manner: the BS transmits with sufficient power, so that an acceptable SINR is reached at the receiver, given the measured interference. When the BS employs both power and modulation level adaptation, it can select the highest modulation level for

which there exists a power in the range of available transmission power levels, such that an acceptable SINR is ensured. The BSs can take turns in performing the allocation based on a staggered protocol [38].

Due to local coordination of users, this allocation method leads to suboptimal solution in terms of channel reuse and achievable user rate per channel. However, if each BS communicates with other BSs through high-speed wire-line or wireless links, it can acquire global network knowledge. Then, it can take appropriate channel allocation and adaptation decisions for users within its cell while considering the impact of these actions on users in other cells. Equivalently, all BSs could pass relevant information to a central controller, which would then arbitrate channel allocation and adaptation actions and pass the outcome of the procedure to BSs.

In a multi-cell system with orthogonal channels, it is desirable to achieve maximal channel reuse and transmit in the highest possible modulation level to users, so that achievable system rate is increased. Transmission power adaptation can be used to adjust interference levels at receivers and aid in achieving high resource reuse. In this chapter, we study the joint problem of channel allocation and transmission adaptation in a multi-cell network with cooperating BSs, in which adaptable physical layer parameters are modulation level and transmission power. Modulation level adaptation is a means of controlling sustainable interference, while transmission power control actively changes interference levels at user receivers. Although illustrated in the context of OFDM, the proposed approach can also be applied in systems which support different multiple access schemes with orthogonal channels.

2.1.1 Related work

The problem of power allocation for a single user across parallel orthogonal channels with additive white Gaussian noise with the objective to maximize the total achievable rate subject to a total power constraint is optimally solved with the water-filling method [39]. The bit allocation in each subcarrier is then determined by the corresponding power allocation. The water-filling solution can also be applied in single-cell multi-user systems with a given set of allocated subcarriers to each user, since in that case power allocation for each user can be studied independently. A different perspective of the single-user problem is studied in [40], where bit and power allocation are performed subject to constraints on certain performance criteria such as bit error probability.

The single-cell multi-user problem with unknown subcarrier assignment to users and different quality of each subcarrier for different users is more difficult, due to the discrete nature of the subcarrier allocation problem. Finding the optimal subcarrier allocation to users and corresponding power and bit allocations for each subcarrier in order to maximize total achievable rate is not straight-forward. In [41], a low-complexity suboptimal algorithm is proposed, which decouples the problem into two sub-problems: (i) find required power and number of subcarriers for each user and (ii) find exact subcarrier and bit allocation. In [42], the discrete subcarrier allocation problem is relaxed into a constrained optimization problem with continuous variables. The problem is shown to belong to the class of convex programming problems, thus allowing the optimal assignment to be found with numerical methods. In [43], the authors consider the dual problem, namely that of finding the optimal subcarrier allocation so as to minimize the total transmitted power and satisfy a minimum rate constraint for each user. The problem is

formulated as an integer programming one and a suboptimal solution is found by using the continuous relaxation. In [44], the relation between the rate maximization problem subject to a power constraint and the power minimization problem subject to a rate constraint is investigated.

In a multi-cell multi-user system, the problem becomes more difficult, even if the assignment of subcarriers to users is predetermined. This is because users in different cells reuse the same subcarriers and cause interference to each other. If the number of cochannel users is relatively large, the interference seen by a user in a subcarrier can be approximated by a Gaussian random variable, according to the central limit theorem. In this case, water-filling could again provide a good solution. However, if this approximation is not valid, water-filling cannot be applied, since the power allocated to a user becomes interference for cochannel users. In addition, if the subcarrier allocation to users is not predetermined, all possible combinations of cochannel users should be checked to determine the best one. In [45] and [46], the authors present heuristic distributed algorithms that are executed independently by each BS and are based on iterative water-filling on a subset of subcarriers and removal of subcarriers in which SINRs are violated. The authors in [47], proposed a heuristic algorithm for joint base station, power and channel allocation with the objective to minimize the number of channels required to provide each user with an acceptable connection. In [32], we considered the problem of channel allocation with modulation and power control in a multi-cell system for generic multiple access schemes with orthogonal channels.

The main focus of power control in literature is on adapting transmission powers for a set of cochannel links, so as to provide the maximum common SINR to users. Following the original centralized algorithm of Zander [20], iterative dis-

tributed algorithms have been proposed in [48, 49]. Qiu *et.al.* [50] studied joint modulation level and power control for a set of cochannel users, with the objective to maximize the total achievable rate and they proposed an iterative algorithm for this purpose. A distributed, suboptimal joint power and rate control algorithm based on Lagrangian relaxation is presented in [51]. In the context of multi-cell systems, Fong *et.al.* [52] consider a system where resources are time slots of a carrier frequency which is available in all cells. In the presence of inter-cell cochannel interference, the problem is to schedule concurrent transmissions of BSs and allocate time slots to users, so as to maximize system capacity. The authors' approach is to identify the main sources of interference in each cell and minimize their impact by applying a special time slot assignment and transmission scheduling for each sector of a cell, the so-called Staggered Resource Allocation algorithm (SRA).

Some of the aforementioned approaches focus on a single channel and a set of cochannel users and attempt to ensure QoS at the physical layer in the form of SINR. In single-cell multi-channel systems, transmission parameters are adjusted empirically for each user, based on channel measurements. A user switches to a different channel only when acceptable SINR cannot be provided with the highest transmission power level or lowest modulation level in the current channel. With joint consideration of channel allocation and transmission parameter adaptation, the user could switch to another channel of better quality and use a higher modulation level.

In multi-cell networks, each BS performs resource allocation and transmission parameter adaptation for the users within its cell, without considering the consequences in other cells. As a result, channel reuse is suboptimal and achievable system rate is decreased. However, if some coordination among BSs is introduced,

channel reuse and transmission parameter adaptation can be studied jointly. The amount of cochannel interference and the susceptibility to it can be controlled by selective insertion of users in a channel and adjustment of transmission parameters. Thus, users can meet their SINR requirements and be maximally "packed" in a channel, so that the total transmission rate in the channel is increased. Furthermore, by appropriate coordination between channel allocation and transmission parameter adaptation, MAC layer QoS of users such as achievable data rates can be more flexibly controlled.

2.1.2 Outline of chapter

The rest of the chapter is organized as follows. In section 2.2, we present the model and assumptions used in our approach. In section 2.3, we characterize the set of achievable rate vectors in one time slot and the achievable rate region when a time-division schedule is employed. In section 2.4, we state the problem and show that it is NP-Complete. The proposed algorithms, together with some practical implementation aspects are described in section 2.5. Optimal assignments for some special cases are derived in section 2.6 and numerical results are shown in section 2.7. Finally section 2.8 concludes this chapter.

2.2 System model

We consider a wireless cellular network of M BSs and K users. Each BS provides coverage to a specific area, its cell, and each user establishes connection with the nearest BS. A time frame is assumed, which is divided in time slots, according to a TDMA scheme. Within each slot of duration T_s secs, each BS employs OFDM

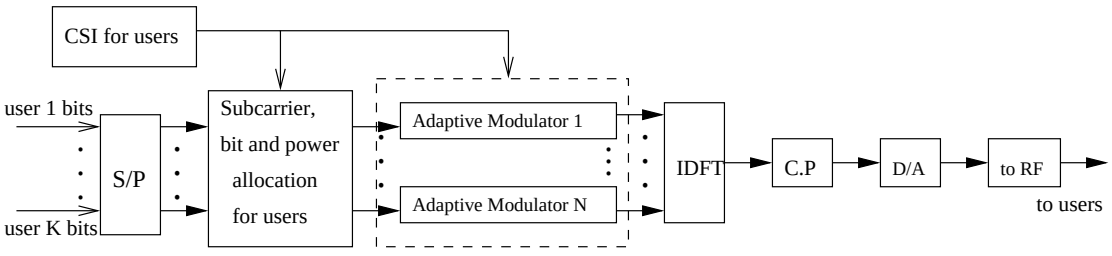


Figure 2.1: Multi-user OFDM transmitter diagram.

transmission with N subcarriers to transmit data to users within its cell. The same set of N subcarriers is used by all BSs. Perfect synchronization is assumed among time slots of TDMA frames in different BSs. Symbol timing synchronization is ensured by appropriate timing estimation. Subcarrier orthogonality is maintained among different subcarriers and different BSs, so that inter-carrier interference (ICI) is not an issue. At each BS, packetized data arrive from higher layer queues and are decomposed into symbol streams before being transmitted to corresponding users in the down-link. The transmission diagram of a multi-user OFDM system for a BS is depicted in figure 2.1.

Each user k is characterized by a bit rate requirement of r_k bits/sec over some time interval $(0, t)$, which consists of several time slots. This requirement is the requested rate by the MAC layer. To achieve this requirement, the user is assigned an OFDM symbol rate (symbols/sec) and a number of bits N_k per OFDM symbol. The latter is given by $N_k = \sum_{n=1}^{N} b_{n,k}$, where $b_{n,k}$ is the number of bits allocated to subcarrier n. These bits constitute the nth subsymbol of user k. User subsymbols can in general consist of different number of bits in different subcarriers, depending on subcarrier quality. The number of bits per subcarrier is selected from a finite L_0-element set of available constellations $\mathcal{M}$. Each subsymbol is normalized to unit power. Assuming that the channel is time-invariant during a time slot duration,

each OFDM symbol of a user in a slot is split into subsymbols over the same set of subcarriers. We will concentrate on subcarrier, bit and power allocation and within one time slot. The rate of user k in a slot, in which a fixed number of S symbols are transmitted, is expressed as

$$r_k = \frac{S}{T_s} \sum_{n=1}^{N} b_{n,k}. \tag{2.1}$$

There exist two versions of the subcarrier assignment problem. When the rate requirements of users are provided as above, the problem is to satisfy these requirements by using the minimum number of subcarriers. When the rate requirements of users are not given, the objective of the allocation algorithm is to maximize the total achievable rate for the K-user system with N subcarriers, namely to transmit the maximum total number of user bits. The two problems are closely related and we refer to them as version I and II respectively. Version I indirectly aims at maximizing the number of users that can be accommodated in the system, while version II aims directly at maximizing total achievable rate.

Clearly, users within the same cell must be assigned to different subcarriers, but users in different cells can reuse the same subcarrier. The link gains $\{G_{ij}^{(n)}\}$ between each BS i and user j in subcarrier n are assumed to be known. They completely characterize the propagation environment between BS i and user j in subcarrier n and take into account path loss, shadowing and multi-path fading. A user j in subcarrier n receives useful signal power from the serving BS and interference from other BSs that transmit in the same subcarrier. In general, the useful and interfering signals are not synchronized in wireless networks. However, in order to simplify the model, we assume symbol-synchronous reception of useful and interfering signals. This is not an unrealistic assumption for indoor environments with relatively small distances between adjacent BSs. The relaxed assumption

that the relative delays of useful and interfering signals do not exceed a symbol duration would also suffice for our model. At the receiver, the signal is sampled at the symbol rate. Assuming that user j is connected to base i_j, the average SINR at the matched filter receiver of user j at subcarrier n is,

$$SINR_j^{(n)} = \frac{G_{i_j j}^{(n)} P_{i_j}^{(n)}}{\sum\limits_{i_k \in \mathcal{B}^{(n)}: k \neq j} G_{i_k j}^{(n)} P_{i_k}^{(n)} + \sigma^2}, \tag{2.2}$$

where $\mathcal{B}^{(n)}$ is the set of BSs that use subcarrier n and $P_i^{(n)}$ is the transmission power of BS i in n. Note that the SINR of a user changes in different subcarriers due to the different impact of multi-path on different subcarrier frequencies. When power control is not employed and all BSs transmit at a common power level P, the powers do not appear in the expression above. We note however that the existence of noise is always implied by restricting the transmission rate to be determined by a finite set of modulation levels.

In our model, we assume that cochannel interference is the prevailing interference type and the noise level is not known. Then, the SINR can be replaced by the signal-to-interference ratio (SIR). Apart from practical implications, this approach eliminates the need for total transmission power constraints. If the noise term is not included in (2.2), the SIR is insensitive to the absolute power levels $\{P_{i_k}\}$ and thus powers can always be adjusted so as to achieve a certain SIR level.

The BER at the output of the detector of a user in a subcarrier should be less than ϵ. With the rationale of subsection 1.2.5, a modulation level of b_i bits per subsymbol can be mapped to a minimum required SIR of γ_i dB, based on (1.6).

2.3 Characterization of achievable rate set

In this section, we characterize the set of achievable rates when modulation level control is used. We consider a system instance with M BSs, N users and subcarrier gains $G_{ij}^{(n)}$ between BS i and user j in subcarrier n. Let S_i be the set of users served by BS i, for $i = 1, \ldots, M$, and let b_j be the modulation level of user j.

Consider a subcarrier n, for which a set of cochannel users needs to be found. An *assignment policy* for a subcarrier is a rule that determines the set of cochannel users and their corresponding modulation levels in that subcarrier. The cardinality of the cochannel user set is at most M, since at most one user from each BS can be included in the subcarrier. An assignment policy consists of the following steps:

- Determination of a BS activation set.
- User selection (at most one user from each activated BS).
- Modulation level (rate) assignment to users.

First, some BSs need to be activated for transmission. A BS activation set is represented by its activation vector $\mathbf{q}$, which is a $M \times 1$ binary vector. The ith component, q_i, corresponds to BS i and equals 1 if BS i belongs to the activation set, otherwise it is 0. Let $\mathcal{S}$ denote the set of all possible BS activation vectors. After determining the BS activation vector, an appropriate user needs to be selected from each BS i. If activation vector $\mathbf{q}$ is given, the interference experienced by any user is known and the user $u_i^*(\mathbf{q}, n)$ that is selected from BS i is the one that achieves the highest modulation level and maintains acceptable SIR, namely user

$$u_i^*(\mathbf{q}, n) = \arg\max_{u_i \in \mathcal{S}_i} b_{u_i(\mathbf{q},n)} = \arg\max_{u_i \in \mathcal{S}_i} \left\{ \max \left[b \in \mathcal{M} : \frac{G_{iu_i}^{(n)}}{\sum_{j \neq i : q_j = 1} G_{ju_i}^{(n)}} \geq f(b) \right] \right\}. \tag{2.3}$$

Clearly, when $q_i = 0$, then no user is selected from BS i and $u_i^*(\mathbf{q}, n) = 0$. When users are selected from each BS according to (2.3), the BS activation vector $\mathbf{q}$ is associated with a modulation (rate) vector $\mathbf{b}^*(\mathbf{q}, n) = (b_{u_1^*(\mathbf{q},n)}, b_{u_2^*(\mathbf{q},n)}, \ldots, b_{u_M^*(\mathbf{q},n)})$. If we repeat the same procedure for all $\mathbf{q} \in \mathcal{S}$ for subcarrier n, we find a set of modulation vectors $\mathcal{X}^{(n)} = \{\mathbf{b}^*(\mathbf{q}, n) : \mathbf{q} \in \mathcal{S}\}$, where each vector in $\mathcal{X}^{(n)}$ corresponds to a BS activation set. The set $\mathcal{X}^{(n)}$ is called *set of achievable rate vectors for subcarrier n*. The optimal assignment policy determines the BS activation vector that leads to the modulation vector with the maximum sum of components over all modulation vectors in $\mathcal{X}^{(n)}$. For multiple subcarriers, the same procedure is applied independently for each subcarrier to determine the BS activation vector that yields the maximum total rate in each subcarrier.

We should note that the optimal assignment policy maximizes the total subcarrier rate in one time slot. When a continuous-time schedule is considered, different BS activation vectors are used in different portions of time in order to achieve certain rates or desired properties of rate vectors such as fairness. In that case, the *achievable rate region for subcarrier n*, $\mathcal{R}^{(n)}$, is defined as

$$\mathcal{R}^{(n)} = \left\{ \mathbf{R} : \mathbf{R} = \sum_{\mathbf{q} \in \mathcal{S}} \tau_{\mathbf{q}} \mathbf{b}(\mathbf{q}, n), \text{ with } \tau_{\mathbf{q}} \geq 0 \text{ and } \sum_{\mathbf{q} \in \mathcal{S}} \tau_{\mathbf{q}} = 1 \right\}, \tag{2.4}$$

where $\tau_{\mathbf{q}}$ denotes the portion of time of a continuous schedule in which BS activation vector $\mathbf{q}$ is used. From (2.4), the achievable rate region is identified as the *convex hull* of the set $\mathcal{X}^{(n)}$ of modulation vectors. It is possible that a different user selection than that in (2.3) is used. In that case, the rate region is formed by time-sharing among different BS activation vectors *and* different user selections from each BS.

When continuous-time schedule is considered for multiple subcarriers, each subcarrier has different link gains and hence it is characterized by its own achievable

rate region. The *achievable rate region for all N subcarriers*, $\mathcal{R}$, is given as

$$\mathcal{R} = \sum_{n=1}^{N} \mathcal{R}^{(n)}, \tag{2.5}$$

where the sum of sets A and B is defined as the set of all vectors that can be written as $\mathbf{a} + \mathbf{b}$ with $\mathbf{a} \in A$ and $\mathbf{b} \in B$. In this work, we are not concerned with time schedules and we concentrate on assignment policies of users to subcarriers within one time slots.

From (2.3), we observe that when the BS activation vector $\mathbf{q}$ is given, it is possible to find the "best" user $u_i^*(\mathbf{q})$ from each BS i, such that $u_i^*(\mathbf{q})$ uses maximum modulation level and thus it is possible to maximize subcarrier rate. This is because the SIR of a user does not depend on modulation levels of users in other BSs. However, when power control is also considered, the SIR of a user depends on powers of all BSs. In addition, even if the selected users from each BS are known, finding BS powers so as to maximize the total subcarrier rate is not straightforward. In section 2.6, an analytical solution is provided for $M = 2$ BSs and continuous rates. In general, for a given BS activation vector, the assignment with the maximum total subcarrier rate is found by checking all $\prod_{i=1}^{M}(|\mathcal{S}_i| + 1)$ possible combinations of user selections from different BSs and by checking all L_0^M possible modulation vectors for each such combination of users. As will be shown in subsection 2.5.1, it is possible to check whether a modulation vector is achievable through a power vector.

2.4 Problem statement

Each user in a cell receives the useful signal from the serving BS through some subcarriers and it receives interference from neighboring BSs that use the same

subcarriers to transmit to other users. A cochannel set of users in a subcarrier is *feasible*, if users simultaneously use the subcarrier and all user SIRs are satisfied. For a given subcarrier, the feasibility of a cochannel user set depends on the number and identities of users through their link gains. Furthermore, it depends on the utilized modulation levels of users, because different modulation constellations are associated with different minimum required receiver SINR values in order to maintain fixed BER and hence they have different amounts of maximum sustainable interference. When power control comes into stage as a means of changing SIRs at user receivers, the feasibility of a cochannel user set also depends on power levels. Finally, cochannel set feasibility depends on the individual subcarrier, due to different link gains of users in different subcarriers. Thus, users that can share one subcarrier, may not be eligible cochannel users in a different subcarrier, or subcarrier reuse may be feasible with different numbers of allocated bits.

When a high modulation level is assigned to a user in a subcarrier, user rate is increased, since more bits are transmitted. If high modulation levels are used, the user needs fewer subcarriers to satisfy rate requirements. Therefore, more users can be accommodated in the system and capacity is increased. However, high modulation levels do not facilitate subcarrier reuse, since they are more vulnerable to interference and thus fewer users can coexist in the same subcarrier. Users that cannot reuse a subcarrier should in general be assigned to different subcarriers and from that point of view capacity is not increased. On the other hand, a low modulation level implies that a small number of user bits is transmitted. The user requires more subcarriers to satisfy rate requirements and thus fewer users can be accommodated in the system. However, low modulation levels favor subcarrier reuse, by allowing more users to be "packed" in the same channel, since they can

sustain more interference. As a consequence, a high modulation level for some users in a subcarrier generates higher rates, but may lead to reduced *total* subcarrier rate due to smaller subcarrier reuse. Low modulation levels yield lower rates but may yield higher total subcarrier rate due to larger subcarrier reuse.

Clearly, there exists a tradeoff between achievable rate per subcarrier and subcarrier reuse. The question that arises is whether there exists a way to perform modulation level control and subcarrier allocation jointly, so as to increase total subcarrier rate and system capacity. In other words, we want to identify the set of cochannel users which results in the maximum total rate in each subcarrier. Ideally, we would like to assign the highest possible modulation level to users and reuse the same subcarrier for as many users as possible. This is feasible if users are close to serving BSs, so that transmissions from other BSs do not cause much interference to them. However, if the locations of users and BSs are such that cochannel interference is an issue, then subcarrier reuse may be feasible only for a subset of users and with certain modulation levels.

2.4.1 Problem complexity

From the discussion of previous sections, it becomes evident that the solution to the problem of maximizing total achievable rate in a subcarrier is determined by the BS activation vector that yields the maximum total rate in the subcarrier. This can be formally stated as follows:

$$\max_{\mathbf{q}\in\mathcal{S}} \sum_{i=1}^{M} b_{u_i(\mathbf{q})} \,, \tag{2.6}$$

subject to the constraint that SIRs at receivers of users that are selected from each BS are acceptable. We now show that the problem is NP-Complete.

Consider a simpler version of the problem with M BSs and one user per BS. Assume that there exists one modulation level b with associated SIR threshold γ and that power control is not used. The BS activation vector that achieves the maximum total rate is the one where the maximum number of BSs are activated, subject to the constraints that the SIRs of users must be acceptable.

We use the $0-1$ Knapsack problem and convert it to an instance of our problem. Let the link gain from each BS to the user in its cell be G and let the link gain between a BS and a user in another cell be 1. Objective (2.6) is written as

$$\max_{\mathbf{q}} \sum_{i=1}^{M} q_i, \tag{2.7}$$

subject to the SIR constraint: $G/(\sum_{i=1}^{M} q_i - G) \geq \gamma$, or equivalently $\sum_{i=1}^{M} q_i \leq G(1+\gamma)/\gamma$, which is identified as a $0-1$ Knapsack problem. Since this problem is known to be NP-Complete [53], our problem is also NP-Complete.

2.5 Proposed heuristic algorithms

Since the enumeration of all BS activation vectors that leads to the optimal solution is of exponential complexity, it is desirable to design heuristic algorithms to construct cochannel sets of users with high total rate per subcarrier. The key idea is to "pack" as many users as possible in a subcarrier, while enabling each user to use high modulation level. Since the objective is to maximize the total achievable rate in the system, it suffices to consider the allocation procedure for each subcarrier separately. The order in which users are inserted in the subcarrier is crucial. The modulation level of a user designates the amount of sustainable cochannel interference and this interference must be kept to a minimum during the insertion procedure. In order to keep the complexity of the algorithms to a reasonable level,

we consider the class of algorithms for which users are sequentially inserted in the subcarrier and no user reassignments are performed. However, we allow modulation level reassignments for cochannel users. In addition, power adaptation will be considered only when modulation adaptation alone does not provide acceptable SIRs. In the sequel, we present two classes of heuristic algorithms that use different preference criteria for the allocation.

2.5.1 Algorithm A

The first class of algorithms uses greedy criteria, such as induced interference to cochannel users, received interference from cochannel transmissions and amount of rate increase to create preferences for the allocation.

Modulation adaptation

At each step of the algorithm, an appropriate user is assigned to a subcarrier and the modulation levels of other users are adjusted, so that acceptable SIRs are ensured. Fix attention to subcarrier n and let $\mathcal{U}^{(n)}$ denote the set of users that are already assigned in n and $\mathcal{B}^{(n)}$ be the set of BSs that transmit to users in $\mathcal{U}^{(n)}$. Let k be the user to be inserted next in n. A user k should use a subcarrier if it has high link gain $G_{i_k k}^{(n)}$, so that it can use a high modulation level. We also consider the interference that is caused by BS i_k of user k to users in $\mathcal{U}^{(n)}$ and the interference caused to k from BSs transmitting to other users in n. Specifically, we take into account the maximum of these two interference values and we define the *Signal-Interference Factor (SIF)* $F_{n,k}$ as follows

$$S_{n,k} = \frac{G_{i_k k}^{(n)}}{\max\left\{ \sum_{j \in \mathcal{U}^{(n)}} G_{i_k j}^{(n)}, \sum_{i_j \in \mathcal{B}^{(n)}} G_{i_j k}^{(n)} \right\}}. \tag{2.8}$$

Among all candidate users, we select the one with the maximum SIF factor. By allowing the least interference increase in the system, future user assignments are also facilitated. Note that when a subcarrier is initially unoccupied, the SIF factor is $S_{n,k} = G^{(n)}_{i_k k}$.

Assume now that user k is tentatively inserted in subcarrier n. Upon insertion, it is possible that k receives enough interference from BSs in $\mathcal{B}^{(n)}$. It is also possible that some users in $\mathcal{U}^{(n)}$ may not sustain the additional interference due to k, so that the SIR thresholds corresponding to assigned modulation levels are violated. In this case, modulation levels of these users need to be reduced, so that users become less susceptible to interference, but user rates are decreased. The addition of a user in a subcarrier is beneficial if subcarrier rate decrease due to users with violated SIRs is less than the rate contribution of the new user, so that finally subcarrier rate is increased. In fact, the most desirable user is the one for which the rate increase is maximized.

In order to formalize these rules, let $b^*_{n,k}$ be the maximum modulation level of user k that leads to acceptable SIR for k upon its insertion in subcarrier n. Let $\mathcal{V}_{n,k} \subseteq \mathcal{U}^{(n)}$ denote the set of users using subcarrier n, for which SIR is not acceptable with the current modulation level after insertion of user k. For each user $m \in \mathcal{V}_{n,k}$, let b^-_m be the modulation level before insertion of user k and b^+_m be the maximum modulation level that ensures acceptable SIR after k is inserted. For subcarrier n and user k, define the *Incremental Rate Factor (IRF)* $T_{n,k}$ as

$$T_{n,k} = b^*_{n,k} + \sum_{m \in \mathcal{V}_{n,k}} (b^+_m - b^-_m). \tag{2.9}$$

Clearly, a user with high IRF is preferable since it leads to high subcarrier rate increase. Note that if k is the first user to be inserted in n, then $T_{n,k} = b_{L_0}$. Efficient user assignment in a subcarrier pertains to insertion of users which cause

least interference to users in $\mathcal{U}^{(n)}$, receive the least interference from BSs in $\mathcal{B}^{(n)}$ and have positive and large rate contribution. To capture these objectives, we define the *Assignment Preference Factor (APF)* $A_{n,k}$ for each subcarrier n and user k as,

$$A_{n,k} = S_{n,k}T_{n,k}. \tag{2.10}$$

Thus, among users which cause or receive the same amount of interference, the one that yields the greatest rate benefit is preferable to join the subcarrier. Moreover, among users which cause the same rate increase, the one with the smallest amount of received or induced interference is inserted in the subcarrier. After user assignment in the subcarrier, the modulation levels are updated and the users belonging in the same cell as the inserted user are not considered for assignment.

Transmission power adaptation

Since the assignment of a user in a subcarrier should not reduce the already achieved subcarrier rate, the sequential assignment algorithm should terminate when $T_{n,k} < 0$ for all remaining users. Furthermore, it may happen that the insertion of a user does not lead to a feasible cochannel user set, even if the lowest modulation level is used for all users. In such cases, modulation adaptation cannot further increase subcarrier rate. While modulation level adaptation essentially adjusts the level of sustainable interference for user receivers in order to maintain an acceptable BER, it does not actively change the SIR level at receivers. Transmission power control can be used together with modulation level adaptation to increase subcarrier reuse and achievable subcarrier rate.

Consider $m \leq M$ cochannel users in a subcarrier. In the sequel, we drop the subcarrier index n. For ease of notation, let $\mathbf{G} = \{G_{ij}\}$ be the $m \times m$ matrix of link

gains from BS i to user j, for $i, j \in \{1, 2, \ldots, m\}$. Let $\mathbf{b} = (b_1, b_2, \ldots, b_m)$ denote the modulation level vector of users and let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_m)$ be the associated SIR threshold vector. Define the BS transmission power vector $\mathbf{P} = (P_1, P_2, \ldots, P_m)$. The SIR of a user j in the cochannel user set is acceptable if

$$SIR_j = \frac{G_{jj}P_j}{\sum_{i=1, i\neq j}^{m} G_{ij}P_i} \geq \gamma_j, \text{ for } j = 1, \ldots, m. \tag{2.11}$$

A modulation vector $\mathbf{b}$ is said to be *achievable* for the cochannel set of m users if there exists a power vector $\mathbf{P}$, such that the SIR constraints that correspond to the modulation levels are satisfied for all m users. In this case, the cochannel user set is called *feasible with respect to* $\mathbf{b}$. Condition (2.11) can be written as

$$P_j \geq \sum_{i=1}^{m} \frac{G_{ij}}{G_{jj}} \frac{\gamma_j}{1+\gamma_j} P_i, \text{ for } j = 1, \ldots, m \tag{2.12}$$

Define a $m \times m$ matrix $\tilde{\mathbf{G}}$ with elements

$$\tilde{G}_{ij} = \frac{\gamma_j}{1+\gamma_j} \frac{G_{ij}}{G_{jj}}. \tag{2.13}$$

Then, condition (2.11) is written in matrix form as

$$\mathbf{P} \geq \mathbf{P}\tilde{\mathbf{G}}. \tag{2.14}$$

The matrix $\tilde{\mathbf{G}}$ is non-negative definite and irreducible. According to the Perron-Frobenius theorem [54], $\tilde{\mathbf{G}}$ has a positive, real eigenvalue λ^*, with $\lambda^* = \max\{|\lambda_i|\}_{i=1}^{M}$, where $\{\lambda_i\}_{i=1}^{M}$ are the eigenvalues of $\tilde{\mathbf{G}}$. The eigenvalue λ^* has an associated eigenvector $\mathbf{P}^*$ with strictly positive entries. Furthermore, the minimum real λ such that the inequality $\lambda\mathbf{P} \geq \tilde{\mathbf{G}}\mathbf{P}$ has solutions $\mathbf{P} > 0$ is $\lambda = \lambda^*$.

In our case, we start by finding the maximum real positive eigenvalue λ^* of $\tilde{\mathbf{G}}$ to guarantee a power vector with positive components. If $\lambda^* \leq 1$, then (2.14)

holds and modulation vector $\mathbf{b}$ is achievable. The power vector that leads to an achievable $\mathbf{b}$ is the eigenvector that corresponds to λ^*.

The purpose of power adaptation is to aid the assignment of a new user in a subcarrier by adjusting power levels of BSs that transmit in the same subcarrier to users. Specifically, an achievable modulation vector which leads to subcarrier rate increase should be found. We consider the tentative assignment of a user k in the subcarrier and check the achievability of modulation vectors for the cochannel set of users, starting from the vector whose entries are equal to b_{L_0}. Each time a modulation vector is not achievable, we decrease the modulation level of one entry and check the vector again. This procedure is repeated until we find an achievable modulation vector with IRF $T_{n,k} > 0$. If such a vector is not found, we set $T_{n,k} = -\infty$ by convention. If an achievable modulation vector with $T_{n,k}$ is found, we compute the SIF of k as,

$$S_{n,k} = \frac{P_{i_k} G^{(n)}_{i_k k}}{\max\left\{P_{i_k} \sum\limits_{j \in \mathcal{U}^{(n)}} G^{(n)}_{i_k j}, \sum\limits_{i_j \in \mathcal{B}^{(n)}} P_{i_j} G^{(n)}_{i_j k}\right\}}, \tag{2.15}$$

where the transmission powers are the entries of the eigenvector associated with the achievable modulation vector. The APF of user k is then computed with (2.10).

2.5.2 Algorithm B

The second class of heuristic algorithms uses the assignment criterion of maximizing the minimum SIR in a subcarrier.

Algorithm B is also based on sequential assignment of users in a subcarrier and it is similar to algorithm A in that it also aims at inserting users that induce high rate benefit in the subcarrier. In that sense, algorithm B uses the IRF factor $T_{n,k}$ given by (2.9). The difference from algorithm A lies in the definition of the SIF

factor. In algorithm A, the assignment was based on a greedy criterion about least additional interference in the subcarrier. In algorithm B, a user assignment in the subcarrier is performed if it maximizes the minimum SIR of users in the subcarrier over all possible user assignments. Since users can have different modulation levels, SIRs of users are scaled by the corresponding SIR thresholds. First, the IRF factors $T_{n,k}$ for each user k are computed. If $T_{n,k} > 0$, the SIF factor is now defined as

$$S_{n,k} = \min\left\{\frac{SIR_{n,k}}{\gamma_{n,k}}, \min_{j \in \mathcal{U}^{(n)}} \frac{SIR_{n,j}}{\gamma_{n,j}}\right\}, \tag{2.16}$$

where $\gamma_{n,k}$ and $\gamma_{n,j}$ are the SIR thresholds corresponding to modulation levels of users k and $j \in \mathcal{U}^{(n)}$. Algorithm B does not simply consider the total induced interference to users, but it also attempts to capture the impact of an assignment on cochannel users, so that the SIR of the user that is closer to the corresponding SIR threshold is maximized over all assignments. Thus, algorithm attempts to increase the number of users with SIRs above certain threshold values. Finally, the APF factor is $A_{n,k} = S_{n,k}T_{n,k}$ and the user with the maximum APF is selected for assignment. When modulation level alone cannot provide further rate benefits, transmission power adaptation is also considered similarly to algorithm A.

2.5.3 Description of Algorithms

The only difference in algorithms A and B is the definition of SIF factors. The main steps for both algorithms can be summarized as follows.

- **STEP 0** : Initially activate modulation level control only.
- **STEP 1** : Consider the first subcarrier n. Initially the list of candidate users $\mathcal{L}$ includes all users.

- **STEP 2** : Compute APF factors $A_{n,k}$ for n and users $k \in \mathcal{L}$.
- **STEP 3** : Select user $k^* \in \mathcal{L}$ with the maximum APF factor and assign it to the subcarrier. Remove k^* and all users served by BS i_{k^*} from $\mathcal{L}$.
- **STEP 4** : Update the APFs and IRFs of users in list $\mathcal{L}$.
- **STEP 5** : If list $\mathcal{L}$ is empty, go to Step 9. Otherwise go to Step 6.
- **STEP 6** : If not all $k \in \mathcal{L}$ have $T_{n,k} < 0$, go to Step 2. If $T_{n,k} < 0 \, \forall \, k \in \mathcal{L}$ and power control is not active, activate power control. Go to Step 7.
- **STEP 7** : For each $k \in \mathcal{L}$, find an achievable modulation vector for users in $\mathcal{U}^{(n)}$ and user k, so that $T_{n,k} > 0$. (Start by all entries equal to b_{L_0} and reduce entries, until an achievable vector is found).
- **STEP 8** : If $T_{n,k} < 0, \forall \, k \in \mathcal{L}$, the assignment for subcarrier n is terminated. Go to step 9. Otherwise go to step 2.
- **STEP 9** : End of assignment for subcarrier n. Proceed to subcarrier $n+1$ and repeat the procedure, until $n = N$.

2.5.4 Practical Considerations

The proposed algorithms are centralized, in the sense that global network knowledge in terms of link gains between all pairs of BSs and users in all subcarriers are required. A possible implementation scenario is depicted in figure 2.2. Each user measures the received useful signal power and interference from serving and neighboring BSs in all subcarriers. This measurement procedure can be performed with known pilot symbols that are split in all subcarriers and are transmitted by

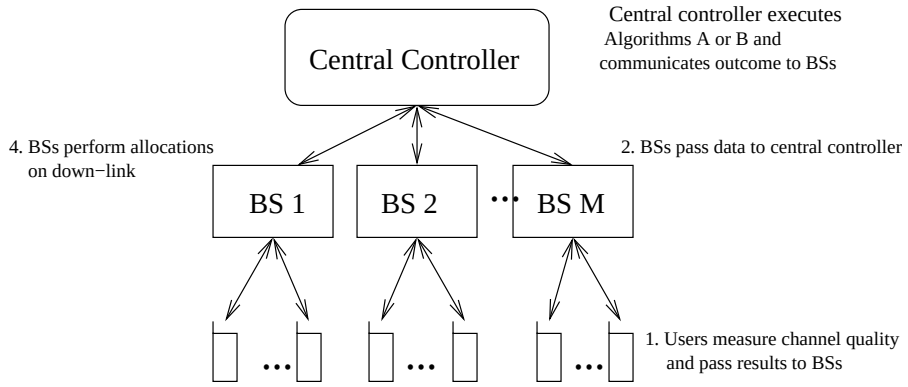


Figure 2.2: Schematic diagram for implementation of algorithms A and B.

BSs in pre-determined, dedicated mini-slots, so that they do not interfere with each other. By measuring the received power of pilot subsymbols in each subcarrier, a user can estimate link gains to all BSs. Then, it passes this information to the serving BS on the up-link. Subsequently, each BS communicates all received link gain data from users in its cell to a central controller. The central controller is a unit with high processing power which is connected to all BSs via high-speed wired or wireless links. The controller executes algorithm A or B and passes the outcome (namely, subcarrier allocation to users, together with modulation level and power control for each subcarrier) to BSs. Each BS takes into account the allocation information and accordingly transmits to users within its cell.

Complexity of algorithms

Finally, a note about the complexity of proposed algorithms. When modulation adaptation is considered, the complexity of algorithms A and B is $O(L_0KM^2)$ per subcarrier for a system with L_0 modulation levels, K users and M BSs. When power control is added, the computationally intensive part of the algorithm is

the determination of an achievable modulation vector. This procedure involves the computation of the eigenvalues of a matrix. In the worst case, the matrix has dimensions $M \times M$ and the eigenvalue computation has complexity $O(M^3)$. Eigenvalue computation may be required up to ML_0 times due to entry reduction of the modulation vector. Thus, the incorporation of power control results in a complexity of $O(L_0KM^5)$ per subcarrier. Such algorithmic complexities are not prohibitive for small- or moderate-sized networks with few BSs which usually arise in practical situations.

2.6 Optimal solution for special cases

We now provide optimal solutions for some special cases of version I and version II of the problem. In version I, where rate requirements of users are provided, the problem is to assign subcarriers to users, such that rate requirements are satisfied with the minimum number of subcarriers. In version II, the problem is to maximize the achievable rate for a set of users.

2.6.1 Problem version I

One modulation level and no power control

We consider a system with $M = 2$ BSs and a set of users, where user k has rate requirements r_k bits/sec. Let U_i be the set of users in BS i, for $i = 1, 2$. Assume that one modulation level b with SIR threshold γ is used and that power control is not employed. We also assume that the set of subcarriers to be allocated to users constitutes a sub-band, so that subcarrier quality in terms of link gain is fixed for a user in all subcarriers. Since $M = 2$, at most two users can share the

same subcarrier. In order to minimize the number of required subcarriers, one has to identify the maximum number of pairs of users from different BSs, so that each pair shares a subcarrier. The number of subcarriers needed for user k in a slot is

$$n_k = \left\lceil \frac{r_k T_s}{Sb} \right\rceil . \tag{2.17}$$

Construct bipartite graph $G = (U \cup V, E)$ as follows. Create one node for each required subcarrier of a user. Thus, $|U| = \sum_{k \in U_1} n_k$ and $|V| = \sum_{k \in U_2} n_k$. An edge (i, j) is added between nodes $i \in U$ and $j \in V$ (denoting subcarriers of users $\alpha \in U_1$ and $\beta \in U_2$ respectively) if SIR thresholds of these users are satisfied, i.e., if

$$\min \left\{ \frac{G_{1\alpha}}{G_{2\alpha}}, \frac{G_{2\beta}}{G_{1\beta}} \right\} \geq \gamma, \tag{2.18}$$

so that these users can use the same subcarrier. A matching $\mathcal{M}$ in a graph G is a subset of edges of G, such that no two edges in $\mathcal{M}$ share the same node. Every edge in $\mathcal{M}$ is called a matched edge. A maximum matching $\mathcal{M}^*$ is a matching of maximum cardinality. As an extension of a theorem stated in [55], we have the following:

Lemma 1 *For one modulation level and no power control, the minimum number of subcarriers required to accommodate users belonging to one of two base stations is equal to the cardinality of a maximum matching in the corresponding bipartite graph plus the number of nodes that are not incident to a matched edge.*

The optimal assignment is as follows. Each edge in $\mathcal{M}^*$ corresponds to two required subcarriers of a pair of cochannel users. Assign each such pair to a separate subcarrier. Then, for each user corresponding to a node that is not incident to a matched edge, consider a separate subcarrier and assign the user to it.

One modulation level and power control

When transmission powers P_1 and P_2 are controllable parameters, the difference from the no-power-control case is the criterion under which two users from different BSs reuse the same channel. The SIRs of cochannel users α and β should satisfy

$$\frac{G_{1\alpha}P_1}{G_{2\alpha}P_2} \geq \gamma \text{ and } \frac{G_{2\beta}P_2}{G_{1\beta}P_1} \geq \gamma. \tag{2.19}$$

By rearrangement of these expressions, we deduce that there exists powers P_1 and P_2 such that SIRs are satisfied, if and only if

$$\sqrt{\frac{G_{1\alpha}G_{2\beta}}{G_{1\beta}G_{2\alpha}}} \geq \gamma. \tag{2.20}$$

Hence, an edge (i, j) is added in the bipartite graph between nodes $i \in U$ and $j \in V$ (denoting subcarriers of users $\alpha \in U_1$ and $\beta \in U_2$ respectively) whenever (2.20) is satisfied. The assignment of users to subcarriers is the same as in the case with no power control. Note that the described approach does not extend to the case of multiple modulation levels, since in that case the number of required subcarriers is not known a priori.

2.6.2 Problem version II

We focus on one subcarrier and we consider two BS-user links. We relax the requirement of discrete modulation (rate) levels and consider continuous rate variables instead. Our goal is to find transmission powers P_1, P_2 and rates b_1 and b_2 so as to maximize the total rate in the subcarrier. The achievable rate with continuous rate variables provides an upper bound on the rate with discrete variables. This problem can be formulated as a non-linear programming problem as follows:

$$\max_{(P_1,P_2,b_1,b_2)} (b_1 + b_2) \tag{2.21}$$

subject to the SIR constraints:

$$\frac{G_{11}P_1}{G_{21}P_2} \geq c(2^{b_1}-1) \;\text{ and }\; \frac{G_{22}P_2}{G_{12}P_1} \geq c(2^{b_2}-1)\; P_i \geq 0,\;\; b_i \geq 0,\; \text{for } i=1,2, \quad (2.22)$$

where $c = -\ln(5\epsilon)/1.5$ as in (1.6). The Lagrangian of the problem is,

$$\begin{aligned}\mathcal{L}(P_1,P_2,b_1,b_2,\lambda_1,\lambda_2) \;&=\; b_1+b_2-\lambda_1\left[c(2^{b_1}-1)G_{21}P_2-G_{11}P_1\right]- \\ &\quad -\lambda_2\left[c(2^{b_2}-1)G_{12}P_1-G_{22}P_2\right], \end{aligned} \quad (2.23)$$

where $\lambda_i \geq 0$, for $i=1,2$ are the Lagrange multipliers. Let $(P_1^*, P_2^*, b_1^*, b_2^*)$ denote the optimal solution. The Karush-Kuhn-Tucker (KKT) conditions are as follows:

$$\left.\frac{\partial\mathcal{L}}{\partial P_i}\right|_{P_i=P_i^*} = 0, \;\text{ and }\; \left.\frac{\partial\mathcal{L}}{\partial b_i}\right|_{b_i=b_i^*} = 0, \;\text{ for } i=1,2 \quad (2.24)$$

$$\lambda_1^*\left[c(2^{b_1^*}-1)G_{21}P_2^*-G_{11}P_1^*\right]+\lambda_2^*\left[c(2^{b_2^*}-1)G_{12}P_1^*-G_{22}P_2^*\right]=0 \quad (2.25)$$

For $(\lambda_1^*, \lambda_2^*) = (1,1)$ we get the solution:

$$P_1^* = \frac{\ln 2}{G_{11}+cG_{12}} \;\;,\;\; P_2^* = \frac{\ln 2}{G_{22}+cG_{21}} \quad (2.26)$$

$$b_1^* = \frac{1}{\ln 2}\ln\left(1+\frac{G_{22}}{cG_{21}}\right) \;\;,\;\; b_2^* = \frac{1}{\ln 2}\ln\left(1+\frac{G_{11}}{cG_{12}}\right). \quad (2.27)$$

2.7 Performance results

2.7.1 Simulation setup

We consider a cellular network of an 8×8 km area with 16 BSs, as illustrated in figure 2.3. Each BS is located in the center of a square grid that represents a cell. The distance between consecutive BSs in the same row or column is 2 km. Users are located in fixed but random positions, uniformly distributed in the area, and each user establishes connection with the closest BS. BSs and users

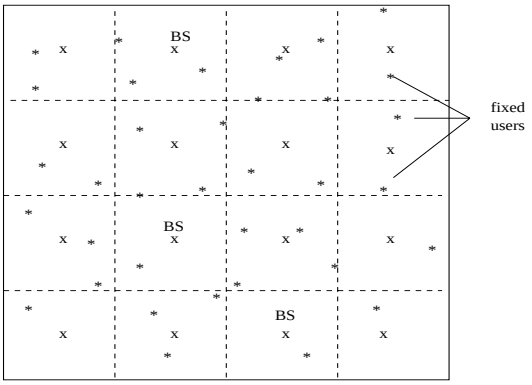


Figure 2.3: The simulated wireless network.

use omni-directional antennas. In order to avoid edge effects, a cell wrap-around technique is used. Each link between a BS and a user is characterized by path loss, shadow fading and multi-path. The path loss causes the received power to decay with distance d from BS according to $1/d^{\kappa}$, where $\kappa = 4$ is the path loss exponent. Shadow fading is modeled by a random variable X that has log-normal distribution with zero-mean and standard deviation $\sigma = 10$ dB. Thus, the received signal power (in dB) for a user at distance d from the BS is,

$$L(d) = L(d_0) + 10 \log X - 10\kappa \log \frac{d}{d_0}, \tag{2.28}$$

where d_0 is known a reference distance and $L(d_0)$ is the received power at d_0. Multi-path fading is modeled with a two-ray model. Each path has a complex gain and a delay. The complex gain is a Gaussian random variable, while the delay is uniformly distributed in $[0, T]$, where T is the symbol duration. The OFDM system has 20 subcarriers. The link gain matrix $\mathbf{G}^{(n)}$ is constructed for all subcarriers by the model above. A target BER of 10^{-3} is assumed for users and the SIR thresholds corresponding to different modulation levels are found for this BER by (1.6).

We focus on subcarrier reuse and compare the performance of different versions

of algorithms A and B in terms of total achievable subcarrier rate, which is captured by the total number of carried bits from cochannel users in each subcarrier. The channel quality is assumed to remain constant within a time slot. Each experiment consisted of the following steps. First, we randomly generate user locations. For each set of user locations, we create a different instance of link gain matrices for each user in each time slot by changing the shadow fading and multi-path and we find the average over $10,000$ such instances. Unless otherwise stated, each such experiment is repeated for 100 randomly generated user location sets. The outcome is the average of these 100 experiments.

2.7.2 Numerical results

The main goal of the simulations is to evaluate and compare the performance of proposed algorithms A and B for subcarrier allocation. It would also be interesting to assess the relative significance of performing modulation and power adaptation. In particular, for each one of algorithms A and B, we consider the following adaptation schemes:

- Modulation level and power control. The algorithms were presented in section 2.5. First, modulation adaptation alone is applied and power control is subsequently activated to assign more users if possible.

- Modulation level control. The algorithm includes only the part of modulation level adaptation and power control is not considered. Namely, the algorithm of subsection 2.5.3 is executed up to step 6.

- Power control. This algorithm applies the described criteria in order to compute the SIF factors of users. However, since one modulation level is

used, it does not include computation of IRF factors. If the SIR threshold for the employed modulation level is γ, the algorithm proceeds as follows. For each user to be inserted in a subcarrier, the feasibility of the resulting cochannel user set is checked. In other words, it is checked whether γ is an achievable common SIR for users. The matrix condition (2.14) is

$$\frac{1+\gamma}{\gamma}\mathbf{P} \geq \mathbf{P}\hat{\mathbf{G}}, \tag{2.29}$$

where matrix $\hat{\mathbf{G}} = \{\hat{G}_{ij}\}$ has elements $\hat{G}_{ij} = G_{ij}/G_{ii}$. By following the same rationale as in subsection 2.5.1, we deduce that if the maximum eigenvalue λ^* of $\hat{\mathbf{G}}$ satisfies $\lambda^* \leq (1+\gamma)/\gamma$, then the modulation level that corresponds to γ is achievable. SIF factors are then computed with powers given by the eigenvector corresponding to λ^*.

The main performance criterion is the average achievable rate per subcarrier, which is equal to the total number of bits of users in a subcarrier. Figure 2.4 illustrates the cumulative distribution function (CDF) of the rate per subcarrier for each of the three adaptation methods above for Algorithm A. Whenever modulation control is included in the adaptation, a scheme with $L = 6$ discrete modulation levels was utilized, while in the scheme with power control, a fixed modulation level (the highest one) was used. For a network with 16 BSs, the maximum achievable rate per subcarrier is 96 bits, since at most one user per BS can be included in a subcarrier. However, the achievable rate is limited by link impairments, such as shadow fading and multi-path. Thus, the achievable rate is usually lower than that value.

We observe that the power control scheme turns out to provide the lowest total rate per subcarrier, whereas the performance of modulation control is significantly

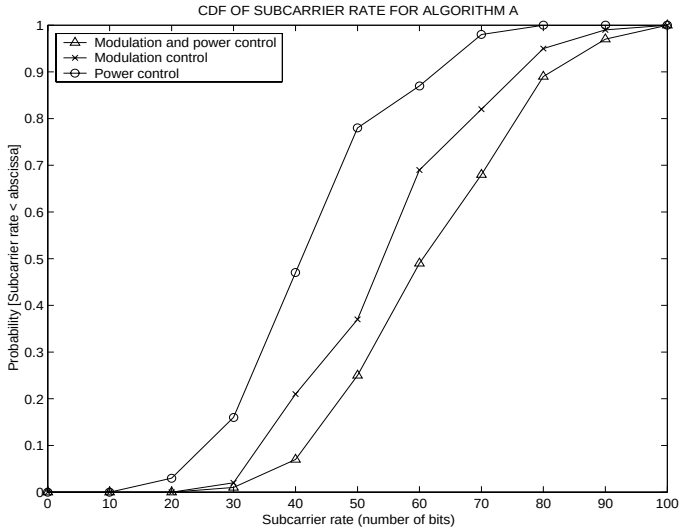


Figure 2.4: Cumulative distribution function of total rate per subcarrier for algorithm A and different adaptation schemes.

better. Joint application of modulation level and power control yields the best performance, since the joint coordination creates appropriate sustainable interference levels and supports higher rates. For example, consider a subcarrier rate value of 60 bits. With modulation level and power adaptation, almost 50% of the subcarriers achieve or exceed this rate, while when only modulation control is considered, this percentage is 30%. For a power control scheme, only 15% of subcarriers have rate higher than 60 bits. The high percentage of subcarriers with high rates in the case of joint modulation and power adaptation indicates that subcarriers are utilized more efficiently for transmission.

The best result in terms of subcarrier rate is therefore achieved by joint modulation and power control. However, the computational complexity of the part of the algorithm when power control is activated is significantly higher than that of modulation control. Recall that several modulation vectors need to be checked

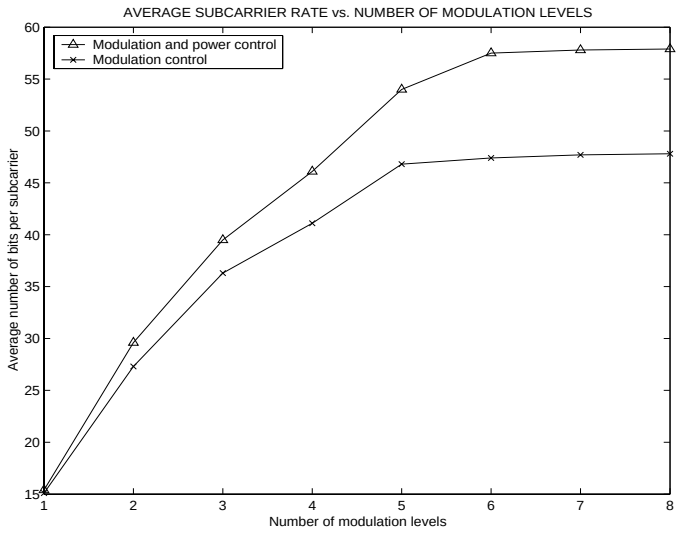


Figure 2.5: Average rate per subcarrier for different number of available modulation levels.

for achievability and each such check involves eigenvalue operations, which are computationally intensive. The computational burden increases with the number of BSs. Thus, whenever the complexity is an issue in an implementable system, modulation level control alone can be used to yield satisfactory performance.

Figure 2.5 shows the average rate per subcarrier as a function of the number of modulation levels. When k modulation levels are utilized, these are the ones with $b_1, b_2, \ldots, b_k$ bits/subsymbol. Simulation results show that the enhancement of an adaptive modulation scheme with power control becomes more beneficial as the number of modulation levels increases up to a certain point. For example, consider the cases of 4 and 5 modulation levels, which can correspond to the case where the highest modulation level is 16-QAM or 32-QAM. The performance gain of joint modulation and power control doubles with respect to that of modulation control with the addition of one modulation level. This can be explained by the fact that

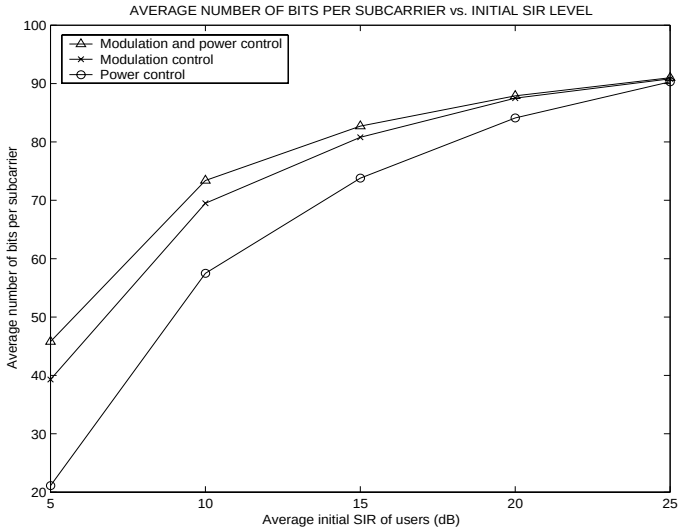


Figure 2.6: Average rate per subcarrier for different initial SIR values.

transmission powers can be controlled such that modulation vectors with higher total rate are achievable.

It can also be deduced that the inclusion of more modulation levels to the system has only marginal contribution to improvement of system performance, since high SIRs are required to maintain fixed BER with high modulation levels. Power control does not have significant impact on system performance when the number of modulation levels increases beyond 6. This can be partly attributed to the fact that the ability of the system to support very high modulation levels depends on the use of high BS transmission powers, which in turn cause excessive interference to other cochannel users. Since the use of additional modulation levels implies additional rate switching complexity, 16-QAM or 32-QAM and lower modulation levels should provide a satisfactory solution in terms of performance and complexity.

In figure 2.6, the performance of the three adaptive schemes is depicted as a function of "user proximity" to BSs. Several thousands of sets of random user positions were generated and for each such set we computed the average user proximity to BSs. User proximity to a BS was mapped to the path loss of the corresponding BS-user path, which in turn depends only on distance from BS. Then, we assumed that a channel is fully loaded, i.e., that all BSs transmit with fixed power and measured the initial SIR, SIR_i^0 of each user i as the ratio of useful signal and interference power

$$SIR_i^0 = \left(\sum_{j=1, j \neq i}^{M} \left(\frac{d_{ii}}{d_{ij}} \right)^4 \right)^{-1} \tag{2.30}$$

where d_{ij} is the distance from user i to the BS in cell j. Random position sets where generated until a sufficient number of scenarios was gathered with certain average initial SIRs. In figure 2.6, a point of initial SIR of x dB in the horizontal axis corresponds to sets of users having average initial SIRs in the range $[x, x+1]$ dB. Thus, a low initial SIR denotes users located relatively far from BSs, or users that are inclined to receive high interference, since they are relatively close to interfering BSs. Simulation results show that adaptive modulation alleviates the effects of interference and that rate performance is significantly better than that with power control. For instance, for an average SINR of 5 dB, the achieved rate per subcarrier for modulation control is twice the rate for power control. This demonstrates the fact that modulation adaptation can be very effective in severe interference environments. Power control alone cannot provide sufficiently good performance, because of the involved SIR balancing concept, which is not profitable in high interference regimes. For milder interference conditions (i.e., higher SINR values), the difference in performance becomes less evident, since all algorithms can combat interference. Joint modulation and power control always

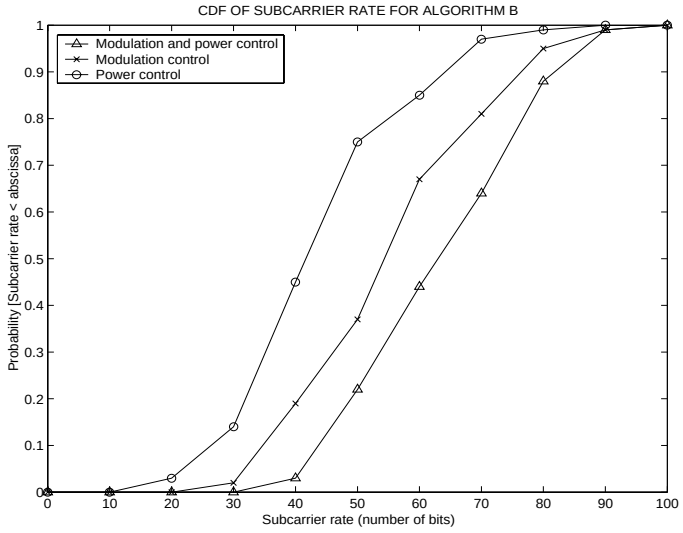


Figure 2.7: Cumulative distribution function of total rate per subcarrier for algorithm B and different adaptation schemes.

achieves the highest rate per subcarrier.

It is also desirable to compare the performance of algorithms A and B, which are based on different principles to perform subcarrier assignment. Figure 2.7 depicts the performance of algorithm B for the three adaptation schemes above. Both algorithms were studied in the same experimental scenarios to allow comparison. The three adaptation schemes exhibit similar trends with these observed in figure 2.4. It is again clear that joint modulation and power control achieves the best performance. Algorithm B is shown to yield significant rate gains compared to algorithm A. In particular, an improvement of $2-4\%$ is achieved in the percentages of subcarriers that achieve or exceed a certain rate and the improvement is more notable when modulation level is used, either alone or with power control. Thus, the percentages of subcarriers that achieve or exceed a rate of 60 bits for the three adaptation schemes are 56%, 33% and 15% respectively. Therefore, algorithm B

provide greater benefits than algorithm A. The explanation lies in the difference of the SIF factors of the two algorithms. Algorithm B takes into account explicitly SIRs of cochannel users and performs the assignment that maximizes the minimum scaled SIR. On the other hand, algorithm A uses a metric that captures the total induced interference of the new user to cochannel users, which may not be efficient in certain cases.

2.8 Conclusion

We applied a cross-layer approach for the problem of subcarrier assignment with modulation and power control which arises in multi-cell multi-user OFDM networks. We defined the framework in which our algorithms take place and characterized the achievable rate region. We identified the complexity of the problem and the need to resort to suboptimal heuristic algorithms. In section 2.6, the optimal solution was found for a system with two BSs by identifying pairs of users that can share a subcarrier and reducing the problem to a maximum matching one. In a network with many BSs and users, the corresponding task would be to identify all possible subsets of users that can share a subcarrier and then consider all possible combinations of modulation levels in order to find the subset which results in maximum subcarrier rate. Clearly, such a procedure becomes intractable for a large system. Therefore some heuristic algorithms with practical value must be devised, which proceed along the lines of the optimal algorithm for the simple case, so that their solution will approximate the optimal one.

Two classes of such algorithms are proposed in section 2.5. The algorithms were based on different criteria to perform user assignment in a subcarrier. The best performance was achieved by algorithm B which each time selects the user

that maximizes the minimum SIR of users in the subcarrier. For each channel assignment problem, the joint adaptation of modulation and power yields the best results, while modulation control alone also performs well. Finally, it was shown that the contribution of multiple modulation adaptation in performance after is marginal after some number of modulation levels.

Chapter 3

Carrier assignment algorithms for OFDM-based networks with channel adaptation

3.1 Introduction

In the previous chapter, we identified and studied the problem that arises from joint consideration of resource allocation and transmission parameter adaptation in a multi-user multi-cell OFDM system. We investigated the impact of modulation and power control on resource reuse which constitutes the primary means of assessing performance of a multi-cell system. We clarified that modulation adaptation adjusts the level of sustainable cochannel interference, while power adaptation can adjust SIRs of users with the objective to maintain an acceptable BER at the receiver. Then, we showed the way in which a synergy can be established between transmission parameter adaptation and resource allocation. The problem reduces to that of activating an appropriate set of BSs and selecting a user from each cell. Transmission from each BS to the corresponding user takes place with controllable

modulation level and power. The emphasis was placed on cochannel set formation for each subcarrier. The packing of a large number of users in each subcarrier and the use of high modulation levels helps in reducing the number of required subcarriers to satisfy certain user rate requirements (according to version I of the problem) or in increasing the total achievable rate (according to version II).

However, this network-wide resource allocation approach has certain deficiencies when considered from the point of view of intra-cell users of a single cell. First of all, the order with which users are assigned in a subcarrier depends also on already assigned cochannel users. The assignment criteria are related to received and induced interference and rate contribution. Thus, a user may not be assigned a subcarrier that is profitable in terms of rate, just because the corresponding transmission causes interference to already assigned cochannel users. Furthermore, due to the nature of the resource allocation problem and its objective, users in the same cell were forced to use different subcarriers in general. Thus, if several users exhibit very good quality in one subcarrier, only one of them is allowed to use the subcarrier and the rest will be assigned to potentially inferior quality subcarriers. It becomes evident that the resource allocation methodology that was used for the multi-cell system is not applicable to a single-cell system, since the resource allocation options for users in the same cell are restricted.

Additional reasons motivate the study of cellular systems on a single-cell basis. In a large-scale system where BSs are not expected to collaborate, each BS acts autonomously and coordinates users within its cell. The BS collects measurement data that are sent by users in the up-link and is responsible for resource allocation and adaptation decisions. In this chapter we focus on the resource allocation problem that comes into stage in single-cell multi-user OFDM systems

with modulation adaptation capabilities. In particular, we investigate the impact of modulation adaptation on subcarrier allocation to users in such multi-carrier systems subject to time resource constraints.

3.1.1 Related work and motivation

Although the issue of adaptive modulation has received considerable attention for the single-link case [56, 57] or the single-channel multi-user case [50], the topic of adaptive modulation in channelized systems remains largely unexplored. A first attempt to consider adaptive modulation in conjunction with time slot allocation has been reported in [58] for a single-carrier system. In that work, the BS receives user measurements about the carrier-to-interference-and-noise ratio at each time slot and subsequently searches for available slots in which a user can use a certain modulation level to support its rate requirements. If an adequate number of slots cannot be found, the modulation level is reduced and a search for more slots that can support this modulation level is initiated.

The problem of resource allocation in the context of OFDM systems has recently attracted much attention. In [59], a scheme that combines OFDM with TDMA for down-link high-rate transmission is presented. The authors consider the issues of efficient organization and flexible allocation of time and frequency resources. The works [42, 60] study the problem of optimal subcarrier allocation to users. They consider the continuous relaxation of the problem, where a subcarrier frequency can be further shared by several users. Among other statements, they conjecture that in the optimal solution, only a few subcarriers will be shared. In [61], the authors present heuristic algorithms to allocate the best set of carriers to each user in terms of channel quality but focus more on relative priorities of users.

In [33] we presented a framework for carrier frequency assignment to users, based on channel quality. The algorithm leads to an efficient allocation, in the sense that each user is assigned to a carrier and occupies the least number of channels.

Most of the aforementioned works that are related to OFDM, focus on subcarrier allocation to users with the objective to maximize the total achievable rate. An underlying time-division scheme is assumed and the resource allocation adaptation is performed in regular time intervals. When the allocation is performed in the frequency domain, each user is assigned a sub-band of subcarriers in one time slot and adaptive modulation is applied in each subcarrier for a user. The same allocation is replicated in subsequent slots. This allocation method is problematic if several users simultaneously have good channel quality in a certain set of subcarriers in a time interval. Then, only one user will be assigned to good quality subcarriers, while the other users may be assigned to lower quality subcarriers. When the allocation is performed in the time domain, each user is assigned a distinct time slot and adaptive modulation is applied in all subcarriers in the slot to transmit data to the user. In such static allocation schemes, the unused subcarriers (as a result of adaptive modulation) within a sub-band or a time slot are wasted and are not assigned to other users. Furthermore, previous works focus on the assignment of subcarriers and do not consider the fact that time resources are also finite and limited. With an appropriate allocation strategy in frequency *and* time domain, resources can be used more efficiently. In this chapter we study the arising issue of resource allocation in a time-slotted OFDM system. We focus on the use of modulation adaptation at the physical layer to create preferences for efficient resource allocation to users under time resource constraints.

3.1.2 Outline of chapter

The rest of the chapter is organized as follows. In section 3.2 we present the model and main assumptions used in our approach and in section 3.3 we provide the formal statement for the problem. In section 3.4 we consider integral user assignment. We characterize the complexity of finding a feasible and an optimal solution and present a heuristic algorithm for the problem. In section 3.5, we study fractional user assignment and present an algorithm that leads to optimal solution for a special case. In section 3.6, our algorithms are categorized in the general framework of heuristics that emanate from Lagrangian relaxation. In section 3.7 we extend our algorithms to the case of time-varying subcarrier quality and derive a meaningful objective when the problem is infeasible. Numerical results are provided in section 3.8. Finally section 3.9 concludes the chapter.

3.2 System model

We consider an OFDM transmission system with N subcarriers and focus on downlink transmission from a single BS to K users in the cell. The N subcarriers constitute a sub-band and are assumed to be part of a system with a total of QN subcarriers, which are organized in Q sub-bands. A time frame of duration T_f secs is assumed, which is divided into C data time slots according to a TDMA scheme. Each time frame includes a portion for control data. Perfect timing synchronization exists among time slots of frames of different subcarriers, so that no ICI among different subcarriers exists.

Data arrive from higher layers and need to be transmitted to different users and the BS needs to utilize time slots and subcarriers for transmission. In order to

simplify the analysis, we assume that subcarriers of a sub-band occupy a contiguous part of spectrum. Therefore, the impact of multi-path channel characteristics (such as path gains and delays) is similar across all subcarriers for a user. Furthermore, we consider very low or no mobility for receivers and surrounding objects, so that the channel coherence time is relatively large and the quality can be regarded as time-invariant within a frame duration. We will extend our treatment for time-varying subcarrier quality in subsection 3.7.1. The amount of cochannel interference experienced at the receiver of a user is assumed to be similar across all time slots of a frame, but the received interference differs in different subcarriers. This situation corresponds to a scenario where neighboring BSs use all time slots but different subcarriers for transmission.

A user i has a bit rate requirement r_i (in bits/sec), which represents the requested rate by the MAC layer. Rate requirements are fixed for one time frame but may change in different frames. To achieve rate requirements, the BS assigns a number of subcarriers and time slots to users. Each subcarrier j that is assigned to user i is modulated by a different number of bits b_{ij}, which constitute the transmitted OFDM subsymbol in j. The number of bits is selected from a finite L_0-element set of available constellations. A given number of subsymbols S can be transmitted in one subcarrier during one slot duration. If the number of time slots occupied by user i in subcarrier j is denoted by α_{ij}, the rate of i in one frame can be expressed as

$$r_i = \frac{S}{T_f} \sum_{j=1}^{N} b_{ij}\alpha_{ij}, \tag{3.1}$$

It is also possible that a user i is assigned to only one subcarrier j. In that case, one modulation level b_i is assigned in all α_i time slots and the user rate is

$$r_i = \frac{S}{T_f} b_i \alpha_i \tag{3.2}$$

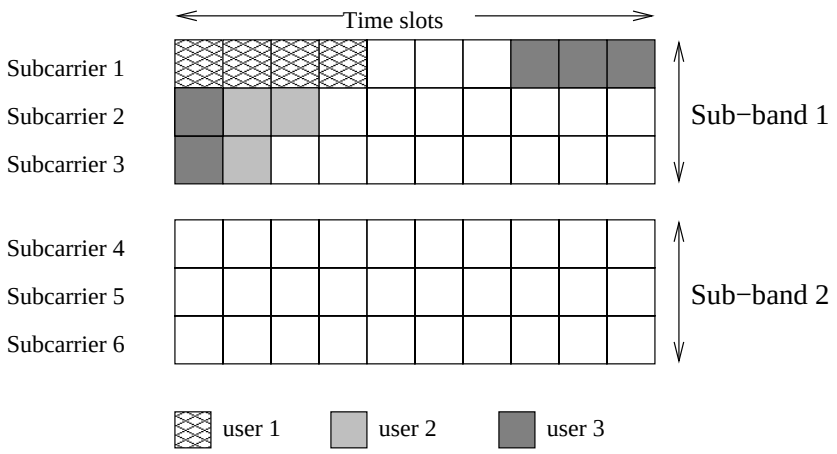


Figure 3.1: Illustrative example of user assignment to subcarriers.

An example with $K = 3$ users, $Q = 2$ sub-bands, $N = 3$ subcarriers per sub-band and $C = 10$ time slots per subcarrier frame is depicted in figure 3.1. User 1 is assigned only to subcarrier 1, while user 2 uses two slots in subcarrier 2 and one slot in subcarrier 3. User 3 occupies slots in all three subcarriers. Then, we say that user 1 is user 1 is *integrally* assigned and users 2 and 3 are *fractionally* assigned.

The BER at the receiver of a user in a subcarrier should be less than ϵ. According to (1.6), a modulation level of b_ℓ bits per subsymbol is associated with a minimum required SINR γ_ℓ. Clearly, for each user and each subcarrier, there exists a maximum modulation level that can be used in the subcarrier by the user, so that an acceptable BER is ensured. Each user measures the useful signal and interference level at each subcarrier with pilot symbols in dedicated time slots at the beginning of the frame and provides the information to the BS in the up-link. Under the assumption of time-invariant subcarrier quality for a frame duration, the BS finds the anticipated SIR per subcarrier for each user and computes the

maximum modulation level that can be supported for each user in each subcarrier.

In the sequel, we use the term "carrier" to refer to an OFDM subcarrier and the term "channel" to refer to a time slot. Unless otherwise stated, the subsequent analysis holds for time-invariant subcarrier quality.

3.3 Problem statement

The link quality experienced by each user depends on propagation factors, such as path loss, shadow fading and multi-path, as well as on cochannel interference at the receiver of the user. The first two propagation factors are independent of subcarrier frequency. With the assumption of the previous section, the effect of multi-path is also independent of the subcarrier for each user. However, interference conditions for a user vary in different subcarriers, due to the different degrees of subcarrier reuse in neighboring BSs. Furthermore, different users perceive different quality for a subcarrier, since user receivers are located in different geographical locations.

The amount of sustainable interference for a user in a channel depends on the modulation level, since the latter determines the minimum required SIR in order to maintain acceptable BER. The number of channels required by the user in order to satisfy rate requirements is also a function of the modulation level, as (3.1) and (3.2) imply. When a high modulation level is assigned to a user in a channel, user rate in the channel increases. As a result, the user will require fewer channels to fulfill its rate requirements. Hence, additional users can be accommodated in the system and capacity is increased. On the other hand, low modulation levels lead to decreased rates for users in channels. Users need more channels in order to satisfy rate requirements and thus fewer users will be accommodated in the system. However, high modulation levels are more vulnerable to interference, require higher

SINR in order to maintain acceptable BER at receiver and therefore can be used only in cases of good channel quality, whereas low modulation levels can be used even in cases of high interference.

In an OFDM system, the allocation of each user in a carrier should ideally entail utilization of minimum number of channels. Equivalently, each user should be assigned to the carrier with the best quality, so that the highest possible modulation level can be used. However, it may happen that most preferable carriers for users are overloaded, in the sense that their available channels cannot accommodate all users. In that case, lower quality carriers need to be utilized for some users, with the expense that more channels (that is, additional bandwidth) will be required in order to satisfy rate requirements of those users. We address the problem of adaptive carrier selection in a time-slotted OFDM system with modulation adaptation capabilities. Our goal is to characterize the structure of the problem, devise efficient algorithms and study the properties of these algorithms.

3.3.1 Problem formulation

We assume that user rate requirements are given and we concentrate on resource assignment algorithms that minimize the number of channels. In particular, the problem we consider is stated as follows:

Problem : *Given a set of users with some rate requirements and given a number of carriers, allocate carriers and channels to users, such that rate requirements are satisfied and the minimum total number of channels are utilized.*

By minimizing the number of utilized channels required to accommodate users, the system is better prepared to accept new users or satisfy additional rate requirements in the future and therefore it responds better to new traffic requirements.

Two cases can be identified with respect to user assignment in carriers:

- Case I : Integral user assignment. Each user is assigned to exactly one carrier and uses channels of this carrier to satisfy its rate requirements.
- Case II : Fractional user assignment. A user can be assigned partially to more than one carriers and uses channels of these carriers to satisfy its rate requirements.

Let α_{ij} be the number of channels required by user i when assigned only in carrier j, for $i = 1, \ldots, K$ and $j = 1, \ldots, N$. From (3.2), α_{ij} can be computed as,

$$\alpha_{ij} = \left\lceil \frac{r_i T_f}{S b_{ij}} \right\rceil, \tag{3.3}$$

where $\lceil x \rceil$ denotes the smallest integer that exceeds x. The long $(NK \times 1)$ vector $\boldsymbol{\alpha} = (\alpha_{ij} : i = 1, \ldots, K, j = 1, \ldots, N)$ specifies completely an instance of the problem. In addition, let x_{ij} denote the *portion (percentage)* of the rate requirements of user i that are satisfied by assignment to carrier j and let $\mathbf{x} = (x_{ij} : i = 1, \ldots, K,\ j = 1, \ldots, N)$ denote the corresponding $(NK \times 1)$ vector. The problem can be stated formally as follows:

$$Z(\mathbf{x}) = \min_{\mathbf{x}} \sum_{i=1}^{K} \sum_{j=1}^{N} \alpha_{ij} x_{ij} \tag{3.4}$$

subject to the constraints:

$$\sum_{j=1}^{N} x_{ij} = 1, \quad i = 1, \ldots, K \tag{3.5}$$

$$\sum_{i=1}^{K} \alpha_{ij} x_{ij} \leq C, \quad j = 1, \ldots, N \tag{3.6}$$

$$x_{ij} \in \{0, 1\}, \quad \text{for all } i \text{ and } j \quad \text{(Case I)} \tag{3.7}$$

$$0 \leq x_{ij} \leq 1, \quad \text{for all } i \text{ and } j \quad \text{(Case II)} \tag{3.8}$$

Constraints (3.5) are the assignment constraints and reveal that user rate requirements should be satisfied by the assignment of each user to subcarriers. Constraints (3.6) are capacity constraints and declare that the capacity of each carrier should not be exceeded. The constraints (3.7) and (3.8) specify the range of values of variables x_{ij}, depending on integral or fractional user assignment. For a problem instance $\boldsymbol{\alpha}$, a user assignment $\mathbf{x}$ is said to be *feasible* whenever it satisfies the constraints above. The user assignment $\mathbf{x}^*$ is *optimal* if $Z(\mathbf{x}^*) \leq Z(\mathbf{x})$ for all feasible assignments $\mathbf{x}$.

3.4 Integral user assignment

3.4.1 Complexity of finding a feasible solution

We consider first the case of integral user assignment, where $\mathbf{x}$ is a binary vector. A first question that arises is that of *feasibility* of an assignment.

(Feasibility Question): *Given N carriers, K users and an instance $\boldsymbol{\alpha}$ of the problem, does there exist a feasible integral user assignment to carriers?*

Let us consider first a simple instance $\mathcal{I}$ of the problem. Assume that for each i, $\alpha_{ij} = \alpha_i$, for $j = 1, \ldots, N$. Thus, a user uses the same number of channels, irrespective of the carrier in which it is allocated. Call each user i an "item" of size α_i and let each carrier of capacity C be a "bin" of size C. Then, it can be shown that the feasibility question is equivalent to the decision version of the bin packing problem.

The optimization version of bin packing is the following:
Given a set of K items with sizes $\alpha_1, \alpha_2, \ldots, \alpha_K$, find the minimum number of bins of capacity C such that all items can be packed into them.

The bin packing problem is known to be NP-hard [53]. Furthermore, the decision version of the bin packing problem,

Given a set of K items with sizes $\alpha_1, \alpha_2, \ldots, \alpha_K$ and an integer N, determine whether it is possible to pack all items in N bins of size C.

is known to be NP-Complete. Since the instance $\mathcal{I}$ of our problem is equivalent to the decision version of bin packing, instance $\mathcal{I}$ is also NP-Complete.

Next, we need to show the NP-Completeness of the more general problem, where the number of utilized channels depends on the carrier in which it is allocated. In order to show that, we use the method of reduction to transform instance $\mathcal{I}$ to an instance $\mathcal{I}'$ of the general problem. Consider first the case of $N = 2$ carriers and K items with sizes $\alpha_1, \alpha_2, \ldots, \alpha_K$ and assume without loss of generality that K is an even number, namely $K = 2\kappa$, for some integer κ. Given the instance $\mathcal{I}$, we construct an instance $\mathcal{I}'$ of the general problem as follows. We have κ users, with $\alpha_{i1} = \alpha_i$, and $\alpha_{i2} = \alpha_{\kappa+i}$, for $i = 1, \ldots, \kappa$. The carrier capacities in $\mathcal{I}'$ are C and $C + \sum_{i=1}^{\kappa}(\alpha_{i2} - \alpha_{i1})$ respectively. Then, instance $\mathcal{I}'$ is equivalent to $\mathcal{I}$, in the sense that a feasible allocation for $\mathcal{I}'$ exists if and only if a feasible allocation for $\mathcal{I}$ exists. Hence, the feasibility problem is proved to be NP-Complete. As a result, there exists no algorithm of polynomial complexity that proves the existence or non-existence of a feasible assignment, unless it involves exhaustive search over all possible assignments of users to carriers.

3.4.2 Complexity of finding an optimal solution

In the previous subsection, we proved that a feasible solution cannot be determined in polynomial time. We now turn our attention to finding the optimal solution of problem (3.4), subject to constraints (3.5)-(3.7). The problem takes the form

of the Generalized Assignment Problem (GAP) [62]. Since the decision version of the problem was shown to be NP-Complete, it is anticipated that the optimization version, namely the GAP problem, is NP-Hard. Indeed, GAP has been proved to be NP-Hard in [63]. Thus, the determination of the optimal solution is possible only with enumeration of all feasible assignments and selection of the one that yields the minimum cost (3.4). Due to the exponential complexity of the procedure, heuristic assignment algorithms of reasonable complexity are sought, which generate feasible assignments and perform close enough to the optimal assignment.

3.4.3 Proposed heuristic algorithm

The purpose of a heuristic algorithm is to determine a feasible assignment which results in a number of utilized channels close enough to that provided by the optimal assignment. The BS collects user measurements about carrier quality and computes the number of channels α_{ij} that each user i requires when assigned to carrier j with (3.3). Based on these values, it considers each user i, sorts parameters $\alpha_{i1}, \ldots, \alpha_{iN}$ in increasing order and constructs a preference list $\mathcal{L}_i$ with the most preferable carriers for assignment to each user i.

A carrier is said to be *overloaded*, if the capacity constraint for the carrier is not satisfied, namely if the carrier capacity is exceeded. A carrier is *under-loaded* when it is not overloaded. The algorithm starts by assigning each user to its best carrier in terms of minimum required number of channels. If after this initial assignment no carrier is overloaded, this is clearly the optimal assignment. On the other hand, if all carriers are overloaded, no feasible assignment exists. The interesting (and most often arising) situation is when there exists a set of overloaded carriers $\mathcal{S}_1$ and a set of under-loaded carriers $\mathcal{S}_2$ after initial assignment.

Fix attention to carriers j and k, where j is overloaded and k is under-loaded after initial user assignment. Users should be transferred from the overloaded (and most preferable) carrier to the under-loaded (and less preferable) one, if there is sufficient capacity in the latter. Users must be transferred so as to induce the minimum additional increase in channel occupancy. For each user i in the overloaded carrier j, we construct a *User-Carrier Transfer Factor (UCTF)* with respect to the tentative transfer of user i from carrier j to k as follows,

$$\Lambda_i\left(j \rightarrow k\right) = \frac{\alpha_{ik}}{\alpha_{ij}}\,, \tag{3.9}$$

where $\Lambda_i\left(j \rightarrow k\right) \geq 1$. This factor captures the transfer "efficiency". Among all candidate users, we transfer the one that causes the minimum inefficiency, i.e., the minimum additional increase in utilized channels. Clearly, user transfers with small UCTF values should take place first. If two or more users have equal UCTF values, ties are broken by index assignment to each user. A feasible solution to the problem is an assignment of each user to a carrier, such that all user requirements are satisfied and no carrier is overloaded.

Consider first the case of $N = 2$ carriers, j and k. Assume that after initial assignment carrier j is overloaded and k is under-loaded. Let $\mathcal{U}_j$ denote the set of users assigned in carrier j. The idea is to select user i_0 in carrier j, such that,

$$i_0 = \arg\min_{i \in \mathcal{U}_j} \Lambda_i\left(j \rightarrow k\right), \tag{3.10}$$

and transfer it to carrier k. User transfers are performed until either both carriers become under-loaded or both become overloaded. In the former case we have a feasible solution and in the latter case no feasible solution exists.

Consider now the general case of $N > 2$ carriers with a set $\mathcal{S}_1$ of overloaded carriers and a set $\mathcal{S}_2$ of under-loaded carriers. Let N_1, N_2 denote the numbers of

overloaded and under-loaded carriers. If $N_1 = 1$ and $N_2 > 1$, we start moving users from that overloaded carrier (say k) to under-loaded ones. In that case, we must select a user i_0 in carrier j and transfer it to an appropriate under-loaded destination carrier k_0, such that

$$(i_0, k_0) = \arg \min_{\substack{i \in \mathcal{U}_j \\ k \in \mathcal{S}_2}} \Lambda_i \left(j \rightarrow k\right). \tag{3.11}$$

If $N_1 > 1$ and $N_2 = 1$, we select a user i_0 from an overloaded carrier j_0 and transfer it to the one under-loaded carrier (say k). In the more general case where $N_1 > 1$ and $N_2 > 1$, there are several overloaded and under-loaded carriers. Then, we need to select a user i_0 in an overloaded carrier j_0 and move it to an under-loaded carrier k_0, so that the minimum number of additional channels is incurred. Namely, we select (i_0, j_0, k_0), such that

$$(i_0, j_0, k_0) = \arg \min_{\substack{j \in \mathcal{S}_1, k \in \mathcal{S}_2 \\ i \in \mathcal{U}_j}} \Lambda_i \left(j \rightarrow k\right). \tag{3.12}$$

User reassignments from overloaded to under-loaded carriers terminate at a stage where all carriers become under-loaded or all become overloaded or when no further reassignments from an overloaded to an under-loaded carrier are possible because they result in at least one carrier being overloaded. In this latter case, the procedure stops when a user cannot be further reassigned to any carrier without at least one carrier being overloaded. There exist rare cases when a reassignment of a user i from an overloaded carrier j to another carrier k leads to an infeasible assignment, while the reassignment of user i' with $\Lambda_{i'}(j \rightarrow k) > \Lambda_i(j \rightarrow k)$ leads to a feasible assignment. The algorithm should take into account this situation, by checking the possibility of reassignment for all users in the overloaded carrier j, when the reassignment procedure seems to terminate. An example where this situation arises is presented in a subsequent section.

For the case of $N = 2$ carriers, the aforementioned carrier assignment algorithm with the consideration of the special case described above finds a feasible assignment, whenever such an assignment exists.

3.5 Fractional user assignment

Consider now the case of fractional user assignment, where a user can be partially assigned to more than one carriers. In that case, $\mathbf{x}$ is a continuous vector with entries $0 \leq x_{ij} \leq 1$. The problem of minimizing Z in (3.4) subject to constraints (3.5), (3.6) and (3.8) is identified as a Linear Programming (LP) problem. The feasible set of solutions is the polytope that is defined by these constraints. The most popular algorithm to solve LP problems is the simplex algorithm [64]. However, due to the fact that the complexity increases fast with the number of variables, we attempt to derive simple algorithms that solve simple cases of the problem with the intention to use them for the design of efficient algorithms for more general cases.

3.5.1 The case of $N = 2$ carriers

Consider the case of $N = 2$ carriers, where each carrier has capacity C time slots. Let the number of required channels for user i be α_i and β_i when the user is allocated to carrier 1 or 2 respectively. The problem is to find vector $\mathbf{x} = (x_i : i = 1, \ldots, K)$, with x_i being the fraction of the request of user i that is assigned to carrier 1, such that the total number of user channels in both carriers is minimized. The problem can be formulated as follows:

$$Z = \min_{\mathbf{x}} \sum_{i=1}^{K} [\alpha_i x_i + \beta_i (1 - x_i)] \tag{3.13}$$

subject to the constraints:

$$\sum_{i=1}^{K} \alpha_i x_i \leq C \tag{3.14}$$

$$\sum_{i=1}^{K} \beta_i \left(1 - x_i\right) \leq C \tag{3.15}$$

$$0 \leq x_i \leq 1, \text{ for } i = 1, \ldots, K. \tag{3.16}$$

The objective function (3.13) can be alternatively written as

$$Z = \sum_{i=1}^{K} \left(\alpha_i - \beta_i\right) x_i + \sum_{i=1}^{K} \beta_i \,. \tag{3.17}$$

Since our goal is to minimize the cost Z, the formulation (3.17) implies the following. When $\alpha_i > \beta_i$, the variable x_i should be very small or ideally zero, so as to induce the smallest increase in cost. This means that user i should utilize carrier 2 as much as possible, since it needs fewer channels in that carrier. Furthermore, when $\alpha_i < \beta_i$, variable x_i should obtain larger values or ideally equal to 1, so as to cause larger reduction in the objective function. Then, carrier 1 should be given preference for user i.

An optimal algorithm for $N = 2$ carriers

There exists an algorithm that achieves the optimal solution for $N = 2$ carriers. Each user is initially assigned to the best carrier. If both capacity constraints are satisfied, this is the optimal assignment, whereas if both capacity constraints are not satisfied, no feasible solution exists. If one of the two constraints is satisfied, users are transferred from the overloaded carrier to under-loaded one, such that they induce the minimum additional increase in number of utilized channels. This is captured by ratios α_i/β_i or β_i/α_i for user i, depending on which carrier is overloaded. For example, if carrier 1 is overloaded and carrier 2 is not, users

are transferred from carrier 1 to carrier 2 in increasing order of ratios β_i/α_i until both constraints are satisfied. The last user in carrier 1 whose reassignment renders carrier 1 under-loaded is assigned partially to both carriers. The fraction of requirements that is maintained in carrier 1 is such that the capacity constraint of carrier 1 is *tightly* satisfied. The remaining portion of requirements is assigned to carrier 2. Clearly, no further reassignments can give a better solution, since additional channels would be required.

Properties of the algorithm

The described algorithm has some interesting properties. First, observe that a feasible solution $\mathbf{x}$ involves the fractional assignment of at most one user. This user is the last one to be reassigned from carrier 1 to carrier 2. The other users are assigned entirely to one of the two carriers. Thus, at most one of K variables x_i are fractional, while the rest $K-1$ are 0 or 1. In addition, if $\alpha_i > \beta_i$ for user i, then the optimal solution can have a non-zero coefficient x_i only if carrier 2 is filled to its capacity. In other words, user i can be assigned to "worse" carrier 1, only if the preferable carrier 2 cannot accommodate more portion of this user. The same rationale cannot be directly applied in finding an optimal algorithm for $N > 2$ carriers. The selection of the appropriate carrier for user reassignment is more complicated, since UCTF factors become products of preference ratios for several users and carriers and rearrangement of previous assignments may be needed.

For $N = 2$ and fractional user assignment, the value of the objective function after the end of our algorithm is the same as that achieved by LP. Therefore, it provides a lower bound for all algorithms for fractional user assignment. Furthermore, for any N, the LP solution provides a lower bound on the number of utilized

channels that are achieved by any algorithm that generates a feasible assignment for the integral user assignment problem.

Resemblance with fractional Knapsack problem

The problem statement and the proposed algorithm that finds the optimal solution for $N = 2$ are similar to those related to the fractional Knapsack problem. The fractional Knapsack problem is

Given a set of K items where each item i has weight w_i and value v_i, find portions x_i of each item, so as to maximize the total value $\sum_i b_i x_i$, subject to a total weight constraint, $\sum_i w_i x_i \leq C$.

There exists a greedy algorithm that solves optimally this problem. Items are selected in decreasing order of ratios v_i/w_i. When the weight constraint is violated for an item, this item is selected fractionally, so that weight constraint is tightly satisfied. Clearly, the UCTF ratios β_i/α_i of our problem correspond to ratios v_i/w_i and users are assigned in increasing order of these ratios (or equivalently, in decreasing order of ratios α_i/β_i), so as to maximize the amount by which the overloaded carrier is unloaded and minimize the amount by which the under-loaded carrier is loaded. The Knapsack in our problem is the initially under-loaded carrier.

3.5.2 Example

We now use an example to demonstrate our arguments. Consider the case $N = 2$ carriers and $K = 8$ users and let the problem instance be described by table 3.1. First, let the capacity of each carrier be $C = 12$. According to the proposed algorithm, each user is allocated to its best carrier. This yields the assignment:

Carrier 1: users 1,4,5,8 Carrier 2: users 2,3,6,7.

User ID (i)	Channels in Carrier 1 (α_i)	Channels in Carrier 2 (β_i)	UCTF Ratio (β_i/α_i)
1	4	13	3.25
2	3	2	0.667
3	6	1	0.167
4	2	4	2.0
5	3	4	1.333
6	5	2	0.4
7	10	5	0.5
8	1	3	3.0

Table 3.1: Parameters for the numerical example.

With this assignment, 10 channels are used in each one of carriers 1 and 2. This is the optimal solution, with objective function value $Z = 20$ channels.

As a second problem instance, assume the same parameters as in table 3.1 but now let the carrier capacities be $C_1 = 6$ and $C_2 = 17$. If each user is assigned to its best carrier, carrier 1 is overloaded and carrier 2 is under-loaded. Hence some of users 1,4,5,8 need to be reassigned to carrier 2. We start from the user with the minimum UCTF, namely user 5. After reassignment, we have

Carrier 1: users 1,4,8 (7 occupied channels).

Carrier 2: users 2,3,5,6,7 (14 occupied channels).

Since carrier 1 is still overloaded, we continue with the user having the minimum UCTF among users 4,5,8, namely user 4. With integral assignment of user 4, we get:

Carrier 1: users 1,8 (5 occupied channels)

Carrier 2: users 2,3,4,5,6,7 (18 occupied channels).

Clearly, the integral assignment of user 4 cannot be carried out, since carrier 2 becomes overloaded. However, if the user with the next smallest UCTF is selected (user 8), we have:

Carrier 1: users 1,4 (6 occupied channels).

Carrier 2: users 2,3,5,7,8 (17 occupied channels).

The algorithm terminates at this point, since no further reassignments are possible. The number of utilized channels is $Z = 23$.

Assume now that fractional user assignment is allowed. In the first step, user 5 is reassigned to carrier 2 as before. However, now user 4 can be fractionally assigned to both carriers. One out of 2 required channel units of user 4 are maintained in carrier 1, but the other 1 unit is reassigned to carrier 2. Hence we have:

Carrier 1: users 1,4 (with fraction 1/2),8 (6 occupied channels)

Carrier 2: users 2,3,4 (with fraction 1/2),5,6,7 (16 occupied channels).

The algorithm terminates at this point. Thus, with fractional assignment, user requirements are satisfied with $Z = 22$ channels.

3.6 Performance bounds

In sections 3.4 and 3.5, we studied the problems of fractional and integral user assignment to carriers with the objective to minimize the total number of utilized channels. The fractional assignment problem was formulated as a LP one. We proposed an algorithm that solves the problem optimally for $N = 2$ carriers. For the general case of $N > 2$, the simplex algorithm can be used to solve the problem.

In this section, our goal is to provide good performance bounds for the inte-

gral assignment problem, which was shown to be NP-hard. In particular, we are interested in good (large) lower bounds in the number of utilized channels. The significance of such bounds is that they can serve as measures for performance evaluation of practical heuristic algorithms. Furthermore, the procedure of deriving the bound may draw guidelines for obtaining good feasible solutions or for designing efficient heuristic algorithms. In section 3.5.1 it was mentioned that if LP is used to solve the fractional assignment problem, the resulting number of utilized channels, Z_{LP}, constitutes a lower bound for the number of channels under integral user assignment. This property stems from the so-called LP relaxation, whereby the integral constraints $\mathbf{x} \in \{0,1\}^{NK}$ are relaxed to constraints $0 \leq x_{ij} \leq 1$, where variables $\{x_{ij}\}$ are continuous.

3.6.1 Lagrangian relaxation

There exist methods which can provide better (larger) lower bounds for minimization problems with integer variables. One of these methods is Lagrangian relaxation [65]. In Lagrangian relaxation, one or more sets of constraints of the original problem are relaxed (eliminated). Each of the relaxed constraints is multiplied by a price (the Lagrange multiplier) and is added to the objective function. The resulting problem without the relaxed constraints is usually easier to solve than the original problem. Given a set of Lagrange multipliers, the solution to the relaxed problem provides a lower bound on the objective function of the original problem. The corresponding Lagrangian dual problem is to select the values of the multipliers so as to maximize the lower bound.

The formulation of the integral assignment problem is stated again here for

convenience

$$Z_I(\mathbf{x}) = \min_{\mathbf{x}} \sum_{i=1}^{K} \sum_{j=1}^{N} \alpha_{ij} x_{ij} \tag{3.18}$$

subject to:

$$\sum_{j=1}^{N} x_{ij} = 1, \quad i = 1, \ldots, K \quad \text{(Assignment constraints)} \tag{3.19}$$

$$\sum_{i=1}^{K} \alpha_{ij} x_{ij} \leq C, \quad j = 1, \ldots, N \quad \text{(Capacity constraints)} \tag{3.20}$$

$$\mathbf{x} \in \{0,1\}^{KN}. \tag{3.21}$$

We can identify two Lagrangian relaxations LR1 and LR2, depending on which constraints are relaxed.

Relaxation LR1 and relation to our algorithm

In relaxation R1 we relax the capacity constraints. Thus, we need to assign each user to exactly one carrier, but the capacity constraints are ignored. For a given Lagrange multiplier vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$, we define

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) &= \sum_{i=1}^{K} \sum_{j=1}^{N} \alpha_{ij} x_{ij} + \sum_{i=1}^{K} \sum_{j=1}^{N} \lambda_j (\alpha_{ij} x_{ij} - C) \\ &= -KC \sum_{j=1}^{N} \lambda_j + \sum_{i=1}^{K} \sum_{j=1}^{N} \alpha_{ij} (1 + \lambda_j) x_{ij}. \end{aligned} \tag{3.22}$$

The original problem is then written as

$$\begin{aligned} Z_I(\mathbf{x}) = \quad & \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) \\ & \text{subject to :} \quad \sum_{j=1}^{N} x_{ij} = 1, i = 1, \ldots, K \\ & \qquad\qquad\quad \mathbf{x} \in \{0,1\}^{KN} \end{aligned} \tag{3.23}$$

and the Lagrangian dual problem is

$$Z_{LD}^{1} = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) \tag{3.24}$$

$$\text{subject to:} \quad \sum_{j=1}^{N} x_{ij} = 1, \ i = 1, \ldots, K \text{ and } \mathbf{x} \in \{0,1\}^{NK}.$$

For given $\boldsymbol{\lambda}$, problem (3.23) is solved by the assignment $\mathbf{x}(\ \boldsymbol{\lambda})$, such that

$$x_{ij^*}(\boldsymbol{\lambda}) = 1, \text{ for } j^* = \arg\min_j \alpha_{ij}(1 + \lambda_j) \tag{3.25}$$

and $x_{ij}(\boldsymbol{\lambda}) = 0$, otherwise. The solution of problem (3.24) can be determined by the sub-gradient method [62, p.173-174].

It turns out that the lower bound provided by LR1 is the same as that provided by LP, i.e., $Z^1_{LD} = Z_{LP}$. LR1 can be used to provide good feasible solutions and defines a class of heuristic algorithms $\mathcal{A}_1$, which is based on user reassignments among carriers. Our proposed algorithm for the integral user assignment falls within this category of algorithms. Initially each user is assigned to the carrier in which it uses the smallest number of channels without any consideration on carrier capacity constraints. Observe that for $\boldsymbol{\lambda} = \mathbf{0}$, the assignment (3.25) to problem (3.23) coincides with our initial user assignment to carriers, where each user is assigned to the best carrier. Furthermore, user reassignment from over-loaded to under-loaded carriers is analogous to Lagrange multiplier updates in the sub-gradient method, which essentially alter the allocations $\mathbf{x}$.

Relaxation LR2 and associated algorithm

In relaxation R2, we relax the assignment constraints. Thus, each carrier has a capacity constraint, but a user can be assigned to more than one carriers. For a given $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$, we have

$$\begin{aligned} \hat{L}(\mathbf{x}, \boldsymbol{\lambda}) &= \sum_{i=1}^{K}\sum_{j=1}^{N} \alpha_{ij} x_{ij} + \sum_{i=1}^{K} \lambda_i (\sum_{j=1}^{N} x_{ij} - 1) \\ &= \sum_{i=1}^{K} \lambda_i + \sum_{i=1}^{K}\sum_{j=1}^{N} (\alpha_{ij} + \lambda_i) x_{ij}. \end{aligned} \tag{3.26}$$

The original problem is then written

$$\begin{aligned} Z_I(\mathbf{x}) = \quad & \min_{\mathbf{x}} \hat{L}(\mathbf{x}, \boldsymbol{\lambda}) \\ & \text{subject to :} \quad \sum_{i=1}^{K} \alpha_{ij} x_{ij} \leq C, \ j = 1, \ldots, N \\ & \qquad\qquad\quad \mathbf{x} \in \{0,1\}^{KN} \end{aligned} \tag{3.27}$$

and the Lagrangian dual problem is

$$\begin{gathered} Z_{LD}^2 = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \min_{\mathbf{x}} \hat{L}(\mathbf{x}, \boldsymbol{\lambda}) \\ \text{subject to: } \textstyle\sum_{i=1}^{K} \alpha_{ij} x_{ij} \leq C, \ j = 1, \ldots, N \text{ and } \mathbf{x} \in \{0,1\}^{NK}. \end{gathered} \tag{3.28}$$

For given $\boldsymbol{\lambda}$, problem (3.27) becomes a set of N Knapsack problems, one for each carrier. Each Knapsack problem can be solved in time $O(KC)$ by using a recursive algorithm [66]. The significance of LR2 lies in the fact that the associated lower bound is higher that that of LP, namely $Z_{LD}^2 \geq Z_{LP}$ [62].

Relaxation LR2 gives rise to another class of algorithms $\mathcal{A}_2$. In analogy to LR1 and class of algorithms $\mathcal{A}_1$, a new family of heuristic algorithms for the integral assignment problem could be designed as follows. Each carrier is treated separately and users are assigned to each carrier as in a Knapsack problem, with the objective to fill carrier capacity and minimize the incurred cost. After initial assignment, there exist three kinds of users: users of set S_1 that are assigned in one carrier, users in set S_2 that are assigned to more than one carrier and users in set S_3 that are not assigned to any carrier. Carrier assignments of users in S_1 should not be changed. The idea is to reassign channels of a user in $i_2 \in S_2$ to a user $i_3 \in S_3$ in a carrier, provided that carrier quality for i_3 is equal or better than quality of i_2. Preference should be given to users in S_2 that use several channels in a carrier and to users in S_3 that use few channels in a carrier so that the number of utilized channels is minimized.

3.7 Further considerations and extensions

3.7.1 Time-varying channel quality

In the previous discussion, we assumed that all channels in a carrier are of the same quality for each user. Based on information about carrier quality, the maximum sustainable modulation level was found for every user in a carrier and thus the number of required channels was computed with (3.3). The user was assigned to the carrier in which it used the minimum number of channels. We now consider the case of time-varying channel quality, where the quality of each individual channel in a carrier changes.

When channel quality changes, different modulation levels can be used in each channel and thus the number of required channels for a user in a carrier cannot be determined a priori. In addition, apart from carrier assignment to users, the additional issue of *channel* assignment within a carrier arises. In order to satisfy rate requirements of a user, the assignment algorithm should specify not only the number of required channels, but also the individual channels that need to be used within a carrier. One could think of a greedy procedure by which the required number of channels for a user in a carrier can be found: the user should be given channels according to the modulation level that it can use in each channel, starting from the channels in which higher modulation levels can be used. However, this approach may result in assignment of the same channels to more than one users after user assignment to carriers. In that case, complicated channel rearrangement procedures are needed in order to ensure that each channel in a carrier will be occupied by at most one user.

The BS knows the quality of each channel for each carrier. For each user i

and channel s in carrier j, let $b_{ij}^{(s)}$ denote the maximum modulation level that user i can sustain when assigned to channel s of carrier j. The modulation level is selected from the finite L_0-element set $\{b_1, b_2, \ldots, b_{L_0}\}$ and depends on channel quality for that user. The number of bits per symbol for each of the modulation levels is assumed to be multiple of the corresponding number of bits per symbol of the minimum (basic) modulation level b_1, i.e., $b_\ell = \ell b_1$, for $1 \leq \ell \leq L_0$. This leads to the conjecture that the rate achieved by a user i in channel s of carrier j if $b_{ij}^{(s)} = \ell_{ij}^{(s)} b_1$ is the same as that when it uses $\ell_{ij}^{(s)}$ channels and in each channel it uses modulation level b_1. In that sense, one channel with modulation level of $b_{ij}^{(s)}$ is equivalent to $\ell_{ij}^{(s)}$ *virtual* channels, each with modulation level b_1 for that user. Then, user i needs

$$\tilde{\alpha}_i = \left\lceil \frac{r_i T_f}{S b_1} \right\rceil \tag{3.29}$$

virtual channels in order to satisfy rate requirements of r_i bits/sec. Since each user i perceives different quality in each channel and carrier, we replace each channel s in a carrier j with $\ell_{ij}^{(s)}$ virtual ones. Then, user i perceives carrier j as having total number of channels (or virtual capacity)

$$C_{ij} = \sum_{s=1}^{C} \ell_{ij}^{(s)}. \tag{3.30}$$

Next, we use the principle of virtual capacity to define preference factors for carrier assignment to users, in accordance to preference factors for the case of time-invariant channel quality. A user should be assigned to the carrier in which it occupies the least *portion* of virtual capacity of the carrier in order to fulfill rate requirements. The portion p_{ij} for a user i and carrier j is

$$p_{ij} = \frac{\tilde{\alpha}_i}{C_{ij}}. \tag{3.31}$$

Each user should initially be assigned to the carrier in which it occupies the least portion of virtual capacity. If overloaded carriers occur after initial assignment, the users need to be reassigned to under-loaded carriers. However, since each carrier is perceived as having different virtual capacity for each user, the notion of overloaded and under-loaded carrier needs to be redefined. A carrier j with a set of users $\mathcal{U}_j$ assigned to it is perceived to be *overloaded for user* $i \in \mathcal{U}_j$ if the total number of virtual channels of users exceeds the virtual capacity of user i. A carrier j is said to be *overloaded* for the user set $\mathcal{U}_j$, if it is overloaded for at least one user in $\mathcal{U}_j$, namely if,

$$\sum_{i \in \mathcal{U}_j} \tilde{\alpha}_i > \min_{i \in \mathcal{U}_j} C_{ij}. \tag{3.32}$$

Finally, a carrier is non-overloaded for a given set of users if it is not overloaded for these users. The algorithm starts by assigning each user to the best carrier. If no carriers are overloaded, this is the optimal assignment. If a carrier is overloaded, however, some users may not perceive the carrier as overloaded. Clearly, the user that needs to be reassigned is selected among those users for which the carrier is perceived to be overloaded, i.e, among set $X_j = \{i \in \mathcal{U}_j : \sum_{m \in \mathcal{U}_j} \tilde{\alpha}_m > C_{ij}\}$. In this case, a user i_0 needs to be reassigned from an overloaded carrier j_0 to a non-overloaded carrier k_0 such that the additional utilized carrier portion is minimal, i.e. such that

$$(i_0, j_0, k_0) = \arg \min_{\substack{j \in \mathcal{S}_1, k \in \mathcal{S}_2 \\ i \in X_j}} \tilde{\Lambda}_i \left(j \to k\right) = \arg \min_{\substack{j \in \mathcal{S}_1, k \in \mathcal{S}_2 \\ i \in X_j}} \frac{p_{ik}}{p_{ij}} \tag{3.33}$$

where p_{ik}/p_{ij} are the UCTF factors of reassignments and S_1, S_2 are the sets of overloaded and non-overloaded carriers.

Remark: In the case of time-varying carrier quality, the BS requires knowledge of channel quality for all channels in all carriers and for all users. Estimation of SIR

for the first channel in each subcarrier for each user can be performed with pilot symbols. When quality changes for different channels within the carrier, the BS can employ prediction and extrapolation techniques to estimate SIR in next slots. In order to accomplish this task, the BS first needs to estimate the instantaneous delay profile in each channel. We do not elaborate on such techniques, which are presented in detail in [19].

3.7.2 Infeasible problem instance

In the algorithm that was presented previously, it was stated that if after initial assignment all carriers are overloaded, then no feasible solution exists. In that case, we need to find a meaningful objective that will specify the action taken by the BS. If the instance of the problem is infeasible, a sensible objective would be to maximize the total rate that can be supported by the system, when all carrier capacity is utilized. This can be expressed as another LP problem as follows:

$$R(\mathbf{x}) = \max_{\mathbf{x}} \sum_{i=1}^{K} \sum_{j=1}^{N} r_i x_{ij} \tag{3.34}$$

subject to the constraints:

$$\begin{aligned} \sum_{j=1}^{N} x_{ij} &= 1, \quad i = 1, \ldots, K \\ \sum_{i=1}^{K} \alpha_{ij} x_{ij} &= C, \quad j = 1, \ldots, N \\ 0 \leq x_{ij} &\leq 1, \quad \text{for all } i \text{ and } j. \end{aligned} \tag{3.35}$$

3.8 Performance results

3.8.1 Simulation setup

We consider a single-cell OFDM transmission system with N subcarriers and focus on subcarrier assignment to users in the cell area. All users have the same rate requirements in bits/sec. Each subcarrier is divided into time slots and is modulated by a number of bits that is selected from a set of $L_0 = 6$ modulation levels. Modulation level i transmits i bits per symbol and is associated with a threshold value γ_i dB, which is calculated by (1.6) for a target BER value of 10^{-3} per slot.

The quality of a subcarrier for a user depends on propagation parameters and interference level. Path loss, shadow fading and multi-path fading are the same across all subcarriers of a user, while interference level differs for every user and subcarrier. The latter represents the amount of activity of neighboring BSs in corresponding subcarriers as perceived by the user. In the simulations, the effect of all parameters above is captured by a composite term I, which differs for each user and subcarrier and is assumed to follow the Gaussian distribution with mean μ dB and standard deviation σ dB. The quality of a subcarrier is the same for all slots. Since rate requirements are equal for all users, the number of slots α_{ij} needed by user i when assigned to subcarrier j depends only on the maximum sustainable modulation level by a user in a subcarrier.

3.8.2 Numerical results

The objective of the simulations is to illustrate the performance of the proposed algorithms and compare it to the derived performance bounds. We consider the following algorithms:

- Best Carrier Selection (BCS) algorithm. This is the presented algorithm in section 3.4.3 for integral user assignment to carriers.

- BCS algorithm with no reassignments of users (BCS-NR). This is the same algorithm, but no user reassignments are performed, once each user is assigned to its best carrier.

- Linear Programming (LP) solution. This corresponds to the case of fractional user assignments and constitutes a lower performance bound over all integral assignment algorithms. It is used as a performance measure for integral user assignment and as a means of assessing performance of fractional user assignment. LP solution is computed by using MATLAB.

- Subcarrier Load Balancing (SLB) algorithm. In SLB, user assignment to carriers is such that carrier loads are as balanced as possible. The SLB algorithm starts by assigning each user to the best carrier, as in the BCS algorithm. Subsequent user assignments are performed so as to minimize the maximum difference in utilized slots in carriers. That is, the appropriate user i is transferred from carrier j to k so as to

$$\min_{i \in \mathcal{U}_j} \max_{(j,k)} |(N_j - N_{ij}) - (N_k + N_{ik})| \tag{3.36}$$

  where N_j, N_k is the total number of slots of users in carriers j, k and N_{ij}, N_{ik} is the number of slots occupied by user i when assigned to carriers j and k respectively.

A problem instance is represented by vector $\boldsymbol{\alpha}$. Different instances may result in a feasible solution or may be infeasible. A first performance measure of interest is the *proportion of feasible solutions* that are achieved by a heuristic algorithm

such as BCS. This measure can be quantified by the following *rate satisfaction ratio*:

$$P = \frac{\text{Total unsatisfied user requirements (bits/sec)}}{\text{Total user requirements (bits/sec)}} = \frac{N_u}{N_t}, \tag{3.37}$$

where N_t represents the sum of all rate requirements of users and N_u is the total unsatisfied user requirements, namely the number of user bits that are not allocated to carriers. These N_u bits/sec belong to users that cannot be accommodated in the subcarrier because capacity is exceeded. Bits of some users remain unallocated whenever a feasible solution does not exist. If the number of feasible solutions is larger, the number of unallocated bits reduces. Thus, this metric attempts to capture the portion of problem instances for which a feasible solution can be derived. Furthermore, the *quality of a feasible solution* is measured by comparing the solution with the optimal solution, namely the LP one. The *efficiency of a feasible solution* is

$$e = \frac{\text{Number of utilized channels from LP}}{\text{Number of utilized channels from feasible solutions}}, \tag{3.38}$$

where $0 < e \leq 1$. A large value for e means that the algorithm performs closer to the optimal solution which is provided by LP.

We consider a static scenario of a system with $N = 10$ subcarriers and $N = 35$ users and study the performance of algorithms with respect to identifying feasible solutions. Figure 3.2 depicts the rate satisfaction ratio as a function of different average SIR ratios. The useful signal power is kept fixed to a value A dB. An average SIR value of μ dB thus corresponds to a situation where the composite term I is a Gaussian random variable with mean A/μ dB and variance (A/σ^2) dB, where $\sigma = 3$ dB. Thus, the SIR per subcarrier takes values over a large enough range and allows different subcarrier qualities for different users. We consider 200

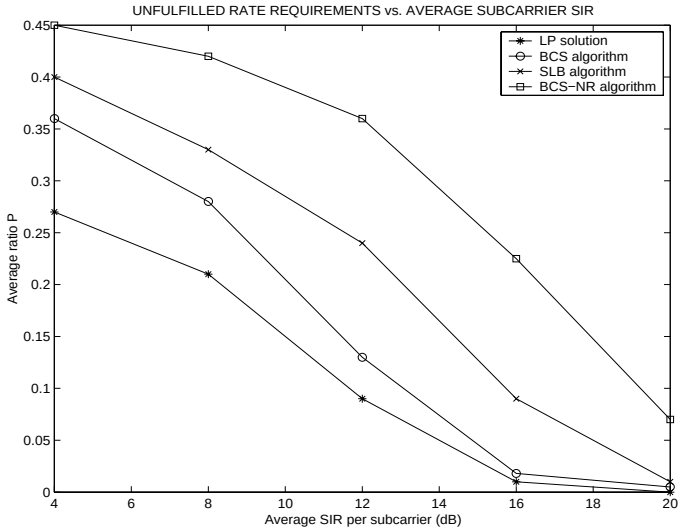


Figure 3.2: Average ratio of unsatisfied user rate requirements for different values of SIR per subcarrier.

scenarios, where each scenario is defined by a different set of interference levels for each user and subcarrier. The results are averaged over these scenarios.

A first observation is that the ability of all techniques to provide feasible solutions increases as the average SIR per subcarrier increases. This occurs since users require fewer time slots to fulfill rate requirements when SIR increases and thus more users can be accommodated in the subcarriers and the percentage of infeasible solutions decreases. As can be observed, the LP solution provides a lower bound on the ratio of unsatisfied user rates P. Since the LP solution corresponds to fractional user assignment, it can be verified that this kind of assignment results in the largest number of feasible solutions. The fractions with which users are allocated to subcarriers are determined by the LP solution. Fractional user assignment is shown to be very efficient for average SIRs larger than 14 dB, in the sense that rate requirements are mostly fulfilled.

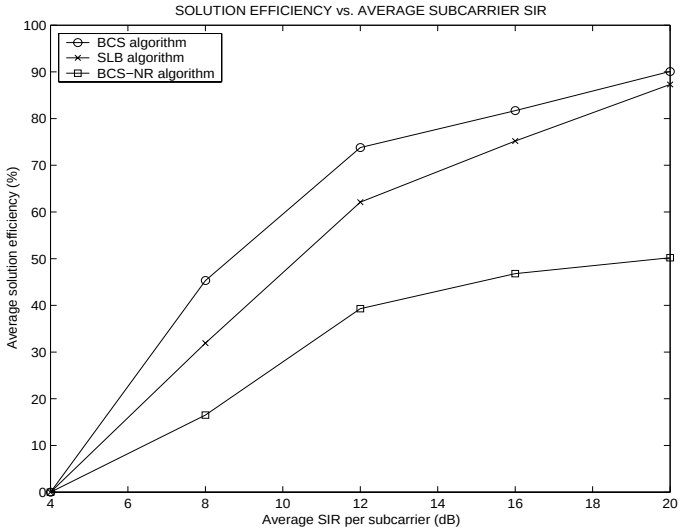


Figure 3.3: Average efficiency of feasible solutions for different values of SIR per subcarrier.

The performance of the proposed BCS algorithm for integral user assignment is also shown in the figure. BCS algorithm always generates a smaller number of feasible solutions than LP. With BCS, fewer users are accommodated in the carrier and higher percentage of user rate requirements remains unsatisfied. For relatively small values of subcarrier SIR (e.g., smaller than 8 dB), the performance of BCS algorithm in terms of unfulfilled rate requirements is inferior to that generated by the LP solution by $30 - 35\%$. When average SIR increases, the performance of BCS approaches that of LP. For adequately high SIRs (greater than 15 dB) both integral and fractional user assignment exhibit practically the same performance. This property is attributed to the fact that for higher SIRs, the required capacity (number of slots) by users reduces, so that a feasible allocation to subcarriers can be derived more easily.

Furthermore, we consider the SLB algorithm, which belongs to the category of heuristics that do not use subcarrier quality to perform the assignment. Due to this feature, the SLB algorithm is effective only in low interference (high SIR) conditions. For smaller SIRs, the differences in quality of different subcarriers incur assignments to inappropriate subcarriers and lead to waste of channels. Finally, we draw the performance curve for the greedy algorithm that simply assigns each user to the best perceived carrier for that user. It becomes evident that user reassignments are very effective, especially for low and moderate SIR cases.

Figure 3.3 illustrates the efficiency of generated feasible solutions, which is captured by its proximity to the LP optimal solution. Feasible solutions for all three heuristics are not generated for SIR = 4 dB. The quality of feasible solutions for BCS improves with increasing SIR. Thus, for moderate SIR values, the BCS solution is within $30 - 40\%$ from the optimal solution while for larger SIRs, it is within $10 - 20\%$. A remarkable conclusion that can be drawn from this figure is that the quality of the solution of the SLB algorithm is close to that of the BCS. Indeed, since we concentrate on feasible solutions and subcarriers are filled almost up to their capacities, a feasible solution both for BCS and SLB algorithms involves an "almost" balanced assignment in different subcarriers. Finally, feasible solutions for algorithm BCS-NR result in at least twice the number of utilized channels compared to the optimal solution.

3.9 Conclusion

We considered the problem of subcarrier assignment to users in a slotted OFDM system with limited time and frequency resources. Our approach placed emphasis on the selection procedure of the optimal subcarrier for each user, subject to

constraints on time resources for each subcarrier. We studied and characterized fractional and integral user assignment to subcarriers and showed that an optimal solution to the former problem can be found even for a simple case. For the integral user assignment, we characterized the complexity of the problem and presented a heuristic algorithm for subcarrier assignment. Our algorithm was categorized in the class of algorithms that stem from Lagrangian relaxation and can serve as an initial step for devising other heuristic methods that fall within the same class of algorithms. Our study focused more on the case of invariable subcarrier quality, but the guidelines for extending our policies to the case of variable subcarrier quality were also provided. Our algorithm results in a satisfactory performance compared to the optimal solution, with regard to the percentage of feasible solutions and the quality of the solution.

Our approach was presented for a sub-band of contiguous subcarriers, such that multi-path characteristics are similar across subcarriers for a user and differences in subcarrier quality for each user are due to different interference levels. Each user can be assigned to one or more subcarriers in the sub-band and time slots within each subcarrier can be utilized for transmission to that user. A similar assignment procedure of users to subcarriers can also be applied for each sub-band. Finally, one or more subcarriers from each sub-band will be used for data symbol transmission to each user, depending on whether integral or fractional user assignment is employed within each sub-band.

Chapter 4

Link adaptation policies for wireless OFDM-based networks

4.1 Introduction

In the previous chapters, we studied cross-layer resource allocation and transmission parameter adaptation issues that arise in the context of OFDM transmission. We concentrated on a snapshot of the system and demonstrated the ways in which resource allocation and transmission parameter adaptation can act collaboratively, with the objective to improve resource utilization and maximize achievable data rates. Our analysis was based on the assumption of perfect channel knowledge at the transmitter. Furthermore, our treatment did not incorporate dynamic channel variation. In such cases, the adaptation regime according to which transmission parameters are adjusted as a response to varying channel conditions is of particular importance.

The ability of a system to provide high data rates is determined to a significant extent by the amount of available channel state information (CSI). In the inherently volatile wireless medium, reliable channel state estimation (CSE) enables the

accurate monitoring of time variations in channel quality, which in turn results in timely adaptation of transmission parameters. Hence, transmission parameters are used, which yield the highest possible instantaneous data rate, while maintaining acceptable BER at the receiver. Furthermore, reliable channel estimation leads to correct resource allocation decisions in the sense that each user is assigned the resources that it perceives as most appropriate.

Irrespective of the multiple access scheme, CSE schemes can be categorized as those being based on physical layer and those based on link layer. A class of physical layer-based channel estimation techniques employs SINR or BER measurements at the receiver and immediate feedback at the transmitter. SINR is measured by sampling the output of the matched filter receiver at the symbol rate, while BER is estimated by observing the output of the detector for a specified time interval. Such measurements do not constitute reliable means of channel estimation in connectionless packet-switched systems, due to the bursty nature of traffic and varying interference level. Furthermore, if link adaptation is carried out based on BER statistics over an observation interval, the system cannot react fast enough to link quality changes, since it takes some time before the change is reflected in the BER statistics.

Another class of channel estimation methods that operate in the physical layer uses pilot symbols. A preamble of known training symbols is used to aid the receiver in identifying channel conditions in terms of link gain. Pilot symbols can also be periodically inserted between transmitted data symbols and the receiver can estimate channel response by interpolation. This latter method is termed pilot-symbol-aided channel estimation and has been studied in [67] for single-carrier systems. Pilot-symbol-aided channel estimation has also been considered for esti-

mating channel quality in OFDM systems. A known time-domain symbol sequence is split into known subsymbols, which provide a means of estimating OFDM subcarrier gain [68, 69]. However, one disadvantage of such techniques may be the significant signaling overhead or the relatively complex signal processing. Moreover, these techniques require knowledge of channel statistics such as delay profile, which cannot be easily determined.

On the other hand, ARQ protocols operate at the link layer and can be employed as channel estimators with a much simpler implementation. In ARQ protocols, information about channel status is provided by the pattern of received positive and negative acknowledgments (ACKs and NACKs) that correspond to transmitted packets. In this chapter, we attempt to create a synergy between the link layer and the physical layer, so as to exploit the simplicity of ARQ protocols and the ability of physical layer for parameter adaptation over a wide range of channel conditions. In particular, our goal is to establish the rule according to which the outcome of the ARQ protocol will trigger physical-layer parameter adaptation in such a way that the transmitter responds better to channel variations.

4.1.1 Related work and motivation

There exist two fundamental techniques for providing reliable and efficient communication over wireless channels: FEC coding schemes and ARQ schemes [7]. In FEC coding schemes, where a fixed number of parity bits are appended to a block of data bits, the throughput is determined by the code rate and it is constant regardless of channel conditions. However, transmission reliability decreases when channel quality degrades and transmission errors occur more often. On the other hand, in ARQ schemes, the throughput depends on the channel status through

the amount of retransmissions, but high reliability is maintained independently of channel state. In order to combine the advantages of both schemes, hybrid ARQ schemes have been proposed. Hybrid ARQ protocols can be distinguished in two main categories: type-I and type-II ARQ. In type-I ARQ schemes, whenever the source receives a NACK for a codeword, it retransmits the same codeword and the receiver attempts to decode it without making use of the original transmission. In type-II ARQ schemes, when the source receives a NACK, it sends additional parity bits for error correction and the receiver attempts to combine the additional provided redundancy with the previously received codeword in order to decode it correctly.

Adaptive hybrid ARQ protocols that use the received ACK/NACK feedback to dynamically control the transmission mode have already been reported in the literature for generic multiple access schemes. In [70] a variable-rate type-I hybrid ARQ scheme is presented, where the code rate is allowed to decrease whenever a NACK is received. In [71] the authors present an adaptive error control scheme with variable-rate codes. The wireless channel is modeled as a finite state Markov chain (MC) and each state corresponds to a packet error rate (PER) and a code rate. The PER is estimated by counting the NACKs in certain observation intervals. For each code C_j, a set of thresholds are computed, such that when this code is used, the system chooses to operate in the state S_i that maximizes the conditional probability $P_{C_j}(r|S_i)$, where r is the NACK counter. This maximization translates to certain inequalities that must be satisfied by r and the thresholds. A similar approach is followed in [14].

The number of successive ACKs and NACKs can be exploited so as to adapt transmission mode according to channel conditions. In [72] the authors consider

the class of ARQ protocols, in which the number of transmitted copies of a data block is varied when a NACK is received. They conjecture that the optimal scheme in terms of throughput efficiency is either sending each block repeatedly until an ACK is received or sending each block a constant (optimum) number of times. More recently, Yao [73] proposed an adaptive GBN ARQ scheme for channels with variable error rates. When the channel is in the high-error-rate state and a number of successive ACKs is received, it switches to the low-error-rate state and transmits one copy of the packet. When the system is in the low-error-rate state and a number of successive NACKs is received, it switches to the high-error rate state and transmits multiple copies of the packet. This work is extended in [74], where the authors present a technique to compute the ACK and NACK thresholds that trigger state transitions. A different route of thought is followed in [75], where the ARQ protocol uses the retransmission history to adapt packet size with the objective to maximize throughput efficiency.

The basic feature of the aforementioned ARQ protocols is that they depend on adaptation of link-layer parameters, such as number of packet copies or packet size. Such adaptation techniques may not fully exploit good channel states to achieve maximum rate, since the range of values over which these parameters are adapted is limited. Furthermore, they may not combat errors effectively in bad channel states, since they do not employ robust enough transmission. On the other hand, when adaptation of physical layer parameters (such as modulation level) is used in ARQ schemes, the aforementioned shortcomings do not exist. Nevertheless, ARQ schemes with adaptive modulation have not been investigated in literature, with the exception of the preliminary work in [76]. In [34] we presented a framework for cooperation of a SR ARQ scheme with modulation and FEC code rate adaptation

for a single link. In that work, link quality was estimated by counting successive ACKs and NACKs.

The ARQ protocols above were studied in the context of single-carrier transmission. The design of ARQ protocols that are suited for OFDM transmission is a largely unexplored topic. In OFDM-based systems, such as IEEE 802.11a and HIPERLAN/2, currently employed ARQ protocols are identical to those used in single-carrier systems [77]. The work in [78] reports an adaptive ARQ scheme for multi-carrier systems that falls within the category of type-II ARQ schemes. Error detection and the request for additional parity bits is performed for each subcarrier. In OFDM, the use of the ACK/NACK feedback to obtain information about channel status presents some novel challenges compared to single-carrier transmission. Each packet symbol is transmitted over parallel subcarriers and each subcarrier has different quality. Ideally, one ACK or NACK should be issued for each separate subcarrier, so that parameter adaptation is performed independently for each subcarrier. This situation corresponds to adaptation on a link basis. However, the ACKs and NACKs at the link layer are usually issued per packet and hence one ACK or NACK comprises several subcarriers. In that case, the ARQ protocol and the associated adaptation mechanisms need to be designed meticulously.

In this chapter, we use the number of successive ACKs and NACKs as a method for estimating channel quality and study the class of transmission parameter adaptation policies that correspond to this method. We start from the case of a single subcarrier and focus on modulation and FEC coding rate adaptation. We use the theory of dynamic programming (DP) and we provide a threshold-based adaptation policy which turns out to be optimal, in the sense that the achievable long-term average throughput per unit time is maximized. Our approach is then extended

to the case of multiple subcarriers, which is more applicable to OFDM.

4.1.2 Outline of chapter

The rest of the chapter is organized as follows. In section 4.2, we present the model and state the assumptions used in our approach. In section 4.3, we consider the problem for a single subcarrier. We formulate the problem of rate adaptation as a Markov decision process (MDP) and prove that the optimal policy has a threshold structure. In section 4.4, the case of multiple subcarriers with similar or different channel qualities is studied. In section 4.5, we outline a heuristic method for providing a suboptimal solution for the case of one subcarrier. Numerical results are illustrated in section 4.6. Finally, section 4.7 concludes this chapter.

4.2 System model

In this section, we describe the adopted model for the single-link case. A link between a transmitter and a receiver corresponds to one subcarrier frequency in the OFDM system. Data arrive from higher layers at the physical layer in the form of a bit stream and need to be transmitted over the link. First, the bit stream enters the FEC encoder, which encodes a k_i-bit data block into a n-bit code word, by appending $n - k_i$ redundant bits. These bits are used by the receiver decoder for error detection or correction. The code rate is $c_i = k_i/n$ and is selected from a set of C code rates $\{k_i/n\}_{i=1}^{C}$. We assume that Reed-Solomon (RS) FEC codes are employed. An (n, k_i) RS FEC code can correct up to $(n - k)/2$ errors. Next, the encoded bit stream is divided into variable-size bit groups, each of which is a subsymbol. Let b_i be the modulation level (in bits per subsymbol), which is

selected from a L_0-element set $\{b_1, \ldots, b_{L_0}\}$. Let s_i be the symbol rate, so that $s_i T_s$ symbols are transmitted in a time slot of duration T_s. Then, the number of bits transmitted in a burst in a time slot duration is $x_i = b_i c_i s_i T_s$. In this work, we assume a fixed symbol transmission rate, so that S symbols are transmitted in a time slot. Then, each pair (b_i, c_i) of modulation level b_i and FEC coding rate c_i is mapped to a rate $r_i = b_i c_i$. The set of available rates is denoted by $\mathcal{R} = \{r_0, r_1, \ldots, r_{N-1}\}$, where $N = CL_0$.

Explicit notification of the utilized modulation level and FEC code rate is provided to the receiver. The integrity of the received burst is checked before the burst is delivered to higher layers at the receiver. The check is performed by computation of the syndrome of the received code word and by decoding of the error detection code. If the burst is found to be correctly received, the receiver acknowledges reception by sending a positive acknowledgment (ACK) back to the transmitter. The transmitter then proceeds to transmission of a new burst. If the burst is erroneously received, a negative acknowledgment (NACK) message is sent, and the transmitter retransmits the same burst. The reverse link from receiver to sender (on which the ACKs and NACKs are sent) is error-free. A selective-repeat (SR) ARQ protocol is used, so that bursts are transmitted continuously and only negatively acknowledged bursts are retransmitted. The reason for the selection of SR is that it provides an upper bound on the achieved throughput for any pure ARQ protocol. Furthermore, SR is selected as the ARQ protocol in OFDM-based HIPERLAN/2 system [79].

Wireless link quality is captured by the average burst SINR γ at the receiver, which characterizes completely the prevailing link conditions during transmission of the burst. A burst transmitted with rate r and received with SINR γ is subject to

error with probability $p_e(r, \gamma)$. For fixed rate r, $p_e(r, \gamma)$ decreases when γ increases. For given SINR γ, $p_e(r, \gamma)$ increases when r increases, since higher rates are more susceptible to errors due to use of high modulation levels or high-rate FEC codes. The time-varying nature of the wireless link is captured via a N-state Markov model, which also accounts for error bursts. For fixed SINR γ, each state S_i is associated with transmission rate r_i and signifies a distinct burst error probability $p_{e,i} = p_e(r_i, \gamma)$. For clarity of presentation, we will assume that $N = 2$, i.e., that there exist two available rates r_0 and r_1, with $r_0 < r_1$.

4.3 Rate adaptation in a single link

4.3.1 Problem statement

When a high rate (i.e. high modulation level or FEC code rate) is used in a burst, more bits are transmitted and therefore throughput is increased. However, high rates render transmitted bursts more susceptible to channel errors. For example, when high-rate FEC codes are used, the burst does not contain many redundant bits and hence it is not well protected from channel errors. When high modulation levels are used, signal points in the constellation diagram become dense and transmission is prone to errors. Thus, more retransmissions may be required in order for a burst to be successfully delivered to the receiver. From that point of view, high rates do not contribute to throughput enhancement, since throughput decreases due to retransmissions.

On the other hand, lower transmission rates convey smaller amounts of information bits on the link but transmitted bits can sustain more channel errors. When lower-rate FEC codes are employed, bursts are better protected from chan-

nel errors since more parity bits are attached. When lower modulation levels are used, signal points in the constellation diagram become sparse. Since transmitted bursts are less error-prone, fewer retransmissions are required.

Clearly, there exists a tradeoff between achievable throughput per transmission and expected amount of retransmissions for a transmission rate. A metric that captures this tradeoff is the throughput at state S_i with rate r_i and SINR γ. This is defined as

$$T_i = T(r_i, \gamma) = r_i[1 - p_e(r_i, \gamma)], \tag{4.1}$$

where term r_i denotes the achievable throughput per transmission and the second term quantifies the effect of retransmissions.

In a time-varying channel with frequent state transitions, the goal is to exploit the feedback of ACKs and NACKs and control the transmission rate, so as to balance the throughput benefit of high rates with the unavoidable retransmissions and ultimately increase throughput. In this study, we focus on the class of adaptation policies that correspond to this specific channel monitoring method with ACKs and NACKs. It is meaningful to study the effect of transmission rate control on throughput over an adequately long time interval. Specifically, the problem that arises is the following:

Problem : *Given a set of transmission rates $\mathcal{R}$ with an achievable throughput and retransmission probability for each rate, and given the pattern of ACKs and NACKs, devise an adaptation policy g, which controls transmission rate based on link quality, such that the long-term average throughput per unit time is maximized.*

4.3.2 Markov Decision Process (MDP) approach

Notation and definitions

The link state is described by a discrete-time Markov chain (MC) $\{X_k\}_{k=0}^{\infty}$, with $X_k = (i_k, j_k, r_k)$, where i_k and j_k is the number of successive ACKs and NACKs respectively until time t_k and r_k is the transmission rate at time t_k. The numbers i_k and j_k, $k = 0, 1, \ldots$, signify the pattern of received ACKs and NACKs. For example, for the ACK/NACK sequence A,A,N,A,N,N,N, we have $i_2 = 2$, $j_2 = 0$, $i_7 = 0$ and $j_7 = 3$. Clearly, at each state X_k, $i_k \cdot j_k = 0$, but $i_k + j_k > 0$, so that either i_k or j_k are zero, but not both. We assume the existence of two large integers, M and M', which specify upper bounds on the successive number of ACKs and NACKs that can be received, so that the state space $\mathcal{X}$ is finite and has size $|\mathcal{X}| = MM'N$.

Let $y_k^{(i)}$ denote the binary decision variable that determines the transmission rate *after* the ACK or NACK at time t_k is received, given that current rate is r_i, with $i = 0, 1$. Thus, when r_0 is currently used, the transmission mode switches to r_1, if $y_k^{(0)} = 1$, and continues to operate at r_0 if $y_k^{(0)} = 0$. Similarly, when the operating rate is r_1, transmission switches to r_0 if $y_k^{(1)} = 1$, or remains at rate r_1 if $y_k^{(1)} = 0$. An *adaptation policy* g is a process $Y = (y_1, y_2, y_3, \ldots)$, where $y_k = y_k^{(0)}$ or $y_k^{(1)}$ is the decision taken at time t_k.

An important subclass of the class of all adaptation policies is the class of stationary policies. A policy is said to be *stationary* if the decision at time t_k depends only on the state of the process at t_k. Let $\mathcal{G}$ denote the set of all stationary policies and consider a policy $g \in \mathcal{G}$. Let $I_{k,i} = I(r_k = r_i)$ be the indicator function, denoting that the rate at time t_k equals r_i. Assume that the ACK or NACK at time t_k is received. Then, depending on the decision taken at t_k, the instantaneous

throughput at t_k with current operating rates r_0 and r_1 for policy g is,

$$T^g_{k,0} = T(r_0,\gamma)[I_{k,0}(1-y^{(0)}_k) + I_{k,1}y^{(1)}_k] \tag{4.2}$$
$$T^g_{k,1} = T(r_1,\gamma)[I_{k,1}(1-y^{(1)}_k) + I_{k,0}y^{(0)}_k].$$

Let $X_k = (i_k, j_k, r_k)$ be the state at time t_k. We define the following operators on states $X = (i, j, r_\ell)$, where current rate is $r_k = r_\ell$ with $\ell = 0$ or $\ell = 1$.

$$\begin{aligned}
A^{(\ell)}_1(0,j,r_\ell) &= (0, j+1, r_\ell) \\
A^{(\ell)}_2(0,j,r_\ell) &= (1, 0, r_\ell) \\
A^{(\ell)}_3(i,0,r_\ell) &= (i+1, 0, r_\ell) \\
A^{(\ell)}_4(i,0,r_\ell) &= (0, 1, r) \\
A^{(\ell)}_5(0,j,r_\ell) &= (0, j+1, r_{\ell\oplus1}) \equiv (0, 0, r_{\ell\oplus1}) \\
A^{(\ell)}_6(i,0,r_\ell) &= (i+1, 0, r_{\ell\oplus1}) \equiv (0, 0, r_{\ell\oplus1}),
\end{aligned} \tag{4.3}$$

where $''\oplus''$ denotes modulo-2 addition. Operators $A^{(\ell)}_1$ and $A^{(\ell)}_5$ are applied when a NACK is received at t_k, while the system operates with rate r_ℓ. With $A^{(\ell)}_1$, the system continues to operate at r_ℓ after processing the NACK, while with $A^{(\ell)}_5$ it switches to another rate. Similarly, operators $A^{(\ell)}_3$ and $A^{(\ell)}_6$ are applied upon reception of an ACK and denote continuing operation with the same rate r_ℓ or switching to a different rate. Operators $A^{(\ell)}_2$ and $A^{(\ell)}_4$ denote the situations when a sequence of NACKs is interrupted by an ACK or a sequence of ACKs is interrupted by a NACK. Note that after rate switching, the ACK/NACK counter is reset.

Transition probabilities and problem objective

Next, we define the transition probabilities and the objective for our problem. When the system operates with rate r_ℓ, the probabilities of a NACK or ACK are

$p_{e,\ell}$ and $(1-p_{e,\ell})$ respectively. Thus, transition probabilities between different states are defined as follows,

$$P(X_{k+1}|X_k, y_k) = \begin{cases} p_{e,1}\,, & \text{if } \; X_{k+1} = A_1^{(1)} X_k \;\text{ and }\; y_k^{(1)} = 0 \\ p_{e,1}\,, & \text{if } \; X_{k+1} = A_5^{(1)} X_k \;\text{ and }\; y_k^{(1)} = 1 \\ p_{e,1}\,, & \text{if } \; X_{k+1} = A_4^{(1)} X_k \\ 1-p_{e,0}\,, & \text{if } \; X_{k+1} = A_3^{(0)} X_k \;\text{ and }\; y_k^{(0)} = 0 \\ 1-p_{e,0}\,, & \text{if } \; X_{k+1} = A_6^{(0)} X_k \;\text{ and }\; y_k^{(0)} = 1 \\ 1-p_{e,0}\,, & \text{if } \; X_{k+1} = A_2^{(0)} X_k. \end{cases} \tag{4.4}$$

The long-term average throughput per unit time for policy g is defined as follows,

$$T^g(x) = \liminf_{n\to\infty} \frac{1}{n} \mathbb{E}_x^g \Big\{ \sum_{k=0}^{n-1} \left(T_{k,0}^g + T_{k,1}^g \right) \Big\} \;\text{ for } x \in \mathcal{X}\,, \tag{4.5}$$

where $\mathbb{E}_x^g \{\cdot\}$ denotes expectation with respect to the probability measure, induced by policy g on the process starting at state x. Therefore, our problem can be formally stated as follows:

$$\max T^g(x) \tag{4.6}$$

$$\text{over all stationary adaptation policies } g \in \mathcal{G},\; x \in \mathcal{X}\;.$$

A policy $g^* \in \mathcal{G}$ is *optimal* in the sense of maximizing long-term average throughput per unit time, if $T^{g^*}(x) \geq T^g(x)$ for all $g \in \mathcal{G}$.

Derivation of the optimal policy for the discounted reward criterion

In order to study the optimization problem (4.6) that involves maximization of long-term average throughput reward, we consider first the corresponding optimization problem associated with the β-discounted reward of horizon n,

$$V_{g,n}^{\beta}(x) = \mathbb{E}_x^g \Big\{ \sum_{k=0}^{n-1} \beta^k \left(T_{k,0}^g + T_{k,1}^g \right) \Big\},\; 0 < \beta < 1. \tag{4.7}$$

Let $V_n^\beta(x)$ be the maximum β-discounted reward of horizon n over all policies $g \in \mathcal{G}$ with initial state $x \in \mathcal{X}$. Since the Markov decision process under consideration has finite state space, the β-optimal reward is achieved by some stationary policy g and satisfies the DP Bellman equation [80]

$$\begin{aligned} V_{k+1}^\beta(x &= (i_{k+1}, j_{k+1}, r_\ell) \\ &= \max_{y_k^{(0)}, y_k^{(1)} \in \{0,1\}} \Big\{ T_{k,0}^g + T_{k,1}^g + \beta\, I_{k,1}\, p_{e,1}\, (1 - y_k^{(1)})\, V_k^\beta(A_1^{(1)} x) \\ &+ \beta\, I_{k,1}\, p_{e,1}\, y_k^{(1)} V_k^\beta(A_5^{(1)} x) + \beta\, I_{k,1}\, p_{e,1}\, V_k^\beta(A_4^{(1)} x) \\ &+ \beta\, I_{k,0}\, (1 - p_{e,0})\, (1 - y_k^{(0)})\, V_k^\beta(A_3^{(0)} x) + \beta\, I_{k,0}\, (1 - p_{e,0})\, y_k^{(0)}\, V_k^\beta(A_6^{(0)} x) \\ &+ \beta\, I_{k,0}\, (1 - p_{e,0})\, V_k^\beta(A_2^{(0)} x) \Big\}. \end{aligned} \tag{4.8}$$

After substituting $T_{k,0}^g$ and $T_{k,1}^g$ from (4.2) and grouping terms together by focusing on coefficients $y_k^{(0)}$ and $y_k^{(1)}$, we have,

$$\begin{aligned} V_{k+1}^\beta(x) &= \max_{y_k^{(0)}, y_k^{(1)} \in \{0,1\}} \Big\{ \Big[T_1 - T_0 + \beta(1 - p_{e,0})(V_k^\beta(A_6^{(0)} x) - V_k^\beta(A_3^{(0)} x)) \Big] I_{k,0} y_k^{(0)} \\ &+ \Big[T_0 - T_1 + \beta p_{e,1}(V_k^\beta(A_5^{(1)} x) - V_k^\beta(A_1^{(1)} x)) \Big] I_{k,1} y_k^{(0)} + \delta \Big\} \end{aligned} \tag{4.9}$$

where δ is a constant, independent of $y_k^{(0)}, y_k^{(1)}$. Hence, we get the following criterion for transition from rate r_0 to r_1:

$$\begin{aligned} \text{If } V_k^\beta(A_3^{(0)} x) - V_k^\beta(A_6^{(0)} x) \le \frac{T_1 - T_0}{\beta(1 - p_{e,0})} &\Longrightarrow y_k^{(0)} = 1 \\ \text{If } V_k^\beta(A_3^{(0)} x) - V_k^\beta(A_6^{(0)} x) > \frac{T_1 - T_0}{\beta(1 - p_{e,0})} &\Longrightarrow y_k^{(0)} = 0. \end{aligned} \tag{4.10}$$

Similarly for the transition from r_1 to r_0 we have

$$\begin{aligned} \text{If } V_k^\beta(A_1^{(1)} x) - V_k^\beta(A_5^{(1)} x) \le \frac{T_0 - T_1}{\beta\, p_{e,1}} &\Longrightarrow y_k^{(1)} = 1 \\ \text{If } V_k^\beta(A_1^{(1)} x) - V_k^\beta(A_5^{(1)} x) > \frac{T_0 - T_1}{\beta\, p_{e,1}} &\Longrightarrow y_k^{(1)} = 0. \end{aligned} \tag{4.11}$$

Consider the transition from r_0 to r_1. Inequalities (4.10) reveal that the transition occurs after a number of successive ACKs are received, which can be considered as an indication that link quality improves. The left hand side of inequalities denotes the difference between throughput efficiencies without and with rate transition. A rate switching occurs if this difference does not exceed a threshold $\tau_0(\gamma)$, which is given by the right hand side of (4.10). Thus, rate switching occurs if operation in current rate is not "efficient enough" in terms of throughput. We observe that the quantity $T_1 - T_0$ in $\tau_0(\gamma)$ can be positive or negative, depending on channel conditions γ and that transition from rate r_0 to r_1 is meaningful only for conditions γ, such that $\tau_0(\gamma) \geq 0$.

The form of the transition criteria (4.10) and (4.11) motivates us to examine the optimality of threshold policies. From the above relations, we observe that if the optimal reward function $V_n^\beta(x)$ is shown to be a concave function of x, then the threshold structure of the optimal policy is evident.

We could not prove the concavity of $V_n^\beta(\cdot)$ by directly using the DP equation. Instead, we demonstrated the concavity of $V_n^\beta(x)$ by formulating a linear programming problem that is equivalent to the MDP problem and by using duality results for it. In the sequel, we state the theorem about the concavity of $V_n^\beta(x)$. The procedure we followed is quite similar to the ones presented in [81, 82], where threshold policies were studied.

Theorem 1 *The function $V_n^\beta(x)$ is concave in x.*

For any $\beta < 1$, the limit $V_\infty^\beta = \lim_{n\to\infty} V_n^\beta(x)$ exists and $V_\infty < \infty$. Indeed, the reward function $V_n^\beta(x)$ is non-decreasing function of x, as can be deduced by definition (4.7). In addition, it is upper bounded as $V_n^\beta(x) \leq (r_0 + r_1)/(1 - \beta)$. Hence, the limit exists.

The minimum cost of the infinite horizon problem is

$$V^{\beta}(x) = \min_{g \in \mathcal{G}} \mathbb{E}_x^g \Big\{ \sum_{k=0}^{\infty} \beta^k \left(T_{k,0}^g + T_{k,1}^g \right) \Big\}. \tag{4.12}$$

Moreover, $V^{\beta}(x)$ is the unique solution of the DP equation (4.8) (for the infinite horizon). From the uniqueness of $V^{\beta}(x)$ and the result above about the existence of the limit $V_{\infty}^{\beta}(x)$, we have that $V^{\beta}(x) = V_{\infty}^{\beta}(x)$. From theorem 1 and the discussion above, we have that $V^{\beta}(x) = V_{\infty}^{\beta}(x)$ is also concave.

The transition criteria (4.10) and (4.11) that were stated for state X_k can also be stated with the optimal reward function $V^{\beta}(x)$. Fix attention to transition from r_0 to r_1. Define now the following operators that denote variations of function $V^{\beta}(\cdot)$ with respect to its arguments:

$$\begin{aligned}
\nabla_r V^{\beta}(i,0,r) &= V^{\beta}(i,0,r_1) - V^{\beta}(i,0,r_0) \\
\nabla_i V^{\beta}(i,0,r) &= V^{\beta}(i+1,0,r) - V^{\beta}(i,0,r) \\
\nabla_{r,i} V^{\beta}(i,0,r) &= [V^{\beta}(i+1,0,r_0) - V^{\beta}(i,0,r_0)] - [V^{\beta}(i+1,0,r_1) - V^{\beta}(i,0,r_1)] \\
\nabla_{i,r} V^{\beta}(i,0,r) &= [V^{\beta}(i+1,0,r_0) - V^{\beta}(i+1,0,r_1)] - [V^{\beta}(i,0,r_0) - V^{\beta}(i,0,r_1)]
\end{aligned}$$

where clearly $\nabla_{r,i} V^{\beta}(i,0,r) = \nabla_{i,r} V^{\beta}(i,0,r)$. Furthermore define $\nabla_{ii} V^{\beta}(i,0,r) = \nabla_i \nabla_i V^{\beta}(i,0,r)$, where $\nabla_{ii} V^{\beta}(i,0,r) \leq 0$, due to concavity of $V^{\beta}(\cdot)$. Also, let $\nabla_{rr} V^{\beta}(i,0,r) = \nabla_r \left(\nabla_r V^{\beta}(i,0,r) \right) = 0$, since there exist only two rates. Now for each (i,r), define the matrix

$$H = \begin{pmatrix} \nabla_{ii} V^{\beta}(i,0,r) & \nabla_{ir} V^{\beta}(i,0,r) \\ \nabla_{ri} V^{\beta}(i,0,r) & \nabla_{rr} V^{\beta}(i,0,r). \end{pmatrix} \tag{4.13}$$

The similarity with the definition of the Hessian matrix for continuous variables is evident. Since the function $V^{\beta}(\cdot)$ is concave, matrix H should be negative semi-definite. A necessary and sufficient condition for that to hold is $\nabla_{r,i} \leq 0$.

Therefore, the following inequality must be satisfied:

$$V^{\beta}(i+1,0,r_0) - V^{\beta}(i+1,0,r_1) \leq V^{\beta}(i,0,r_0) - V^{\beta}(i,0,r_1) \tag{4.14}$$

Then, the left hand side of (4.10) is monotone non-increasing function of i_k, the number of consecutive ACKs. Thus, $V^{\beta}(A_3^{(0)}x) - V^{\beta}(A_6^{(0)}x)$ changes the relation of its value compared to the value of $\tau_0(\gamma_0,\gamma_1)$ at most once during reception of ACKs. We deduce that there must exist a number $i_{\beta}^{*}(\gamma)$, (i.e., a number of successive ACKs), such that the rate switching criterion (4.10) is transformed to the following threshold form:

$$y_k^{(0)} = \begin{cases} 1, & \text{if} \quad i \geq i_{\beta}^{*}(\gamma) \\ 0, & \text{if} \quad i < i_{\beta}^{*}(\gamma) \end{cases} \tag{4.15}$$

By using similar arguments, it can be shown that the optimal adaptation policy for switching from r_1 to r_0 is also of threshold type, where the threshold $\tau_1(\gamma) = j_{\beta}^{*}(\gamma)$ now represents a number of successive NACKs. The optimal policy is

$$y_k^{(1)} = \begin{cases} 1, & \text{if} \quad j \geq j_{\beta}^{*}(\gamma) \\ 0, & \text{if} \quad j < j_{\beta}^{*}(\gamma) \end{cases} \tag{4.16}$$

The transition from r_0 to r_1 is meaningful only when link conditions improve significantly, so that the high throughput potential of r_1 is exploited without frequent retransmissions. Then, the number of successive ACKs is used as an indication that link conditions improve. On the other hand, the transition from r_1 to r_0 is performed when link conditions deteriorate and the increased number of retransmissions causes the throughput to degrade. In such cases, it is preferable to switch to rate r_0, so as to mitigate the detrimental effect of retransmissions. The successively received NACKs are used as a means of detecting deteriorating link conditions.

Optimal policy for the average reward criterion

We derived the optimal policy for the β-discounted reward criterion and we proceed into characterizing the optimal policy for the average reward criterion (4.23) under which the problem was initially stated. We achieve that by using standard results about the relationship of the β-discounted reward problem and the average reward problem [80]. The optimal policy for the average reward problem about transition from r_0 to r_1 is,

$$y_k^{(0)} = \begin{cases} 1, & \text{if} \quad i \geq i_\beta^*(\gamma) \\ 0, & \text{if} \quad i < i_\beta^*(\gamma). \end{cases} \tag{4.17}$$

We now provide the outline of the proof to our argument. In the β-discounted reward problem, the threshold j_β^* depends on β and on the difference $V^\beta(i+1,0,r_0) - V^\beta(i+1,0,r_1)$. In the average reward problem, the threshold depends on the difference $h(i+1,0,r_0) - h(i+1,0,r_1)$, where $h(x) = \lim_{n\to\infty}(V^{\beta_n}(x) - V^{\beta_n}(0,0,0))$, for $x \in \mathcal{X}$, for some sequence $\beta_n \to 1$. The above limit exists and therefore $h(\cdot)$ is well defined [80, p.95-96]. Furthermore, $h(\cdot)$ inherits the structure of $V^\beta(x)$ and therefore it is also concave. The proof is concluded by using similar arguments as those used for the β-discounted problem. By using a similar line of thought, we can deduce the optimal policy for the average reward problem and the transition from r_1 to r_0.

Generalization for multiple rates

The generalization of the rate adaptation policy above for the case of $N > 2$ available rates $\{r_0, r_1, \ldots, r_{N-1}\}$ is straightforward. The system state X_k at time t_k is $X_k = (i_k, j_k, r_k)$, where i_k, j_k are the number of successive ACKs and NACKs at time t_k and r_k is the current rate. If the current rate is $r_i \in \{r_0, \ldots, r_{N-2}\}$ and

the number of contiguous ACKs exceeds a threshold, the system switches to the next higher rate r_{i+1}. On the other hand, if the current rate is $r_i \in \{r_1, \ldots, r_{N-1}\}$ and the number of successive NACKs exceeds a threshold, the system transitions to the next lower rate r_{i-1}.

4.4 Rate adaptation for multiple links

4.4.1 Problem statement

In the previous section, we studied the issue of rate adaptation for a single link that corresponds to an OFDM subcarrier. It was assumed that subsymbols carried by this subcarrier are detected independently from other subcarriers. The receiver checks the integrity of the burst by decoding the error detection code that is appended in the data bits by the transmitter. Depending on the outcome of the decoder, an ACK or NACK is generated for that subcarrier. Rate adaptation is subsequently applied for that subcarrier. The single-link model with ACK/NACK feedback per subcarrier results in accurate tracking of link conditions and allows optimal rate adaptation, according to the described threshold policy. It also can be readily applicable in cases where the bit stream corresponding to each subcarrier is independently encoded. Although the single-link model captures the intuition behind rate adaptation and can be applied in describing the adaptation mechanism for parameters that are employed on a subcarrier basis (such as modulation level) it may be inadequate in describing an OFDM system with adaptation mechanisms that may be applied in the time domain (such as FEC coding).

The latter case is depicted in figure 4.1. The bit stream is encoded with an FEC code and is divided into bit groups, where each group constitutes an OFDM

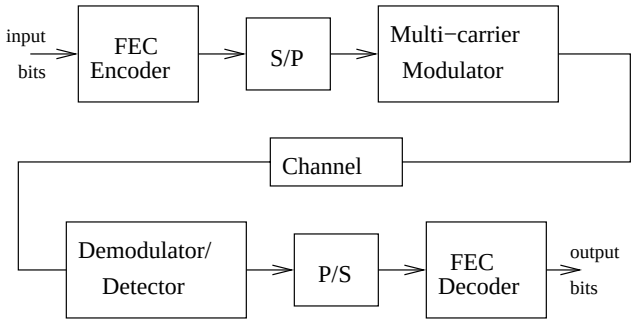


Figure 4.1: Schematic diagram of an OFDM transmission system with FEC encoding.

symbol. Each OFDM symbol is divided into variable-size bit subgroups, the subsymbols. OFDM subsymbols are transformed into time-domain samples via IDFT and are transmitted in the channel. At the receiver, subsymbols are reconstructed from received time samples. After parallel-to-serial conversion, the outcome of the decoder is affected by bits of all subsymbols that are transmitted in corresponding subcarriers. The difference now is that only one ACK or NACK is generated, depending on the output of the decoder and this ACK/NACK corresponds to *all* subcarriers. It is therefore meaningful to construct a model that comprises simultaneous transmission over several subcarriers (links). This multi-link model describes an ARQ protocol at the link layer, where ACKs and NACKs are issued per burst, while the burst is transmitted in parallel over several subcarriers.

4.4.2 Special case: Multiple links of same quality

First, we study the special case of Q subcarriers of the same quality. The subcarriers correspond to a sub-band of contiguous frequencies with similar interference levels, so that the transmitted burst experiences similar quality across all sub-

carriers. The state of the system at time t_k is again captured by the number of successive ACKs and NACKs, as well as the rate which is used in subcarriers. Since all subcarriers have the same quality, the same rate is used in all subcarriers.

A rate adaptation policy is a sequence of decisions about rate switching at each time instant. The difference from the single-link case that was studied in section 4.3 is in the definition of throughput, which is now

$$T(r_i, \gamma) = Q r_i (1 - p_e(r_i, \gamma))^Q. \tag{4.18}$$

The transition probabilities between different states are now given by

$$P(X_{k+1}|X_k, y_k) = \begin{cases} 1-(1-p_{e,1})^Q, & \text{if} \quad X_{k+1} = A_1^{(1)} X_k \ \text{ and } \ y_k^{(1)} = 0 \\ 1-(1-p_{e,1})^Q, & \text{if} \quad X_{k+1} = A_5^{(1)} X_k \ \text{ and } \ y_k^{(1)} = 1 \\ 1-(1-p_{e,1})^Q, & \text{if} \quad X_{k+1} = A_4^{(1)} X_k \\ (1-p_{e,0})^Q, & \text{if} \quad X_{k+1} = A_3^{(0)} X_k \ \text{ and } \ y_k^{(0)} = 0 \\ (1-p_{e,0})^Q, & \text{if} \quad X_{k+1} = A_6^{(0)} X_k \ \text{ and } \ y_k^{(0)} = 1 \\ (1-p_{e,0})^Q, & \text{if} \quad X_{k+1} = A_2^{(0)} X_k. \end{cases} \tag{4.19}$$

For the multi-link case, an ACK while the system is at rate r_0 is issued whenever all bursts of corresponding links are received without error, i.e., it is issued with probability $(1-p_{e,0})^Q$. On the other hand, a NACK while the system operates at rate r_1 is issued when at least one burst of a link is received in error. Thus, the probability of NACK is $1-(1-p_{e,1})^Q$. By following the same rationale as in section 4.3, we can show that the policy that maximizes long-term average throughput per unit time has a threshold structure.

Let θ and ξ be the ACK and NACK thresholds for the single-link case and let θ_Q and ξ_Q be the corresponding thresholds for the case of Q links. In the multi-link case, the probability of ACK decreases and the expected time to receive a

given number of ACKs increases, as compared to the single-link case. Hence, in the multi-link case, we should have $\theta_Q \leq \theta$, so that the multi-link and single-link systems have the same response time to improving link conditions. Similarly, since the probability of NACK increases in the multi-link case, it should be $\xi_Q \geq \xi$ so that the response time to deteriorating link conditions is the same for the single- and multi-link cases.

4.4.3 Extension to multiple links of different quality

We now assume that the user burst is transmitted over Q subcarriers, where each subcarrier is characterized by different link quality. For clarity of exposition, we again assume that two rates r_0, r_1 are available, with $r_0 < r_1$. Hence, a different transmission rate can be used at each subcarrier. The system state X_k at time t_k is $X_k = (i_k, j_k, \mathbf{r}_k)$, where i_k and j_k are the number of successive ACKs and NACKs j_k at time t_k and $\mathbf{r}_k = (r_k^1, r_k^2, \ldots, r_k^n)$ is the rate vector that specifies the utilized rate at each subcarrier n, with $r_k^n \in \{r_0, r_1\}$. The throughput in that case is,

$$T(\mathbf{r}_k) = \Big(\sum_{n=1}^{Q} r_k^n \Big) \prod_{n=1}^{Q} (1 - p_{e,k}^n), \tag{4.20}$$

where $p_{e,k}^n \in \{p_{e,0}, p_{e,1}\}$. The state space is again assumed to be finite and has size $MM'2^Q$, where M, M' are upper bounds on the numbers of contiguous ACKs and NACKs that can be received and 2^Q is the number of possible rate vectors.

The challenging problem that arises in the case of multiple links with different quality is to determine the rate vector that corresponds to rates of all Q individual subcarriers, given the received ACK/NACK feedback about aggregate transmission over all Q subcarriers. When several successive NACKs are received, the individual subcarriers that incur incorrect reception due to deteriorating link conditions

cannot be identified. An incorrect decision about rate reduction in a subcarrier which is not responsible for NACKs, will result in throughput loss. Similarly, the decision of the transmitter to increase or maintain the rate of a subcarrier which actually causes NACKs will also lead to losses. Similar situations of throughput losses arise when incorrect decisions are taken while successive ACKs are received.

The first question is with regard to the policy that maximizes the long-term average throughput per unit time. Each possible rate vector is associated with a total rate, which is given by the sum of its entries. We start by sorting rate vectors in increasing order of total rates. Note that more than one rate vectors can have the same total rate. Thus, for $Q = 3$ subcarriers and two available rates r_0, r_1 there exists one vector of total rate $3r_0$, three vectors of total rate $2r_0 + r_1$, three vectors of total rate $2r_1 + r_0$ and one vector of rate $3r_1$. Let S_0, S_1, S_2, S_3 denote the states that correspond to these four total rate values. A possible adaptation policy could be as follows. When the system operates in state S_i, for $i = 0, 1, 2$ and it receives a certain number of ACKs, it switches to state S_{i+1} with higher total rate. When the system operates in state S_i, $i = 1, 2, 3$ and it receives a certain number of NACKs, it transitions to state S_{i-1} with lower total rate.

An important arising issue is that of determining the single rates in each subcarrier. For example, if the system is in state S_1 with current rate vector (r_0, r_0, r_1) and transition to state S_2 occurs, the new rate vector could be (r_1, r_1, r_0), (r_1, r_0, r_1) or (r_0, r_1, r_1). Since knowledge about individual subcarrier qualities is not available, the appropriate rate vector in state S_2 can be decided only by heuristic methods. One of the three alternative rate vectors could be initially assigned. A NACK feedback would indicate that link conditions do not improve, in which case other rate vectors of the same total rate can be tried. On the other hand, if the system

operates in state S_2 and a certain number of NACKs is received, all rate vectors of total rate $2r_0 + r_1$ should be checked prior to transition to state S_1.

4.5 Heuristic determination of thresholds for the single-link case

The link adaptation policy that maximizes long-term average throughput per unit time involves identification of thresholds $i^*_\beta(\gamma)$ and $j^*_\beta(\gamma)$ that trigger transitions. These thresholds depend on β and on link conditions γ. In a wireless link with time-varying quality, the optimal threshold values will also vary. Furthermore, due to the bursty nature of traffic and time-varying interference level, link conditions γ cannot be reliably estimated. Therefore, the system cannot distinguish the transmission mode that is more efficient in terms of throughput. As a result, the accurate determination of thresholds that are used in the adaptation policy becomes problematic. In this section, we present a practical but suboptimal method to compute the thresholds. Our approach uses a finite-state MC model to compute the expected throughput for rate switching between r_0 and r_1 as a function of the thresholds for fixed link conditions γ. Then, an optimization problem is formulated, that captures the relative proximity of the expected throughput and the ideal one over a wide range of link conditions. The resulting threshold values are independent of link conditions γ.

Let the ACK and NACK thresholds be $i^*_\beta = \theta$ and $j^*_\beta = \xi$ respectively. The system that describes the transition between transmission modes with rates r_0 and r_1 in link conditions γ is modeled as a finite-state Markov chain (figure 4.2). States G_i and B_i denote good and bad link conditions. Rate r_1 is used in good

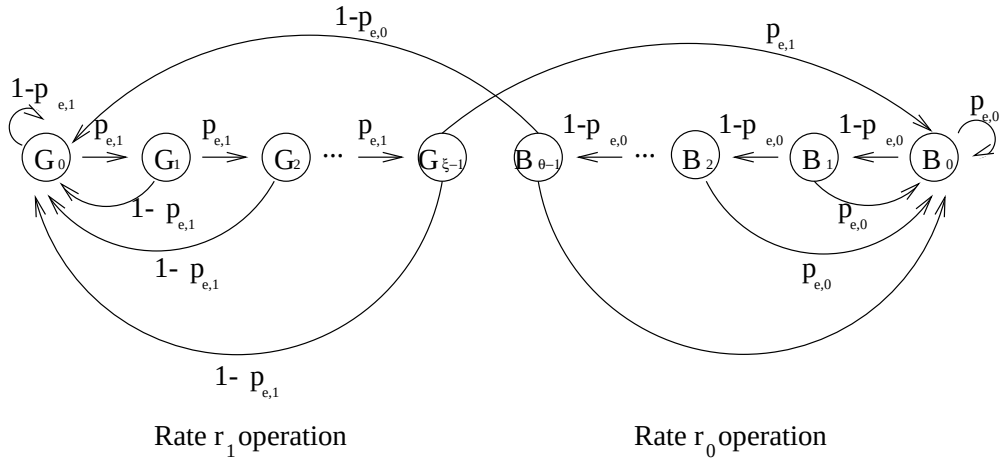


Figure 4.2: Markov chain model for state transitions.

link conditions and rate r_0 is used in bad link conditions. The corresponding burst error probabilities are $p_e(r_1, \gamma) = p_{e,1}$ and $p_e(r_0, \gamma) = p_{e,0}$. When the transmitter uses rate r_1 and has received i consecutive NACKs, it is in state G_i, $0 \leq i \leq \xi - 1$. The transmitter assumes that link conditions change from good to bad ones upon receiving ξ consecutive NACKs and then it switches to rate r_0. If the sequence of NACKs is interrupted by an ACK, it returns to G_0. On the other hand, when the transmitter uses rate r_0 and has received i successive ACKs, it is in state B_i, $0 \leq i \leq \theta - 1$. The transmitter perceives link conditions as changing from bad to good ones, when it receives θ successive ACKs and then it switches to rate r_1. If the series of ACKs is interrupted by a NACK, it returns to B_0. This system is modeled by a $(\theta + \xi)$-state MC.

The steady-state probability $\pi = (\pi_0^G, \pi_1^G, \ldots, \pi_{\xi-1}^G, \pi_0^B, \pi_1^B, \ldots, \pi_{\theta-1}^B)$ can then be easily obtained. Of particular interest are the steady-state probabilities of the good and bad states

$$\pi^G = \sum_{i=0}^{\xi-1} \pi_i^G = \frac{\sigma_1(\theta, \xi)}{\sigma_1(\theta, \xi) + \sigma_2(\theta, \xi)} \tag{4.21}$$

$$\pi^B = \sum_{i=0}^{\theta-1} \pi_i^B = \frac{\sigma_2(\theta,\xi)}{\sigma_1(\theta,\xi)+\sigma_2(\theta,\xi)}$$

where

$$\begin{aligned} \sigma_1(\theta,\xi) &= (1-p_{e,1}^{\xi})\, p_{e,0}, (1-p_{e,0})^{\theta} \\ \sigma_2(\theta,\xi) &= (1-p_{e,1}) \left(1-(1-p_{e,0})^{\theta}\right) p_{e,1}^{\xi}. \end{aligned} \tag{4.22}$$

The steady-state distribution of the MC specifies the portion of time when the system stays in different states. Define the corresponding throughput efficiencies as $T_0(\gamma) = r_0(1-p_{e,0}(\gamma))$ and $T_1(\gamma) = r_1(1-p_{e,1}(\gamma))$. Then, the average throughput of the system for thresholds (θ,ξ) is

$$\overline{T}(\gamma) = T_0(\gamma)\pi^B + T_1(\gamma)\pi^G. \tag{4.23}$$

Given link conditions γ, the system should ideally operate at the rate that guarantees the maximum throughput. Thus, the ideal throughput for link conditions γ is given by

$$T_{ideal}(\gamma) = \max\Big\{r_0(1-p_{e,0}(\gamma)), r_1(1-p_{e,1}(\gamma))\Big\}. \tag{4.24}$$

We follow the rationale outlined in [74] to derive threshold values that are independent of γ. To this end, we consider M discrete SINR values $\{\gamma_i\}_{i=1}^{M}$. For each value of γ_m, we compute $p_{e,i}(\gamma_m)$ analytically and then we find the values $\overline{T}(\gamma_m)$ and $T_{ideal}(\gamma_m)$. Our objective is to find threshold values (θ,ξ) such that the average throughput approximates the ideal one in the range of SINR values. If we consider the squared error criterion, this objective can be written as

$$\min_{\theta,\xi} \mathcal{E} = \sum_{m=1}^{M} \left(\overline{T}(\gamma_m) - T_{ideal}(\gamma_m)\right)^2, \tag{4.25}$$

where the thresholds (θ, ξ) appear in the steady-state distributions π^G, π^B of $\overline{T}(\gamma)$. This is an unconstrained optimization problem and can be solved with numerical methods.

We now attempt to get some intuition about the solution to this problem. Fix attention to $T_0(\gamma)$ and $T_1(\gamma)$. First, note that $p_e(r, \gamma)$ is increasing function of r and thus $1 - p_{e,0}(\gamma) > 1 - p_{e,1}(\gamma)$. For a certain range of values of γ, defined as $\{\gamma : (1 - p_{e,0}(\gamma))/(1 - p_{e,0}(\gamma)) > r_1/r_0\}$, we have $T_0(\gamma) > T_1(\gamma)$. For large values of γ, when $p_{e,i} \approx 0$, we have $T_1(\gamma) > T_0(\gamma)$. By using this heuristic argument and the monotonicity of $T_i(\gamma)$, for $i = 0, 1$, we deduce that the curves corresponding to $T_0(\gamma)$ and $T_1(\gamma)$ should have the form depicted in figure 4.3. That is, there exists a cross-over point γ^*, at which the two curves intersect. Rate r_0 leads to higher throughput for $\gamma < \gamma^*$, due to the fact that retransmissions have significantly negative effect for rate r_1. Rate r_1 performs better for good link conditions, where the high throughput potential is fully exploited. Clearly, the performance curve of the ideal throughput T_{ideal} should be the envelope of the curves of $T_0(\gamma)$ and $T_1(\gamma)$. Thus, the objective function in (4.25) achieves its minimum value, 0, when $T_{ideal}(\gamma) = T_0(\gamma)$, for $\gamma \leq \gamma^*$ and $T_{ideal}(\gamma) = T_1(\gamma)$, for $\gamma \geq \gamma^*$. Taking into account (4.23), we have that the steady-state distribution of bad and good states should be $(\pi^B, \pi^G) = (1, 0)$ for $\gamma \leq \gamma^*$ and $(\pi^B, \pi^G) = (0, 1)$ for $\gamma \geq \gamma^*$. Recall now the definitions (4.22) in terms of σ_1, σ_2. Observe that

$$\begin{aligned} \lim_{\xi \to \infty} \sigma_1(\theta, \xi) = p_{e,0}(1 - p_{e,0})^{\theta}, \quad \lim_{\theta \to \infty} \sigma_1(\theta, \xi) = 0 \\ \lim_{\theta \to \infty} \sigma_2(\theta, \xi) = p_{e,1}^{\xi}(1 - p_{e,1}) \quad \lim_{\xi \to \infty} \sigma_2(\theta, \xi) = 0. \end{aligned} \tag{4.26}$$

In order for the steady-state distributions above to hold, we should have

$$\theta \to \infty, \text{for } \gamma \leq \gamma^* \text{ and } \xi \to \infty, \text{for } \gamma \geq \gamma^*. \tag{4.27}$$

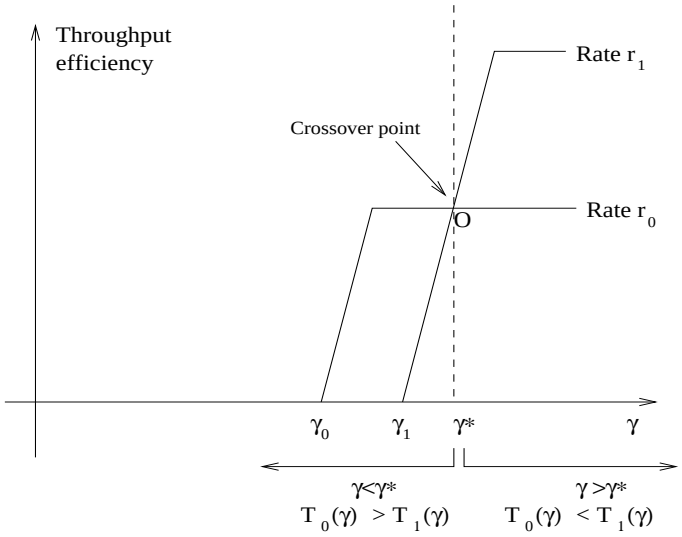


Figure 4.3: Illustrative example for throughput curves for rates r_0 and r_1.

This can be explained intuitively as follows. For $\gamma \leq \gamma^*$, rate r_0 is more preferable than r_1 in terms of throughput. Thus, transitions from r_1 to r_0 should be facilitated, while transitions from r_0 to r_1 should be limited. A small NACK threshold ξ encourages transitions from r_1 to r_0 and a large ACK threshold θ discourages transitions from r_0 to r_1. Ideally, we should have $\theta = \infty$ and $\xi = 1$. On the other hand, for $\gamma \geq \gamma^*$, it is desirable for the system to operate at rate r_1. Thus, transitions from r_0 to r_1 should be encouraged, while transitions from r_1 to r_0 should be prohibited. A small ACK threshold θ favors transitions from r_0 to r_1 and a large NACK threshold ξ discourages transitions from r_1 to r_0. Again, in the ideal case, it should be $\theta = 1$ and $\xi = \infty$.

Finally, we note that this heuristic method can be used in order to compute the ACK and NACK thresholds for multiple links with same or different link quality.

4.6 Simulation results

4.6.1 Simulation settings

First, we consider a single link between one sender and one receiver which represents transmission over a single subcarrier frequency. The goal of our simulation study is to evaluate the performance of the proposed rate adaptation policies and quantify the impact of different parameters on system performance. The primary issue is the determination of thresholds that will be used in the adaptation algorithm. The exact threshold values could be determined from the MDP problem by the policy iteration algorithm [80]. However, these thresholds would be impractical, since they depend on link conditions γ. Since link conditions constantly change, thresholds cannot be accurately estimated. Hence, we apply the heuristic method of the previous section for computing threshold values independent of γ.

We study the performance of the proposed link adaptation algorithm for a system where the controllable parameters are modulation level and FEC code rate. When an M-QAM modulation scheme with $b = \log_2 M$ bits/symbol is used and energy-per-bit-to-noise ratio at the receiver is $E_b/N_0 = \gamma$ dB, the BER can be approximated as [83]

$$BER \simeq \frac{\left(1 - \frac{1}{2^b}\right)}{b} Q\left(\sqrt{\frac{6\, b\, \gamma}{(2^b)^2 - 1}}\right), \tag{4.28}$$

where $Q(x) = 1/\sqrt{2\pi} \int_x^\infty \exp\left(-t^2/2\right) dt$ is the Q-function. The probability of byte error is then $p_b = 1 - (1 - BER)^8$. We encode the bit stream with Reed-Solomon (RS) codes. A (n, k) RS FEC code is constructed as follows. An amount of $n - k$ redundant bytes is appended to k information bytes, so that a FEC block of n protected bytes is created. A (n, k) FEC RS code can correct up to $t = (n - k)/2$

bytes in error. The probability P_F that a FEC block is erroneously decoded is

$$P_F = 1 - \Pr\left(\text{number of bytes in error} \leq t\right) = 1 - \sum_{j=0}^{t} \binom{n}{j} p_b^j (1-p_b)^{n-j}. \quad (4.29)$$

A transmitted burst contains a number of FEC blocks. In order to keep the burst size almost the same regardless of the utilized modulation level or FEC code rate, a different number w of FEC blocks are included in the burst. For instance, for higher modulation levels and/or higher FEC coding rates, more FEC blocks are included. The probability that a burst is received in error is

$$\begin{aligned} p_e &= 1 - \Pr\left(\text{no FEC block is received in error}\right) \\ &= 1 - (1 - P_F)^w \ . \end{aligned} \quad (4.30)$$

We start our study by formulating and solving the optimization problem (4.25). For this problem, it was shown that the pair of ACK/NACK thresholds that minimizes the error between expected and ideal throughput lies in the infinite space. However, large threshold values are clearly impractical, since they lead to extremely slow response to varying link conditions. Hence, we restrict ourselves to finite suboptimal thresholds. These are derived by setting a small positive value for the objective function $\mathcal{E}$ and terminating the algorithm when this value is achieved.

Next, the computed thresholds are provided as input to the OPNET discrete event simulator [84]. For each experiment, we fix the value of E_b/N_0 and compute the probability of burst error p_e for the utilized modulation level and FEC code rate. Potential rate transitions occur according to the MC model depicted in figure 4.2 with the computed thresholds θ and ξ. Link conditions are assumed to remain unchanged. The SR ARQ protocol is adopted and the ACKs and NACKs are received error-free. Experiments cover the range of $5 - 15$ dB.

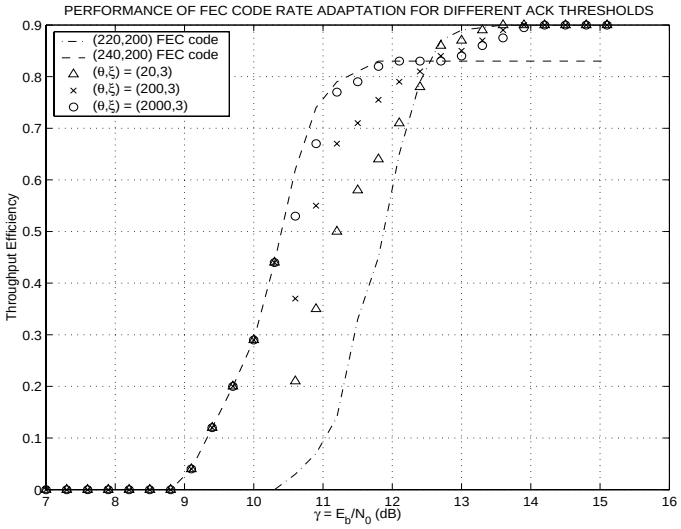


Figure 4.4: FEC code rate adaptation: Throughput efficiency for different values of the ACK threshold.

4.6.2 Numerical results

We study the performance of rate adaptation for the following cases:

- Case A: A system with fixed $(220, 200)$ RS FEC code and controllable modulation level between 8-QAM and 16-QAM.
- Case B: A system with fixed 16-QAM modulation level and controllable FEC code rate, by using a $(220, 200)$ and a $(240, 200)$ code.

The performance measure is normalized throughput, where the normalization is with respect to a system that uses 16-QAM modulation and a code of rate 1. We first study the performance of FEC code rate adaptation with 16-QAM modulation. In figure 4.4, we first plot the throughput efficiencies of the two FEC codes as a function of γ, where each code is used for the entire experiment with no rate adaptation. The existence of a cross-over value γ^* can be observed. Then, we

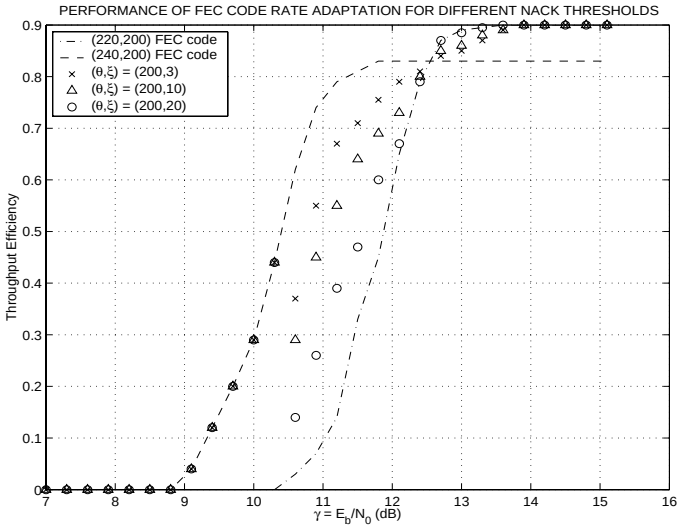


Figure 4.5: FEC code rate adaptation: Throughput efficiency for different values of the NACK threshold.

plot the throughput of three adaptive schemes with fixed NACK threshold $\xi = 3$ and different ACK thresholds $\theta = 20$, 200 and 2000 respectively. A higher ACK threshold yields better performance for $\gamma < \gamma^*$, since it reduces the amount of transitions from the low-rate code to the high-rate one, or equivalently it encourages operation in the low-rate code, which can sustain more interference. Thus, system performance in this region approximates that of the low-rate code. In region $\gamma > \gamma^*$, a higher ACK threshold leads to lower throughput, since the system is forced to wait more until it switches from the low-rate to the high-rate code. In addition, we observe that the threshold values $(\theta, \xi) = (2000, 3)$ cause the throughput to approximate the envelope of the individual throughput curves more closely.

The rate adaptation algorithm was applied for values of γ, where both codes have non-zero throughput. Thus, for $\gamma < 10.3$ dB, where the $(220, 200)$ code has zero throughput, the system operates with the $(240, 200)$ code. This convention rule is

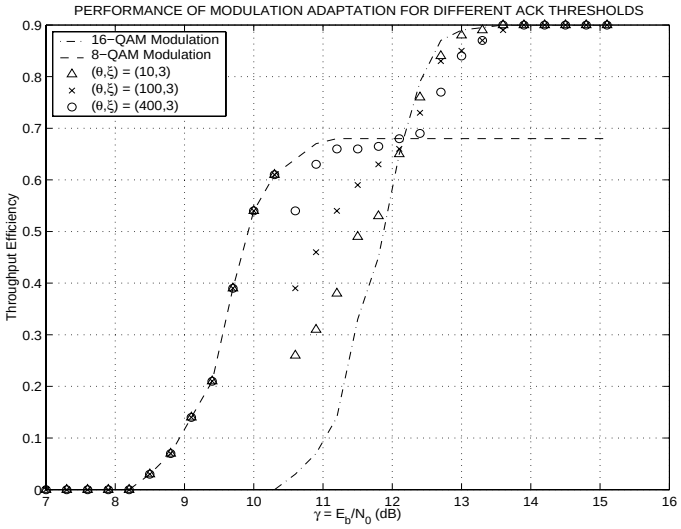


Figure 4.6: Modulation level adaptation: Throughput for different values of the ACK threshold.

used in all subsequent simulations as well.

In figure 4.5, we depict the performance of FEC code rate adaptation with fixed ACK threshold $\theta = 200$ and different NACK thresholds $\xi = 3$, 10 and 20. A lower NACK threshold is more preferable for $\gamma < \gamma^*$, since it facilitates transitions from the $(220, 200)$ to the $(240, 200)$ code and hence increases the portion of time when the lower-rate code is used. For $\gamma > \gamma^*$, a higher NACK threshold is required, so that the system does not transition to the low-rate code often enough and operates with the $(220, 200)$ code. Overall, it can be observed that the curve corresponding to $\theta = 200$, $\xi = 3$ yields the closest approximation to the envelope.

In figures 4.6 and 4.7, the case of modulation adaptation with a fixed-rate $(220, 200)$ FEC code is considered. In figure 4.6 the ACK threshold was varied and the NACK threshold is fixed, while in figure 4.7, the NACK threshold was changed and the ACK threshold is fixed. We deduce that the throughput curves

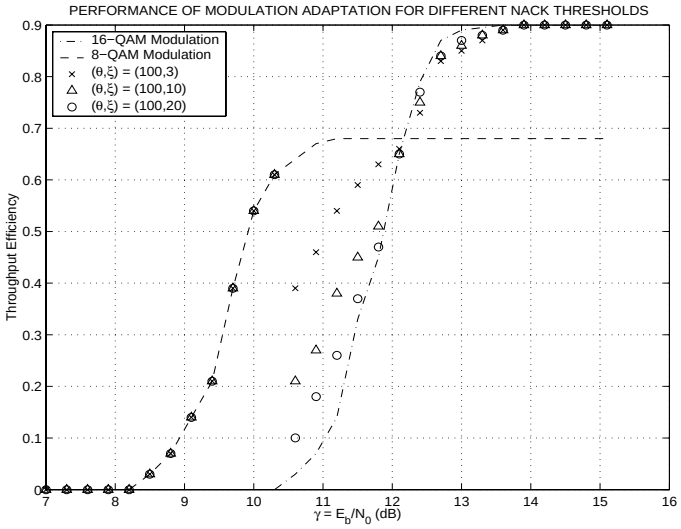


Figure 4.7: Modulation level adaptation: Throughput for different values of the NACK threshold.

for different ACK and NACK threshold values exhibit similar trends to those observed for the case of FEC code rate adaptation. However, for the same NACK threshold, the ACK threshold for the case of modulation adaptation is lower than the corresponding threshold for FEC code rate adaptation. Hence, for $\xi = 3$, ACK threshold values $\theta = 400$ and $\theta = 2000$ are most appropriate for modulation adaptation and FEC code rate adaptation respectively.

The threshold values that were used in the simulations above were not the optimal ones. The purpose of these simulations is to demonstrate the impact of the different magnitudes of ACK and NACK thresholds on system performance. The optimal threshold values that are generated from the solution to optimization problem (4.25) are depicted in table 4.1.

A first conclusion that can be drawn pertains to the relative values of ACK and NACK thresholds that achieve satisfactory performance. Clearly, a high through-

NACK threshold (ξ)	ACK threshold (θ_A) (Case A)	ACK threshold (θ_B) (Case B)
1	41	7
2	322	55
3	2594	258
4	15224	378
5	43527	634
6	77709	983

Table 4.1: ACK threshold values for rate adaptation of cases A and B.

put is achieved if the ACK threshold θ is significantly higher that the NACK threshold ξ. A small NACK threshold (of the order of some NACKs) implies that in the case of link deterioration the system must respond fast and decrease the transmission rate. A high ACK threshold (of the order of hundreds or thousands of ACKs) means that a more conservative policy needs to be adopted when link conditions improve, in the sense that a decision about rate increase is taken only when a large number of ACKs is received.

Such large ratios of (θ/ξ) can be explained if we consider the slope of throughput efficiency curves that correspond to different transmission rates. Whenever the system operates with a high rate and link quality deteriorates, the transition to the lower rate must be performed as fast as possible, since throughput decreases very fast (observe for example the large negative slope of the throughput curve for the high-rate FEC code at the region of $10 - 12$ dB). If rate adaptation is not timely enough, a large throughput loss is incurred. On the other hand, when link conditions improve and the system needs to switch to a higher rate, the rate

(ξ)	θ for $Q=1$	θ for $Q=2$	θ for $Q=5$	θ for $Q=10$	θ for $Q=20$
1	41	25	14	10	8
2	322	114	42	25	16
3	2594	504	94	44	27
4	15224	2241	205	70	37
5	43527	7788	457	105	48
6	77709	17831	1014	153	61

Table 4.2: ACK threshold values θ for FEC code rate adaptation as a function of NACK threshold ξ for different number of utilized subcarriers, Q.

switching decision is taken after an adequately large number of ACKs is received. This conservative policy aims at minimizing the risk of an incorrect decision. The nature of the policy is due to the fact that the rate improvement from operation at a higher rate is not large enough to justify a fast and premature decision.

A second observation that can be made is with regard to the relative values of thresholds for different kinds of rate adaptation. In particular, the ACK threshold for modulation level adaptation is smaller than the corresponding value for FEC code rate adaptation, if the value of NACK threshold is fixed. For modulation adaptation, the performance difference between different transmission modes is larger than the corresponding difference for FEC code rate adaptation. Therefore, the system becomes more willing to increase transmission rate, since the throughput benefit from such a transition would be large.

Next, we expand our simulation studies to the multi-link case, where each link represents one OFDM subcarrier. We consider the situation where subcarriers have the same quality, so that the same transmission parameters are used in each

of them. We focus on the case of FEC code rate adaptation and compute the ACK threshold values for different number of subcarriers Q. The results are illustrated in table 4.2. The ACK threshold value that determines transition to higher rates decreases as the number of subcarriers increases. Therefore, when transmission is performed by using more subcarriers, the system should respond faster to improving link conditions. Again, this can be attributed to the large performance gain (which increases with increasing number of subcarriers) in the case of transition to higher rates.

Finally, a remark about the values of ACK and NACK thresholds is in order. For the case of FEC code rate adaptation, the threshold ratios (θ/ξ) for the single-link case are of the order of several hundreds or even thousands, which may seem unrealistic. However, in the multi-link case that reflects OFDM transmission, the threshold ratios are of the order of tens or hundreds. When modulation level adaptation is applied in the multi-link case, the threshold values are anticipated to decrease even more. These results demonstrate the significance of rate adaptation in OFDM transmission as a means of improving system performance.

4.7 Conclusion

In this chapter, we focused on a simple link monitoring method and studied the class of rate adaptation policies that correspond to this method. The notion of a link between a transmitter and a receiver is considered. Burst transmission and rate adaptation take place over that link. Starting from the single-link case which represents one subcarrier in OFDM transmission, we showed that the optimal policy is of threshold type and presented a suboptimal heuristic methodology to compute threshold values. We also demonstrated the nature of the adaptation policy. It

should be conservative in rate increase and fast in rate decrease. We specified the impact of thresholds and different transmission rates on performance. Next, we expanded our policy to the multi-link case which represents OFDM transmission over several subcarriers. We considered the cases of subcarriers with similar or different quality. For subcarriers of similar quality, we studied the effect of number of subcarriers on threshold and observed that the ACK/NACK feedback may prove a valuable tool in performing timely rate adaptation.

Chapter 5

Adaptive resource allocation in OFDM-based wireless networks with smart antennas

5.1 Introduction

Smart antennas constitute perhaps the most promising means of increasing capacity in wireless systems. The deployment of a smart antenna array with several antenna elements opens up the spatial dimension within a single cell and enables the use of SDMA. This multiple access technique allows many intra-cell users to be served simultaneously by the same conventional channel. Within each channel, the beamformer has to form one beam for each user that is assigned in the channel. The radiation pattern of the beam that corresponds to each user is adjusted, so that nulls are placed in the directions of interference and the main lobe is steered to the direction of the desired user. Along with beamforming, transmission power control is used as a means of adjusting the interference levels at different receivers. The objective is to find beamforming vectors and powers so as to ensure

an acceptable SINR at each receiver.

However, the existence of smart antennas at the physical layer raises significant issues for channel allocation at the MAC layer, since smart antennas control the intra-cell channel reuse pattern to a certain extent. In this chapter, we attempt to address and study some of these issues when the underlying multiple access scheme has orthogonal channels. The emphasis is placed on OFDM transmission, which presents some novel characteristics when considered in the spatial dimension.

5.1.1 Related work and motivation

Adaptive beamforming and power control in a single channel have received considerable attention in the literature. In [85], the authors propose an iterative algorithm for joint transmit power control and receive beamforming for a set of cochannel links for the up-link direction. The algorithm converges to a feasible solution of powers and beamforming vectors, if there exists one, and this solution minimizes total transmitted power over all feasible power allocations and beamforming vectors. However, a weak point of the approach is that the algorithm cannot detect infeasible solutions that cause divergence if the SINR requirements of cochannel links cannot be supported. The same authors in [86] present an iterative algorithm for joint power control and beamforming for the down-link. The problem is transformed to an equivalent problem of transmit power control and receive beamforming in the up-link and the technique outlined in [85] is applied. The same principle is used in [87], in order to derive transmit powers and receive beamforming vectors for a set of cochannel users for each subcarrier of an OFDM system. A low-complexity technique that includes transferring the beamforming from the frequency to the time domain is also proposed.

The feasibility issue of the power minimization problem has also attracted significant attention. In fact, this problem is closely linked with the problem of finding beamforming vectors and powers so as to maximize the minimum SIR of a set of users. The solution to the latter balancing problem determines the range where the former problem has a feasible solution [88]. For the problem of SIR balancing, there exists an iterative algorithm that always converges to the maximum common SIR for a set of cochannel links [89, 90].

Down-link beamforming for power minimization in a single-cell system is studied in [91]. The authors state the related non-linear programming problem and propose low-complexity methods to construct feasible approximations to the optimal solution. The basic idea is to decouple the problems of finding the beam orientation and the transmission power, compute a beam for each user separately and attempt to find a feasible power vector by solving a linear system of equations. In [92], the authors study beamforming for a single-user OFDM system with multiple antennas at the transmitter and the receiver, with the objective to maximize receiver SINR. The work in [93] studies the problem of cochannel user separation in the up-link of an OFDM/SDMA system and proposes filtering and successive interference cancellation algorithms to efficiently distinguish user symbols.

An information-theoretic treatment of beamforming for a single-user channel is presented in [94]. Among other results, the paper states that for correlated fading between different antennas, the beamforming vector that maximizes capacity and average user SNR is associated with the maximum eigenvector of the vector channel covariance matrix. From an information-theoretic point of view, the problem of capacity maximization in a multiple-antenna channel is equivalent to that of determining the input covariance matrix. Beamforming is a special transmission

strategy where the input covariance matrix has rank one. An overview of results regarding capacity-achieving strategies for different kinds of available CSI can be found in [95] and the references therein. For perfect CSI, beamforming achieves channel capacity in the information-theoretic sense [30]. A capacity-achieving technique for the multi-user multi-antenna broadcast channel has been recently proposed in [96]. The technique is based on a transmission strategy known as Costa precoding, according to which users are sorted and beamforming for user k is performed by treating users 1 through $k-1$ as noise and by considering only the interference from the rest of the users. This method is shown to have an one-to-one correspondence with up-link beamforming and successive decoding, which is known to achieve capacity for the multiple-access channel.

The common characteristic of these approaches is that they concentrate on a single channel and do not study the impact of SDMA on channel allocation. Some attempts towards identifying this impact are only recently reported. In [97, 98], the authors describe heuristic algorithms for time slot allocation in a SDMA/TDMA system with the objective to increase capacity, while [99] presents a framework for joint time slot allocation and packet scheduling based on packet transmission deadlines for a SDMA system. In [35], we considered the joint problem of subcarrier allocation, transmission rate control and beamforming in an OFDM/SDMA system. The problem was addressed for a system with and without channel reuse. In the former case, a methodology for constructing cochannel user sets with high total subcarrier rate was outlined. In the latter case, where each subcarrier was used by at most one user, beamforming was considered as an additional dimension to enhance user SINR. We proposed suboptimal heuristics for channel allocation, with the objective to maximize total achievable system rate and provide QoS to

users in the form of minimum rate guarantees.

With the exception of these works, channel allocation in the context of OFDM or other multiple access scheme has hitherto been studied independently from user spatial separation through SDMA and channel reuse. Intra-cell channel reuse is suboptimal and is usually based on static cell sectorization [52] or beam switching methods, which do not fully capture user mobility, channel dynamics and traffic load variations. Related research on beamforming has mostly focused on beam adaptation for each cochannel link in a single channel, so as to ensure an acceptable SINR at each receiver. Thus, in a multi-channel system, beam adaptation of users is performed independently in each channel, without any consideration of its impact on other allocated channels or on user QoS at the MAC layer. In the case of OFDM, since user spatial channel characteristics vary in different subcarriers, certain sets of users are eligible to reuse certain subcarriers, while others are not. A particular allocation of users to subcarriers affects the total achievable system rate and the degree to which QoS is ensured for each user. Therefore, it is important that the appropriate cochannel user sets be identified for each subcarrier. This is turn translates to finding beamforming vectors and powers so as to support a cochannel user set with acceptable SINR.

Therefore, an appropriate strategy at the BS is necessary, such that the issues of channel allocation and user spatial separability can be studied jointly. The extent to which users are spatially separable depends on transmit beamforming, power control and selective user assignment in subcarriers, so that an acceptable SINR level is ensured at each user receiver. We study the joint problem of intra-cell channel allocation and transmit beamforming for a single-cell OFDM/SDMA system. We propose heuristic algorithms to assign users to subcarriers, while

appropriately adjusting beam directions and transmission powers. We study the class of greedy algorithms with assignment criteria such as minimum induced or received interference and minimum SIR per subcarrier, as well as the class of SIR balancing algorithms.

5.1.2 Notational remarks

A few words about the notation in this chapter before we proceed. Vectors and matrices are shown with boldface letters. The cardinality of set $\mathcal{X}$ is denoted as $|\mathcal{X}|$. Superscripts $*$, T and H denote conjugate of a complex number, transpose and conjugate transpose of a vector or matrix, and $\|\mathbf{u}\|$ is the ℓ_2-norm of complex vector $\mathbf{u} = (u_1, \ldots, u_n)^T$, i.e., $\|\mathbf{u}\| = \sqrt{\sum_{i=1}^{n} |u_i|^2}$. The dominant generalized eigenvector of matrix pair $(\mathbf{A}, \mathbf{B})$, $\mathbf{u}_{max}(\mathbf{A}, \mathbf{B})$, is the normalized eigenvector that corresponds to the largest positive eigenvalue of eigenproblem $\mathbf{Ax} = \lambda\mathbf{Bx}$. When $\mathbf{A}$, $\mathbf{B}$ are symmetric and positive definite, this is equivalent to the eigenproblem $\mathbf{Cy} = \lambda\mathbf{y}$, with $\mathbf{C} = \mathbf{L}^{-1}\mathbf{A}(\mathbf{L}^{-1})^H$ and $\mathbf{y} = \mathbf{L}^H\mathbf{x}$, where $\mathbf{L}$ is a non-singular lower triangular matrix that appears in the Cholesky decomposition of $\mathbf{B}$, which is $\mathbf{B} = \mathbf{L}\mathbf{L}^H$ [100].

5.1.3 Outline of chapter

The rest of the chapter is organized as follows. In section 5.2, we present the adopted model and assumptions in our approach and in section 5.3 we state the problem and demonstrate the associated tradeoffs. The proposed algorithms are presented in section 5.4 for single-rate transmission and an optimal solution for a simple special case is provided. These notions are extended to the case of multi-rate transmission in section 5.5. In section 5.6 numerical results are illustrated. Finally section 5.7 concludes the chapter.

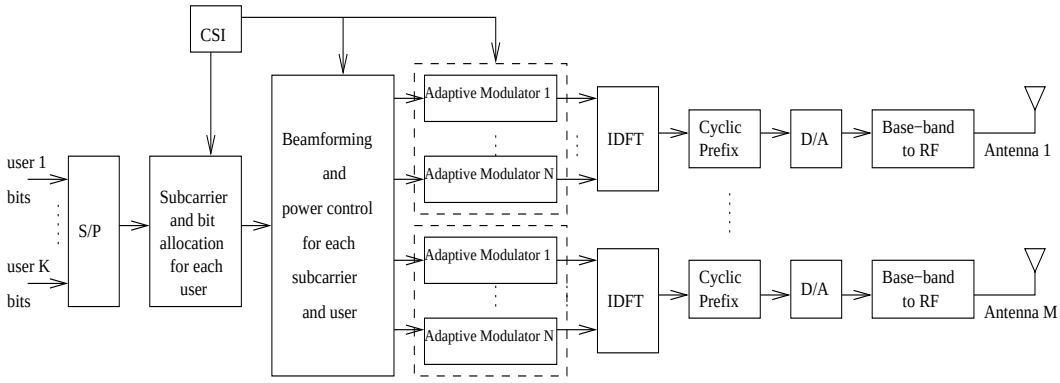


Figure 5.1: Block diagram of a multi-user OFDM/SDMA transmitter.

5.2 System model

We consider down-link OFDM transmission with N orthogonal subcarriers from a base station (BS) to K users in its cell. The BS is equipped with a uniform linear array of M antennas, while each receiver has an omni-directional antenna. The diagram of a multi-user OFDM/SDMA transmitter is depicted in figure 5.1.

An underlying slotted transmission scheme is again assumed. Packetized user data arrive from higher layers and are decomposed into bit streams before being transmitted to the corresponding users. Channel quality is assumed to remain constant for the duration of one time slot. Each user k has a minimum rate requirement of r_k bits/sec over some time interval $(0, t)$, which consists of several time slots. This requirement denotes the QoS that the MAC layer requests from the physical layer. Different number of bits $b_{n,k}$ of user k can be assigned in each subcarrier n. If bit allocation is replicated for each of the S transmitted symbols of user k in the slot, the rate of k in a slot is given by (2.1). We will concentrate on subcarrier, bit and power allocation and beamforming within a time slot. For single-rate transmission and assuming constant rate allocation across time slots,

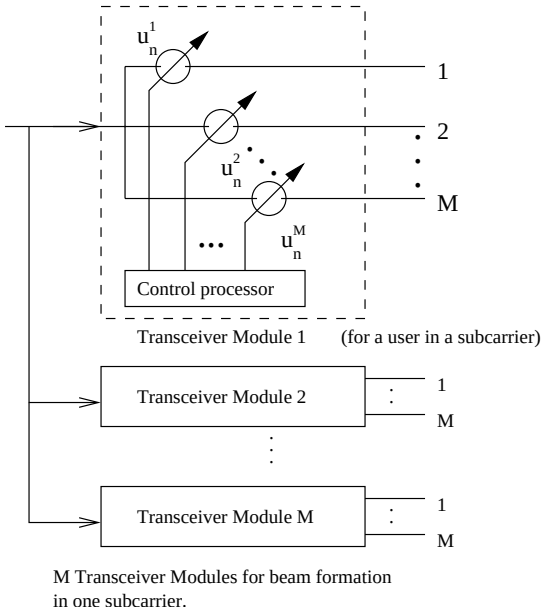


Figure 5.2: The structure of M transceiver modules for one subcarrier, n.

r_k can be mapped to a minimum number of required subcarriers n_k in a slot.

User bits enter the subcarrier allocation module. This module determines the cochannel sets of users for different subcarriers and allocates bits of users to subcarriers. Beamforming and power control is subsequently performed for each user that is allocated to a subcarrier. Under SDMA, the base station can form at most M beams and transmit to at most M out of K users simultaneously in each subcarrier. A beam $\mathbf{u}_{n,k} = (u^1_{n,k}, u^2_{n,k}, \ldots u^M_{n,k})^T$ is formed by a dedicated transceiver and a power $P_{n,k}$ is assigned to user k in subcarrier n. Beams are normalized to unit power, i.e., $\|\mathbf{u}_{n,k}\| = 1$. We assume that M transceivers (beamformers) exist for each subcarrier, so that a separate beam can be formed for each one of the M users that can be separated in each subcarrier. A set of M transceivers in a subcarrier is shown in figure 5.2. The subcarrier and bit allocation and beamforming operations are interdependent and also depend on available CSI at the BS.

Next, user bits are forwarded into M parallel modules of N modulators. Each modulator modulates the corresponding subcarrier with bits of users that are allocated to that subcarrier. Each user symbol in the stream is then transformed into N time domain samples with IDFT. After the cyclic prefix addition and D/A conversion, continuous signals of users are transmitted in parallel from the M antennas.

By following the rationale of section 1.4.1, we can characterize the multi-path channel from each antenna m to each user k by defining complex gains $\beta_{k,\ell}$, time delays $\tau_{k,\ell}$ and angles $\theta_{k,\ell}$ for the ℓth path of user k. The spatial signature of user k in subcarrier n is

$$\mathbf{a}_{n,k} = \sum_{\ell=1}^{L} \xi_{k,\ell}^{*}(n)\mathbf{v}_n(\theta_{k,\ell}) \tag{5.1}$$

and the spatial covariance matrix of user k in subcarrier n is

$$\mathcal{H}_{n,k} = \sum_{\ell=1}^{L} A_{k,\ell}\, \mathbf{v}_n(\theta_{k,\ell})\mathbf{v}_n^H(\theta_{k,\ell}), \tag{5.2}$$

where $A_{k,\ell}$ is the variance of the complex gain $\beta_{k,\ell}$.

At the receiver of each user k, the composite signal is down-converted and digitized and the time samples are transformed into subsymbols with DFT. The received signal at subcarrier n of user k is

$$R_{n,k} = \sqrt{P_{n,k}}\left(\mathbf{a}_{n,k}^H\mathbf{u}_{n,k}\right) d_{n,k} + \sum_{\substack{j\in\mathcal{U}^{(n)}\\ j\neq k}} \sqrt{P_{n,j}}\left(\mathbf{a}_{n,k}^H\mathbf{u}_{n,j}\right) d_{n,j} + z_{n,k}, \tag{5.3}$$

where $\mathcal{U}^{(n)}$ is the set of users that use subcarrier n. The receiver of user k is aware of spatial channel characteristics of k and treats other signals as noise. The expected SINR at the output of the matched filter at subcarrier n of user k is

$$SINR_{n,k} = \frac{P_{n,k}\left(\mathbf{u}_{n,k}^H\mathcal{H}_{n,k}\mathbf{u}_{n,k}\right)}{\sum\limits_{\substack{j\in\mathcal{U}^{(n)}\\ j\neq k}} P_{n,j}\left(\mathbf{u}_{n,j}^H\mathcal{H}_{n,k}\mathbf{u}_{n,j}\right) + \sigma^2}\ . \tag{5.4}$$

In our model, we assume that cochannel interference is the prevailing type of interference and that the noise level is not known at the transmitter. Then, the SINR can be replaced by the signal-to-interference ratio (SIR) and total power constraints are not required. A similar assumption was adopted in chapter 2. We assume that CSI about all users is available at the BS. Deterministic CSI involves exact knowledge of angular and multi-path characteristics for each path and each user, which is difficult to obtain in practice. CSI in terms of spatial covariance matrices of users in all subcarriers is more common and can be obtained by using the method outlined in section 1.4.1.

The BER at the output of the detector of a user in a subcarrier should not exceed ϵ. With the rationale of subsection 1.2.5, each modulation level of b_i bits per subsymbol is mapped to a minimum required SIR γ_i as in (1.6). If one modulation level b is used, one corresponding SINR threshold γ is defined.

5.3 Problem statement

SDMA allows intra-cell reuse of a subcarrier by multiple users. Two or more users are called *spatially separable* in a subcarrier if they simultaneously receive useful signals in the subcarrier and there exist beamforming vectors and powers for each user such that the SIR requirements at corresponding receivers are satisfied. For a given subcarrier, spatial separability depends on the number and identities of individual users through spatial covariance matrices of users, which in turn capture angular and multi-path characteristics of user channels. If multiple transmission rates are employed, spatial separability of users also depends on the number of bits that constitute user subsymbols in the subcarrier. This is because different number of bits are associated with different modulation levels. These are in turn

associated with different required minimum SIR values for an acceptable BER and hence they have different amounts of maximum sustainable interference. In addition, beamforming vectors and transmission powers affect interference levels and SIRs of all receivers and thus affect spatial separability. Finally, user separability depends on each individual subcarrier: users that can share one subcarrier, may not be eligible cochannel users in a different subcarrier, or subcarrier reuse may be feasible with different numbers of allocated user bits. The dependence of spatial separability on subcarriers is attributed to the fact that angular and multi-path characteristics of users are reflected differently in different subcarrier frequencies.

Each user in a subcarrier experiences cochannel interference from transmissions to other users in the subcarrier. When a large number of bits is used for transmission to a user in a subcarrier, the rate for that user is increased and the user needs fewer subcarriers to satisfy certain rate requirements. Thus, more users can be accommodated and capacity is increased. However, larger numbers of bits for users in a subcarrier render spatial separability more difficult, since the maximum sustainable amount of interference is decreased and hence fewer users can reuse the same subcarrier. Non-separable users should in general be assigned to different subcarriers and from that point of view system capacity is not enhanced. On the other hand, with a small number of bits per subcarrier and thus lower assigned rate, a user needs more subcarriers to satisfy rate requirements and thus fewer users can be accommodated in the system. However, a small number of assigned bits facilitates spatial separability of more users, since cochannel transmissions are less sensitive to cochannel interference. From the discussion above, it is not clear whether a small number of users with high rates or a large number of users with low rates yields higher rate in a subcarrier.

The arising issue is whether there exists a way to perform subcarrier allocation and user spatial separation jointly so as to maximize the total number of bits per subcarrier. This problem is equivalent to identifying cochannel sets of users with large total number of bits per subcarrier. Ideally, each subcarrier should have a large number of spatially separated users and a large number of bits. This is possible if users are spatially well separable. For a single LOS path, spatial separability is easier when users are well separated in angle. For the more general case, spatial signatures of users should not be highly correlated and spatial covariance matrices of users and beamforming vectors should be such that users do not induce much interference to each other.

The identification of the cochannel user set that achieves maximum subcarrier rate is a hard optimization problem. First, an appropriate subset of spatially separable users must be identified. The cardinality of the spatially separable cochannel user set is limited by the number of antennas M. Then, beamforming vectors and powers must be computed for these users, so that SIRs at receivers are above the SIR thresholds that correspond to the assigned numbers of bits. The problem is that the SIR at a receiver depends on beamforming vectors and powers of all other users. The enumeration of all possible user assignments in a subcarrier is of exponential complexity. In addition, even if the cochannel set of users is given, the computation of beamforming vectors and powers that maximize the sum of user SIRs is a highly non-linear problem.

The discussion above necessitates the adoption of suboptimal heuristic algorithms for constructing cochannel sets of spatially separable users with appropriate beamforming vectors and powers. In the sequel, we consider three heuristic algorithms for subcarrier allocation in the context of OFDM/SDMA. The first two

algorithms fall within the category of greedy heuristics, but utilize different criteria for assignment of users to subcarriers, namely minimum induced or received interference to or from other users and minimum SIR per subcarrier. The third algorithm follows a different approach in subcarrier allocation and tries to maintain the highest possible common SIR in each channel by jointly adapting beamforming vectors and powers.

In section 5.4 these algorithms are presented for single-rate transmission and emphasis is placed on the construction of cochannel user sets with appropriate beamforming and power control. The presented principles are extended to the multi-rate case in section 5.5.

5.4 Single-rate transmission: Proposed heuristic algorithms

The key idea of the proposed algorithms is to assign users to appropriate subcarriers so that user minimum rate requirements are satisfied and total system rate is increased. If minimum rate requirements were not included in the formulation, it would suffice to consider the allocation procedure separately for each subcarrier. Because of these requirements, it is important to assign the appropriate users to subcarriers, such that future user allocations are facilitated.

5.4.1 Algorithm A

The first class of algorithms utilizes the greedy criterion of minimum induced and received interference to or from cochannel users. In order to keep the complexity to a reasonable level, we consider algorithms for which users are sequentially inserted

in the subcarrier and no user reassignments are performed. However, we allow beamforming adjustment for cochannel users. In algorithms A and B, power control is considered only when beamforming alone is insufficient in providing the required SIRs for users.

Beamforming vector adaptation

The basic goal is to form large cochannel sets of spatially separable users in each subcarrier. At each step of the algorithm, an appropriate user is assigned to a subcarrier and beamforming vectors of other users are adjusted, so that acceptable SIRs are ensured. An inserted user in a subcarrier should induce the least cochannel interference to users that are already assigned in that subcarrier and should receive the least interference from those users.

Fix attention to subcarrier n and let k denote the user to be inserted next in the subcarrier. Let $\mathcal{U}^{(n)}$ denote the set of users that are already assigned in n. Let $\mathbf{u}_{n,j}$ and $P_{n,j}$ be the beamforming vector and power for user $j \in \mathcal{U}^{(n)}$. Insertion of user k in n creates a new interference instance for cochannel users in n. Thus, beamforming vectors may need to be recomputed, so as to maintain acceptable SIRs. For each user $j \in \mathcal{U}^{(n)}$, we define the ratio of desired power generated by beam $\mathbf{u}_{n,j}$, over interference power which is caused to other cochannel users, including the new user k in subcarrier n. In fact, we are interested in the maximum value of this ratio, $\Psi_{n,k}^{(j)}$ over all directions $\mathbf{u}_{n,j}$,

$$\Psi_{n,k}^{(j)} = \max_{\mathbf{u}_{n,j}} \frac{\mathbf{u}_{n,j}^H \mathcal{H}_{n,j} \mathbf{u}_{n,j}}{\mathbf{u}_{n,j}^H \Big(\sum\limits_{\substack{i \in \mathcal{U}^{(n)} \\ i \neq j}} \mathcal{H}_{n,i} + \mathcal{H}_{n,k} \Big) \mathbf{u}_{n,j}}, \text{ subject to } \|\mathbf{u}_{n,j}\| = 1. \tag{5.5}$$

The vector $\mathbf{u}_{n,j}^*$ that maximizes this ratio is the dominant generalized eigenvector of matrix pair $\Big(\mathcal{H}_{n,j}, \sum_{i \in \mathcal{U}^{(n)}, i \neq j} \mathcal{H}_{n,i} + \mathcal{H}_{n,k} \Big)$ and it is found with the method

outlined at the end of section 5.1. We also compute the corresponding ratio for user k that is tentatively inserted in subcarrier n

$$\Psi_{n,k} = \max_{\mathbf{u}_{n,k}} \frac{\mathbf{u}_{n,k}^H \mathcal{H}_{n,k} \mathbf{u}_{n,k}}{\mathbf{u}_{n,k}^H \Big(\sum_{j \in \mathcal{U}^{(n)}} \mathcal{H}_{n,j} \Big) \mathbf{u}_{n,k}}, \text{ subject to } \|\mathbf{u}_{n,k}\| = 1. \tag{5.6}$$

The denominator of this ratio reflects the interference caused by user k to other cochannel users. Again, the vector $\mathbf{u}_{n,k}^*$ that maximizes this ratio is the dominant generalized eigenvector of matrix pair $\Big(\mathcal{H}_{n,k}, \sum_{j \in \mathcal{U}^{(n)}} \mathcal{H}_{n,j}\Big)$. With the beamforming vectors $\mathbf{u}_{n,k}^*$ and $\mathbf{u}_{n,j}^*$, $j \in \mathcal{U}^{(n)}$, we evaluate the SIRs for user k and users $j \in \mathcal{U}^{(n)}$.

Power adaptation

If SIRs of some users do not exceed the minimum required SIR γ, we fix the computed beamforming vectors and activate power control. Given a cochannel user set and their beamforming vectors, the question is whether there exist powers so that all SIRs exceed γ. For each subcarrier n, let i, j be indices of users in that subcarrier. Define $\mathbf{U}_n$ as the ensemble of computed beamforming vectors for users in n, i.e., $\mathbf{U}_n = \big\{\mathbf{u}_{n,k} : k \in \mathcal{U}^{(n)}\big\}$. Then, we define the $(|\mathcal{U}^{(n)}| \times |\mathcal{U}^{(n)}|)$ matrix $\mathbf{A}(\mathbf{U}_n)$. The (i,j)-th element of $\mathbf{A}(\mathbf{U}_n)$ specifies the cochannel interference caused by the beam of the jth user to the receiver of the ith user of subcarrier n, normalized by the useful signal power of i. That is,

$$\mathbf{A}(\mathbf{U}_n)[i,j] = \begin{cases} \frac{\mathbf{u}_{n,j}^{*H} \mathcal{H}_{n,i} \mathbf{u}_{n,j}^*}{\mathbf{u}_{n,i}^{*H} \mathcal{H}_{n,i} \mathbf{u}_{n,i}^*}, & \text{if } i \neq j \\ 1, & \text{if } i = j. \end{cases} \tag{5.7}$$

Define also a $(|\mathcal{U}^{(n)}| \times 1)$ vector $\mathbf{P}_n = (P_{n,i} : i \in \mathcal{U}^{(n)})$, which contains the powers of all users in $\mathcal{U}^{(n)}$. Then, the requirement $SIR_{n,i} \geq \gamma$ for all users $i \in \mathcal{U}^{(n)}$ can be

written in a matrix form as,

$$\frac{1+\gamma}{\gamma}\mathbf{P}_n \geq \mathbf{A}(\mathbf{U}_n)\mathbf{P}_n\,. \tag{5.8}$$

The matrix $\mathbf{A}(\mathbf{U}_n)$ is non-negative definite and irreducible. From the Perron-Frobenius theorem, it has a positive, real eigenvalue $\lambda_{max}(\mathbf{A}(\mathbf{U}_n)) = \max\{|\lambda_i|\}_{i=1}^{|\mathcal{U}^{(n)}|}$, where λ_i, $i = 1, \ldots, |\mathcal{U}^{(n)}|$ are the eigenvalues of $\mathbf{A}(\mathbf{U}_n)$. The eigenvalue $\lambda_{max}(\mathbf{A}(\mathbf{U}_n))$ has an associated eigenvector with strictly positive entries. Furthermore, the minimum real λ such that the inequality $\lambda\mathbf{P}_n \geq \mathbf{A}(\mathbf{U}_n)\mathbf{P}_n$ has solutions $\mathbf{P}_n > 0$ is $\lambda = \lambda_{max}(\mathbf{A}(\mathbf{U}_n))$. In our case, we start by finding the maximum real positive eigenvalue of $\mathbf{A}(\mathbf{U}_n)$ to request the existence of a power vector with positive entries. If

$$\lambda_{max}(\mathbf{U}_n) \leq \frac{1+\gamma}{\gamma}\,, \tag{5.9}$$

then (5.8) holds and SIR level γ is achievable. The power vector that leads to an achievable γ is the eigenvector that corresponds to $\lambda_{max}(\mathbf{A}(\mathbf{U}_n))$.

Next, we define an assignment preference factor (APF) $\Phi_{n,k}$ for subcarrier n and user k. First, the beamforming vector and power must yield strong desired signal for user k. Furthermore, all beams and powers should be such that the interference caused by user k to other users, as well as the interference on k by other users is low. These requirements are captured by ratio

$$\Phi_{n,k} = \frac{P_{n,k}(\mathbf{u}_{n,k}^{*H}\mathcal{H}_{n,k}\mathbf{u}_{n,k}^{*})}{\max\left\{P_{n,k}\mathbf{u}_{n,k}^{*H}\Big(\sum_{j\in\mathcal{U}^{(n)}}\mathcal{H}_{n,j}\Big)\mathbf{u}_{n,k}^{*},\ \sum_{j\in\mathcal{U}^{(n)}}P_{n,j}(\mathbf{u}_{n,j}^{*H}\mathcal{H}_{n,k}\mathbf{u}_{n,j}^{*})\right\}}\,. \tag{5.10}$$

Clearly, if power control is not activated (when all SIRs exceed γ after initial beam computations with (5.5) and (5.6)), the ratios $\Phi_{n,k}$ do not include powers.

At each step of the algorithm, $\Phi_{n,k}$s are computed for all subcarriers n for which a user insertion leads to acceptable SIRs and for all users k that have not

satisfied minimum rate requirements n_k. Among assignments that yield acceptable SIRs for users, we select the one with the maximum preference factor $\Phi_{n,k}$. After each assignment, the rate of user k is updated. When a user reaches n_k, it is not considered for assignment until all users reach their minimum rate requirements. If the cardinality of the cochannel user set reaches M for a subcarrier, this subcarrier is not considered for user assignment. The algorithm terminates when no further assignments are possible to any channel.

5.4.2 Algorithm B

The second class of heuristic algorithms is based on the criterion of maximizing the minimum SIR in a subcarrier. In Algorithm A, we prefer a user that causes and receives the least interference to and from cochannel users. By following this greedy approach of least incremental interference, we aimed at inserting as many users as possible in subcarriers. In algorithm B, a user assignment in a subcarrier is performed if it maximizes the minimum SIR of users in the subcarrier over all possible user assignments. By this assignment, we intend to facilitate future assignments and ultimately increase the number of users with SIR above threshold γ. Thus, the APF factors are now defined as

$$\Phi_{n,k} = \min\left\{SIR_{n,k}, \min_{j \in \mathcal{U}^{(n)}} SIR_{n,j}\right\}. \tag{5.11}$$

5.4.3 Description of algorithms A and B

The only difference in algorithms A and B is the definition of APF factors. The main steps for both algorithms can be summarized as follows:

- **STEP 0** : Initialize list of candidate users $\mathcal{L}$ that have not achieved mini-

mum rate requirements. Initially all subcarriers are included in the list $\mathcal{C}$ of candidate subcarriers.

- **STEP 1** : Compute APF factors $\Phi_{n,k}$ for all subcarriers $n \in \mathcal{C}$ and all users $k \in \mathcal{L}$. For each pair (n, k), start with beamforming vector adaptation and activate power control if required.

- **STEP 2** : Select pair (n^*, k^*) with maximum APF factor and perform the assignment. If $|\mathcal{U}^{(n^*)}| = M$, remove subcarrier n^* from $\mathcal{C}$. If minimum rate requirements are reached for user k^*, remove k^* from $\mathcal{L}$.

- **STEP 3** : Update APFs for users in $\mathcal{U}^{(n)}$ and user rates.

- **STEP 4** : If $|\mathcal{U}^{(n)}| = M$ for all n, or if no further assignments are possible, go to step 6. Otherwise, go to step 5.

- **STEP 5** : If list $\mathcal{L}$ is not empty, go to step 1. If it is empty, consider all users again in $\mathcal{L}$, before going to step 1.

- **STEP 6** : End of algorithm.

The computationally intense part of the algorithms is the computation of the dominant generalized eigenvectors for the APF factors. For each user, it involves Cholesky decomposition of a $M \times M$ matrix and calculation of the maximum eigenvalue of an appropriate matrix, as outlined at the end of section 5.1. Both these procedures are known to be of complexity $O(M^3)$. The selection of each assignment is performed in $O(NKM^3)$ time and each algorithm needs $O(N^2KM^4)$ time.

5.4.4 Algorithm C

Algorithms A and B perform greedy sequential assignment of users in subcarriers based on different criteria. Recall that beamforming vector and power adaptation were decoupled, since fixed beams were used to find feasible powers. Algorithm C follows a different approach, in the sense that it attempts to provide the maximum common SIR for users in a subcarrier. A salient feature of Algorithm C is that it performs joint adaptation of beamforming vectors and powers in order to obtain the desirable common SIR.

Consider again the set of users $\mathcal{U}^{(n)}$ and define the $(|\mathcal{U}^{(n)}| \times |\mathcal{U}^{(n)}|)$ matrix $\mathbf{B}(\mathbf{U}_n)$ with elements

$$\mathbf{B}(\mathbf{U}_n)[i,j] = \begin{cases} \mathbf{u}_{n,j}^H \mathcal{H}_{n,i} \mathbf{u}_{n,j} & \text{if } i \neq j \\ 0, & \text{if } i = j. \end{cases} \tag{5.12}$$

Thus, $\mathbf{B}(\mathbf{U}_n)$ is the interference matrix between users in subcarrier n. Define also the diagonal matrix

$$\mathbf{D} = \text{diag}\left\{ \frac{1}{\mathbf{u}_{n,i}^H \mathcal{H}_{n,i} \mathbf{u}_{n,i}} : i \in \mathcal{U}^{(n)} \right\} \tag{5.13}$$

An instance in which all users achieve a common SIR γ_c in the *down-link* by using the ensemble of beamforming vectors $\mathbf{U}_n$ and power vector $\mathbf{P}_n$ is described by the set of equations,

$$\mathbf{DB}(\mathbf{U}_n)\mathbf{P}_n = \frac{1}{\gamma_c}\mathbf{P}_n \,. \tag{5.14}$$

Thus, γ_c is a reciprocal eigenvalue of matrix $\mathbf{DB}(\mathbf{U}_n)$. Matrix $\mathbf{DB}(\mathbf{U}_n)$ has the same properties as $\mathbf{A}(\mathbf{U}_n)$ with respect to the existence of an eigenvector $\mathbf{P}_n$ with positive entries. Therefore, we have $1/\gamma_c = \lambda_{max}(\mathbf{DB}(\mathbf{U}_n))$. The maximum possible common SIR, γ_c^*, is therefore

$$\gamma_c^* = \frac{1}{\min\limits_{\mathbf{U}_n} \lambda_{max}(\mathbf{DB}(\mathbf{U}_n))} \,. \tag{5.15}$$

We now consider the corresponding problem of beamforming and power control for the same users in the *up-link*. In that case, the controllable parameters are transmission powers of users in the up-link and beamforming vectors. It can be verified that the instance in which all users achieve a common SIR $\tilde{\gamma}_c$ in the up-link by using an ensemble of beamforming vectors $\tilde{\mathbf{U}}_n$ and transmit power vector $\tilde{\mathbf{P}}_n$ is described by the set of equations,

$$\mathbf{D}\mathbf{B}^T(\tilde{\mathbf{U}}_n)\tilde{\mathbf{P}}_n = \frac{1}{\tilde{\gamma}_c}\tilde{\mathbf{P}}_n \tag{5.16}$$

and the maximum possible common SIR $\tilde{\gamma}_c^*$ is

$$\tilde{\gamma}_c^* = \frac{1}{\min\limits_{\tilde{\mathbf{U}}_n} \lambda_{max}(\mathbf{D}\mathbf{B}^T(\tilde{\mathbf{U}}_n))}\,. \tag{5.17}$$

For the relationship between the down-link problem (5.15) and the up-link problem (5.17), the following properties have been proved in [89, 90]:

Property 1 *For a given set of beamforming vectors* $\mathbf{U}_n$*, it is* $\lambda_{max}(\mathbf{D}\mathbf{B}(\mathbf{U}_n)) = \lambda_{max}(\mathbf{D}\mathbf{B}^T(\mathbf{U}_n))$*.*

Property 2 *The up-link and down-link problems have the same solution in terms of maximum achievable common SIR, i.e., it is* $\gamma_c^* = \tilde{\gamma}_c^*$*.*

Property 3 *The beamforming vectors of corresponding users that solve the down-link problem (5.15) and the up-link problem (5.17) are the same, namely* $\mathbf{U}_n^* = \tilde{\mathbf{U}}_n^*$*.*

Property 4 *For the following iterative algorithm (Algorithm I), the sequence of eigenvalues* $\lambda_{max}^{(t)}$ *is monotonically decreasing with the iteration number* t *and the algorithm converges to a minimum, which is related to the maximum common SIR through (5.15) and (5.17).*

ALGORITHM I

- **STEP 1**: Set $t = 0$. Start with arbitrary initial beamforming vectors $\mathbf{U}_n^{(0)}$.
- **STEP 2**: $t \leftarrow t + 1$. For given $\mathbf{U}_n^{(t)}$, solve the following eigen-problem for the *uplink*:

$$\mathbf{D}\mathbf{B}^T(\mathbf{U}_n^{(t)})\mathbf{P}_n^{(t)} = \lambda_{max}^{(t)}\mathbf{P}_n^{(t)} \,. \tag{5.18}$$

- **STEP 3**: For the computed $\mathbf{P}_n^{(t)}$, solve a set of *decoupled* generalized eigen-problems

$$\mathbf{u}_{n,k} = \arg\max_{\mathbf{u}_{n,k}} \frac{\mathbf{u}_{n,k}^H \mathcal{H}_{n,k}\mathbf{u}_{n,k}}{\mathbf{u}_{n,k}^H \mathcal{R}_{n,k}(\mathbf{P}_n^{(t)})\mathbf{u}_{n,k}^H}, \text{ subject to: } \|\mathbf{u}_{n,k}\| = 1, \ \forall k \in \mathcal{U}^{(n)}. \tag{5.19}$$

  where

$$\mathcal{R}_{n,k}(\mathbf{P}_n^{(t)}) = \sum_{\substack{j \in \mathcal{U}^{(n)} \\ j \neq k}} P_{n,j}^{(t)} \mathcal{H}_{n,j} \,. \tag{5.20}$$

- **STEP 4**: With the computed $\mathbf{U}_n^{(t)}$, go to step 2. Continue until convergence.

Observe that at step 3 the quantity to be maximized is the SIR of user k in the up-link. The beamforming vectors $\mathbf{U}_n^*$ at the end of the algorithm are the required down-link beams. If $\lambda_{max}^* = \lambda_{max}(\mathbf{D}\mathbf{B}^T(\mathbf{U}_n^*))$ is the eigenvalue at the end of the algorithm, the corresponding down-link power vector is given by the eigenvector of $\mathbf{B}(\mathbf{U}_n^*)$ that corresponds to λ_{max}^*.

We now proceed to the description of algorithm C. For a set of users $\mathcal{U}^{(n)}$, let $\gamma_{c,n}$ denote the maximum common SIR, as it is computed by applying algorithm I. For each user $k \in \mathcal{U}^{(n)}$, let $\gamma_{c,n}(k)$ denote the common SIR of remaining users when k is removed from subcarrier n. Again, $\gamma_{c,n}(k)$ is found by Algorithm I, after we delete the kth row and column of $\mathbf{B}^T$. The main steps of algorithm C are as follows:

- **STEP 0** : Start by assigning all K users in each subcarrier n.
- **STEP 1** : Run algorithm I for each subcarrier. Outcome is a vector of common SIRs, $\boldsymbol{\gamma_c} = (\gamma_{c,1}, \ldots, \gamma_{c,N})$,
- **STEP 2** : If $\gamma_{c,n} \geq \gamma$ for all n, desirable SIRs are achieved in all subcarriers. Go to step 6. Otherwise, go to step 3.
- **STEP 3** : For each $k \in \mathcal{U}^{(n)}$, compute $\gamma_{c,n}(k)$.
- **STEP 4** : Select pair (n^*, k^*) with maximum $\gamma_{c,n}(k)$ and remove user k^* from subcarrier n^*.
- **STEP 5** : Update user rates. If a user reaches minimum rate requirements, do not consider it for further removal. If $\gamma_{c,n} \geq \gamma$ for some subcarrier n, do not remove more users from this subcarrier. Go to step 2.
- **STEP 6** : Algorithm is terminated.

5.4.5 Solution for a special case

We now consider the special case of $K = 2$ users in a subcarrier for single-rate down-link transmission. We assume that $M \geq 2$. Our objective is to find the maximum common achievable SIR γ_c^* of users and the beamforming vectors and powers that achieve this SIR. Let $\mathcal{H}_i$, $\mathbf{u}_i$ and P_i be the spatial covariance matrix, beamforming vector and power of user i, for $i = 1, 2$. Start with initial beamforming vectors $\mathbf{u}_i^{(0)}$. In the first iteration in step 2 of Algorithm I, we find $\lambda_{max}^{(1)}$ as a function of $\mathcal{H}_i$ and $\mathbf{u}_i^{(0)}$ and the power ratio $\mu^{(1)} = P_2/P_1$. In step 3, we find beamforming vectors $\mathbf{u}_1 = \mathbf{u}_{max}(\mathcal{H}_1, \mathcal{H}_2)$ and $\mathbf{u}_2 = \mathbf{u}_{max}(\mathcal{H}_2, \mathcal{H}_1)$. In the second

iteration, we have

$$\lambda_{max}^{(2)} = \sqrt{\lambda_{max}(\mathcal{H}_1, \mathcal{H}_2)\lambda_{min}(\mathcal{H}_1, \mathcal{H}_2)} \tag{5.21}$$

and power ratio $\mu^{(2)} = \sqrt{\lambda_{max}(\mathcal{H}_1, \mathcal{H}_2)/\lambda_{min}(\mathcal{H}_1, \mathcal{H}_2)}$, where $\lambda_{max}(\mathcal{H}_1, \mathcal{H}_2)$ and $\lambda_{min}(\mathcal{H}_1, \mathcal{H}_2)$ are the maximum and minimum generalized eigenvalues of $(\mathcal{H}_1, \mathcal{H}_2)$. These do not change in subsequent iterations. Thus the maximum common SIR is

$$\gamma_c^* = \frac{1}{\sqrt{\lambda_{max}(\mathcal{H}_1, \mathcal{H}_2)\lambda_{min}(\mathcal{H}_1, \mathcal{H}_2)}} \tag{5.22}$$

with beamforming vectors $\mathbf{u}_1, \mathbf{u}_2$ and power ratio given above.

5.5 Extensions to multi-rate transmission

In the previous section, we presented three heuristic algorithms for the case of single-rate transmission. We now extend our approach to the case of multi-rate transmission, where different rates can be assigned to users, due to assignment of different number of bits in a subcarrier. We start by stating spatial separability conditions for a set of users.

5.5.1 Spatial separability conditions

Consider first the case of single-rate transmission. Recall that a cochannel set of users is called spatially separable with respect to SIR threshold γ if there exist beamforming vectors and powers such that all user SIRs exceed γ. From the discussion in the previous subsections, we have the following:

Corollary 1 *A cochannel set of users is spatially separable with respect to rate b associated with SIR threshold γ if and only if $\lambda_{max}^* \leq 1/\gamma$, where λ_{max}^* is the outcome of algorithm I.*

We now consider multi-rate transmission. Consider $m \leq M$ cochannel users in a subcarrier n. Let $\mathbf{b} = (b_1, b_2, \ldots, b_m)$ denote the rate vector of users and let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_m)$ be the associated SIR threshold vector.

A rate vector $\mathbf{b}$ is said to be *achievable* for the cochannel set of m users in subcarrier n if there exist beamforming vectors $\mathbf{U}_n$ and a power vector $\mathbf{P}_n$, such that the SIR constraints that correspond to user rates are satisfied for all m users, i.e., when $SIR_i \geq \gamma_i$, for $i = 1, \ldots, m$. Then, the set of users is called spatially separable with respect to rate vector $\mathbf{b}$. It can be verified that the SIR requirements are written in a matrix form as

$$\mathbf{P}_n \geq \hat{\mathbf{D}}\mathbf{B}(\mathbf{U}_n)\mathbf{P}_n, \tag{5.23}$$

where $\mathbf{B}(\mathbf{U}_n)$ is defined as in (5.12) and diagonal matrix $\hat{\mathbf{D}}$ now includes the SIR thresholds that correspond to a rate vector and is defined as

$$\hat{\mathbf{D}} = \text{diag}\left\{ \frac{\gamma_i}{\mathbf{u}_{n,i}^H \mathcal{H}_{n,i} \mathbf{u}_{n,i}} : i = 1, \ldots, m \right\}. \tag{5.24}$$

If $\lambda_{max}^*(\mathbf{b})$ is the eigenvalue at the end of algorithm I (where now $\hat{\mathbf{D}}$ is used instead of $\mathbf{D}$), then the following is true:

Corollary 2 *A cochannel set of users is spatially separable with respect to rate vector* $\mathbf{b}$ *if and only if* $\lambda_{max}^*(\mathbf{b}) \leq 1$.

5.5.2 Multi-rate transmission

When multi-rate transmission is used, user allocation to subcarriers should be performed based on rate benefit criteria as well. In particular, each time a new user is assigned to a subcarrier, the rate of this subcarrier should be increased. However, it may happen that upon insertion of a new user k, the SIRs of some

users decrease and hence the rates of these users need to be reduced so as to maintain acceptable BER. The assignment of a user k in a subcarrier is beneficial if the total rate in the subcarrier after insertion of the new user exceeds the rate before insertion of k.

Assume now that user k is assigned to subcarrier n. Assume that after beamforming vector and power adaptation, $b^*_{n,k}$ is the rate of user k that leads to acceptable SIR of k upon insertion of k in subcarrier n. For each user $j \in \mathcal{U}^{(n)}$, let b_j^- be the rate before insertion of user k and b_j^+ be the rate after k is inserted. Then, define the *Incremental Rate Factor (IRF)* $T_{n,k}$ as follows,

$$T_{n,k} = b^*_{n,k} + \sum_{j \in \mathcal{U}^{(n)}} (b_j^+ - b_j^-). \tag{5.25}$$

Clearly, a user with high IRF is preferable since it leads to high rate increase in the channel. The purpose of our algorithm should be to aid the insertion of a new user in subcarrier n, by adjusting beamforming vectors and powers of all users in n such that an achievable rate vector that leads to channel rate increase is found. However, the problem of determining beamforming vectors and powers in order to maximize the total subcarrier rate after insertion of a new user is a highly non-linear optimization problem.

For that reason, we can consider the tentative assignment of a user k to the subcarrier and check the achievability of rate vectors for the cochannel set of users, starting from the vector whose entries equal the maximum rate (modulation level) b_{L_0}. The achievability of a given rate vector can be checked by using the methodology and the corollary in subsection 5.5.1. Each time a rate vector is found not to be achievable, we decrease the rate of one entry and try again. This procedure is repeated until we find an achievable rate vector with IRF $T_{n,k} > 0$. If such a vector is not found, we set $T_{n,k}$ to $-\infty$ by convention. The algorithm performs user in-

sertions by selecting the pair (n^*, k^*) with the maximum IRF factor until $T_{n,k} < 0$ for all pairs of subcarriers and users. Alternatively, the achievable rate vectors can be used to define metrics such as the one in (2.16). In addition, IRF factors $T_{n,k}$ can be considered jointly with APF factors $\Phi_{n,k}$ in appropriately defined metrics.

5.6 Simulation results

5.6.1 Simulation setup

We consider a single-cell OFDM system with $K = 10$ users that are uniformly distributed in the cell area. The BS is equipped with an antenna array with M elements with $\delta = \lambda/2$. Each receiver has an omni-directional antenna. For illustrative reasons, we restrict ourselves to a system with $N = 10$ available subcarriers and single-rate transmission. Due to single-rate transmission, minimum rate requirements of users are normalized by subcarrier rate and can be considered as equivalent to a minimum number of subcarriers that need to be utilized by the user. Thus, each user k needs to use at least $x_k = 3$ subcarriers.

The received power decays with distance d from the base station as d^{-4}. For each link corresponding to an antenna and a user receiver, multi-path fading is simulated with an 2-ray model. The angle of the first path, θ_1 is uniformly distributed in $[0, 2\pi]$, while the angle of the second path, θ_2 deviates from θ_1 by a random amount, uniformly distributed in $[0, 0.1\pi]$. The complex gain of each path is an independent log-normal random variable with standard deviation $\sigma = 6\,\text{dB}$, which accounts for shadow fading. The spatial covariance matrices of users are determined by using this model.

5.6.2 Comparative results

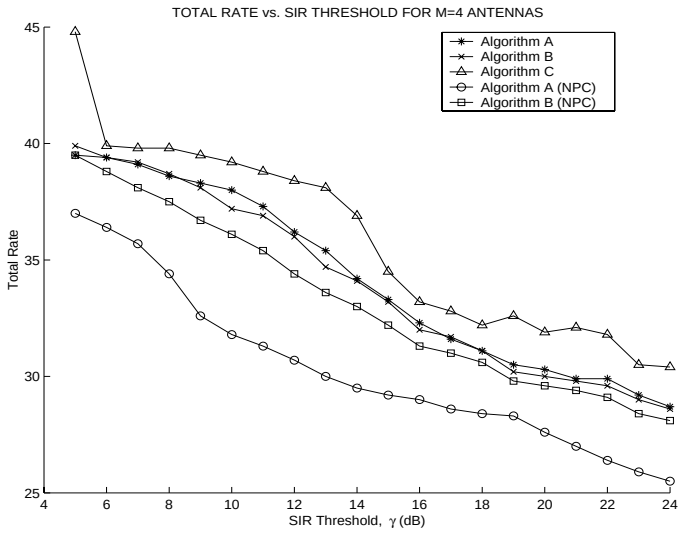


Figure 5.3: Total achievable system rate vs. SIR threshold for $M = 4$ antennas.

The primary objective of the simulations is to evaluate and compare the performance of proposed algorithms A, B and C and the different alternatives for beamforming and power control. It is also desirable to assess the performance benefit of power control in algorithms A and B. Hence, we present results for these algorithms with and without power control (NPC).

The first performance metric is the total achievable system rate in terms of total number of utilized subcarriers at the end of the algorithm. The second metric is the total residual rate, which is defined as the additional required rate so that users reach their minimum rate requirements. Clearly, an algorithm is more preferable if it yields high system rate and low total residual rate. Results were averaged over several random experiments with different channel conditions. The observed fluctuations in the plots are due to minimum rate requirements of users. When these are omitted, the derived curves are expected to be smoother.

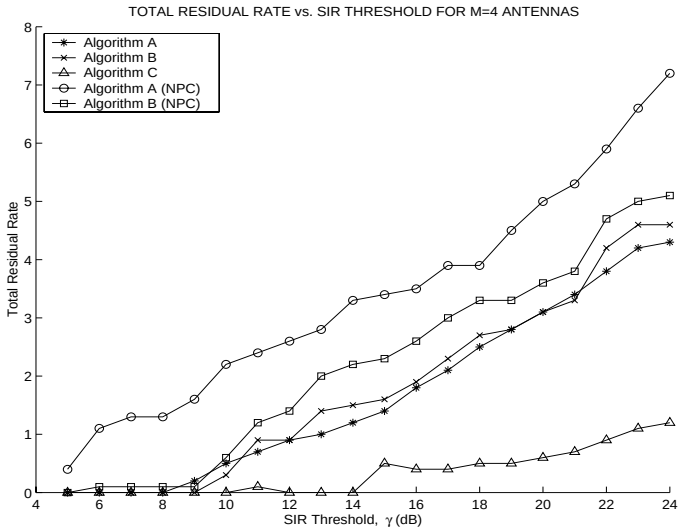


Figure 5.4: Total residual rate vs. SIR threshold for $M = 4$ antennas.

In figure 5.3, the total system rate is depicted as a function of the SIR threshold γ for an OFDM/SDMA system with $M = 4$ antennas. A high SIR threshold corresponds to a more stringent BER requirement. Algorithm C achieves the best performance for the entire range of values of γ, while algorithm A always performs slightly better than algorithm B. Furthermore, power control seems to provide significant rate benefits when incorporated in algorithm A. Thus, for moderate values of γ (in the range $10 - 15$dB), rate improvements of about $20 - 25\%$ can be achieved by power control in algorithm A, while the corresponding rate difference with power control in algorithm B is only $5 - 10\%$. In addition, the performance of algorithm B with no power control is relatively close (within $5 - 10\%$) to that of algorithm A with power control. This seems to suggest that algorithm B with no power control could be implemented in situations where reduced algorithmic complexity is a prerequisite. For large values of γ (e.g., $\gamma > 17$dB), the three of the four alternatives of algorithms A and B result in similar performance.

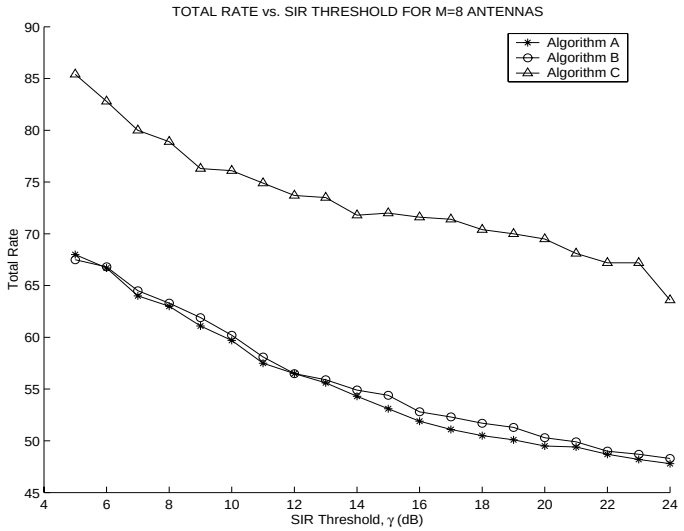


Figure 5.5: Total rate vs. SIR threshold for $M = 8$ antennas.

In figure 5.4, we depict the performance of the algorithms with respect to the total residual rate metric. Algorithm C yields much better performance than all other techniques. Minimum rate requirements of users are always satisfied for $\gamma \leq 14$dB and a very small portion of user requirements remains unsatisfied even for larger γ. Algorithms A and B are again shown to result in similar performance. Finally, the same trends are illustrated in figure 5.5 for $M = 8$ antennas. Algorithm C again leads to rate benefits of about 25% for moderate values of γ and about 40% for higher values of γ, compared to the other algorithms. For $M = 8$, the total residual rate was zero for all algorithms, which verifies the capacity improvement with more antenna elements.

By comparing the achievable rate for $M = 4$ and $M = 8$ antennas, we deduce that algorithms A and B with $M = 8$ antennas yield rate of only $30 - 35\%$ more than algorithm C with $M = 4$ antennas. At the same time, algorithm C achieves almost double rate for $M = 8$ antennas. This observation justifies the claim that

system performance depends drastically on physical layer methods and appropriate channel allocation techniques at the MAC layer. Our results also indicate that the SIR balancing algorithm C that involves joint adaptation of beamforming vectors and powers always outperforms greedy algorithms A and B, where the computation of beamforming vectors and powers is decoupled.

5.7 Conclusion

We attempted to capture the impact of SDMA on channel allocation in an OFDM system, which is characterized by orthogonal channels. The single-rate case is studied in detail and extensions for the multi-rate case are proposed. For the single-rate case, we present three heuristic algorithms for joint channel allocation, beamforming and power control. The first two algorithms use greedy assignment criteria and decouple the operations of beamforming and power control. The third one is based on SIR balancing for the assignment and uses joint beamforming and power control. Performance results demonstrate that this specific combination of SIR balancing assignment with joint beamforming and power adaptation yields significantly better performance than other algorithms.

Chapter 6

Adaptive channel allocation in OFDM-based smart antenna systems with limited transceiver resources

6.1 Introduction

In the previous chapter, we considered an OFDM/SDMA system under the assumption of unlimited transceiver resources at the BS, where a transceiver is perceived as a communication unit which is used to set up a distinct beam. According to our approach, a separate beam could be formed by a dedicated transceiver for each user in a spatially separable cochannel set in a subcarrier. Thus, for a system with N subcarriers and M antennas, the existence of NM transceiver modules was a prerequisite, since at most M users can be separated in a subcarrier.

6.1.1 Related work and motivation

The issue of limited transceiver resources at the BS has not been addressed in literature, primarily because the use of limited transceivers has not been appropri-

ately motivated. First of all, some studies concentrate on single-channel multi-user scenarios, where the assumption that the number of transceiver modules equals or exceeds the number of users M in the spatially separable cochannel user set can be considered to be valid. Thus, for a system with M transmit antennas, a separate beam can be formed by a dedicated transceiver for each user in the spatially separable cochannel set. This assumption has been adopted for example in [89, 91]. A second category of studies considers multi-channel multi-user systems and focuses on time division multiplexing. In each time slot, each transceiver forms a beam for each of the (at most) M spatially separable users and the allocation potentially changes in subsequent time slots. Thus, the single-channel case applies here as well and at most M transceivers are required in each slot. Such scenarios are described in [97, 98].

OFDM systems present some novel challenges with respect to resource allocation. Owing to the fact that channel allocation is performed in the frequency domain and because of the different impact of subcarrier frequencies on spatial channel characteristics of a user, a different beam may need to be formed in each subcarrier to ensure acceptable SIR. In single-user OFDM systems, the required transceivers must be at least as many as the subcarriers and this reasonable assumption is implied in [92].

In multi-user OFDM systems, the different spatial characteristics of users and the different perceived subcarrier quality for each user necessitate the use of a separate beam for each user in a spatially separable cochannel user set of each subcarrier. For a system with N subcarriers and M antennas, NM transceivers may need to be active at the same time. Depending on the values of N and M, this number can be of the order of several hundreds. However, high implementation

complexity and infrastructure cost, physical space inadequacy or specifications on maximum induced interference to neighboring locations and users may impose limitations on the number of beams that can be formed at a BS. These situations arise more often in WLANs, WPANs or other indoor environments. The limitation in the number of formed beams subsequently affects channel allocation, since channel assignment to users and user clustering into a limited number of beams formed by corresponding transceivers are interrelated issues. The efficiency of channel assignment to users depends on channel reuse, which in turn is determined by beam formation and by allocation of users and channels to different transceivers. With an appropriate combined assignment strategy at the base station, these issues can be jointly addressed.

In this chapter, we investigate the impact of smart antennas with limited transceiver resources on MAC layer channel allocation in an OFDM system. Our objective is to increase system capacity and provide minimum rate guarantees to users. We propose heuristic algorithms to assign spatially separable users in the same channels and distribute users and channels within available transceivers, while appropriately adjusting beam patterns by transmit beamforming. The criteria for assignment capture spatial characteristics of users, induced interference by beam patterns and beam cross-correlation properties. The main goals of our study are to evaluate the benefits of this cross-layer approach in terms of achievable system rate and identify the tradeoffs that are associated with resource (channel and transceiver) limitations.

6.1.2 Outline of chapter

The rest of the chapter is organized as follows. In section 6.2 we provide the adopted model for our approach. In section 6.3, we present the problem, outline the rationale of our approach and describe the proposed algorithms. In the same section, we provide some extensions to our approach and examine a special case of the problem. Section 6.4 contains numerical results and section 6.5 concludes this chapter. For notational remarks of this chapter, refer to subsection 5.1.3.

6.2 System model

The down-link of a single-cell system with K users is considered. The BS is equipped with a uniform linear antenna array of M elements and uses single-rate OFDM transmission with N orthogonal subcarriers and b_0 bits per utilized subcarrier.

An underlying slotted transmission scheme is again assumed. Packetized user data arrive from higher layers and are decomposed into bit streams before being transmitted to the corresponding users. Channel quality is assumed to remain constant for the duration of one time slot. Each user k is characterized by minimum rate requirements r_k (in bits/sec) that need to be satisfied by the assignment algorithm over some time interval $(0, t)$, which consists of several time slots. We will concentrate on subcarrier, transceiver and power allocation as well as beamforming within a time slot. For single-rate transmission and assuming constant rate allocation across time slots, r_k can be mapped to a minimum number of required subcarriers n_k in a slot. The block diagram of an OFDM/SDMA transmitter with limited transceiver resources is depicted in figure 6.1.

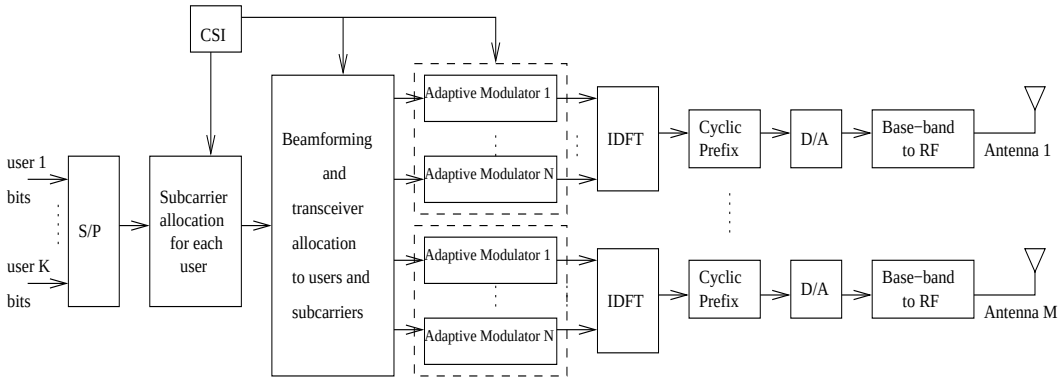


Figure 6.1: Block diagram of a multi-user OFDM/SDMA transmitter with limited transceiver resources.

The structure of the transmitter is similar to that depicted in figure 5.1 for unlimited transceiver resources. User bits again enter the subcarrier allocation module, which determines cochannel sets of users in different subcarriers. Next, beamforming is performed. The difference in the case of limited number of transceivers is that there exist only C transceivers and each of them can form a beam $\mathbf{u}_c = \left(u_c^1, \ldots, u_c^M\right)^T$, for $c = 1, \ldots, C$. Beams are normalized, i.e., $\|\mathbf{u}_c\| = 1$ and power control is not considered. A set of C transceivers is depicted in figure 6.2. Users and subcarriers are then appropriately allocated to transceivers. Beam computation, user assignment to transceivers and subcarrier allocation to users are interdependent procedures as will be shown in the sequel.

Clearly, different users that are allocated to the same transceiver (i.e, users covered by the same beam in space) must use different subcarriers. Furthermore, if two or more users use the same subcarrier, they must be assigned to different transceivers. The output of the module is then forwarded into M parallel modules of N adaptive modulators and the transmission procedure is identical to that for the system with unlimited number of transceivers in chapter 5. The adopted multi-

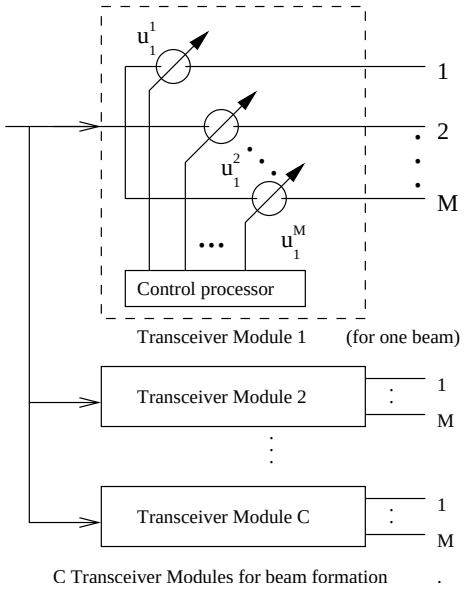


Figure 6.2: The structure of C transceiver modules.

path channel model is also similar to that described in previous chapters.

Fix attention to user k that uses subcarrier n and receives the useful signal from beam $\mathbf{u}_c$ which is formed by transceiver c. A user transmits its data through different subcarriers. It can also use several transceivers, provided that different subcarriers are activated for transmission in different transceivers. This restriction is needed, so that user signals are distinguished at the receivers. Let $d_{n,k}$ denote the transmitted subsymbol of k in subcarrier n. The received signal for user k at subcarrier n is

$$R_{n,k}^{c} = (\mathbf{a}_{n,k}^{H}\mathbf{u}_c)d_{n,k} + \sum_{\substack{b=1 \\ b\neq c}}^{C}\sum_{\substack{j=1 \\ j\neq k}}^{K}(\mathbf{a}_{n,k}^{H}\mathbf{u}_b)d_{n,j}\,. \tag{6.1}$$

The first term denotes the useful power of user k, while the second term captures cochannel interference in subcarrier n, caused by signals transmitted to other users in other beams. Define again the spatial covariance matrix of user k in subcarrier n as $\mathcal{H}_{n,k}$. The expected received power in subcarrier n of user k due to transmission

towards user $j \neq k$ in beam $b \neq c$ is $\mathbb{E}\{|\left(\mathbf{a}_{n,k}^H \mathbf{u}_b\right) d_{n,j}|^2\} = \mathbf{u}_b^H \mathcal{H}_{n,k} \mathbf{u}_b$, where it was assumed that subsymbols are normalized to unit power. We assume again that the major limitation in the system is cochannel interference rather than noise, so that receiver SINR is approximated by SIR. The expected SIR at the output of the matched filter receiver at subcarrier n of user k is

$$SIR_{n,k}^c = \frac{\mathbf{u}_c^H \mathcal{H}_{n,k} \mathbf{u}_c}{\sum_{\substack{b=1 \\ b \neq c}}^{C} \mathbf{u}_b^H \mathcal{H}_{n,k} \mathbf{u}_b}. \tag{6.2}$$

We observe that the SIR depends only on beams $\{\mathbf{u}_b\}_{b=1}^C$ that use subcarrier n for transmission and not on the identities of individual cochannel users.

We assume that estimates of the spatial covariance matrices of users are available at the BS. The BER at the output of the detector of a user in a subcarrier must not exceed ϵ as usual. With the rationale of subsection 1.2.5, the modulation level of b_0 bits per subsymbol is mapped to a minimum required SIR γ as in (1.6).

6.3 Channel allocation in OFDM/SDMA systems with limited transceiver resources

6.3.1 Problem statement

Two or more users are called *spatially separable* in a subcarrier if they simultaneously use the same subcarrier and there exist beamforming vectors, one for each user, such that the minimum SIR values at corresponding receivers are satisfied. As in the case of unlimited transceiver resources, spatial separability for a given subcarrier depends on spatial covariance matrices of users, which describe angu-

lar and multi-path channel characteristics of users. User separability also varies depending on the individual subcarrier.

Beamforming vectors also affect spatial separability. Each beamforming vector corresponds to one of the C beams that are formed by the C transceivers at the base station. Clearly, users that are illuminated by the same beam must use different subcarriers. In addition, two or more users in different beams may or may not use the same subcarrier, depending on the amount of induced cochannel interference by the beams. The latter is a function of beam orientations and of the spatial and multi-path channel characteristics of users. Each user that is assigned to a transceiver and a subcarrier receives cochannel interference from beams of other transceivers that use the same subcarrier to transmit to other users. Users and subcarriers must be associated with transceivers and beams must be computed for each transceiver, so that user SIRs for all subcarriers are acceptable and total achievable rate is increased.

The question that arises is whether there exists a joint strategy to perform subcarrier and transceiver allocation to users, as well as transceiver beam adaptation, so as to maximize the number of assigned users per subcarrier and ultimately increase system capacity. If the beamforming vectors are known, the problem reduces to that of finding maximum cardinality cochannel user sets for each subcarrier. For each subcarrier, the cardinality of a spatially separable cochannel user set is limited by the number of antennas, M. Identifying the maximum cardinality cochannel set is equivalent to finding the maximum clique in an appropriately defined graph, which is an NP-Complete problem [53]. When beamforming vectors are controllable, the problem becomes even more challenging.

Consider first the case of unlimited number of transceivers. First, a large set of

spatially separable users must be identified for each subcarrier. Then, beamforming vectors for these users must be computed, so that receiver SIRs are acceptable. The problem is that SIR at a receiver depends on beamforming vectors of other users. The enumeration of all possible user assignments in a subcarrier is of exponential complexity. In addition, even if the cochannel user set is given, the computation of beamforming vectors that lead to acceptable SIRs is not straightforward. When the constraint on the number of available transceivers comes into picture, our goal should be to reduce the number of (at most NM) initially formed beams to C. This can be done by sequentially unifying two or more beams into single beams, until the desired number of C beams is reached. Note that the NM initial beams correspond to the situation where M users can be separated in each subcarrier.

6.3.2 Proposed approach

The presented heuristic algorithms consist of two stages. In the first stage, users are assigned to subcarriers and beams are computed, assuming that the number of transceivers is unlimited. In the second stage, sets of formed beams are sequentially replaced with new single beams.

The first stage of the algorithm

The basic idea in the first stage is to create large cochannel sets of spatially separable users in each subcarrier. An appropriate user is sequentially assigned to a subcarrier and beamforming vectors of cochannel users are adjusted, so that acceptable SIRs are ensured. Inserted users should cause the least interference to users that are already assigned in the subcarrier and should receive least interference from them, so that future user assignments are facilitated. The procedure at

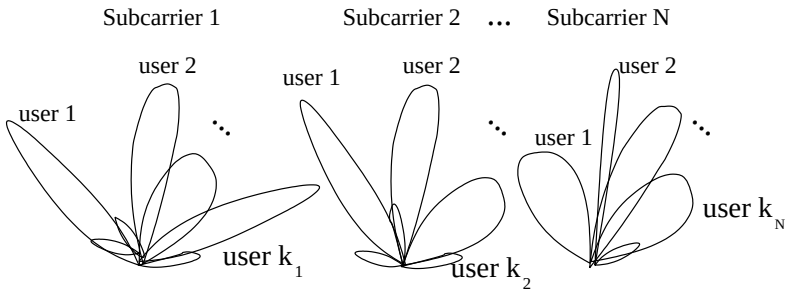


Figure 6.3: The beams for assigned users in each subcarrier after the first stage of the algorithm.

the first stage is thus similar to that in the beamforming vector adaptation portion of algorithm A in chapter 5 and it is briefly outlined here as well.

Let $\mathcal{U}^{(n)}$ be the set of users that are already assigned in subcarrier n and let k be the user to be inserted next in the channel. For each user $j \in \mathcal{U}^{(n)}$, we compute the ratio $\Psi_{n,k}^{(j)}$ as in (5.5). We also compute the ratio $\Psi_{n,k}$ for user k by using (5.6). With beamforming vectors $\mathbf{u}_{n,k}^{*}$ and $\mathbf{u}_{n,j}^{*}$, $j \in \mathcal{U}^{n}$, we evaluate the SIRs for user k and users $j \in \mathcal{U}^{(n)}$. If all SIRs exceed threshold γ, we proceed by defining an assignment preference factor (APF) $\Phi_{n,k}$ for subcarrier n and user k, similar to that in (5.10) but without the powers. This factor captures the requirement that the assigned user should have high desired signal and that it should cause and receive least interference to or from other users in n. Thus

$$\Phi_{n,k} = \frac{\mathbf{u}_{n,k}^{*H} \mathcal{H}_{n,k} \mathbf{u}_{n,k}^{*}}{\max\left\{ \mathbf{u}_{n,k}^{*H} \left(\sum_{j \in \mathcal{U}^{(n)}} \mathcal{H}_{n,j} \right) \mathbf{u}_{n,k}^{*} \, , \, \sum_{j \in \mathcal{U}^{(n)}} \mathbf{u}_{n,j}^{*H} \mathcal{H}_{n,k} \mathbf{u}_{n,j}^{*} \right\}} \, . \tag{6.3}$$

The assignment with maximum $\Phi_{n,k}$ is preferable over all users k and subcarriers n. User insertion in a subcarrier continues until no further assignment leads to acceptable user SIRs. This procedure is performed for all N subcarriers. At the end of the first stage of the algorithm there will be $\sum_{n=1}^{N} k_n$ beams, where $k_n \leq M$

is the number of users that are allocated in subcarrier n. A pictorial view of the situation at this point is shown in figure 6.3.

The second stage of the algorithm

In the second stage, the goal is to reduce the number of beams to C, while maintaining high subcarrier reuse, so as to achieve high total rate. Based on the assignment criterion from the first stage, beams $\{\mathbf{u}_{n,k}\}$ result in large subcarrier reuse and low cochannel interference. If these beams are unified, the new beams are more likely to maintain desirable properties of old beams.

It is clear that only beams in different subcarriers can be combined to one new beam, since the new beam cannot serve two users in the same subcarrier. We will consider for unification only pairs of beams, so as to reduce complexity. At each iteration of the unification algorithm, the key idea is to select the appropriate pair of beams from different subcarriers and replace it with a single beam that encompasses users in the initial beams. Different criteria for selection of beam pairs and subsequent computation of the new beam can be applied.

Fix attention to beams $(b_k, b_\ell) \equiv (k, \ell)$ that belong to subcarriers n and m respectively and have beamforming vectors $\mathbf{u}_{n,k}$ and $\mathbf{u}_{m,\ell}$. For now, assume that each beam covers one user, as is the case after termination of the first stage of the algorithm. Thus, assume that users k and ℓ are covered by beams b_k and b_ℓ respectively. Note that b_k and b_ℓ may even happen to cover the same user in subcarriers n and m. Our objective is to replace beams $\mathbf{u}_{n,k}$ and $\mathbf{u}_{m,\ell}$ with a new beam $\mathbf{u}_c$.

The rationale for the selection of a beam pair is to *combine beams of different subcarriers with similar orientations*. Then, the desirable properties of original

beams are more likely to be maintained for the new beam as well, in the sense that SIRs of users in these subcarriers will be high and cochannel interference to users residing in other beams will be low. Furthermore, since the objective is to reduce the number of beams as much as possible, our algorithm should combine beams of the same user in different subcarriers, so that users finally use several subcarriers in only one beam. Such beams also have similar orientations, since they depend on spatial covariance matrices of users which do not vary significantly in neighboring subcarriers. The algorithm selects the pair of beams (k^*, ℓ^*) with the minimum Euclidean distance among all beam pairs, i.e., it selects the pair

$$(k^*, \ell^*) = \arg\min_{(k,\ell)} \|\mathbf{u}_{n,k} - \mathbf{u}_{m,\ell}\|^2 . \tag{6.4}$$

Note that $\|\mathbf{u}_{n,k} - \mathbf{u}_{m,\ell}\|^2 = \|\mathbf{u}_{n,k}\|^2 + \|\mathbf{u}_{m,\ell}\|^2 - 2\Re(\rho_{k\ell})$, where $\rho_{k\ell} = \mathbf{u}_{n,k}^H \mathbf{u}_{m,\ell}$ is the cross-correlation between beam vectors $\mathbf{u}_{n,k}$ and $\mathbf{u}_{m,\ell}$ and $\Re(\cdot)$ denotes the real part of a complex number. For normalized beams, (6.4) reduces to

$$(k^*, \ell^*) = \arg\max_{(k,\ell)} \Re(\rho_{k\ell}) . \tag{6.5}$$

Next, the new beam $\mathbf{u}_c$ that replaces beams $\mathbf{u}_{n,k}$ and $\mathbf{u}_{m,\ell}$ must be calculated. In the sequel, we present two methods for computation of the new beam.

Approach A: Maximum new/old beam cross-correlation

The new beam vector $\mathbf{u}_c$ should have the least Euclidean distance from beam vectors $\mathbf{u}_{n,k}$ and $\mathbf{u}_{m,\ell}$, or equivalently, it should have high cross-correlation with these beams. Thus, $\mathbf{u}_c$ is computed as the solution to the following optimization problem:

$$\max_{\mathbf{u}_c} \ \Re\left[\mathbf{u}_c^H(\mathbf{u}_{n,k} + \mathbf{w}_{m,\ell})\right], \text{ subject to } \|\mathbf{u}_c\| = 1 . \tag{6.6}$$

By applying the Lagrangian multiplier method, we get the optimal solution

$$\mathbf{u}_c^* = \frac{\mathbf{u}_{n,k} + \mathbf{u}_{m,\ell}}{\sqrt{2(1 + \Re(\rho_{k\ell}))}} . \tag{6.7}$$

After computing $\mathbf{u}_c^*$, we (tentatively) replace $\mathbf{u}_{n,k}$ and $\mathbf{u}_{m,\ell}$ with $\mathbf{u}_c^*$ and evaluate the SIRs of users k, ℓ and of users in $\mathcal{U}^{(n)}$ and $\mathcal{U}^{(m)}$. Note that only users in subcarriers n and m are affected by this beam replacement. If all SIRs exceed γ, we replace beams k, ℓ with beam $\mathbf{u}_c^*$ and proceed to the selection of the next beam pair. If SIRs of some users do not exceed γ, some existing beams (and thus, users in these beams) in subcarriers n and m must be removed, so that cochannel interference to users in these subcarriers is reduced and SIRs increase. However, the elimination of a beam (user) in subcarrier n or m results in system rate decrease of b_0 bits. Thus, the number of removed beams should be kept as low as possible and hence appropriate beams for removal must be selected.

Let $V(k,\ell)$ be the set of users in subcarriers n and m (where the initial beams k and ℓ belonged) whose SIR is less than γ after computing new beam $\mathbf{u}_c^*$ and replacing beam pair (k,ℓ). Assume now that a user $\kappa \in V(k,\ell)$ is removed (together with its beam). If $\kappa \in \mathcal{U}^{(n)}$, then SIRs of users $j \in \mathcal{U}^{(n)}$ change to new values

$$SIR_j(\kappa) = \frac{\mathbf{u}_{n,j}^H \mathcal{H}_{n,j} \mathbf{u}_{n,j}}{\sum\limits_{\substack{i \in \mathcal{U}^{(n)} \\ i \neq \{j,\kappa,k\}}} \mathbf{u}_{n,i}^H \mathcal{H}_{n,j} \mathbf{u}_{n,i} + \mathbf{u}_c^{*H} \mathcal{H}_{n,j} \mathbf{u}_c^*} \,. \tag{6.8}$$

Note that user k is not included in the sum above, since beam $\mathbf{u}_{n,k}$ is removed, but its presence is implied in beam $\mathbf{u}_c^*$. Similarly, if $\kappa \in \mathcal{U}^{(m)}$, SIRs of users $j \in \mathcal{U}^{(m)}$ are affected. We choose to remove the beam b_{κ^*} (user κ^*) that leads to maximization of the minimum SIR of remaining users in the two subcarriers. Thus,

$$\kappa^* = \arg \max_{\kappa \in V(k,\ell)} \min_{\substack{j \in \mathcal{U}^{(n)} \cup \mathcal{U}^{(m)} \\ j \neq \kappa}} SIR_j(\kappa) \,. \tag{6.9}$$

By eliminating the user that maximizes the minimum SIR, we intend to keep SIRs high enough and thus increase the number of users with acceptable SIRs. The process of beam elimination according to criterion (6.9) continues until all SIRs of

users are acceptable. Then, the algorithm proceeds to the selection of the next pair of beams based on criterion (6.5) and the procedure terminates when the number of beams is reduced to C.

Approach B: Maximum signal strength/minimum induced interference

According to a second criterion, the new beam $\mathbf{u}_c$, which is the outcome of unification of beams $\mathbf{u}_{n,k}$ and $\mathbf{u}_{m,\ell}$, must lead to high desired signal strength at receivers of users k and ℓ that were covered by the original beams. It should also cause low interference to other users in subcarriers n and m. We are interested in finding the beam $\mathbf{u}_c^*$ that maximizes the following ratio

$$Z(k,\ell) = \max_{\mathbf{u}_c} \frac{\mathbf{u}_c^H \left(\mathcal{H}_{n,k} + \mathcal{H}_{m,\ell}\right) \mathbf{u}_c}{\mathbf{u}_c^H \Big(\sum\limits_{\substack{j \in \mathcal{U}^{(n)} \\ j \neq k}} \mathcal{H}_{n,j} + \sum\limits_{\substack{j \in \mathcal{U}^{(m)} \\ j \neq \ell}} \mathcal{H}_{m,j} \Big) \mathbf{u}_c}, \text{ subject to } \|\mathbf{u}_c\| = 1. \quad (6.10)$$

After computing $\mathbf{u}_c^*$, SIRs of users are calculated and users are sequentially eliminated according to (6.9), until acceptable SIRs are ensured.

6.3.3 Description of the algorithm

The main steps of the general algorithm can be summarized as follows:

- **STEP 1**: Run the first stage of the algorithm. Find a beam $\mathbf{u}_{n,k}$ for each user k in a spatially separable cochannel user set in subcarrier n.
- **STEP 2**: For each pair of beams (k,ℓ) in different subcarriers, compute cross-correlation $\rho_{k,\ell}$. Select pair (k^*,ℓ^*) with maximum cross-correlation.
- **STEP 3**: Find new beam $\mathbf{u}_c^*$ with approach A or B above.
- **STEP 4**: Perform the elimination process based on (6.9), until all user SIRs exceed γ. Unify beams k and ℓ.

- **STEP 5**: If number of remaining beams is C, terminate the algorithm. Else, go to step 2 and repeat the procedure.

The complexity of finding generalized eigenvectors of a $M \times M$ matrix is $O(M^3)$. The first stage of the algorithm involves generalized eigenvector computation for cochannel users for all NK possible assignments and for each of the (at most) NM user insertions and thus it has complexity $O(N^2KM^4)$. The second stage involves selection of the pair of beams with maximum cross-correlation (complexity $O(N^2M^2)$), computation of new beam (complexity $O(1)$ for approach A and $O(M^3)$ for approach B), elimination of users (complexity $O(M^2)$) and beam merging (complexity $O(\log(NM))$). Thus, the second stage has complexity $O\left(N^2M^2\log(NM)\right)$ for approach A and $O((N^2M^2 + M^3)\log(NM))$ for approach B.

6.3.4 Further considerations and extensions

Unification of beams with more than one users

As the algorithm progresses, one or both of the beams that are selected for unification will not include just one user in one subcarrier, as was the case in the previous subsection. A beam may contain several subcarriers of a user, or users with different subcarriers. These beams are the outcome of an earlier merging process in the algorithm. The algorithm should be modified to deal with these cases as well.

Consider a beam pair (k, ℓ) with beamforming vectors $\mathbf{u}_k$ and $\mathbf{u}_\ell$. Let beam k contain users $k_1, \ldots, k_t$, where user k_i resides in subcarrier n_i, $i = 1, \ldots, t$ and let beam ℓ contain users $\ell_1, \ldots, \ell_s$, where ℓ_i uses subcarrier m_i, $i = 1, \ldots, s$. The problem is again to compute a new beam $\mathbf{u}_c^*$ that will replace beams k and ℓ.

When approach A is applied, $\mathbf{u}_c^*$ depends only on vectors $\mathbf{u}_k$ and $\mathbf{u}_\ell$ and not on the users that reside in the beams. Thus, $\mathbf{u}_c^* = (\mathbf{u}_k + \mathbf{u}_\ell)/\sqrt{2(1 + \Re(\rho_{k,\ell}))}$,

similarly to (6.7). When approach B is considered, some changes in (6.10) are required. The new ratio must consider that the new beam $\mathbf{u}_c^*$ should yield high desired signal power for all $t+s$ users within beams k and ℓ and should cause low interference to other users in subcarriers n_i, $i=1,\ldots,t$ and m_i, $i=1,\ldots,s$. The following changes are needed in the definition of $Z(k,\ell)$:

$$\mathcal{H}_{n,k} \text{ and } \mathcal{H}_{m,\ell} \quad \text{become} \quad \sum_{i=1}^{t} \mathcal{H}_{n_i,k_i} \text{ and } \sum_{i=1}^{s} \mathcal{H}_{m_i,\ell_i}$$

$$\sum_{\substack{j\in\mathcal{U}^{(n)}\\ j\neq k}} \mathcal{H}_{n,j} \quad \text{becomes} \quad \sum_{i=1}^{t} \sum_{\substack{j\in\mathcal{U}^{(n_i)}\\ j\neq k_i}} \mathcal{H}_{n_i,j}$$

$$\sum_{\substack{j\in\mathcal{U}^{(m)}\\ j\neq \ell}} \mathcal{H}_{m,j} \quad \text{becomes} \quad \sum_{i=1}^{s} \sum_{\substack{j\in\mathcal{U}^{(m_i)}\\ j\neq \ell_i}} \mathcal{H}_{m_i,j}.$$

Next, SIRs for users in beams k and ℓ are computed. If all SIRs exceed γ, we replace $\mathbf{u}_k$ and $\mathbf{u}_\ell$ with the computed $\mathbf{u}_c^*$ and proceed to the selection of the next beam pair. If some SIRs of users in some subcarriers are not acceptable, some users that use the same subcarriers need to be eliminated. Let $\mathcal{X}$ be the set of users in beams k and ℓ, i.e., $\mathcal{X} = \left\{\cup_{i=1}^{t}\mathcal{U}^{(n_i)}\right\} \cup \left\{\cup_{i=1}^{s}\mathcal{U}^{(m_i)}\right\}$. Again, let $V(k,\ell)$ denote the set of users with unacceptable SIR. Similarly to (6.8), let $SIR_j(\kappa)$ be the SIR of user $j \in \mathcal{X}$ if user $\kappa \in V(k,\ell)$ is removed. The criterion for removal of a user is again that of maximizing the minimum SIR for remaining users, i.e.,

$$\kappa^* = \arg \max_{\kappa\in V(k,\ell)} \min_{\substack{j\in\mathcal{X}\\ j\neq\kappa}} SIR_j(\kappa), \ . \tag{6.11}$$

is removed. Note that only users and not beams are removed at each step of the procedure. However, if all users that belonged to a beam are gradually eliminated to create acceptable SIRs for used subcarriers, that beam is finally removed from the system.

Minimum rate requirements for users

If minimum rate requirements n_k for each user k are considered, the described methods need to be modified. First, assume that each beam contains one user in one subcarrier and that merging has been performed. If SIRs of some beams (users) are violated, some beams need to be removed, until SIRs are acceptable. During this process, users must continue to satisfy their minimum rate requirements after each beam elimination. Thus, if τ_k is the rate of user k before a beam elimination, the condition $\tau_\kappa - n_\kappa \geq 1$ must be added to criterion (6.9), so that elimination of beam κ and subsequent rate reduction of user κ by one subcarrier do not cause violation of n_κ. The same condition should be added in (6.11).

Extensions to the algorithm

In step 2 of the algorithm, the pair of beams for merging was selected according to a maximum cross-correlation criterion. Then, the new beam was computed by using approach A or B. We now explain a more efficient but computationally intensive method for beam selection. Assume that a new beam $\mathbf{u}_c^*$ is computed with (6.10). If SIRs of some users are not acceptable, some users need to be removed. After removing a user with criterion (6.9) or (6.11), we can compute a new beam $\hat{\mathbf{u}}_c^*$ with (6.10). Clearly, $\hat{\mathbf{u}}_c^*$ differs from $\mathbf{u}_c^*$, since the denominator of (6.10) now does not include the removed user. If SIRs are not satisfied, another user is removed and a new beam is computed. The procedure terminates when acceptable SIRs are found for all users.

6.3.5 Optimal solution for a special case

We consider the case of $C = 2$ transceivers with fixed beams $\mathbf{u}_1$ and $\mathbf{u}_2$. Each user is assigned to a transceiver and uses subcarriers in the corresponding beam. We assume that the set of subcarriers constitutes a sub-band, within which the spatial covariance matrix $\mathcal{H}_k$ for each user k is fixed. The problem is to satisfy given rate requirements x_k for each user k (where x_k denotes number of required subcarriers for single-rate transmission) and use the minimum number of subcarriers.

Let U_i be the set of users in beam $\mathbf{u}_i$, for $i = 1, 2$. At most two users from different transceivers can reuse the same subcarrier. Then, we have to find the maximum number of such user pairs, where each pair occupies a subcarrier. The problem is equivalent to finding a maximum matching on a bipartite graph. A bipartite graph $G = (U \cup V, E)$ is constructed as follows. One node for each required subcarrier of a user is added to the graph. Thus, $|U| = \sum_{i \in U_1} x_i$ and $|V| = \sum_{i \in U_2} x_i$. An edge (i, j) is added between nodes $i \in U$ and $j \in V$ (which denote subcarriers of users $\alpha \in U_1$ and $\beta \in U_2$ respectively) if SIRs of these users exceed γ, i.e., if

$$\min \left\{ \frac{\mathbf{u}_1^H \mathcal{H}_\alpha \mathbf{u}_1}{\mathbf{u}_2^H \mathcal{H}_\alpha \mathbf{u}_2} , \frac{\mathbf{u}_2^H \mathcal{H}_\beta \mathbf{u}_2}{\mathbf{u}_1^H \mathcal{H}_\beta \mathbf{u}_1} \right\} \geq \gamma \, . \tag{6.12}$$

The assignment that minimizes the number of required subcarriers is as follows. We start by finding the maximum matching $\mathcal{M}^*$. Each edge in $\mathcal{M}^*$ corresponds to a cochannel pair of users. Assign each such pair to a separate subcarrier. Then, for each user corresponding to a node that is not incident to a matched edge, consider a new subcarrier and assign the user to it. The minimum number of subcarriers to satisfy rate requirements of users equals $|\mathcal{M}^*|$ plus the number of nodes that are not incident to a matched edge.

6.4 Simulation results

6.4.1 Simulation setup

We simulate a single-cell system with $K = 15$ users that are uniformly distributed in the cell area. An antenna array with M elements and $\delta = \lambda/2$ is employed. The BS uses OFDM transmission at 5 GHz. For illustrative reasons, we restrict ourselves to a system with $N = 10$ subcarriers. The received power decays with distance d from the BS as d^{-4}. For each link corresponding to an antenna element and a receiver, multi-path fading is simulated with an L-ray model. The angle of each path is uniformly distributed in $[0, \pi]$. The delay between paths is uniformly distributed in $[0, T]$. The complex gain of each path is an independent log-normal random variable with standard deviation $\sigma = 6$ dB, which accounts for shadow fading. Results were averaged over 100 random experiments with different channel conditions and user locations.

6.4.2 Comparative results

The objective of the simulations is to evaluate and compare the performance of the proposed approaches to our problem. It is also desirable to quantify the impact of different parameters on performance. First, we do not consider minimum rate requirements and focus on achievable system rate. In order to have a fair comparison for the proposed heuristics, the following approaches are simulated:

- Approach A: The first stage of the algorithm is initially executed. At the second stage, the beam pairs are selected for unification based on criterion (6.5). The new beam is computed with (6.7). Next, beams are sequentially eliminated, based on (6.9) or (6.11), until SIRs of remaining users exceed γ.

- Approach B: The first stage of the algorithm is again executed and the beam pair selection is based on (6.5). The new beam is computed with (6.10). After beam elimination with (6.9) or (6.11), a new beam is calculated again with the modified ratio in (6.10). This iterative process of beam elimination and new beam computation terminates when user SIRs are acceptable.

The performance metric is average subcarrier throughput, which is defined as the number of assigned users per subcarrier. In figure 6.4 average subcarrier throughput is illustrated as a function of the number of available transceivers (beamformers) for $M = 4$ antennas, for different multi-path scenarios and SIR threshold $\gamma = 10\,\text{dB}$. For $M = 4$, we observe that for the same multi-path channel conditions (number of paths, L), approach B always performs better than approach A. This performance benefit is attributed to the iterative nature of approach B, where beam vectors are continuously updated, as opposed to approach A, where beam vectors are computed once. Different criteria were also utilized for beam computation in the two approaches. For $L = 1$, the difference in the performance of the two approaches is almost fixed and independent of the number of transceivers C. Approach B yields almost 25% higher rate than approach A. For $L = 2$, the difference decreases as the number of transceivers increases. For relatively small number of transceivers, approach B outperforms A by almost 20%, while for larger values of C, approach B is better than A by about 4%. An important observation is that the resulting throughput with $L = 2$ is larger than that for $L = 1$ for both approaches A and B, due to the additive effect of multi-path.

The most significant observation from figure 6.4 is that performance improves as the number of transceivers increases, but it does not improve after a certain number of transceivers C^*. This means that the system has reached its spatial

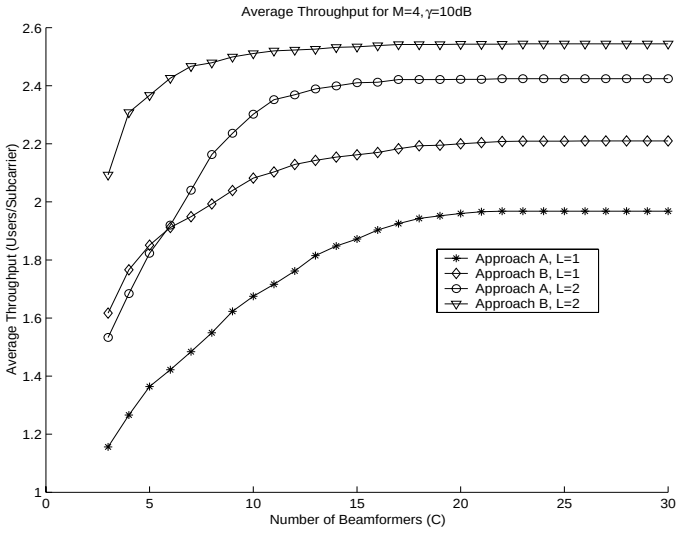


Figure 6.4: Average throughput vs. number of transceivers for approaches A and B, for multi-path with $L = 1$ and $L = 2$ paths and $M = 4$ antennas.

separability performance limits and cannot accommodate more users in the same subcarrier. For example, for approaches A and B and a LOS path, it is $C_A^* = 17$ and $C_B^* = 13$ respectively with limiting throughput of 1.9 and 2.18 users per subcarrier. As the number of paths increases to $L = 2$, the corresponding values become $C_A^* = 12$ and $C_B^* = 9$ with limiting throughput of 2.4 and 2.5 users per subcarrier respectively. Similar conclusions can be drawn from figure 6.5 for $M = 8$ antennas. It can be observed that the limiting throughput values increase and the number of transceivers for reaching this limit also increases. Thus, for approaches A and B and a LOS path, it is $C_A^* = 26$ and $C_B^* = 22$ with limits 2.25 and 2.7 users per subcarrier. For $L = 2$, the exact values of C_A^* and C_B^* cannot be deduced from the figure, but the limiting throughput is about 3 and 3.5 users per subcarrier respectively. It can be observed that the performance benefit of approach B over approach A increases with increasing number of antennas and

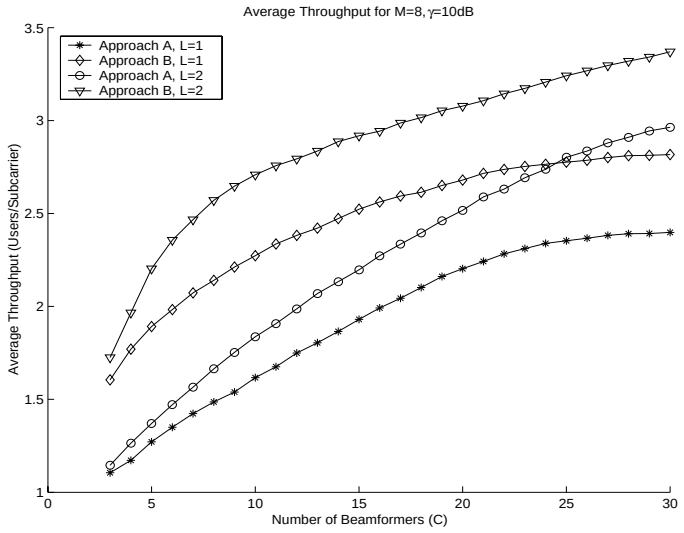


Figure 6.5: Average throughput vs. number of transceivers for approaches A and B, for multi-path with $L = 1$ and $L = 2$ paths and $M = 8$ antennas.

decreases with increasing number of paths. Furthermore, the number of required transceivers beyond which no performance improvement is anticipated, increases in proportion with the number of antennas.

When minimum rate requirements for users come into play, a meaningful performance measure is the residual rate (throughput) of users. This is defined as the number of additional subcarriers that a user needs so as to satisfy minimum throughput requirements. Clearly, an algorithm is more efficient if it yields low total residual rate. We assume that the minimum number of required channels is uniformly distributed in $\{1, 2, 3, 4, 5\}$. In figure 6.6, the total residual throughput of users is shown as a function of the number of transceivers C, for $M = 4$ antennas, with $\gamma = 10\,\mathrm{dB}$ and $L = 2$. The residual throughput for both approaches reduces as the number of transceivers increases and it can be seen that approach B performs better than approach A, when $C < 15$ transceivers. However, both

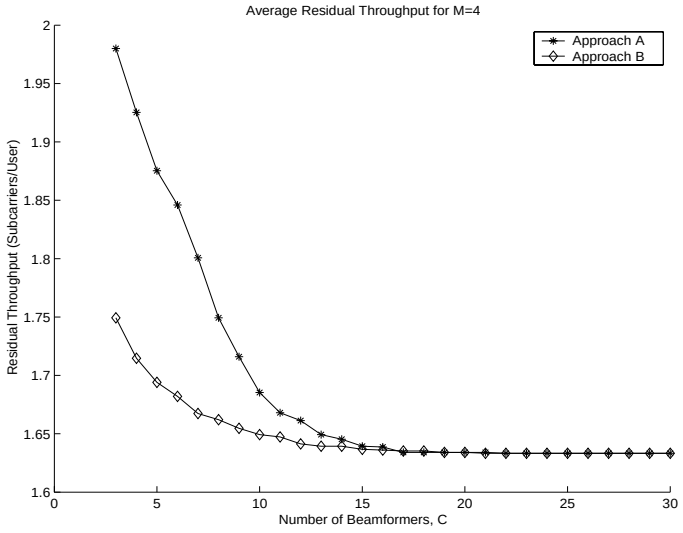


Figure 6.6: Residual throughput vs. number of transceivers for approaches A and B, for $M = 4$ antennas.

approaches have the same performance for $C > 15$ and no further reduction in residual throughput is observed. This is another indication that the system has reached its performance limits. For $M = 8$ antennas, the corresponding performance limit was $C = 31$ transceivers. These performance limits are comparable with the limits deduced by figures 6.4 and 6.5.

Finally, we evaluate the performance of the greedy assignment method at the first stage of the algorithm. This constitutes a meaningful methodology for performing subcarrier assignment and beamforming when transceiver resources are unlimited. Since approaches A and B are not used in the first stage, we are only interested in the impact of multi-path on performance. In figure 6.7, we plot the average throughput as a function of the SIR threshold γ for different multi-path conditions, where a high γ corresponds to a stringent BER requirement. The throughput is reflected in the average number of spatially separable users per sub-

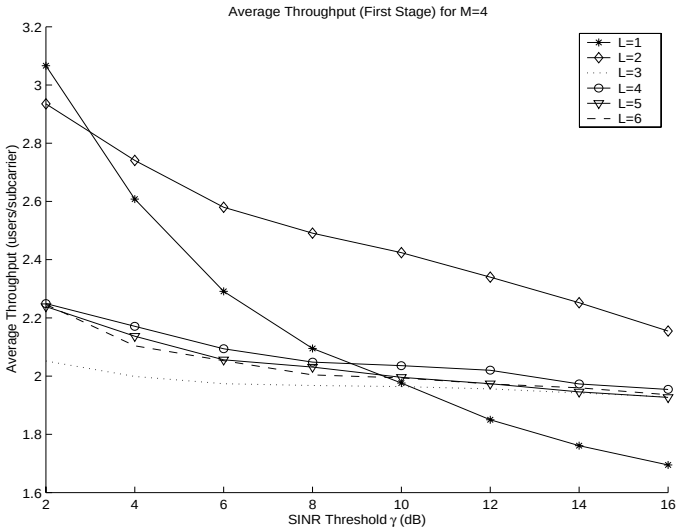


Figure 6.7: Average throughput vs. SIR threshold, for unlimited number of transceivers and $M = 4$ antennas.

carriers. We observe that for $L = 1$, throughput decays almost exponentially with increasing γ, while when $L = 2$ the rate of decay is smaller. This is another evidence of the fact that transmission over multi-path channels can lead to improved performance. For larger number of paths, e.g., $L = 3, 4, 5$ or 6, only minor differences in performance could be observed. However, average throughput for $L > 1$ is superior to that for $L = 1$ when $\gamma > 10$dB. In the limit of large SINR thresholds, two users are separable on average in a subcarrier for $L = 3, 4$ and 5.

Although in a realistic system the number of subcarriers N will be larger, subcarrier reuse will depend on spatial properties of users, beamforming and the resource (subcarrier and transceiver) allocation policy. Similar tendencies are thus anticipated in a large-scale system, with more subcarriers and users. Our results manifest the necessity for a sophisticated system design, so as to provide QoS to users and improve system performance. For a given BER requirement at the re-

ceiver and a given number of antennas, there exists a crucial number of transceivers C^*, beyond which no further performance benefits can be anticipated with regard to total system rate or total residual rate. Viewed differently, the number of transceivers can be made as small as C^* with no performance loss.

6.5 Conclusion

We addressed the joint problem of space division multiplexing and channel allocation in an OFDM-based system with limited transceiver resources. We identified the particular characteristics of this coupled resource assignment problem and we proposed heuristics for subcarrier and transceiver assignment to users, as well as adaptive beamforming, so as to increase total achievable system rate and provide QoS to users in the form of minimum rate guarantees. Our primary goal was to identify the impact of smart antennas and limited transceiver resources on MAC layer channel allocation. Our approach is novel, in the sense that the transceivers are not perceived as independent servers, but cause cochannel interference to users served by other transceivers in the same subcarrier. The proposed heuristics attempt to capture the intuition behind this composite assignment problem.

Our results indicate that the method which employs iterative beam computation based on maximum signal strength and minimum interference performs remarkably well. Moreover, there exists a crucial number of transceivers, beyond which performance cannot be improved. Subcarrier reuse and incurred throughput losses at the second stage of the algorithm quantify the impact of smart antennas and limited transceiver resources on the performance of MAC layer channel allocation. The proposed policies can thus serve as benchmarks and the illustrated plots can provide useful design criteria for real systems.

Chapter 7

Conclusion and future work

7.1 Summary of contributions

Several resource allocation problems were considered in this dissertation in the context of OFDM-based wireless broadband networks. The joint problem of channel allocation, modulation and power control for a multi-cell multi-user OFDM network was considered in chapter 2. Specifically, we focused on the impact of modulation and power control on subcarrier reuse, which constitutes a measure of user capacity. We described a framework within which base stations can cooperate in order to derive the best assignment policy in terms of total user rate. We characterized the set of achievable rate vectors and the rate region of the system and demonstrated the complexity of the problem. Next, we proposed two classes of centralized algorithms that lead to efficient resource utilization in terms of total achievable system rate. The first class of algorithms utilized criteria such as minimum induced and received interference and rate contribution of an assignment, while the second class was based on increasing the minimum SIR in a subcarrier. Both classes of algorithms were implemented with modulation control and with

or without power control. In consistency with already available results for the single-channel case, our results demonstrate that the best performance in terms of achievable system rate is obtained when both modulation level and power are controllable parameters. However, the use of modulation control alone leads to satisfactory performance with significantly lower complexity. The assignment algorithm that attempts to maximize the minimum SIR is shown to achieve the best performance out of the two classes of algorithms.

In chapter 3, the problem of subcarrier assignment for intra-cell users subject to time resource constraints was studied. Unlike the previous chapter, we adopted a different approach, aiming at using modulation control as a means of creating preferences for subcarrier assignment. Our objective was to assign time and frequency resources to users, such that users satisfy rate requirements by using the minimum number of channels. We presented the framework within which our approach is applicable, by organizing resources into sub-bands of subcarriers and by focusing on subcarrier assignment within a sub-band. For the case of time-invariant subcarrier quality, we studied integral and fractional user assignment, whereby a user is assigned exclusively to one subcarrier or can be partially assigned to more than one subcarriers. For fractional user assignment, we formulated the problem as a linear programming one and presented an algorithm that achieves the optimal solution for the special case where the sub-band consists of two subcarriers. For integral user assignment, we characterized the complexity of finding an optimal or a feasible solution and proposed a heuristic algorithm for subcarrier assignment to users. The algorithm was based on initial assignment of users to the best subcarriers and subsequent reassignments based on a minimum additional capacity increase criterion. Our approach was extended to the case of time-varying subcar-

rier quality by using the notions of virtual channel and virtual capacity. Finally, we utilized the method of Lagrangian relaxation to obtain performance bounds for the integral user assignment and we showed that our heuristics belong to the general class of algorithms that stem from Lagrangian relaxation. The performance of our heuristics was shown to be very satisfactory compared to the optimal solution, with regard to the percentage of feasible solutions and the quality of the solutions.

The synergy between link-layer ARQ protocols and physical layer parameter adaptation in the context of OFDM was considered in chapter 4. Controllable parameters were FEC code rate and modulation level. We considered a simple channel monitoring method that was based on counting successive ACKs and NACKs and defined a throughput metric that captures the effects of the transmission policy on the achievable rate and the retransmissions. For the single-user single-subcarrier case, we formulated the transmission parameter adaptation problem as a Markov decision process one and showed that the policy that maximizes the long-term average throughput per unit time is of threshold type. The optimal policy is simple and intuitive. Transmission rate should be increased whenever the number of successive ACKs exceeds a threshold and should be decreased whenever the number of successive NACKs exceeds a threshold. After identifying the difficulty of applying this policy in a time-varying wireless link, we devised a methodology to heuristically estimate the threshold values independently of channel conditions. Next, we extended this approach to the case of multiple subcarriers, which is more applicable to OFDM. For subcarriers of the same quality, we showed that the optimal policy is again of threshold type. For subcarriers of different quality however, additional issues arise. The impact of several parameters on system performance was quantified. We demonstrated the nature of the rate adaptation policy. It should be

conservative for rate increase and it should respond fast for rate decrease. We also noted the differences when adapting transmission between different rate pairs. The performance of the adaptation policies with the computed suboptimal thresholds was shown to be very close to the ideal one. For OFDM systems with multiple subcarriers, the ACK/NACK feedback may prove a valuable tool for performing rate adaptation, since threshold ratios are shown to be of the order of a few tens or hundreds.

In chapter 5, the impact of smart antennas on MAC layer resource allocation for a single-cell multi-user OFDM system with unlimited transceiver resources was studied. We obtained a detailed transmission model for OFDM/SDMA and demonstrated the exact impact of spatial and multi-path parameters on channel quality. We started from the case of single-rate transmission and proposed algorithms for joint channel allocation, beamforming and power control with the objective to increase channel reuse and provide minimum rate guarantees to users. The first two algorithms utilized criteria such as minimum additional interference or minimum SIR in a subcarrier to perform user assignment. In these algorithms, power control was activated only when necessary. The third algorithm aimed at providing the highest common SIR per channel by simultaneous adaptation of beamforming and power control. This last algorithm was the most efficient in terms of providing the highest total rate and the least residual rate, suggesting that SIR balancing algorithms with joint consideration of beamforming and power adaptation are more preferable. Our ideas were also extended to the case of multi-rate transmission and conditions for rate achievability were obtained, which could provide the guidelines for the design of efficient heuristic algorithms.

Finally, in chapter 6, we addressed the resource allocation issue that arises in

OFDM-based smart antenna systems with limited transceiver resources. We identified the constraints associated with this coupled assignment problem. Users that are assigned in the same transceiver should use different subcarriers and users that are assigned in the same subcarrier and different transceivers may interfere due to the beam patterns of corresponding transceivers. We presented some meaningful heuristics for transceiver and subcarrier assignment to users, as well as beamforming. The algorithms consisted of two stages. In the first stage, the assignment was performed under no constraints in the number of transceivers. In the second stage, the allocation was adjusted to the limited number of transceivers by unifying beams with criteria such as beam cross-correlation and interference. The best performance is obtained by a beam unification algorithm that iteratively performs beam computation and elimination of users with unacceptable SIRs. Moreover, we found that there exists a crucial number of beamformers, beyond which system performance cannot be further improved. This number specifies the performance limits of the system with respect to spatial separation capabilities. These results could provide very useful design criteria for practical systems.

7.2 Further extensions

In chapter 2, the problem of joint channel allocation, modulation and power control was addressed and the objective was to maximize total achievable system rate. We recognize that the algorithm may treat users with poor channel conditions unfairly, in the sense that these users may not achieve high rates and may not satisfy their QoS requirements. Therefore, it would be interesting to address the problem within the framework of fairness in rate allocation. Fairness could be incorporated in the model by introducing minimum rate guarantees or by defining some other fairness

criteria for rate allocation. The issue of providing fair rate allocations with OFDM-based transmission is a challenging one because of the different dimensions in which rate allocation takes place and the better rate granularity. In a single-cell system, rate allocation is performed in the frequency domain, with bit allocation of users across subcarriers, as well as in the time domain with appropriate user scheduling methods. In a multi-cell system, additional degrees of freedom are BS activation sets and user selection from different cells.

Our policies for resource allocation on a session basis can also be viewed in the context of rate- and power-aided admission control. Whenever the number of user sessions that request service increases or link conditions deteriorate, the system should guarantee the existence of a feasible allocation of subcarriers, powers and modulation levels to admitted sessions through an appropriate admission control policy. If no feasible assignment of the aforementioned parameters exists that enables existing sessions and the new session to be accommodated with the desired QoS, the admission control policy should include a rule for down-grading QoS of some users.

In our analysis, we considered a snapshot model of the system and concentrated on efficient resource allocation to users. Thus, we assumed the existence of infinite-length buffers that handle session traffic at BSs and we did not consider the impact of the physical layer adaptation on buffer dynamics of the system. Furthermore, no particular arrival or channel variation patterns were adopted. However, transmission rate adaptation presents some novel challenges in the case of finite-length buffers. Since different transmission rates extract different amounts of bits from the buffer, the objective of maximizing the total achievable system rate while maintaining bounded buffer lengths becomes meaningful. The metrics that describe

amounts of interference and rate increments should be enhanced with novel ones that capture the requirement for appropriate buffer management. In a single-cell system, the objective is to devise subcarrier and bit allocation policies for users, such that buffer lengths do not grow without bound and the total achievable rate is maximized. For a multi-cell system, the corresponding task would be to identify composite policies for scheduling BS activation sets and allocating users within BSs to subcarriers, with the same objective as in the single-cell case.

Finally, the proposed heuristic algorithms are centralized and are executed by a central agent that has global network information. Although such algorithms can be applicable in small- or moderate-scale systems, their complexity is prohibitive for larger-scale systems. An interesting topic for investigation would be to devise distributed versions of such algorithms, which could reduce the amount of coordination between BSs and would be easier to implement in real time. Distributed algorithms would be executed independently in each cell. As a first step towards this direction, single transmitter-user pairs and a single subcarrier frequency could be considered. By performing interference measurements, a user can choose when to enter the channel and with which rate, or when to defer from entering the channel, aiming at maximizing its own rate benefit. The structure of this problem may require the adoption of utility function models for users and may lead to iterative algorithms.

With respect to the subcarrier assignment problem of chapter 3, interesting issues arise if we consider power control on a subcarrier basis together with modulation adaptation. Power control adds a new dimension to the problem, since it changes the conditions for feasible subcarrier allocation to users. In particular, it would be of interest to study the problem of fractional user assignment in con-

junction with a total power constraint for each user over all utilized subcarriers. Power control and subcarrier assignment can also be employed jointly to solve or mitigate a transmission impairment that arises in OFDM, namely that of high peak-to-average power. This issue arises in the transmitted signal in the time-domain, due to the fact that each time sample is a superposition of transmitted subsymbols in different subcarriers, each multiplied by a different power factor. Selective user assignment to subcarriers and simultaneous power adaptation per subcarrier could help resolve this issue by maintaining a low peak-to-average-power ratio (PAPR). In addition, the constraint in PAPR could be incorporated into the original problem of subcarrier assignment to users with the objective of minimizing the number of utilized channels. This would constitute another aspect of a cross-layer approach, where the MAC layer action of resource allocation considers a physical layer constraint about PAPR.

An immediate further extension of the link adaptation policies of chapter 4 is the multi-user case, where each user occupies a set of subcarriers and needs to satisfy some rate requirements. A subcarrier allocation policy specifies the number and identities of subcarriers of these users and the rate at each subcarrier. Given the ACK/NACK feedback per user, an adaptation policy may now consist of combined rate adaptation and subcarrier allocation to users. For example, a policy could be to decrease rate in some subcarriers and assign some new subcarriers to a user, or to increase rate in some subcarriers of another user and remove other subcarriers from that user. It would be worthwhile to understand the structure of such adaptation policies.

A further direction for the problem considered in chapter 5 is to elaborate in the case of multi-rate transmission and assess the performance of different heuristic

algorithms that could be designed along the lines of that chapter. Another perspective of the problem would be to extend our treatment to multi-cell systems, where a user is characterized by a different spatial signature or spatial covariance matrix with respect to each BS. BS assignment can balance traffic loads, alleviate interference and improve system performance when combined with appropriate beamforming and power control. As a first step, single-channel one-dimensional multi-cell systems could be considered, where a user can be assigned to one of at most three surrounding BSs. The identification of meaningful objectives and heuristics and the incorporation of channel allocation as another dimension to improve performance are some of the issues that warrant further investigation.

Finally, we draw the analogy between the addressed assignment problem in chapter 6 and the corresponding scheduling problem that arises at the packet level. User packets arrive at buffers and need to be transmitted according to a scheduling policy that can be applied in each time slot. Each one of the C transceivers can be viewed as a server that serves user packets. In a given subcarrier frequency, up to M user queues can be served by the servers with SDMA, where M is the number of antennas. These users form a user activation set for that subcarrier. However, since the formed beams may interfere with each other, there exist certain constraints on the eligible user activation sets for scheduling. A scheduling policy at each slot consists of determining such user activation sets for each subcarrier. Furthermore, it can be shown that the employment of n subcarriers by a user for splitting each packet symbol is equivalent to transmission of n reference-length packets from the corresponding user queue. Thus, the arising issue is that of forming user activation sets for each subcarrier by considering the number of subcarriers per user and the associated impact on queue lengths. Several ideas and stabilizing scheduling

policies that have been proposed in literature for simple systems could be extended to such generalized scheduling problems.

7.2.1 Extensions to other multiple access schemes

Although the treatment of all the resource allocation problems was presented for OFDM, the basic principles can be applied to other multiple access schemes as well. In TDMA, the channels that are allocated to users are orthogonal time slots and resource allocation is performed on a slot basis. Assuming that channel quality remains invariant over several slots, the difference in TDMA is that the quality of different channels over which the allocation is performed is the same for a user. Hence, a user is characterized by a fixed gain or fixed spatial covariance matrix across all channels for single-antenna and multi-antenna transmission respectively. In a CDMA system with a pool of deterministic codes, different users use different codes to modulate and transmit their symbols. These codes can in general be non-orthogonal to each other due to non-zero pairwise cross-correlations. Therefore, a user that utilizes a code receives cochannel interference by other users that use the same code, as well as inter-channel interference from other utilized codes which are correlated with that code. For single-antenna transmission, codes can be reused by users in different cells, while with multi-antenna directional transmission, they can be reused even by users in the same cell. A unified framework that encompasses operation of TDMA, OFDM and CDMA with SDMA is included in [36].

In OFDM, transmission rate adaptation was achieved by modulating each sub-carrier with a different number of bits. The inherent tradeoff between high rate and sustainable amount of cochannel interference was identified. That is, a high modulation level yields high rate per channel, but renders channel reuse more dif-

ficult, since it requires higher SINR (and hence, lower interference) to maintain an acceptable BER. In CDMA, transmission rate can be varied by adapting the spreading gain (number of chips) for a code or by adapting the modulation level per user symbol. Similar tradeoffs can be identified here as well. For example, codes with low spreading gain achieve higher rates, but they usually have higher cross-correlation with other codes and they are associated with lower SIRs. A first attempt to tackle the problem of allocation of variable spreading-gain deterministic codes with the objective to maximize total system rate and provide minimum rate guarantees to users was recently published in [101].

7.2.2 Extensions to higher layers

The purpose of this dissertation is to address some of the issues that arise from a unified consideration of the physical and the MAC layer. In the previous subsections, we also addressed some topics for future investigation with respect to physical layer considerations on scheduling problems.

The common denominator of all presented problems is that they are all concerned with single-hop transmission from a transmitter to one or more receivers. Then, the objective is to find a channel assignment in terms of cochannel user sets, in conjunction with appropriate adaptation of transmission parameters, such that acceptable link quality is ensured. In multi-hop networks, this channel assignment represents a set of transmitter-receiver pairs (links) that are activated for transmission at the same channel, so that transmission conflicts are avoided. This information is passed to the routing layer, which needs to determine routing paths to forward traffic from the source to the destination. In that case, the routing algorithm and the metrics used for routing decisions should capture the effects

of channel assignment and transmission parameter adaptation, so that messages are routed subject to end-to-end QoS guarantees, such as delay and throughput. Cross-layer protocol design that embraces the physical, MAC and routing layers is another broad research area that is open for future investigation.

BIBLIOGRAPHY

[1] V.H. MacDonald, "The cellular concept", *The Bell Systems Technical Journal,* vol.58, no.1, pp.15-43, 1979.

[2] S. Nanda, K. Balachandran and S. Kumar, "Adaptation techniques in wireless packet data services", *IEEE Communications Magazine,* vol.38, no.1, pp.54-64, Jan. 2000.

[3] IEEE 802.16 WG, "Air Interface for Fixed Broadband Wireless Systems", *IEEE Standard for Local and Metropolitan Area Networks,* Dec. 2001.

[4] IEEE 802.11 WG, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHz Band", *Supplement to IEEE 802.11 standard,* Sept. 1999.

[5] IEEE 802.11 WG, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-speed Physical Layer in the 2.4 GHz Band", *Supplement to IEEE 802.11 standard,* Sept. 1999.

[6] ETSI, "Broadband radio access networks (BRAN); HIPERLAN Type 2 technical specification, Part I: physical layer", Oct. 1999.

[7] S. Lin and D.J. Costello and M.J. Miller, "Automatic-repeat request error-control schemes", *IEEE Communications Magazine,* vol.22, no.12, pp.5-16, Dec. 1984.

[8] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey", *IEEE Personal Communications,* vol.3, no.3, pp.10-31, June 1996.

[9] K.N. Sivarajan, R.J. McEliece and J.W. Ketchum, "Channel assignment in cellular radio", *Proceedingss IEEE 39th Vehicular Technology Conference,* vol.2, pp.846-850, 1989.

[10] C.W. Sung and W.S. Wong, "Sequential packing algorithm for channel assignment under cochannel and adjacent-channel interference constraint", *IEEE Transactions on Vehicular Technology,* vol.46, no.3, pp.676-686, Aug. 1997.

[11] R. Mathar and J. Mattfeldt, "Channel assignment in cellular radio networks", *IEEE Transactions on Vehicular Technology,* vol.42, no.4, pp.647-655, Nov. 1993.

[12] D. Everitt and D. Manfield, "Performance analysis of cellular mobile communication systems with dynamic channel assignment", *IEEE Journal on Selected Areas in Communications,* vol.7, no.8, pp.1172-1180, Oct. 1989.

[13] P.A. Raymond, "Performance analysis of cellular networks", *IEEE Transactions on Communications,* vol.39, no.12, pp.1787-1793, Dec. 1991.

[14] B. Vucetic, "An adaptive coding scheme for time-varying channels", *IEEE Transactions on Communications,* vol.39, no.5, pp.653-663, May 1991.

[15] M.B. Pursley and S.D. Sandberg, "Variable-rate coding for meteor-burst communications", *IEEE Transactions on Communications,* vol.37, no.11, pp.1105-1112, Nov. 1989.

[16] J.G. Proakis, *Digital Communications,*Mc Graw-Hill, 2000.

[17] N. Morinaga, M. Nakagawa and R. Kohno, "New concepts and technologies for achieving highly reliable and high-capacity multimedia wireless communication systems", *IEEE Communications Magazine,* vol.37, no.1, pp.34-40, Jan. 1997.

[18] A.J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels", *IEEE Transactions on Communications,* vol.45, no.10, pp.1218-1230, Oct. 1997.

[19] T. Ue, S. Sampei, N. Morinaga and K. Hamagushi, "Symbol rate and modulation level-controlled adaptive modulation/TDMA/TDD system for high-bit rate wireless data transmission", *IEEE Transactions on Vehicular Technology,* vol.47, no.4, pp.1134-1147, Nov. 1998.

[20] J. Zander, "Performance of optimum transmitter power control in cellular radio systems", *IEEE Transactions on Vehicular Technology,* vol.41, no.1, pp.57-62, Feb. 1992.

[21] T. Keller and L. Hanzo, "Adaptive multicarrier modulation: a convenient framework for time-frequency processing in wireless communications", *Proceedings of the IEEE,* vol.88, no.5, pp.611-640, May 2000.

[22] J.S. Chow, J.C. Tu and J.M. Cioffi, "A discrete multitone transceiver system for HDSL applications", *IEEE Journal on Selected Areas in Communications,* vol.9, no.6, pp.895-908, Aug. 1991.

[23] T. Keller and L. Hanzo, "Adaptive modulation techniques for duplex OFDM transmission", *IEEE Transactions on Vehicular Technology,* vol.49, no.5, pp.1893-1905, Sept. 2000.

[24] K. Sheikh, D. Gesbert, D. Gore and A. Paulraj, "Smart antennas for broadband wireless access networks", *IEEE Communications Magazine,* vol.37, no.11, pp.100-105, Nov. 1999.

[25] G. Okamoto, *Smart antenna systems and wireless LANs,* Kluwer Academic Publishers, Boston, 1999.

[26] H. Sampath, S. Talwar, J. Tellado, V. Erceg and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: Design, performance and field trial results", *IEEE Communications Magazine,* vol.40, no.9, pp.143-149, Sept. 2002.

[27] V. Tarokh, N. Seshadri and A.R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction", *IEEE Transactions on Information Theory,* vol.44, no.2, pp.744-765, March 1998.

[28] L.C. Goddara, "Applications of antenna arrays to mobile communications, Part I: Performance improvement, feasibility and system considerations", *Proceedings of the IEEE,* vol.85, no.7, pp.1031-1060, July 1997.

[29] L.C. Goddara, "Applications of antenna arrays to mobile communications, Part II: Beam-forming and direction-of-arrival considerations", *Proceedings of the IEEE,* vol.85, no.8, pp.1195-1245, Aug. 1997.

[30] A. Narula, M.D. Trott and G.W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays", *IEEE Transactions on Information Theory,* vol.45, no.7, pp.2418-2433, Nov. 1999.

[31] R.A. Monzingo and T.W. Miller, *Introduction to adaptive antenna arrays,* Wiley, 1980.

[32] I. Koutsopoulos and L. Tassiulas, "Channel state-adaptive techniques for throughput enhancement in wireless broadband networks", *Proceedings of IEEE INFOCOM,* vol.2, pp.757-766, 2001.

[33] I. Koutsopoulos and L. Tassiulas, "Carrier assignment algorithms in wireless broadband networks with channel adaptation", *Proceedings of IEEE International Conference on Communications,* vol.5, pp. 1401-1405, 2001.

[34] I. Koutsopoulos and L. Tassiulas, "Link adaptation policies for wireless broadband networks", *Proceedings of IEEE Global Telecommunications Conference,* vol.1, pp. 572-576, 2001.

[35] I. Koutsopoulos and L. Tassiulas, "Adaptive resource allocation in SDMA-based wireless broadband networks with OFDM signaling", *Proceedings of IEEE INFOCOM,* vol.3, pp.1376-1385, 2002.

[36] I. Koutsopoulos, T. Ren and L. Tassiulas, "The impact of Space Division Multiplexing on resource allocation: A unified approach", *accepted for publication in IEEE INFOCOM,* 2003.

[37] J. C.-I. Chuang, "Performance issues and algorithms for dynamic channel assignment", *IEEE Journal on Selected Areas in Communications,* vol.11, no.6, pp.955-963, Aug. 1993.

[38] J. C.-I. Chuang and N.R. Sollenberger, "Spectrum resource allocation for wireless packet access with application to advanced cellular internet service", *IEEE Journal on Selected Areas in Communications,* vol.16, no.6, pp.820-829, Aug. 1998.

[39] T.M. Cover and J.A. Thomas, *Elements of information theory,* Wiley, 1991.

[40] T.J. Willink and P.H. Wittke, "Optimization and performance evaluation of multicarrier transmission", *IEEE Transactions on Information Theory,* vol.43, no.2, pp.426-440, March 1997.

[41] H. Yin and H. Liu, "An efficient multiuser loading algorithm for OFDM-based broadband wireless systems", *IEEE Global Telecommunications Conference,* vol.1, pp.103-107, 2000.

[42] W. Yu and J.M. Cioffi, "FDMA capacity of Gaussian multi-access channels with ISI", *IEEE Transactions on Communications,* vol.50, no.1, pp.102-111, Jan. 2002.

[43] C.Y. Wong, R.S. Cheng, K. Ben Letaief and R.D. Murch, "Multiuser OFDM with adaptive subcarrier, bit and power allocation", *IEEE Journal on Selected Areas in Communications,* vol.17, no.10, pp.1747-1758, Oct. 1999.

[44] L. Hoo, J. Tellado and J.M. Cioffi, "Discrete dual QoS loading algorithms for multicarrier systems", *IEEE International Conference on Communications,* vol.2, pp.796-800, 1999.

[45] C. Yih and E. Geraniotis, "Adaptive modulation, power allocation and control for OFDM wireless networks", *11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications,* vol.2, pp.809-813, 2000.

[46] H.-J. Su and E. Geraniotis, "A distributed power allocation algorithm with adaptive modulation for multi-cell OFDM systems", *5th IEEE International Symposium on Spread Spectrum Techniques and Applications,* vol.2, pp.474-478, 1998.

[47] S. Papavassiliou and L. Tassiulas, "Improving the capacity of wireless networks through integrated channel base station and power assignment", *IEEE Transactions on Vehicular Technology,* vol.47, no.2, pp.417-427, May 1998.

[48] J. Zander, "Distributed cochannel interference control in cellular radio systems", *IEEE Transactions on Vehicular Technology,* vol.41, no.3, pp.305-311, Aug. 1992.

[49] G.J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence", *IEEE Transactions on Vehicular Technology,* vol.42, no.4, pp.641-646, Nov. 1993.

[50] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems", *IEEE Transactions on Communications,* vol.47, no.6, pp.884-894, June 1999.

[51] S.-L. Kim, Z. Rosberg and J. Zander, "Combined power control and transmission rate selection in cellular networks", *IEEE 50th Vehicular Technology Conference - Fall,* vol.3, pp.1653-1657, 1999.

[52] T. Fong, P. Henry, K. Leung, X. Qiu and N. Shankaranarayanan, "Radio resource allocation in fixed broadband wireless networks", *IEEE Transactions on Communications,* vol.46, no.6, pp.806-817, June 1998.

[53] M.R. Garey and D.S Johnson, *Computers and intractability: A guide to the theory of NP-completeness,* Freeman, New York, 1979.

[54] R.A. Horn and C.R. Johnson, *Matrix Analysis,* Cambridge University Press, MA 1999.

[55] S. Papavassiliou and L. Tassiulas, "Joint optimal base station and power assignment for wireless access", *IEEE/ACM Transactions on Networking,* vol.4, no.6, pp.857-872, Dec. 1996.

[56] W.T. Webb and R. Steele, "Variable rate QAM for mobile radio", *IEEE Transactions on Communications,* vol.43, no.7, pp.2223- 2230, July 1995.

[57] A. Goldsmith, "Adaptive modulation and coding for fading channels", *Proceedings of the IEEE Information Theory and Communications Workshop,* pp.24-26, 1999.

[58] T. Ikeda, S. Sampei and N. Morinaga, "TDMA-based adaptive modulation with dynamic channel assignment for high-capacity communications systems", *IEEE Transactions on Vehicular Technology,* vol.49, no.2, pp.404-412, March 2000.

[59] J. Chuang and N. Sollenberger, "Beyond 3G: Wideband wireless data access based on OFDM and dynamic packet assignment", *IEEE Communications Magazine,*, vol.38, no.7, pp.78-87, July 2000.

[60] W. Rhee and J.M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation", *Proceedings of IEEE 51st Vehicular Technology Conference,* vol.2, pp.1085 -1089, 2000.

[61] V. Tralli, A. Vaccari, R. Verdone and O. Andrisano, "Adaptive time and frequency resource assignment with COFDM for LMDS systems", *IEEE Transactions on Communications,* vol.49, no.2, pp.235-238, Feb. 2001.

[62] L.A. Wolsey, *Integer Programming,* Wiley 1998.

[63] M.L. Fisher, R. Jaikumar and L.N. Van Wassenhove, "A multiplier adjustment method for the generalized assignment problem", *Management Science,* vol.32, no.9, pp.1095-1103, Sept. 1986.

[64] D.G. Luenberger, *Introduction to linear and non-linear programming,* Addison-Welsey, 1984.

[65] M.L. Fisher, "The Lagrangian relaxation method for solving integer programming problems", *Management Science,* vol.27, no.1, pp.1-18, Jan. 1981.

[66] T.H. Cormen, C.E. Leiserson and R.L. Rivest, *Introduction to Algorithms,* McGraw Hill, 1990.

[67] J.K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels", *IEEE Transactions on Vehicular Technology,* vol.40,no.4, pp.686-693, Nov.1991.

[68] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems", *IEEE Transactions on Vehicular Technology,* vol.49, no.4, pp.1207 -1215, July 2000.

[69] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems", *IEEE Transactions on Signal Processing,* vol.49, no. Issue: 12 , Dec. 2001 Page(s): 3065 -3073

[70] S.D. Sandberg and M.B. Pursley, "Retransmission schemes for meteor-burst communications", *Proceedings of 9th annual International Phoenix conference on Computers and Communications,* pp.246 -253, 1990.

[71] M. Rice and S.B. Wicker, "Adaptive error control for slowly varying channels", *IEEE Transactions on Communications,* vol.42, no.2/3/4, pp.917-926, Feb.-April 1994.

[72] H. Brunnel and M. Moeneclaey, "On the throughput performance of some continuous ARQ strategies with repeated transmissions", *IEEE Transactions on Communications,* vol.34, no.3, pp.244-249, March 1986.

[73] Y.-D. Yao, "An effective GBN-scheme for variable-error-rate channels", *IEEE Transactions on Communications,* vol.43, no.1, pp.20-23, Jan. 1995.

[74] A. Annamalai and V.K. Bhargava, "Analysis and optimization of adaptive multicopy transmission ARQ protocols for time-varying channels", *IEEE Transactions on Communications,* vol. 46, no.10, pp.1356-1368, Oct. 1998.

[75] E. Modiano, "An adaptive algorithm for optimizing the packet size used in wireless ARQ protocols", *Wireless Networks,* vol.5, no.4, pp.279-286, July 1999.

[76] M. Naijoh, S. Sampei, N. Morinaga an Y. Kamio, "ARQ schemes with adaptive modulation/TDMA/TDD systems for wireless multimedia communication services",*Proceedings of the 8th International Symposium on Personal, Indoor and Mobile Radio Communications,* vol.2, pp.709-713, 1997.

[77] H. Li, G. Malmgren and M. Pauli, "Performance comparison of the radio link protocols of IEEE802.11a and HIPERLAN/2", *Proceedings of IEEE 52nd Vehicular Technology Conference,* vol.5, pp.2185-2191, Fall 2000.

[78] E. Ayanoglu, "Adaptive ARQ/FEC for multitone transmission in wireless networks", *Proceedings IEEE Global Telecommunications Conference,* vol.3, pp.2278 -2283, 1995.

[79] H. Li, J. Lindskog, G. Malmgren, G. Miklos, F. Nilsson and G. Rydnell, "Automatic repeat request (ARQ) mechanism in HIPERLAN/2", *Proceedings of IEEE 51st Vehicular Technology Conference,* vol.3, pp.2093-2097, Spring 2000.

[80] S. Ross, *Introduction to stochastic dynamic programming*, Academic Press, 1983.

[81] S. Papavassiliou, L. Tassiulas and P. Tandon, "Meeting QoS requirements in a cellular network with reuse partitioning", *IEEE Journal on Selected Areas in Communications,* vol.12, no.8, pp.1389-1400, Oct. 1994.

[82] Z. Rosberg, P.P. Varaiya and J.C. Walrand, "Optimal control of service in tandem queues", *IEEE Transactions on Automatic Control,* vol.27, no.3, pp.600-610, June 1982.

[83] B. Sklar, *Digital Communications: Fundamentals and Applications*, Prentice-Hall, 2001.

[84] OPNET Modeler, *3rd Millenium Technologies,* http://www.mil3.com.

[85] F. Rashid-Farrokhi, L. Tassiulas and K.J. Ray Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays", *IEEE Transactions on Communications,* vol.46, no.10, pp.1313-1324, Oct. 1998.

[86] F. Rashid-Farrokhi, K.J.Ray Liu and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems", *IEEE Journal on Selected Areas in Communications,* vol.16, no.8, pp.1437-1450, Oct. 1998.

[87] M. Olfat, K.J.Ray Liu and F. Rashid-Farrokhi, "Low complexity adaptive beamforming and power allocation for OFDM over wireless networks", *Proceedings IEEE International Conference on Communications,* vol.1, pp.523-527, 1999.

[88] M. Schubert and H. Boche, "Solvability of coupled downlink beamforming problems", *Proceedings of IEEE Global Telecommunications Conference,* vol.1, pp. 614-618, 2001.

[89] M. Schubert and H. Boche, "A unifying theory for uplink and downlink multiuser beamforming", *Proceedings of International Zurich Seminar on Broadband Communications, Access, Transmission, Networking,* pp. 27.1-27.6, 2002.

[90] H. Boche and M. Schubert, "SIR balancing for multiuser downlink beamforming - A convergence analysis", *Proceedings of IEEE International Conference on Communications,* vol.2, pp.841-845, 2002.

[91] C. Farsakh and J.A. Nossek, "Spatial covariance based downlink beamforming in an SDMA mobile radio system", *IEEE Transactions on Communications,* vol.46, no.11, pp.1497-1506, Nov. 1998.

[92] K.-W. Wong, R.S.-K. Cheng, K. Ben Letaief and R.D. Murch, "Adaptive antennas at the mobile and base stations in an OFDM/TDMA system", *IEEE Transactions on Communications,* vol.49, no.1, pp.195-206, Jan. 2001.

[93] P. Vandenameele, L. Van Der Perre, M.G.E. Engels, B. Gyselinckx and H.J. De Man, "A combined OFDM/SDMA approach", *IEEE Journal on Selected Areas in Communications,* vol.18, no.11, pp.2312-2321, Nov. 2000.

[94] S.A. Jafer and A. Goldsmith, "Beamforming capacity and SNR maximization for multiple antenna systems", *Proceedings 53rd Vehicular Technology Conference,* vol.1, pp.43-47, Spring 2001.

[95] S.A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback", *Submitted to IEEE Transactions on Wireless Communications.*

[96] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality", *submitted to IEEE Transactions on Information Theory,* July 2002.

[97] F. Shad, T.D. Todd, V. Kezys and J. Litva, "Dynamic slot allocation (DSA) in indoor SDMA/TDMA using a smart antenna basestation", *IEEE/ACM Trans. Networking,* vol.9, no.1, Feb. 2001.

[98] R. Kuehner, T.D. Todd, F. Shad and V. Kezys, "Forward link capacity in smart antenna base stations with dynamic slot allocation", *IEEE Transactions on Vehicular Technology,* vol.50, no.4, pp.1024-1038, July 2001.

[99] H. Yin and H. Liu, "An SDMA protocol for wireless multimedia networks", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,* vol.5, pp. 2613 -2616, 2000.

[100] G.H. Golub and C.F. Van Loan, *Matrix computations,* Johns Hopkins University Press, Baltimore, 1996.

[101] U.C. Kozat, I. Koutsopoulos and L. Tassiulas, "Dynamic code assignment and spreading gain adaptation in synchronous CDMA wireless networks", *Proceedings of International Symposium on Spread Spectrum Techniques and Applications,* 2002.