Dynamic Resource Allocation in Packet-Oriented Multi-Cell OFDMA Systems

vorgelegt von
Diplom-Ingenieur
Mathias Bohge
aus Berlin

von der Fakultät IV – Elektrotechnik und Informatik –
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
– Dr.-Ing. –

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr. Tansu Alpcan
Berichter: Prof. Dr.-Ing. Adam Wolisz
Berichter: Prof. Dr.-Ing. Ralf Lehnert
Tag der wissenschaftlichen Aussprache: 02. Dezember 2010

Berlin 2010
D 83

Acknowledgments

Throughout my dissertation years, a lot of people supported me professionally, as well as personally. Without their help, it would have been much harder to finish this thesis – if it would have been finished at all.

First of all, I'd like to thank Prof. Adam Wolisz for his scientific mentoring and financial certainty through all the years. Of course, there is much more to thank for – especially lots of funny situations and mails – but I'd rather do that in person. I also thank Prof. Ralf Lehnert for his support and speed in reviewing this thesis. And Prof. Hommel and Prof. Armin Zimmermann, as well as all the colleagues from the DFG Graduiertenkolleg "GK Magsi" for many great ideas and good times. Definitely, I have to greatfully thank my friend Prof. James Gross for many years of fruitful cooperation, some really nice publications and many glad hours. Another "thank you" goes to my project partners at Ericsson Aachen/Stockholm, and particularly to Dr. Michael Meier. A very special "thank you" goes out to all the TKN folks, especially to our sectretaries Petra, Heike and Sonja, to our staff Sven, Mali, George and Peter, and of course to all my dear PhD colleagues. Particularly, I want thank Daniel (Willi) Willkomm and Sven (Sviety) Wiethölter for hundreds of year of studying, living, partying, and PhDing together. Let's continue doing number two and three. Maybe together with Marc (Marci) Löbbers (thank you for even more parties and showing that going into industry can be fun, too). An additional thank you for proof-reading this thesis goes out to Jan (Saw-Hiw Haver) Hauer, and Thomas (no nick name) Menzel. And to Niels for caring about the coffee - without drinking coffee (and eating meat). And to Onur for sitting on the other side of the hallway and "happa". Moreover, I'd like to thank Dr. Hans-Florian Geerdes, Dr. Andreas Eisenblätter, and Dr. Uli Türke for the cooperation on computing adaptive power profiles. And the Eurecom folks, Prof. Raymond Knopp and Dr. Florian Kaltenberger for the OFDMA traces. Another big "thank you" goes out to my students who helped me in doing my research. Particularly, I'd like to thank Martin Renwanz, Farshad Naghibi, Lana Abdelkarim, and Michael Döring for the valuable inputs to this thesis.

On the personal side, I want to thank my dear friends Dr.-Ing. Klaus Bürgel and his lovely wife Ingrid, Björnie, Paul, Janina, Chris, Kiwi, Chryz, Aga, Klaus, Erina, Uwe, Bijan, Karsten, Uli, der Meise, dem Wolff, Elfi, Peter, Familie Raskop, Familie Preidt, der Jugend: Almut, Julian, Henni, Kathanrina, Denise, Jona, Ele, Amelie (Cookie), Flip (der kleine Lenkdrachen), Felix, Adrian (Addi), Judith, Lydia, Timo und Patty. Last but not least, I'd like to thank my familiy, Moni and Dipl.-Ing. Mani Bohge, Mimi, Doron, Theo and Eleah Wohlfeld, Moni Neelsen, as well as Olaf and Mia. And, of course, my wife Frauke.

Zusammenfassung

Die Art der Nutzung von Mobiltelefonen hat sich im letzten Jahrzehnt entscheidend verändert. Während vor zehn Jahren Telefongespräche von relativ geringer Bitrate den Hauptanteil des Datenverkehrs in zellulären Mobiltelefonnetzwerken ausgemacht hat, dominieren heute hochbitratige Multimedia-Anwendungen den Netzwerkverkehr. In Zuge dieser Entwicklung und der Integration mit dem Internet, haben moderne zelluläre Systeme eine paketorientierte Architektur erhalten. Außerdem hat der immerwährende Bedarf an höheren Datenraten die Entwicklung neuartiger Übertragungs- und Medienzugriffsverfahren erfordert. Die derzeit am weitest verbreitete kombinierte Übertragungs- und Medienzugriffstechnik ist Orthogonal Frequency Division Multple Access (OFMDA). OFDMA-basierte Systeme spalten die vorhandene Systembandbreite in schmalbandige parallele Subkanäle auf. Ein hochratiger Bitstrom wird am Sender in mehrere Ströme von geringer Bitrate gesplittet und über die Subkanäle verteilt gesendet. Dabei können unterschiedliche Subkanäle gleichzeitig unterschiedlichen Nutzern zugeteilt werden. Verschiedene Forschungsarbeiten aus den letzten Jahren zeigen die Möglichkeit einer Verbesserung der Systemeffizienz auf, wenn die Übertragungsressourcen, also die Übertragungsleistung und die Subkanalzuordnung, dynamisch zugeteilt werden. Jedoch haben entsprechende Mechanismen eine erhöhte Systemkomplexität und einen zusätzlichen Signalisierungsaufwand zur Folge. Außerdem wurden sie ursprünglich für Systeme entwickelt, die sich von paketorientierten zellulären Systemen in drei Punkten grundsätzlich unterscheiden: in der Übermittlung der Daten in Paketform, in Art und Umfang des notwendigen Signalisierungsaufwandes und bezüglich der besonderen Interferenz-Situation in zellulären Systemen. Die vorliegende Arbeit widmet sich den grundsätzlichen Fragestellungen, die auftreten, wenn entsprechende dynamische Mechanismen im Downlink von OFDMA-basierten zellulären Systemen zum Einsatz kommen sollen. Insbesondere werden im Rahmen dieser Arbeit mathematische Optimierungsmodelle entwickelt, durch die das Potential des Einsatzes verschiedener dynamischer Mechanismen in solchen Systemen bestimmt werden kann. Es wird gezeigt, dass durch den Einsatz verschiedener dynamischer Techniken am Daten- und Signalisierungskanal der Durchsatz, die Fairness zwischen Teilnehmern, sowie die Übertragungszuverlässigkeit entscheidend verbessert werden können; und dass durch eine zellübergreifend optimierte Ressourcenvergabe auf sonst übliche Interferenzunterdrückungsmechanismen verzichtet werden kann. Da es mit der heute üblichen Hardware in entsprechenden Systemen nicht möglich ist, die Ressourcenverteilungsprobleme optimal zu lösen, werden anschließend an die Gewinnpotentialbestimmung der einzelnen Ansätze heuristische Algorithmen entwickelt. Diese Heuristiken sind in der Lage, die Optimierungsprobleme sub-optimal, aber effizient zu lösen. Im einzelnen werden ein kom-

binierter Subkanal-Zuweisungs-/Packet-Auswahl-Algorithmus, ein dynamischer Leistungszuweisungs- und ein dynamischer Kodierungsalgorithmus für den Signalisierungskanal, sowie ein zellübergreifender Leistungs-Koordinierungsalgorithmus vorgestellt. Zusammenfassend zeigt die vorliegende Arbeit die potentiellen Gewinne, die durch den Einsatz dynamischer Mechanismen in paketorientierten OFDMA-basierten zellulären System erzielbar sind auf, und stellt heuristische Algorithmen zur effizienten Implementierung der dynamischen Ansätze zur Verfügung.

Abstract

Over the last decade, the scope of cellular networks has been shifted from low bit rate voice delivery to high bit rate multimedia services. Consequently, cellular systems have evolved from the circuit switched to the packet oriented system architecture, and an integration with the Internet has taken place. The perpetual need for higher data rates in such systems has caused the development of efficient transmission and medium access techniques. The currently most popular one among them is orthogonal frequency division multiple access (OFDMA). In general, OFDMA systems split the overall system bandwidth into a number of low data rate sub-channels. Each high data rate stream arriving at the transmitter is split into several low data rate streams that are transmitted in parallel over a sub-set of the available sub-channels, where distinct subsets can be assigned to different active terminals. Lately, it has been shown that the system's transmission capacity and quality can be significantly improved, if the sub-channels and the available transmission power are dynamically allocated among the terminals according to their instantaneously experienced channel quality. These performance gains, however, come at the cost of significant signaling overhead and increased system complexity. Moreover, the dynamic mechanisms have originally been developed for systems that differ from packet-oriented cellular systems in three major points: the packet nature of the traffic, the quality and quantity of the required control signaling data, and the particular interference situation of cellular systems. This thesis faces the challenges that arise from these three differences when deploying the dynamic mechanisms in the downlink of packet-oriented cellular OFDMA based systems. In particular, mathematical optimization problems are developed in order to determine the potential of the individual dynamic approaches. It is shown that by dynamically assigning the transmission resources, the system's throughput performance, its inter-terminal fairness, as well as the transmission reliability can be significantly improved, if data and control channel resources are assigned dynamically. Furthermore, a high potential to increase the system performance is observed, if neighboring cells jointly optimize their resource utilization, instead of applying legacy interference mitigation techniques. In order to cope with the real world equipment's limited computational power, which makes optimally solving the mathematical optimization problems in real-time impossible, individual heuristic algorithms are developed to sub-optimally solve them: a combined data channel packet scheduling/resource allocation algorithm, a control channel dynamic power allocation algorithm, a control channel dynamic coding algorithm, and an inter-cell power coordination algorithm. In summary, this thesis demonstrates the potential of using dynamic mechanisms in packet-oriented cellular OFDMA based networks and presents means to efficiently deploy them in real world systems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

OVER the last ten years the scope of cellular systems has changed significantly. While in the former days cellular phones have mainly been used for telephone calls and short messages, nowadays the devices offer any kind of multimedia services, such as video telephony, audio and video streaming, or arbitrary application (app) download and execution. In order to better fit the needs of these new multimedia network services, cellular network standards have been evolving from the classical circuit switched network type (as it is the case with the Global System for Mobile Communications (GSM) standard) to its packet oriented counterpart (*e. g.,* the general packet radio service (GPRS) standard). Today, the packet switched architecture is a key point in the design of upcoming cellular system standards.

As the growing number of multimedia services goes along with a growing number of customers that use these services, there is a perpetual need for ever higher data rates in cellular systems. The orthogonal frequency division multiplexing (OFDM) transmission technology is the per se transmission standard for modern digital communication systems as it allows high data rates at comparatively low transmitter and receiver complexity. The general idea of OFDM is to split the overall system bandwidth into a number of low data rate sub-channels. Each high data rate stream arriving at the transmitter is split into several low data rate streams that are transmitted in parallel over the sub-channels. At the receiver, the high data rate stream is then reassembled. Due to its stability against various kinds of interference, OFDM is especially appreciated as technical basis for wireless communication systems. An OFDM based system is referred to as orthogonal frequency division multiple access (OFDMA) system, if it applies the medium access technique frequency division multiple access (FDMA) for multi-user handling. In an OFDMA system, the data of a single user is transmitted on a subset of the available sub-channels, while at the same time, the data of other users is transmitted on other (distinct) sub-channel subsets. By adapting the transmission parameters, such as the sub-channel transmission power, the sub-channel modulation type and coding rate, the user/sub-channel assignments, and/or the number of sub-channels per subset to the individual user needs and instantaneous sub-channel quality conditions, the performance in terms of throughput, robustness and/or power efficiency of OFDMA transmissions can be greatly improved. These performance gains, however, come at the cost of an increased computational complexity of the transmitter and the receiver, and significant signaling overhead that needs to be exchanged on a separate *control channel*. Systems that

apply these mechanisms are commonly referred to as *dynamic OFDMA* systems. Even though dynamic OFDMA based cellular standards have been around for several years, there are still open system design questions, and there is no commercially working dynamic OFDMA based cellular system to date. This thesis is concerned with answering these open questions in cellular dynamic OFDMA system design. In particular, it faces the challenges that arise when adopting well known dynamic mechanisms in the downlink of cellular OFDMA based systems.

Challenges The major differences between cellular systems of consideration and the systems for which dynamic OFDM(A) mechanisms have originally been developed (namely streaming systems such as digital audio broadcasting (DAB) and digital video broadcasting (DVB), or connection oriented point-to-point systems such as Digital Subscriber Line (DSL) systems) lie in (1) the packet nature of the traffic, (2) the quality and quantity of the required control signaling data, and (3) the particular interference situation of cellular systems. The research questions discussed in the following chapters arise from these three differences:

1. *Are special packet-aware resource allocation techniques necessary in packet-oriented OFDMA systems?* The data destined for users of modern cellular systems is delivered in packets, which might occasionally arrive back-to-back, but most of the time are delivered in bursts of packets resulting in periods of very high data rates on the one hand, and periods of zero data transmission on the other hand. Obviously, this yields to situations, where in a certain cell there is no data to be transmitted to the currently best situated user. It is thus expected that the potential of throughput gains due to deploying dynamic mechanisms is smaller in packet-oriented networks than in streaming systems, where full buffers can be assumed at all times. Moreover, large amounts of transmission resources might be wasted, if a subset of sub-channels is assigned to a terminal, for which only one small packet is enqueued. So far, mathematical models to solve the resource assignment optimization problem do not consider packet buffering, *i. e.,* they do not match the resources assigned to a single terminal to the data to be transmitted to it. There are heuristics that influence standard optimization problems by excluding terminals with empty buffers and adding weights according to the buffer fill levels. It is however, not clear, how close the performance of these approaches gets to optimality, and whether specialized packet-aware assignment strategies would be much more complex and perform significantly better.

2. *Does reliable control data transmission require a disproportional split of the available transmission resources in favor of the control channel redundancy?* The control signaling effort in a dynamic OFDMA based cellular system consumes a significant part of the available system resources. The signaling data contains information about which terminal is assigned which sub-channels and/or which modulation and coding types are used. It is mandatory that each connected terminal is able to decode the control data, even if it is exposed to strong interference from the surrounding cells. Hence, control channel reliability is a key issue in cellular dynamic OFDMA system design. Strong coding on the control channel might be necessary, which in turn reduces the resources available for user data transmissions. In existing mathematical models, this trade-off is not considered.

3. *Can any performance gain be expected, if neighboring cells jointly optimize their usage of transmission resources?* Interference mitigation in cellular systems is mandatory, as a high number of users might transmit in close vicinity to each other, even though they are connected to different cells. A standard interference coordination technique is to divide the overall system bandwidth in frequency bands and to use distinct frequency bands in neighboring cells following certain frequency reuse patterns. The reuse pattern scheme, however, is static and mostly adjusted before the system is put into operation, and, thus, known to be inefficient. An alternative approach is to allow the usage of the complete spectrum in each cell and to integrate the interference coordination and the resource allocation task at each base station. There is lack of mathematical models to evaluate the potential gains of such an approach.

To answer the questions regarding the usage of dynamic mechanisms in packet-oriented cellular OFDMA systems above, new optimization problems need to be formulated and solved. Obtaining the optimal solutions, however, is in most cases a computationally complex task, such that it is impossible to get them in real-time. Consequently, a new challenge arises each time a certain dynamic mechanism shows the potential to improve the system performance:

4. *A low complexity mechanism* needs to be developed, in order to achieve the potential gains under real world constraints, such as limited computational power, quick solution obsolescence, and restricted power supply.

Only if the performance gains of these low complexity mechanisms are shown to be significant enough to justify the additional computational complexity, the according dynamic OFDMA technique is considered a valid candidate for packet-oriented cellular OFDMA systems.

Contributions Within this thesis, mathematical optimization problems are developed in order to answer the research questions posed above. The new optimization problems are mainly obtained by refining well known dynamic OFDMA mathematical optimization approaches (the control channel related problem of Chapter 5 is an exception). All formulated optimization problems are solved in numerous different instances in order to obtain meaningful results regarding the potential of the related dynamic OFDMA techniques. As all of them will be shown to have a strong potential to improve packet-oriented cellular system performance, individual heuristic algorithms are derived from the optimization problem formulations. In particular, four low complexity algorithms to implement dynamic mechanisms in real world systems are presented:

- a combined packet scheduling / resource allocation algorithm,
- a control channel dynamic power allocation algorithm,
- a control channel dynamic coding algorithm, and
- an inter-cell power coordination algorithm.

Summing up, this thesis demonstrates the potential of using dynamic mechanisms in packet-oriented cellular OFDMA based networks and presents means to efficiently deploy them in real

world systems. While in principle all presented mechanisms are applicable to the up- and down-link direction of such systems, the up-link direction faces additional significant challenges that are not in the scope of this thesis, such as synchronization issues, CSI signaling, and problems related to the peak-to-average ratio of the up-link transmission signal. By facing the above challenges and providing solutions to the connected problems, however, all significant cellular down-link related dynamic OFDM research issues are considered in this thesis.

Structure The general structure of this thesis is as follows: First, some background information on the considered technologies is provided. Then, the thesis challenges are further discussed, and a detailed system model is presented. The main part of the thesis is devoted to answering the research questions posed above, where each question is covered in an individual chapter. In each of these chapters at least one according low complexity algorithm is presented. The thesis ends with a conclusions and outlook chapter that summarizes the results and provides some links to research follow ups. The detailed thesis structure is as follows:

Chapter 2 – Background: An introduction to digital data transmission using orthogonal frequency division multiplexing is provided. Moreover, it is explained, how diversity can be exploited by dynamically distributing the resources among the terminals of an OFDMA system.

Chapter 3 – Scope of the thesis: The challenges faced in this thesis are further discussed, the system model is introduced, simulation scenarios are defined and some information about the hardware in use is given.

Chapter 4 – Packet Traffic Aware Resource Allocation: An answer to the question '*Are special packet-aware resource allocation techniques necessary in packet-oriented OFDMA systems?*' is given by formulating and solving an according optimization problem. A heuristic combined packet and resource scheduling algorithm is presented.

Chapter 5 – Optimized Control/Data Channel Capacity Split: An answer to the question '*Does reliable control data transmission require a disproportional split of the available transmission resources in favor of the control channel redundancy?*' is provided. Two heuristic control channel adaptation algorithms (dynamic coding and dynamic power allocation) are presented.

Chapter 6 – Interference Mitigation for Frequency Reuse 1: An answer to the question '*Can any performance gain be expected, if neighboring cells jointly optimize their usage of transmission resources?*' is found. An according inter-cell power coordination algorithm is presented.

Chapter 7 – Conclusions and Outlook: The overall work is concluded. The answers to the research questions are summarized. In addition, future work opportunities are pointed out.

A list of the author's publications, the acronyms used in this thesis, the bibliography, and an index are provided as thesis appendices.

Chapter 2

Background

THE purpose of this chapter is to review some basics on digital transmission schemes, wireless channel characteristics, orthogonal frequency division multiplexing (OFDM), as well as dynamic resource allocation in OFDM systems in order to ease the understanding of this thesis. The following discussion is mostly intended for readers with no or very little background on transmission in wireless networks and aims in providing an intuitive but still sufficiently detailed introduction to the most important concepts.

2.1 Digital Communication Basics

The goal of a telecommunication system is to transmit a message that was produced by a source over a channel to a sink. The message must not be modified, since messages are defined as information that must not be altered.

A *signal* is the physical representation of a message, *e. g.,* by dynamically changing acoustic pressure, brightness or voltage characteristics. If the representation is not an electrical signal, converters might be necessary to transmit the message on a given telecommunication system. Examples for converters on the sender side include microphones to capture voice, and video-cameras to capture images; and on the receiver side loudspeakers, and monitors. An *analog signal* represents a message that originates from a continuous source, *e. g.,* human speech. A *digital signal* consists of sequence of symbols out of a finite set. Using an analog-to-digital converter (ADC), it is possible to substitute an analog signal by a digital signal. Analog-to-digital converting (*digitizing*) implies the task of sampling the analog signal (mostly at equidistant points in time), and quantizing the sample values to predefined values that can be represented by a set of symbols. Throughout this thesis, digital signals are considered. It is assumed that for each message that has to be transferred there is an appropriate digital representation – in the following referred to as *data stream* $d(kT)$.

Within a digital communication system, digital representations of messages are sent from a source to a sink. For analyzing purposes, the connecting channel can be subdivided into *transmission instances* (see Figure 2.1): *digital channel*, *baseband channel* and *physical channel*. Note that the digital channel consists of the physical channel and a number of sub-systems:

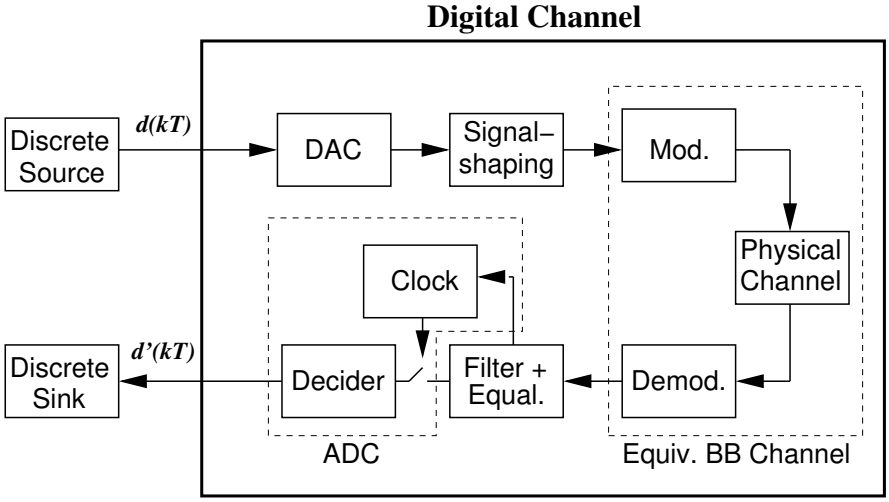
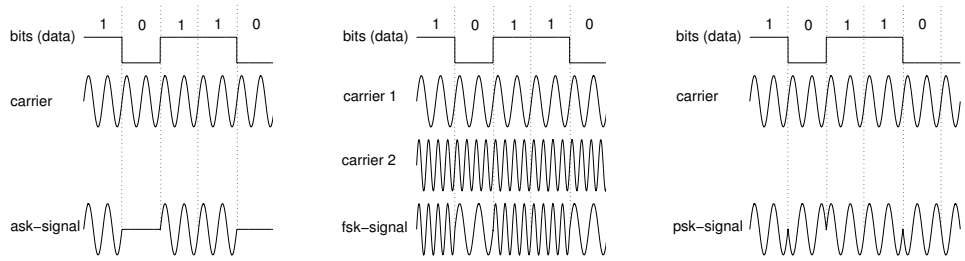


Figure 2.1: a digital communication system including the digital channel

The *digital-to-analog converter (DAC)* converts the data stream into physical pulses that are then shaped to form the actual sending signal. By the use of *signal shaping*, the system is able to adapt to the frequency response of the physical channel that (if not assumed to be ideal) acts as a filter. In a baseband transmission system, the shaped signal is directly sent over the physical channel to the receiver that mainly consists of an input-filter with equalizer and an ADC. The input filter suppresses the impact of linear distortion and noise and regains the clock pulses for the ADC that recovers the original data stream from the received signal. If the transmission was successful, the recovered data stream $d'(kT)$ equals the original data stream $d(kT)$. However, in modern communication systems the baseband transmission technique is rarely used. There are several reasons to deploy modulation when sending a signal over a given channel, *e. g.,*:

- most channels have preferable transmission characteristics in a higher frequency band than in the baseband (even though the attenuation is stronger, it is more uniform over a certain bandwidth),
- the transmission does not support direct current (DC) portions of the baseband signal due to system components as transformers, as well as
- frequency multiplexing possibilities.

In a communication system that features modulation, a *modulator (MOD)* at the sender side is responsible to move the signal spectrum into a higher frequency range. On the other side of the channel, a *demodulator (DEMOD)* down-converts the signal-spectrum to baseband. Hence, all filtering and digital/analog converting can be done as in a baseband transmission system. Therefore, the combination of modulator, physical channel, and demodulator is referred to as *equivalent baseband channel*.



(a) amplitude shift keying (ask) (b) frequency shift keying (fsk) (c) phase shift keying (psk)

Figure 2.2: a comparison of binary modulation types

Signal modulation

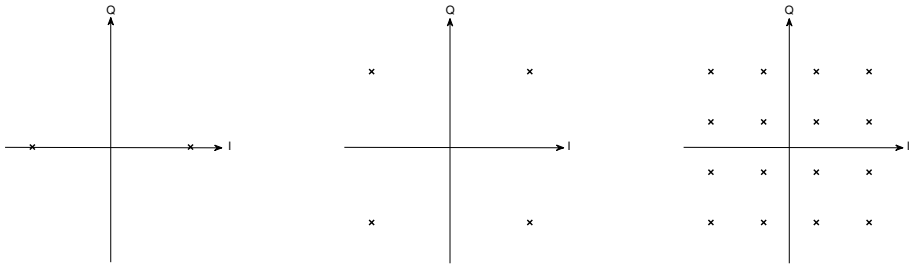
In a multilevel modulation system, each symbol represents m bits. Thus, there is a need for at least 2^m different symbols in the system. The special case with $m = 2$ bits is usually referred to as *binary* modulation. In a binary modulation system, each symbol represents exactly one bit of the data stream. Hence, the bit rate equals the symbol rate. Note that two different symbols suffice to create a representing sequence for each possible data stream, which renders binary modulation the less complex and most robust kind of modulation.

Binary modulation types

There are three basic types of binary modulation that can be deployed to a data stream, in order to convert it into a higher frequent transmission signal [1]: amplitude shift keying (ASK), frequency shift keying (FSK), and phase-shift keying (PSK) (shown in Figure 2.2). In each case, two signal levels (symbols) are required that correspond to binary 1s and 0s respectively. Generally, the modulation operation consists in modifying a high frequent single-frequency audio-tone called the *carrier signal*

$$c(t) = A_{\mathrm{c}} \cdot \cos(\omega_{\mathrm{c}} t + \varphi_{\mathrm{c}}) \tag{2.1}$$

w.r.t. the given data stream. The parameter A_{c} is the carrier amplitude, ω_{c} is the carrier frequency in radians per second, and φ_{c} is the carrier phase shift. If ASK is used, the carrier amplitude A_{c} is a function of the data stream $d(kT)$. ASK is easy to implement, yet involves high bit error probability (BEP) values. This simplest form of ASK is referred to as on off keying (OOK). If FSK is deployed, ω_{c} is a function of the data stream, and A_{c} is a constant. Considering binary FSK, this essentially means that the system switches between two carrier signals of carrier frequencies $\omega_{\mathrm{c},1}$ and $\omega_{\mathrm{c},2}$. With PSK, the frequency and the amplitude of the carrier signal are kept constant while the carrier phase shift φ_{c} is a function of the data stream, as shown in Figure 2.2c. To point out that a binary modulation type is considered, the binary modulation PSK is mostly referred to as binary phase shift keying (BPSK). BPSK delivers the smallest possible BEP value of the binary modulation types ($p_{bit,ASK} \geq p_{bit,FSK} \geq p_{bit,BPSK}$ for a given channel), yet it is the most complicated type to be implemented.



(a) BPSK (b) QPSK (c) 16QAM

Figure 2.3: phase diagram examples

Multilevel modulations types

The idea of shifting the carrier phase φ_c w.r.t. to the data-stream can be extended to more than just two different phase shifts. Similarly, there are multilevel modulation alternatives of the other two basic binary modulation types, multilevel amplitude shift keying (M-ASK) and multilevel frequency shift keying (M-FSK). In an M-ASK system, all M symbols have the same phase, yet differ in the amplitude. Consequently, all symbol-representing points of an M-ASK phase diagram are located on the in-phase axis, most likely with equidistant spacing. Each M-FSK symbol utilizes its very own carrier. Thus, an M-FSK system features M different carrier frequencies, and it can't be illustrated by a phase diagram that corresponds to a single carrier frequency. However, since M-ASK has extremely high energy requirements and the deployment of M-FSK goes along with high bandwidth claims, both schemes are rarely used [1].

If a PSK system can distinguish between four different carrier phase shifts, it is called a quadrature phase shift keying (QPSK) system. Figure 2.3b shows the matching QPSK *phase diagram*. A phase diagram illustrates the set of permitted symbols of a given modulation type in the complex plane. The position of a complex symbol $d = a + jb$ within the complex plane is determined by its in-phase component a (its value on the in-phase axis I) and its quadrature component b (its value on the quadrature axis Q). In Figure 2.3b it can be seen that the four symbols of a QPSK system have the same amplitude while the phase difference between the symbols is $90°$. For higher multi-level constellations, a combination of ASK and PSK is mostly used. This type of modulation is known as quadrature amplitude modulation (QAM). An example QAM phase diagram (also referred to as *16-point constellation*) is shown in Figure 2.3c. The illustrated QAM system features 16 different symbols. Hence, the number of bits per symbol is $log_2(16) = 4$. The impact of the rather high symbol error probability (SEP) of an M-QAM can be reduced by means of coding, *e. g., Gray codes*. Gray codes map the M QAM symbols to the M possible sequences of $log_2(M) = m$ bits, such that symbol errors that are due to a mix-up of neighboring points in the phase diagram (the most appearing error at the receiver) lead to a minimum of bit errors [2].

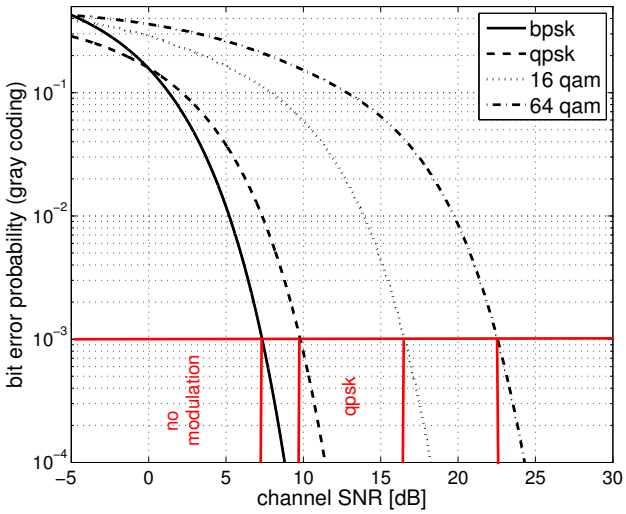


Figure 2.4: adaptive modulation w.r.t. the instant channel SNR

Adaptive modulation

The symbol error probability of a received signal directly depends on the signal-to-noise ratio (SNR) (*cf.* next section) that the receiver experiences on the given channel. The SNR is determined by (1) the transmit power, (2) the modulation, (3) noise and (4) the channel attenuation. While the first two can be arbitrarily chosen, the latter two change according to a stochastic process. In order to stabilize the SEP, the transmission power and modulation type can be chosen, such that the channel attenuation's impact is equalized.

Systems that adapt the modulation type according to the instant SNR are referred to as adaptive modulation systems. Figure 2.4 shows the decision SNR thresholds of an adaptive system that switches between BPSK, QPSK, 16-QAM, and 64-QAM in order to stick to an error probability of 0.001: ∼7dB to deploy BPSK, ∼9.5dB to deploy QPSK, ∼16.5dB to deploy 16-QAM, and ∼22.5dB to deploy 64-QAM. If the instant SNR value is less than 7dB no modulation is deployed, hence energy is saved. The frequency with which a wireless system switches the modulation scheme mainly depends on the wireless channels characteristics, which are shortly reviewed in the following section.

Multi-carrier modulation

A multi carrier modulation (MCM) (also referred to as *multi-channel modulation* or *multi-channel communication*) system divides the available bandwidth into a number of sub-channels, on which data can be delivered in parallel. Note that if a system of bandwidth B with a maximum data rate of r is divided into S sub-channels, the maximum data rate per sub-channel is r/S – *i. e.,* the maximum system rate is not increased. However, as the channel responses per sub-channel are usually better than the channel response of a regular non-MCM system (also

referred single carrier modulation (SCM) system) featuring the same bandwidth, more efficient transmission is possible [2]. This gain comes, of course, at the cost of higher transmission complexity.

2.2 Wireless Channel Characteristics

The are several books on how to analyze and model the wireless channel, among which I found "Mobile Radio Communications" by R. Steele and L. Hanzo [3] to be the most valuable. In the following paragraphs I quickly review those parts that are most relevant to this thesis. For more detailed information the interested reader is referred to the original book.

The most important difference between a wireless channel and a solid connection (guided medium) is the wireless channel's unreliable behavior. Particularly, its attenuation may vary within small frequency ranges and short time spans. Moreover, due to scatterers that are usually present in the communication environment, the transmitted signal follows *multiple different paths*, *i. e.,* multiple copies of the transmitted signal coexist in the wireless channel, each following a path i of individual length l_i. Hence, for each signal copy there is a different delay value $\tau_i = \frac{l_i}{c}$ (c is the speed of light) that describes its traveling time to the receiver. The range between the delay of the slowest and fastest traveling signal copy is called the *delay spread* $\Delta\sigma$. In addition, each signal copy experiences a different attenuation due to different obstacles on its path and its individual path length.

The lack of stability and multi-path propagation turn *reliable wireless communication* into a difficult task. In order to successfully implement wireless communication systems, thus, exact analysis of the signal influencing effects present on the wireless channel is necessary. In order to facilitate this analysis, the equivalent baseband channel introduced in Section 2.1 can be further subdivided into the *modulation channel*, the *radio channel*, and the *propagation channel* (see Figure 2.5) [4]. Each instance experiences different *signal-influencing effects*:

The propagation channel is influenced by the linear and reciprocal *multiplicative effects* path loss, shadowing and fading, which result in the *attenuation* h of the transmitted signal. The radio channel consists of the propagation channel and the transmit and receive antennas. It is influenced by the non-reciprocal *additive* effects noise and interference (electromagnetic waves stemming from other communicating devices). Moreover, the transmitted signal is scaled by an antenna-dependent factor. The modulation channel consists of the radio channel plus all radio frequency (RF) system components. While amplifying the received signal, the devices damage the signal by introducing additional noise coming from electric circuits. In contrast to the linear and reciprocal multiplicative effects above, the non-reciprocal additive effects can't be compensated. In the following, a short discussion on how to model the impact of the different signal-influencing effects introduced above is provided.

2.2.1 Multiplicative effects

For analyzing and modeling purposes it is useful to further distinguish between different multiplicative (attenuating) effects. The total attenuation h is a product of three different effects that

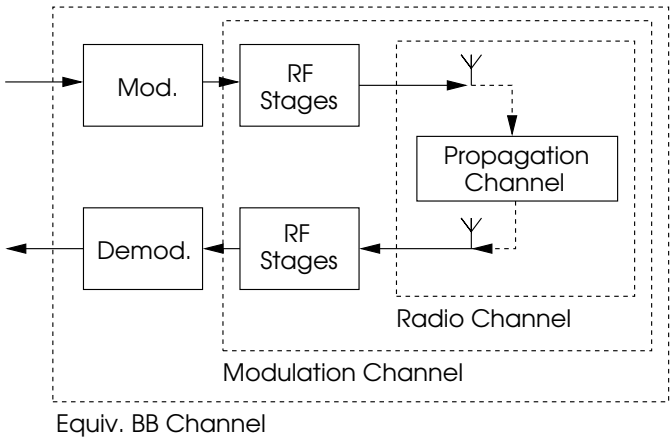


Figure 2.5: instances of a wireless channel

have an impact on the transmitted signal - path loss, shadowing, and fading:

$$h = h_{PL} \cdot h_{SH} \cdot h_{FA} \tag{2.2}$$

Path loss The path loss effect h_{PL} is an abstraction of the propagation phenomena that depend on deterministic environmental factors, such as the transmission medium, the frequency and the distance between transmitter and receiver. In most scenarios none of these factors varies significantly in short time spans, therefore its importance lies in larger time scales representing seconds or minutes. The term *free space path loss* describes the attenuation that effects a signal propagating in free space over a distance d between two antennas assuming line of sight (LOS) (*i. e.,* there are no obstacles or reflecting surfaces on the direct connection line between the antennas). It can be exactly calculated using the Maxwell equations [5]. The result shows that the received power is inversely proportional to the square of the distance - which is not very surprising since in free space the radiated energy propagates equally in every direction and the wave can be seen as sphere of increasing radius. As the surface increases with the square of the radius while the total energy over the sphere must stay the same, the energy per unit surface decreases at the inverse rate. However, the received power is also inversely proportional to the square of the frequency.

In the context of aeronautical communication, the assumptions for free space path loss scenario described above are valid. Considering cellular communication systems, however, these assumptions are unrealistic since in the latter communication usually takes place close to the reflecting earth surface. The so called *two ray model* is a simple model that adds a reflecting surface to the LOS scenario. It uses the height of the transmitter and receiver antennas to calculate a second single ray that is reflected by the surface on its way from the transmitter to the receiver where it sums up with the LOS ray to a resulting ray. It was shown that the model fits quite well the actual path loss in LOS environments featuring communication close to a reflecting surface like open fields, open water or even highways. However, the assumptions of the two ray model are not realistic in urban, suburban, indoor or underwater environments that don't necessarily

feature LOS and usually consist of many reflectors and scatterers [6].

For these reasons, empirical and semi-empirical models have been developed to calculate the path loss between a transmitter and a receiver at a specific distance d in specific environments for different frequencies [7]. The models are usually of the form

$$h_{PL} = \frac{p_{rx}}{p_{tx}} = \frac{K}{d^{\alpha}} \,, \tag{2.3}$$

where the reference loss K and the path loss exponent α are fitted to measurement results in the environment. The empirical models require extensive measurements, whereas the semi-empirical use a mix of empirical and theoretical data. For every new area, calibration measurements are required to calculate correction factors for the general models.

Shadowing The shadowing effect h_{SH} (sometimes also referred to as *slow fading* effect) is an abstraction of the sum of several propagation phenomena such as *reflection*, *diffraction*, *refraction*, *scattering*, and *absorption*. Measurements in a static environment show that the received signal power is not deterministic, but varies due to the objects in and around the signal path. These Gaussian distributed stochastic, location dependent differences in the measured signal power in relation to the theoretical value calculated by the path loss formulas are reflected by the shadowing effect h_{SH}. Since averaging many received signal power values at the same distance yields the exact value given by the path loss, the shadowing effect h_{SH} can be modeled as a Gaussian distributed random variable with zero mean given by

$$p(h_{SH}) = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{sh}}^2}}\, e^{-\frac{h_{SH}^2}{2\sigma_{\mathrm{sh}}^2}} \,, \tag{2.4}$$

where σ_{sh}^2 is the shadowing variation expressed in dB.

Fading The fading effect h_{FA} (sometimes also referred to as *fast fading* effect) is an abstraction of the propagation phenomena that are due to the multipath signal propagation explained above. The reception of multiple copies of the transmitted signal, each having followed a different path, results in an interference pattern at the receiver, where at some points the signals interfere constructively and at others destructively.

If no element in the environment is moving, the channel is said to be *time invariant*, *i. e.,* at all times fading has the same frequency varying impact $\bar{G}(f)$ on the signal. In contrast, if any kind of movement is encountered, the channel is *time variant*. In a case where only the receiver is moving (*e. g.,* in a cellular scenario, where the base station is never moving), the path lengths $l_i(v,t)$ change as a function of the receiver's speed v and the time t. If the scatterers are moving, too, one has to consider the relative receiver-scatterer speed for each fading path. The influence that a moving receiver exerts on the signal copy traveling on path i is referred to as the *Doppler effect*: it is the change in the observed carrier frequency f_{c} of a signal copy, caused by a constant change in the effective path length l_i between the transmitter and the receiver and the changing angel of the impinging ray. The *Doppler shift* is the difference between the original

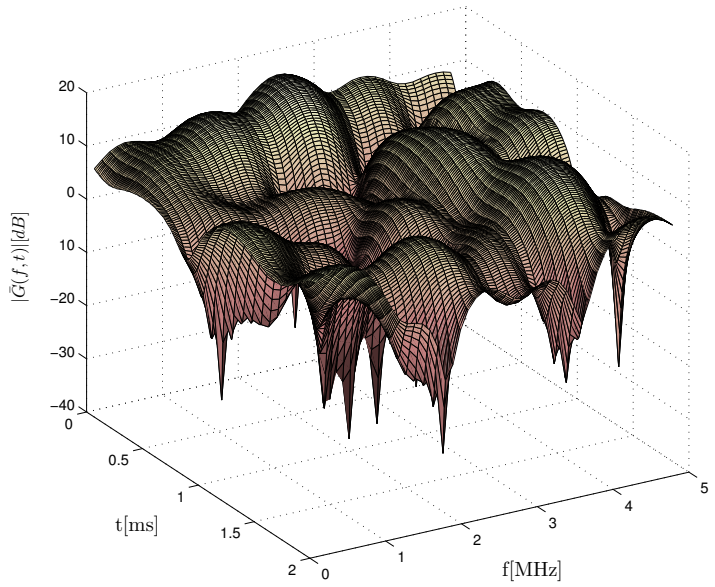


Figure 2.6: fading impact on the received signal power in frequency and time

value of f_c and the experienced frequency at the receiver. The range between the minimal and maximal Doppler Shift in the environment is called the *Doppler spread*. As a result, the receiver experiences a varying signal strength along with a continously changing delay spread through the changing interference pattern. Therefore the received signal fades – sometimes very quickly. In a time variant channel, fading is responsible for the most rapid and intensive variations in signal strength (amplitude) and phase which are experienced on a small time scale, *i. e.,* they are measured in milli- and microseconds, depending on the relative velocity between the transmitter and the receiver. In this case, it has a time and frequency varying, complex valued impact $\bar{G}(t, f)$ on the signal, where $h_{FA} = |\bar{G}(t, f)|$ as shown in Figure 2.6.

Assuming a great number of uncorrelated paths and no LOS (no direct path dominates indirect paths), fading can be modeled by choosing the amplitudes of the arriving signal copies according to the Rayleigh distribution. If there is a direct path, the amplitudes are Ricean distributed. Moreover, if the angles of the incoming signals are uniformly distributed, their power density spectrum can be assumed to have a Jakes-like shape, as well as an exponential power delay profile [8]. The according fading models are mostly referred to as *Ricean fading* and *Rayleigh fading* models. They output fading values for a given environment that is determined by the maximum velocity of its objects $v_{\max}$, the system's center frequency f_c, the environment propagation parameters, the object locations, and an representative initial delay spread value $\Delta\sigma$.

2.2.2 Additive effects

The two major additive effects that potentially distort a signal are *noise* and *interference*. Both effects are non-reciprocal and, thus, can't be compensated. Consequently, it is a major goal in the design of wireless communication systems to keep their impact as small as possible.

The amount of additive effects in a digital communication system is best shown using an *eye pattern* (the name is chosen because of its resemblance to the human eye) [2]. For example, Figure 2.7 illustrates the on off keying eye patterns that would be shown on an oscilloscope for a received signal without (Figure 2.7a) and with (Figure 2.7b) the distorting impact of noise and interference. For on off keying signals, we can display the received signal $r(t)$ on the vertical input with a horizontal sweep rate set at $1/T$. Note that the transmission quality is the better the more open the eye is.

Noise

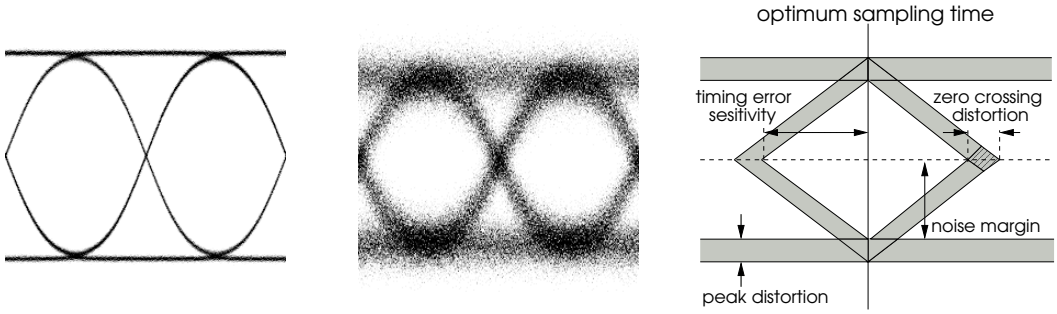
Noise in general refers to *unintended* electromagnetic radiation (*i. e.,* it is not intended to carry information) experienced at the receiver of interest. It may come from arbitrary sources, such as atmospheric disturbances, electronic circuits or man made machinery. Noise is always present.

Thermal Noise Thermal noise is due to the movement of charged particles inside electronic components existent in every receiver system – and is therefore unavoidable [9]. Most of these electronic devices are Gaussian noise sources, which have a spectrum that is constant over almost the complete frequency range. The kind of noise produced by these sources is called white in analogy to white light, which contains all frequencies, too. White noise is uncorrelated. In real electronic devices of limited bandwidth, the noise is filtered with the signal. So the noise is no longer white at the device's output, but low-pass filtered, taking the shape of the filter's transfer function. The values to calculate the effective noise power of a certain device depend on its working temperature and receiving bandwidth. They can be found in its data-sheets.

Man-made Noise Man-made noise is produced by machines while they are operated by human beings. The actual noise is radiated energy which is not supposed to be generated. One example are the electromagnetic waves that appear when an electronic device is turned on or off. The characterization of man-made noise is very complicated since many of the parameters that have to be measured are specific to the environment [5].

Interference

Interference in general refers to *intended* electromagnetic radiation (*i. e.,* it is intended to carry information), which is not destined for the receiver of interest. In other words, signal waves intended for third party systems, or waves that have originally intended to be transmitted on a different frequency channel of a multi carrier modulation system are referred to as *interfering waves*. The interfering waves are correlated with the receiving signal waves because they have the same or nearly the same frequency. In digital communication systems, each physical pulse



(a) ideal (b) distorted (c) effects of ISI on eye opening [2]

Figure 2.7: a comparison of OOK eye pattern with and without the impact of additive effects

of form $g(t)$ representing a symbol in the data sequence $\{d_n\}$

$$v(t) = \sum_{n=0}^{\infty} d_n g(t - nT) \ , \tag{2.5}$$

with a band-limited frequency response $G(f)$ for $|f| > B$ (with B being the system bandwidth) can be viewed as a separate wave that might interfere at the receiver with other waves coming from the same or a different source. The presence of interference depends on the environmental situation, as well as on the system itself.

Inter-symbol interference (ISI) If interfering pulses are part of the same sequence of data coming from the same source, we refer to it as inter-symbol interference (ISI) [2]. Interference among different symbols of a single sequence is due to the fact that in the real world data is always transmitted over bandwidth limited channels having a frequency response $C(f)$ limited *e. g.,* to $|f| \leq B$. Consequently, in the time domain, the received signal implies a convolution of the pulse's and the channel's responses

$$r(t) = \sum_{n=0}^{\infty} d_n \int_{-\infty}^{\infty} g(\tau) c(t - nT - \tau) d\tau + \eta \ , \tag{2.6}$$

where η represents additive white gaussian noise. The convolution term is responsible for the overlapping of consecutive symbols in the time domain. We assume that the receiver filters the received signal with a filter response $x(t)$ and samples this signal at times $t = kT + \tau_o$, where τ_o is the channel's propagation delay. The sequence of received samples $y(k)$ can then be written as

$$y(k) = d_k + \sum_{\substack{n=0 \\ n \neq k}}^{\infty} d_n x_{k-n} + \nu_k \ , \tag{2.7}$$

where the term d_k represents the desired information symbol at the kth sampling instant, ν_k is the sampled AWGN and the sum term represents the ISI. Note that ISI distorts the position of

the zero-crossings and causes a reduction in the signal's eye opening, which renders the system more sensitive to synchronization errors (Figure 2.7c).

Inter-channel interference (ICI) The term inter-channel interference (ICI) (also referred to as inter carrier interference) refers to the interaction of waves that stem from adjacent sub-bands of the spectrum but belong the the same transmission of a multi carrier modulation system. Means to avoid this kind of interference include the usage of so called guard bands (small parts of the spectrum in which no power might be radiated) between adjacent sub-bands as shown in Figure 2.8.

Co-channel interference (CCI) In contrast to ICI, in the case of co-channel interference (CCI) the waves stem from different transmitters using the same or overlapping parts of the spectrum. The interfering power due to CCI is

$$\text{CCI}: \quad \sigma^2_{\xi,\text{cci}} = \sum_{\forall\, j} h_j^2\, p_j \ , \tag{2.8}$$

where j is an arbitrary transmitter within the receiver's interference range, h_j is the attenuation between j and the receiver, and p_j is the power with which j transmits in the receiver's band.

One of the most effective means to combat CCI is *regulation*: the spectrum is divided in to licensed parts, on which only user holding the license are allowed to transmit with a transmit power smaller than a *given maximum*. Another option – especially within a certain part of the spectrum – is *frequency planning*: the part of the spectrum is subdivided into sub-bands and the space is sub-divided into regions. In each region transmission is allowed only in a single sub-band, where the sub-band/region distribution is such that regions that are located in the interference range of each other are never allocated the same sub-band. The number of necessary sub-bands depends on the frequency reuse pattern. In cellular systems, the regions are referred to as *cells*, which are again subdivided into *sectors*. A typical frequency reuse pattern in a cellular system is $3/4$ (used *e. g.,* in the Global System for Mobile Communications (GSM) standard [10]): the overall band is divided into 4 sub-bands that are distributed among the cells. Each cell subdivides its band again, such that the transmission in each sector are performed in a different part of the spectrum. Consequently, only $1/12$th of the band is used for transmission in each sector.

One of the most important reasons for CCI is a *lack of regulation* as is *e. g.,* present in the so called Industrial Scientific and Medical (ISM) bands. In these parts of the spectrum, everybody is allowed to radiate power according to certain criteria (power threshold, duty cycle, *etc.*) with any kind of device. An example for an ISM device is a wireless local area network (WLAN) base station. Note that microwave ovens radiate power in the same frequency band, resulting in the fact that you might experience significant WLAN problems if your house mate is preparing a soup. Also, different WLAN base stations that are in each others interference range experience CCI if they are set to utilize the same frequency channel. Another reason for CCI is the weather: during periods of abnormally high-pressure, signals which would normally exit through the atmosphere can instead be reflected by the troposphere. The signal, thus, travel much further

than intended and causes interference to transmitters that are usually out of range. In addition, poor frequency planning might be a reason for CCI.

Using a combination of sophisticated transmission and power control mechanisms, however, it is today possible to design systems that are non-susceptible to CCI, *i. e.,* it is possible for the systems to differentiate between the receiving signal and the interfering signals [11]. Consequently, if such systems are used in the cellular context, the transmissions in each cell can be done in the same frequency band (referred to as frequency reuse 1 (FR1) scenario). Currently, the most famous representative of such systems is the code division multiple access (CDMA) based Universal Mobile Telecommunications System (UMTS) [12].

2.2.3 Signal-to-noise ratio

The signal-to-noise ratio (SNR) is a measure to quantify a transmission channel's quality. Particularly, it gives an insight on how much the receiving signal has been corrupted by noise. In order to determine the quality of a signal that has been received subject to instantaneous channel conditions, the signal's transmission power p needs to be related to the additive (the attenuation h) and multiplicative effects (the noise power σ_η^2) it experienced on its way to the receiver:

$$\text{SNR}: \quad \gamma_{\text{snr}} = \frac{h^2 p}{\sigma_\eta^2} \tag{2.9}$$

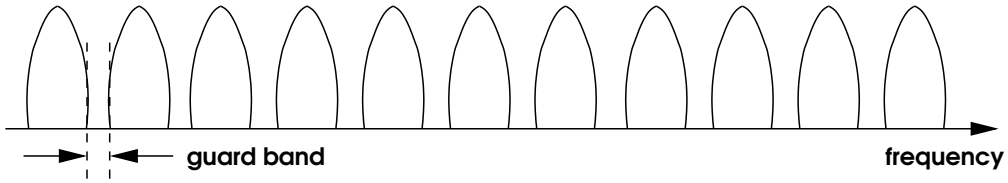
The SNR is mainly used in systems, where interference has a comparatively low impact on the signal quality. In such systems, the bit error probability and symbol error probability can be determined depending on the SNR and the chosen modulation type either using a closed formula (*cf.* [2] for a comprehensive list of SEP and BEP formulas) or by means of a look-up table that has been obtained by measurements. In both cases, the dependency of BEP and SEP on the SNR can be depicted as a graph – as shown *e. g.,* in Figure 2.4. Note that due to the variation of the multiplicative and additive effects the signal-to-noise plus interference ratio of a certain transmission $\gamma_{\text{snr}}(f,t)$ varies as a function of the transmission frequency f and time t. The period of time in which a determined instant SNR value stays more or less constant is called *coherence time*, the frequency range for which the SNR value is valid is referred to as *coherence bandwidth* [13].

2.2.4 Signal-to-noise-plus-interference ratio

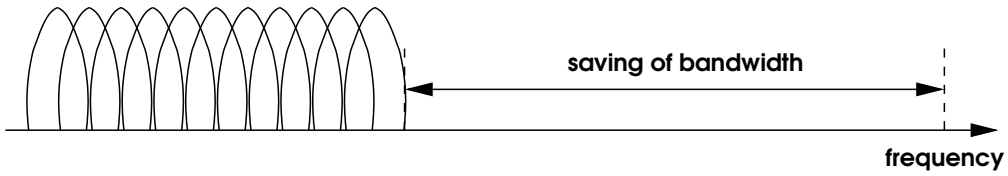
The signal-to-noise plus interference ratio (SNIR) is a measure to quantify a transmission channel's quality, if the presence of interference (expressed by the interference power $\sigma_{\xi,}^2$) is non negligible, but assumed to account for transmission errors. The BEP and SEP of such systems can be determined as described for the non interference limited systems described above, but using the SNIR:

$$\text{SNIR}: \quad \gamma = \frac{h^2 p}{\sigma_{\xi,+}^2 \sigma_\eta^2} \tag{2.10}$$

instead of the SNR. As it is the case with the SNR, the signal-to-noise plus interference ratio of a certain transmission $\gamma(f,t)$ varies as a function of the transmission frequency f and time t.



(a) conventional FDM spectrum



(b) OFDM spectrum

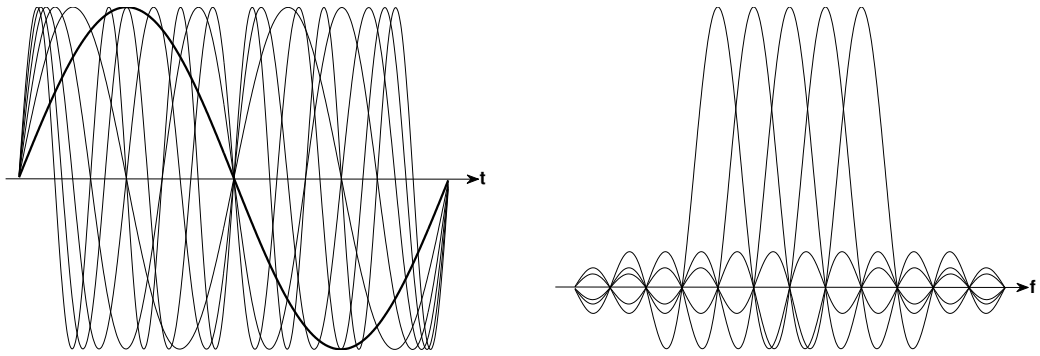
Figure 2.8: OFDM gets rid of the guard bands by adopting orthogonal sub-carriers

2.3 OFDM Basics

Wireless transmissions are limited by interference and noise. The need for ever higher data rates involves very short symbol times in conventional single carrier modulation systems, which makes them especially susceptible to inter-symbol interference. In recent years, orthogonal frequency division multiplexing (OFDM) has become a very popular ISI-proof alternative. The general idea of OFDM dates back several decades. It was first descibed in the 1960's [14, 15]. A long time passed until the first working systems could be presented, and end-user-applications did not appear on the mass market before the 1990's. Today, OFDM has gained widespread interest as the technical basis for many telecommunication applications, as the Digital Subscriber Line (DSL) [16], or digital audio broadcasting (DAB) [17], or the closely related digital video broadcasting (DVB) [18]. It serves also as the transmission scheme of the WLAN standards IEEE802.11a/g [19] and ETSI Hiperlan2 [20], as well as in the wireless metropolitan area network (MAN) standard IEEE802.16 Worldwide interoperability for Microwave Access (WiMAX) [21] and the UMTS successor Long Term Evolution (LTE) [22].

2.3.1 Modulating and demodulating OFDM signals

OFDM is a multi carrier modulation system (see Section 2.1). MCM systems make use of frequency division multiplexing (FDM), *i. e.,* they split the system's bandwidth into S narrow-band sub-channels, also known as sub-carriers [23], on which data is transmitted in parallel at a low rate. The S sub-carriers are equidistantly distributed on the frequency axis. In a conventional FDM system, the total frequency band is divided into non-overlapping sub-channels (as shown in Figure 2.8a) in order to avoid inter-channel interference. Each sub-carrier is modulated separately before the channels are frequency multiplexed. To further reduce the risk of cross-talk between the individual sub-carriers and enable the use of conventional filters and demodulators as receivers, guard-bands separate the sub-channel spectra. Thus, the non-ideal FDM technique leads to an inefficient use of the available bandwidth.



(a) OFDM sub-carriers in time domain... (b) ...and frequency domain.

Figure 2.9: sub-carrier orthogonality leads to cross-talk elimination

OFDM has been invented to increase the efficiency of spectrum usage by omitting guard bands and overlapping sub-channels (as shown in Figure 2.8b). The problem with overlapping channels is that the impact of inter-channel interference on the transmission quality is very large. Fortunately, in [14] the following has been observerd: in order to avoid ICI even with overlapping sub-channels, the sub-channel spectra have to be placed such that the spectrum of each channel has zero crossings exactly where its adjacent channels have their maxima. An according sub-channel spectrum placing is obtained, if the sub-carrier frequencies are *orthogonal* to each other. In the case of orthogonal sub-carriers, ICI can completely be eliminated. A special system property accounts for the orthogonality: each sub-carrier has exactly an integer number of cycles in the symbol interval T_s and the number of cycles differs by exactly one (cf. Figure 2.9a). The multiplexed spectra of the orthogonal sub-carriers are shown in Figure 2.9b. It can be seen that due to orthogonality, there is no cross-talk from other channels at the center frequency of each sub-carrier. Therefore, it is theoretically possible to exactly recover the transmitted data.

OFDM transmission based on sub-carrier orthogonality

Achieving sub-carrier orthogonality is a mathematically complex problem. In the 1960s, only some very expensive military machinery was equipped with OFDM technique [24]. Multiple banks of sub-carrier oscillators and coherent demodulators were necessary to implement the OFDM theory at the time it was invented. Several evolutionary steps in the technical and mathematical progress had to be taken, before the OFDM technique was cheap enough to enter the broad market. In 1971, Weinstein and Ebert applied the discrete Fourier transform (DFT) to parallel data transmission systems as part of the modulation and demodulation process [25]. In 1981, Hirosaki realized an OFDM system that features multiplexed QAM modulated sub-carriers using DFT [26]. When the fast Fourier transform (FFT) was invented as an efficient implementation of the DFT, completely digital implementations could be built around special-purpose FFT hardware. In the 1990s, the very large scale integration (VLSI) technology made

high-speed, large-size FFT chips commercially affordable. Today, both transmitter and receiver are implemented using efficient FFT techniques that reduce the number of operations from S^2 in DFT to $S \log S$, as it was first suggested in [27]. In the following, a short introduction on how QAM modulation in combination with DFT/FFT is used to generate OFDM symbols is given (large parts have been extracted form [23] and [28]).

DFT/FFT based QAM-OFDM symbol generation In general, an OFDM signal $s(t)$ consists of the sum of modulated signals X_s that are send in parallel on the orthogonal sub-carriers:

$$s(t) = \sum_{i=0}^{S-1} X_i^{(t)} = \sum_{i=-\frac{S}{2}}^{\frac{S}{2}-1} X_{i+\frac{S}{2}}^{(t)} \tag{2.11}$$

In the case of QAM-modulation, each complex modulation symbol $d_s = a_s + jb_s$ is represented by an in-phase component a_s and a quadrature component b_s. To obtain the modulated QAM signals X_s, the components modulate the quadrature modulators $\cos \omega_s t$ and $\sin \omega_s t$ respectively:

$$\begin{aligned} X_s^{(t)} &= a_s \cos \omega_s t + b_s \sin \omega_s t \\ &= \Re\left[(a_s + jb_s)(\cos \omega_s t - j \sin \omega_s t)\right] \\ &= \Re\left[d_s \exp(j\omega_s t)\right] \end{aligned} \tag{2.12}$$

Consequently, the OFDM signal can be calculated as:

$$s(t) = \Re\left[\sum_{i=-\frac{S}{2}}^{\frac{S}{2}-1} d_{i+\frac{S}{2}} \exp(j\omega_s t)\right] \tag{2.13}$$

Recall that in order to assure orthogonality among the sub-carriers each sub-carrier has exactly an integer number of cycles in the OFDM symbol interval T_s and the number of cycles differs exactly by one. Accordingly, the carriers of the signals $X_s^{(t)}$ have to offset by exactly $\omega_0 = 2\pi/T_\mathrm{s}$. Equation 2.11 can thus be rewritten as follows:

$$s(t) = \Re\left[\sum_{i=-\frac{S}{2}}^{\frac{S}{2}-1} d_{i+\frac{S}{2}} \exp(j2\pi(f_\mathrm{c} - \frac{i+0.5}{T_\mathrm{s}})t)\right] , \tag{2.14}$$

where f_c is the system's carrier frequency around which the sub-carriers are arranged. Alternatively, omitting the actual carrier frequency, the complex baseband notation can be used [23]. Taking into account that the OFDM signal $s(t)$ consists of a concatination of single OFDM symbols of length T_s, we can finally write the complex baseband notation of the OFDM symbol at $t = t_s$ (start time):

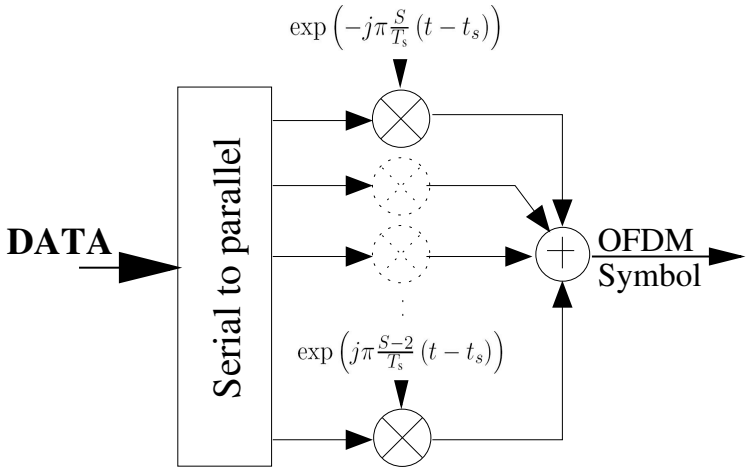


Figure 2.10: data flow in an OFDM modulator

$$s(t) = \begin{cases} \sum_{i=-\frac{S}{2}}^{\frac{S}{2}-1} d_{i+\frac{S}{2}} \exp\left(j2\pi\frac{i}{T_s}(t-t_s)\right) & \text{iff} \quad t_s \leq t \leq t_s + T_s \\ 0 & \text{iff} \quad t < t_s \wedge t > t_s + T_s \end{cases} \tag{2.15}$$

An according OFDM baseband modulator is shown in Figure 2.10: arriving data in terms of modulation symbols are split by a serial-to-parallel converter and then distributed among the S oscillators. A fundamental problem with the orthogonality requirements above is that it translates into a need of S highly synchronized oscillators. Fortunately, it can be shown that taking the inverse discrete Fourier transform of the original block of S QAM symbols and then transmitting the IDFT coefficients serially is exactly equivalent to the operations required by the OFDM transmitter. This is not very surprising, as the IDFT formula comprises exactly the summation of exponential function terms that is necessary to form the OFDM signal:

$$s(n) = \sum_{i=0}^{N-1} d_i \exp\left(j2\pi\frac{in}{N}\right) \tag{2.16}$$

Here, time t is replaced by the sample-number n. In practice, this transform can be implemented very efficiently by the inverse fast Fourier transform (IFFT) using the radix-4 algorithm [29]. Using FFT/IFFT techniques, an OFDM modulator can be built as easy as shown within the simple OFDM transmission sketch in Figure 2.11. It mainly consists of an serial-to-parallel converter that splits the given PSK or QAM modulated high-speed data stream into S parallel low-rate data-streams, an IFFT block with at least S complex in- and outputs, a parallel-to-serial converter, and a digital-to-analog converter (DAC) that forms an electromagnetic wave out of the IFFT output data.

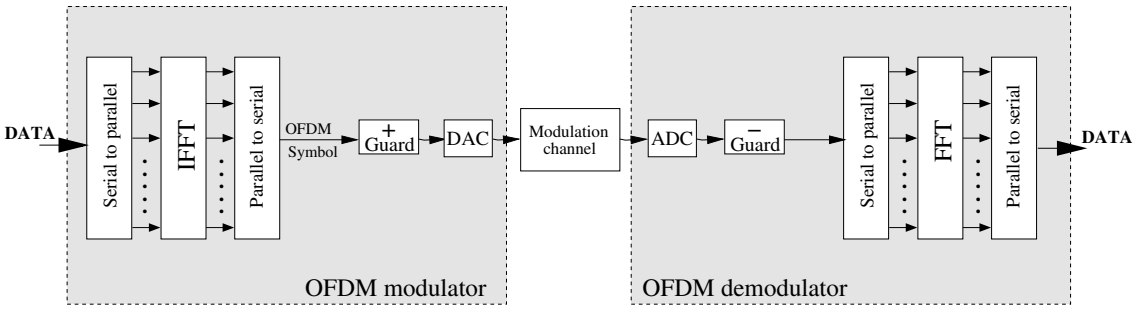


Figure 2.11: a simple IFFT/FFT based OFDM transmission sketch

OFDM symbol demodulation In general, an OFDM receiver demodulates a sub-carrier s by first down-converting the signal with a frequency of s/T_s and then integrating it over T_s seconds. From Equation (2.17) it can be seen, that the result of this operation is the desired modulation symbol of sub-carrier s scaled by the constant factor T_s:

$$r(t) = \int_{t_s}^{t_s+T_\mathrm{s}} \exp\left(-j2\pi\frac{s}{T_\mathrm{s}}\left(t-t_s\right)\right) s(t)dt \tag{2.17}$$

$$= \int_{t_s}^{t_s+T_\mathrm{s}} \exp\left(-j2\pi\frac{s}{T_\mathrm{s}}\left(t-t_s\right)\right) \sum_{i=-\frac{S}{2}}^{\frac{S}{2}-1} d_{i+\frac{S}{2}} \exp\left(j2\pi\frac{i}{T_\mathrm{s}}\left(t-t_s\right)\right) dt \tag{2.18}$$

$$= \sum_{i=-\frac{S}{2}}^{\frac{S}{2}-1} d_{i+\frac{S}{2}} \int_{t_s}^{t_s+t_s} \exp\left(j2\pi\frac{i-s}{T_\mathrm{s}}\left(t-t_s\right)\right) dt = d_{s+\frac{S}{2}}T_\mathrm{s}$$

Recall that the frequency difference $(i-s)/T_\mathrm{s}$ produces an integer number of cycles for each sub-carrier s within the time-domain integration interval T_s. Accordingly, the integration over T_s in Equation 2.17 ideally results in zero for all sub-carriers apart from sub-carrier s. Note, however, that for these operations not only the receiving oscillators need to be highly synchronized, but synchronization with the oscillators on the transmission side has to be assured, too.

Again, fortunately this is not the case for IFFT/FFT based OFDM transmission systems. Figure 2.11 shows a simplified schematic of such a system. Instead of demodulating the single sub-carriers, the received wave is sampled by means of an analog-to-digital converter. The samples are used as input for an FFT block, which outputs a scaled version of the modulation symbol per sub-carrier. Since the FFT and IFFT algorithms are almost similar, it is possible to build OFDM transceivers that are based on one single FFT integrated circuit (IC).

Guard insertion: the cyclic prefix By using the techniques described above, OFDM systems transmit data on orthogonal sub-channels, which in theory makes it insusceptible to inter-channel interference. Moreover, by choosing the OFDM sub-channel bandwidth smaller than the systems coherence bandwidth (see Section 2.2.4), the individual sub-carriers theoretically experience *frequency flatness*, and thus inter-symbol interference can be avoided. In practice,

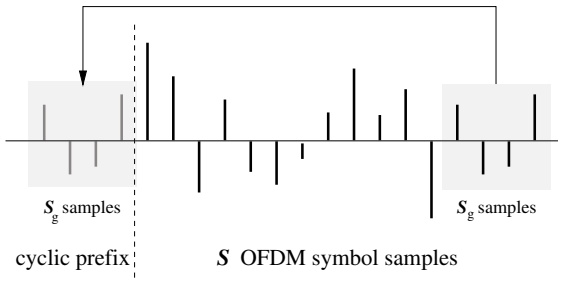


Figure 2.12: the cyclic prefix of an OFDM symbol consists of copies of the last S_g samples

however, frequency flatness cannot completely be achieved. Hence, each sub-carrier has a certain frequency response, which modifies the signals going through it individually. This results in a frequency selective behavior, *i. e.,* transmission symbols are deformed, as they consist of different frequency components that are differently attenuated and delayed. As a consequence, ISI cannot completely be avoided in practice. Unfortunately, the presence of ISI endangers the system's orthogonality, and a lack in orthogonality in an OFDM system in turn yields inter-channel interference.

To further deal with preserving the sub-carrier orthogonality and combat the problem of ISI and ICI, OFDM systems place a guard interval of length T_g at the beginning of each OFDM symbol, the so called *cyclic prefix (CP)* [23]: the last S_g of the S original OFDM symbol samples (the output of the IFFT block) are copied and placed right before the first OFDM symbol sample as shown in Figure 2.12. This way, the overlapping of successive OFDM symbols does not translate in ISI, if the overlapping in terms of samples is smaller than S_g. Moreover, by copying the last S_g samples instead of just using an empty guard (*i. e.,* stopping the transmission for the duration of T_g), it is guaranteed that none of the delayed signal copies (due to multi-path, *cf.* Section 2.2) features a phase transition within the original S OFDM symbol samples received on the dominant path, if the cyclic prefix length T_g is greater than the multi path delay spread $\Delta\sigma$. In other words, each sub-carrier has in any case an integer number of cycles within one FFT interval and, thus, orthogonality is preserved. Since the cyclic prefix provides no new information, it results in an overhead in time which reduces the system efficiency by a factor of $T_g/T_s + T_g$. This overhead, however, is not significant if the number of original OFDM samples *i. e.,* the number of sub-carriers) is large compared to the number of guard samples.

2.3.2 Orthogonal frequency division multiple access (OFDMA)

If the transmission technique OFDM in a multi-user system is combined with frequency division multiple access (FDMA), it is usually referred to as orthogonal frequency division multiple access (OFDMA) system. In contrast to an OFDM-time division multiple access (TDMA) system, where a terminal uses all available sub-carriers for a limited amount of time, in an OFDMA system a single terminal might use a part of the spectrum (*i. e.,* one or several sub-carriers) at any point in time. The resource distribution is thus done in the frequency domain, rather than in the time domain. In other words, in an OFDMA system several terminals transmit or receive in parallel using distinct parts of the system bandwidth.

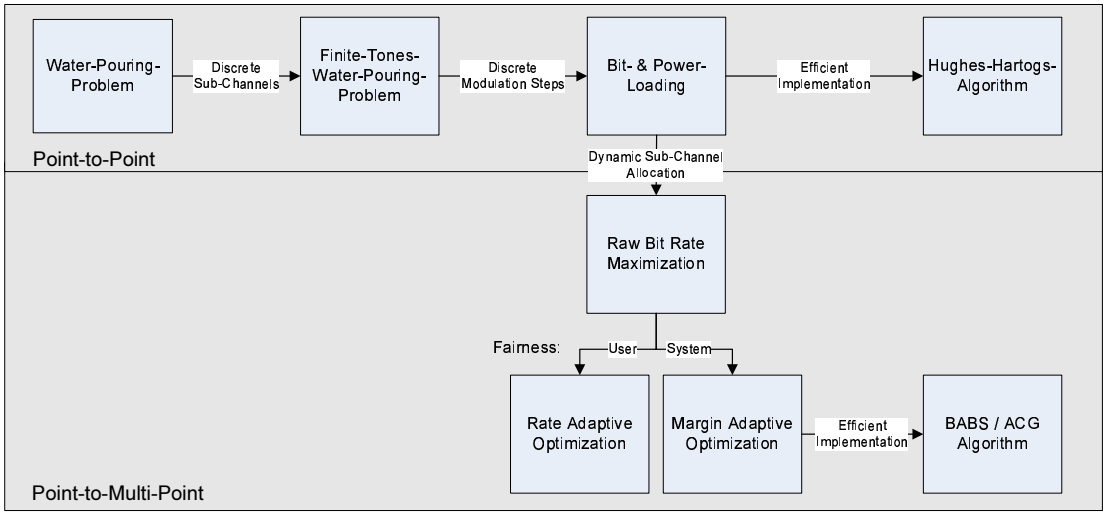


Figure 2.13: an overview of the cross-layer optimization approaches presented in this section

2.4 Dynamic Resource Allocation Optimization in OFDM

Throughout this section, transmitter schemes that adapt to any channel variation are referred to as *dynamic*. Potential adaptation mechanisms comprise the modulation type adaptation, transmission power adaptation, as well as the sub-carrier assignment selection per terminal. In contrast, schemes which do not adapt to channel variations are referred to as *static*.

The general idea to adapt OFDM transmissions to the actual channel conditions is motivated by the fact that different sub-carriers experience different attenuation conditions (if their spacing in frequency is larger than the coherence bandwidth, Figure 2.14). If this frequency-selective behavior is assumed to stay constant for some time span (*i. e.,* the attenuation of each sub-carrier stays constant for the considered time duration), one might ask the question: "Does it make sense for the transmitter to *adapt* to the frequency-selective attenuation of the channel in order to transmit data *better* (*i. e.,* faster, more reliable, etc.)?"

In order to enable adaptive transmissions, the physical layer (PHY) and the medium access control layer (MAC) of the according systems have to cooperate. Particularly, the channel state information and the user requirements (in terms of data rate and/or its actual queuing situation) have to be jointly evaluated. An according dynamic system approach is thus mostly referred to as cross-layer approach. Recently, a lot of research effort has been spent on cross-layer system design. It has been shown that cross-layer mechanisms potentially provide significant performance gains for various systems. In this section several aspects of cross-layer system optimization regarding wireless multi-carrier systems are reviewed. Basic optimization models are discussed and selected heuristic approaches realizing cross-layer policies by the means of dynamic resource allocation are presented (see Figure 2.13). Two specific areas are treated separately: Models and dynamic approaches for single transmitter/receiver pairs – *i. e.,* a point-to-point communication scenario – as well as models and approaches for point-to-multi-point communication scenarios – such as the down-link of a wireless cell.

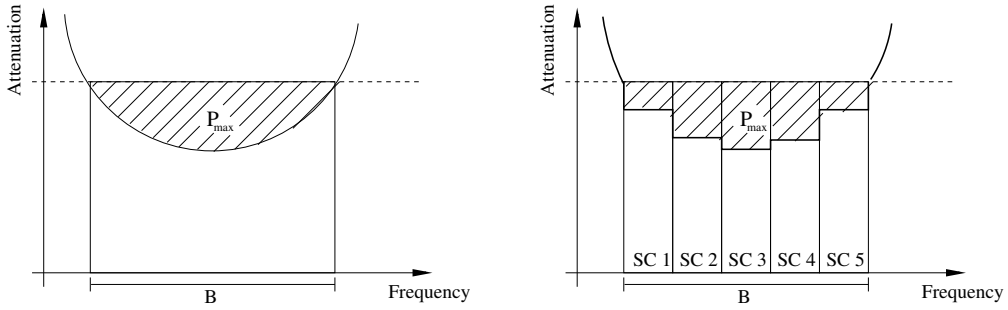


Figure 2.14: principle of information theory's "water-filling" theorem and its application to a five sub-carrier OFDM system

2.4.1 Dynamic schemes for point-to-point communications

First, results with respect to the adaptation of transmit power and modulation types for OFDM systems if a transmitter is communicating with a single receiver are reviewed. Information theory, and in particular the *water-filling theorem* [30], provides an important answer to the question on how to jointly optimize these two: in general, knowing the transfer function of a channel, its capacity can be found (where *capacity* is defined as the maximum bit rate at which data can be transmitted with an arbitrary small bit error probability). According to the theorem, the channel's capacity is achieved by adapting the transmit power to its transfer function. Roughly speaking, given a certain power budget more transmit power is applied to frequencies experiencing a lower attenuation. Thus, given the transfer function, the optimal power distribution is similar to inverting the transfer function and pouring a liquid, *i. e.,* power, into the shape (see Figure 2.14). Consequently, the scheme was termed *water-filling*. The higher the variance of the transfer function (assuming a constant average attenuation) the higher is the resulting capacity. Hence, a flat transfer function delivers the lowest capacity for a certain power budget and average attenuation.

Apart from the fact that the *optimal* power distribution is somewhat computationally difficult to obtain, the mathematical derivations of the water-filling theorem cannot be applied directly to OFDM systems due to two reasons:

1. The water filling theorem assumes continuous frequency attenuation functions. In OFDM systems, usually one attenuation value per sub-carrier is available, yielding a discrete (sampled) version of the attenuation function (as shown in Figure 2.14). In other words, water filling requires "systems" featuring an unlimited number of sub-carriers of infinitely small bandwidth, which is impractical.

2. The water filling theorem is based on a continuous relationship between the allocated power and the achievable capacity. Since in real systems only a finite set of modulation types is available, the resulting power allocations per sub-carrier differ from the water-filling ones.

As a consequence, in order to leverage the water-filling benefits in OFDM systems, a discrete version of the scheme is necessary.

Point-to-point system assumptions To evolve from the continuous nature of the water filling theorem to a discrete version, let us assume a system of bandwidth B with a discrete number of sub-carriers S, each of which featuring a sub-carrier bandwidth $\Delta f = B/S$ [31]. The instant sub-carrier states at time t are represented by a vector of signal-to-noise plus interference ratio values $\gamma^{(t)} = (\gamma_1^{(t)} \dots \gamma_S^{(t)})$, where SNIR value $\gamma_s^{(t)}$ depends on sub-carrier s's instant attenuation and its transmit power share (*cf.* Equation (2.10)).

Finite tones water-filling

Using Shannon's capacity formula, a consequent transformation of the capacity problem for the case of S sub-carriers is given by the following formulation:

$$\max_{\gamma^{(t)}} \sum_s \Delta f \log_2 \left[1 + \gamma_s^{(t)}\right] \tag{2.19}$$

$$\sum_{\forall s} p_s^{(t)} \leq p_{\max} \tag{2.20}$$

Equation (2.19) states that the capacity is obtained by optimally distributing the transmit power among the sub-carriers, where $\gamma_s^{(t)}$ increases with sub-carrier s's power share. Thus, with an infinite amount of transmission power, an infinite capacity would theoretically be possible. However, note that in this case the power distribution is subject to a total power budget (Equation (2.20)). The combination of optimization goal (2.19) and this total power constraint forms a non-linear, continuous optimization problem, which is referred to as *finite tones water-filling problem* [32]. It can be solved analytically by applying the technique of Lagrangian multipliers [33].

Solving the finite tones water-filling problem delivers continuous rate-shares for discrete sub-carriers. To take a further step towards discrete water-filling, the real-valued rate shares need to be replaced by whole-numbered bit assignments.

Loading algorithms

Only a fixed amount of modulation types is available for sub-carrier data transmission in OFDM systems. Thus, for those systems Shannon's formula, as it is used in (2.19), is not a valid option to translate a sub-carrier's state into its rate-share. Instead, modulation assignments from a finite set need to be derived from the channel states. Denote by $F(.)$ the adequate function that delivers the rates of the available modulation types $m_s = F\left(\gamma_s^{(t)}, P_{\text{err}}\right)$ with respect to the SNIR and a predetermined target error probability P_{err}. Note that $F(.)$ is a piece-wise constant function over the SNIR. Substituting Shannon's formula by $F(.)$ in (2.19) leads to the following optimization formulation:

$$\vartheta = \max_{\gamma^{(t)}} \sum_s F\left(\gamma_s^{(t)}, P_{\text{err}}\right) . \tag{2.21}$$

In combination with total power constraint (2.20) this optimization goal specifies the *bit rate maximization* integer programming problem [34]. Solving the problem results in optimal power- and modulation-type per sub-carrier choices with respect to the total power budget. In general,

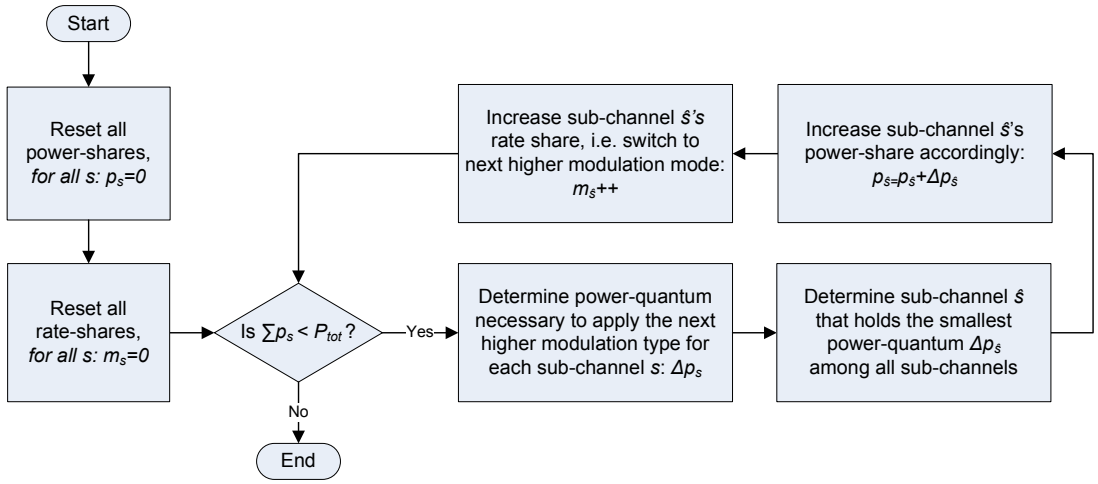


Figure 2.15: principle of the Hughes-Hartogs loading algorithm

integer programming problems are difficult to solve. Fortunately, simple greedy algorithms already yield optimal solutions for this class of problems (see below). Note that maximizing the bit rate is only one possible metric. Another option is minimizing the transmit power for a given rate [35], or minimizing the bit error probability for a certain rate and power budget [36]. Further algorithms are presented in [37–40].

In general, such algorithms are referred to as *loading algorithms*. The group of loading algorithms can be subdivided into *bit-* and *power-loading algorithms*. Bit-loading algorithms adapt the number of bits transmitted per sub-carrier according to the sub-carrier states. Correspondingly, power-loading algorithms adapt the transmit power. However, as in most cases the number of bits is adapted together with the transmit power, both schemes and their combination are referred to as loading algorithms in the following. The earliest loading algorithm has been proposed by Hughes-Hartogs [41]. Its principle is quite simple (*cf.* flow-chart in Figure 2.15): for each sub-carrier, calculate the amount of power required to transmit data with the lowest modulation type. Then, the sub-carrier that requires the least amount of power is selected, the amount of power is allocated to it and the required additional power for applying the next higher modulation type is calculated for this sub-carrier (while the total power budget is decreased by the allocated amount). The algorithm terminates if no more transmit power is available. It determines for a discrete amount of modulation types the optimal power allocation with respect to the target bit error probability while maximizing the data rate. Hence, it solves the *bit rate maximization problem*. Note that the same scheme can also be used to determine the optimal power allocation in order to minimize the transmit power subject to a rate constraint (the *margin maximization problem*). In this case the algorithm simply runs until the target data rate is reached. Although the Hughes-Hartogs algorithm does not enumerate all feasible solutions, the required amount of steps is quite high. For example, assume the M modulation steps to differ by one bit. Then, for transmitting a total of 1000 bits the algorithm will have to perform 1000 iterations. Therefore, faster schemes reaching the optimal or near-optimal power allocation have been of interest. For example, Chow *et al.* [35] presented a faster loading algorithm in order

to minimize the transmit power while maintaining a required data rate. They propose to start with an *equal* power distribution and then alter this distribution in order to reach the required rate. Many further bit-loading approaches have been presented. For an extensive discussion on different approaches see [42].

2.4.2 Dynamic schemes for point-to-multi-point communications

In this section, results regarding the application of dynamic mechanisms in point-to-multi-point scenarios are reviewed. Particularly, the down-link of a single cell belonging to an OFDMA (see Section 2.3.2) based cellular system featuring a base station and J terminals is considered. In this setup, we are mainly interested in an effect referred to as *multi-user diversity*: as the terminals' locations are distributed over the complete cell area, the same sub-carrier s is likely to be in different quality states for different terminals. In other words, the multi-user communication scenario is characterized by a spatial selectivity of the sub-carrier states. The reason for this spatial selectivity is the fact that the fading process is, in general, statistically independent for different terminals, as long as their receive antennas are separated considerably (by a minimum spacing of one wavelength).

Point-to-multi-point system assumptions In the following it is assumed that time is slotted into frames, and that the base station knows the SNIR values $\gamma_{j,s}^{(t)}$ per terminal j and sub-carrier s. The set of all instant SNIR values forms a $J \times S$ matrix $\mathbf{\Gamma}^{(t)}$. Based on that knowledge, a dynamic algorithm at the base station generates *disjunctive sets of sub-carriers* assigned to each terminal. The specific assignment $x_{j,s}^{(t)}$ of sub-carrier s to terminal j at time t is a variable of the system, where

$$x_{j,s}^{(t)} = \begin{cases} 1, & \text{if } s \text{ is assigned to } j \text{ at } t \\ 0, & \text{if } s \text{ is not assigned to } j \text{ at } t. \end{cases}$$

The set of all assignment variables $x_{j,s}^{(t)}$ forms the binary assignment matrix $\mathbf{X}^{(t)}$. Based on the power-rate function $F(.)$, for each terminal/sub-carrier combination $<j \times s>$ one out of the M modulation types is selected depending on the instant SNIR-value $\gamma_{j,s}^{(t)}$. Recall that the SNIR value depends on the current channel state, as well as on the transmission power share the base station assigns to terminal j on sub-carrier s. All information about the power, sub-carrier and modulation to be used per sub-carrier and terminal is collected in so called assignment sets. The base station informs each terminal of its next assignment set before it starts the payload data transmission on a separate control channel. The sets are assumed to be valid for a single frame-time.

Multi-user sum rate maximization

Regarding the single cell system setup described above, a straightforward optimization approach is to maximize the overall bit rate of the cell per down-link phase, where the SNIR values and

the assignment matrix are the system variables:

$$\max_{\mathbf{\Gamma}^{(t)},\mathbf{X}^{(t)}} \sum_{\forall j} \sum_{\forall s} F\left(\gamma_{j,s}^{(t)}, P_{\text{err}}\right) x_{j,s}^{(t)} \,. \tag{2.22}$$

In combination with the *disjunctive sets* constraint

$$\sum_{\forall j} x_{j,s}^{(t)} \leq 1 \qquad \forall s \tag{2.23}$$

and total power constraint (2.20) optimization goal (2.22) forms the *multi-user sum rate maximization problem*. Again, the total power constraint limits the overall transmit power (as in the case of the finite tones water-filling problem (2.19)). The disjunctive sets constraint is specific to the multi-user scenario: it limits the assignment of each sub-carrier to at most one terminal at a time. As in the case of the finite tones water-filling problem, an integer optimization problem is encountered. However, in this case it is required to find the optimal power allocation (and thus SNIR values) plus deciding on the allocation variable $x_{j,s}^{(t)}$ for each terminal/sub-carrier pair. Fortunately, as in the finite tones water-filling case, the resulting integer programming problem can be solved easily by a greedy algorithm described in [43].

The *multi-user sum rate maximization*, however, exhibits a fairness issue, as terminals in good positions (*e. g.,* close to the base station) are always favored when it comes to the sub-carrier distribution. As a consequence, some terminals experience high transmission delays for packets, if they receive anything at all. This is due to optimization goal (2.22) that aims at maximizing the raw cell-throughput, *i. e.,* the sum-rate of the cell for each down-link phase. Alternatively, different optimization goals can be formulated that account for intra-cell fairness. Fairness among the terminals, however, comes in any case at the cost of a decreased sum-rate throughput of the cell.

Rate adaptive optimization / max-min optimization

The goal of *rate adaptive optimization* is to guarantee that *each* terminal obtains a rate of at least ϵ in the down-link. Since this bound is mainly determined by the worst situated terminal (it requires the most resources among all terminals to achieve ϵ), rate adaptive optimization is also referred to as *max-min optimization*. By maximizing ϵ, while at the same time guaranteeing that each terminal's rate is larger than that, the maximum equal rate per terminal is achieved:

$$\begin{aligned} \max_{\mathbf{\Gamma}^{(t)},\mathbf{X}^{(t)}} \quad & \epsilon \\ \text{s. t.} \quad & \sum_{s} F\left(\gamma_{j,s}^{(t)}, P_{\text{err}}\right) x_{j,s}^{(t)} \geq \epsilon \quad \forall j \,. \end{aligned} \tag{2.24}$$

Note, that the total power constraint (2.20) and the disjunctive sets constraint (2.23) are again part of the overall problem formulation (for the complete mathematical formulation see [44] or [45–47]). The problem with this optimization approach lies in the fact that a very badly situated user drastically reduces the overall system performance. Moreover, in many cases achieving the same rate per terminal is not really useful, as each terminal j most probably has its specific data rate requirement depending on the application in use.

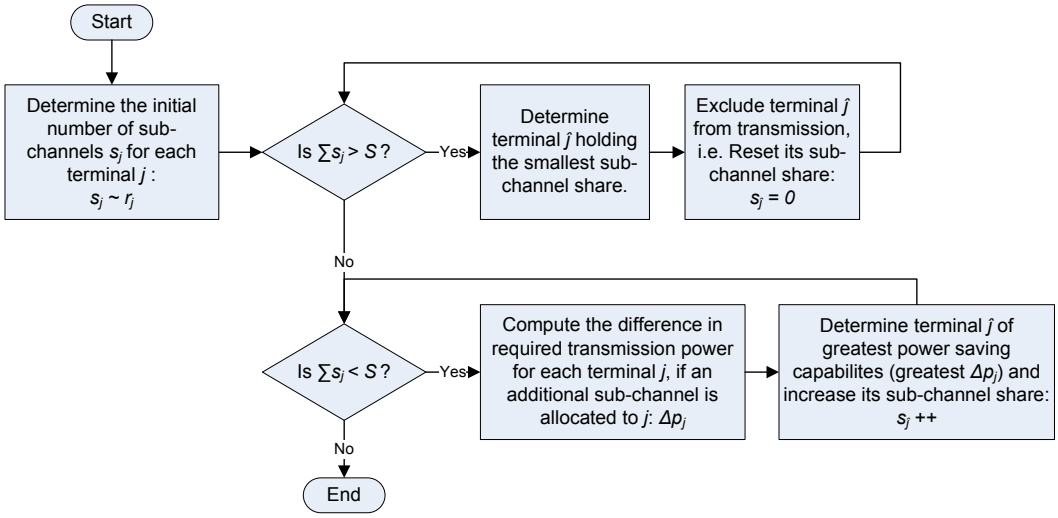


Figure 2.16: principle of the Bandwidth Assignment Based on SINR (BABS) algorithm

Margin adaptive optimization

The objective of this *margin adaptive optimization* approach is to minimize the overall transmit power (the sum over the individual power shares per sub-carrier $p_s^{(t)}$), while guarantying the individual rate requirements:

$$\begin{aligned} \min_{\mathbf{\Gamma}^{(t)},\mathbf{X}^{(t)}} \quad & \sum_s p_s^{(t)} \\ \text{s. t.} \quad & \sum_s F\left(\gamma_{j,s}^{(t)}, P_{\text{err}}\right) x_{j,s}^{(t)} \geq r_j^{(t)} \quad \forall j \ . \end{aligned} \tag{2.25}$$

The disjunctive sets constraint (2.23) is again part of the overall problem formulation (for the complete mathematical formulation see [48] or [49–56]).

2.4.3 Generating optimal and sub-optimal solutions

The three optimization problems presented above belong to the group of integer programming (IP) problems. IPs are in general known to be difficult to solve. Although the amount of possible solutions is finite, finding the optimal solution remains a difficult task, possibly requiring a "brute force" enumeration and comparison of all feasible solutions. In fact, the margin and rate adaptive optimization problems have been proved to be NP-complete [42]. As a consequence, a significant computational overhead can be expected at the base station to solve them *optimally*. Hence, a lot of research work has been spent on developing schemes that deliver optimal or nearly optimal solutions at low cost. Most of these proposals for solving the rate or margin adaptive optimization problem belong to one of three different methods: relaxation, problem splitting, and heuristics.

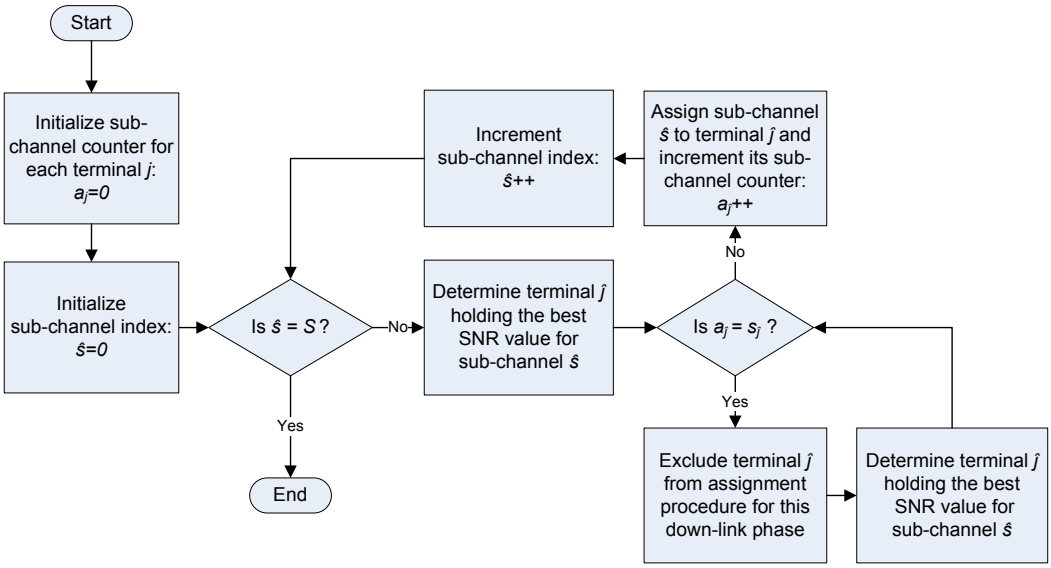


Figure 2.17: principle of the Amplitude Craving Greedy (ACG) algorithm

Relaxation The first method is to relax the integer constraint on the bit- or sub-carrier assignments. Thus, for calculation purposes each sub-carrier is allowed to carry a non-integer amount of bits and can be assigned to multiple different terminals during one down-link phase. By relaxing the integer constraint on the rate and margin adaptive optimization problems, both become linear programming (LP) problems, which can be solved efficiently. However, after solving the relaxed problem, the LP solution has to be reevaluated as only integer solutions are feasible from a system's point of view. Usually, this is done by reassigning the sub-carriers to the terminals with the largest non-integer fraction. This approach has first been presented by Wong *et al.* [48]. It serves as comparison basis for multiple later studies on the margin-adaptive problem.

Problem Splitting Following the second proposal, the optimization problem is split into two less complex problems [53]. First, the *number* of sub-carriers $s_j^{(t)}$ each terminal needs to fulfill its rate requirements is determined (referred to as sub-carrier allocation). Then, the specific sub-carrier/terminal pairs are generated (*i. e.,* the best matching sub-carriers are selected per terminal). This can be done efficiently by the use of state-of-the-art matching algorithms [34].

Heuristics A third common approach is to solve the rate or margin adaptive problem by heuristics that are mostly based on sorting procedures. One such approach is presented by Kivanc *et al.* in [53]. It is the heuristic realization of the analytical two-step approach presented above. Resource Allocation (determining the number of sub-carriers each terminal should receive) is done using the greedy bandwidth assignment based on SNIR (BABS) algorithm (shown as flowchart in Figure 2.16). Once the resource allocation is determined for each terminal, the specific assignment of the sub-carriers is done by the amplitude craving greedy (ACG) algorithm (Figure 2.17). An overview of further heuristics can be found in [42].

2.4.4 Peformance gain

The generation of optimal solutions, and even the generation of sub-optimal solutions for the dynamic assignment problems require significant additional computational effort at the transmitter. In the following the performance results of several different approaches are reviewed to answer the question whether this effort is justifiable.

Point-to-point communications In [57] Czylwik compares the performance of an single carrier modulation (SCM) system with frequency equalization to the performance of an OFDM system with static and adaptive modulation. For different wireless channels Czylwik finds an improvement of around 2 dB when switching from the SCM system to the static OFDM system (both applying the same modulation type), and a further improvement of around 4 dB when switching from the static OFDM system to the OFDM system with adaptive modulation. In case of no line-of-sight, the performance gains are much larger, around 8 dB for switching from the SCM system to static OFDM and around 5 dB when switching from the static to the adaptive OFDM. Similar results advocating adaptive modulation were found by Rohling *et al.* in [58].

In a further study [59], Czylwik investigates the performance difference (in terms of bit error probabilities) between adaptive modulation (with static power allocation) and adaptive loading (variable power and bit allocation). For all considered channels (two different ones, relying on measurements) static OFDM is significantly outperformed (by about 5 dB by OFDM with adaptive modulation. However, the difference between adaptive modulation and adaptive loading is rather small, around 1 dB. This indicates that the computational expense related to adapting the bit *and* power allocation is not worth the performance gain achieved, at least for these channel characteristics studied. Similar results were found by Barreto *et al.* in [60] for an IEEE 802.11a-like system applying adaptive modulation and adaptive loading. In [61] the authors find that the biggest advantage for a coded, interleaved system is obtained by performing adaptive modulation and not by adaptive loading of bits and power (adaptive modulation achieves a 3 dB gain versus a static system, while additional power adaption only adds 1 dB more). In [62] the same authors show that soft decision decoding yields a much larger performance gain than adaptive power allocation.

Point-to-multipoint communications In the case of point-to-multi-point communications, sub-carriers as well as power and bits can be assigned dynamically to different terminals according to the chosen approach (*e. g.,* margin- or rate adaptive). Two qualitatively different performance gain cases are considered: the potential gain when generating the optimal solutions off-line, as well as the real-time achievable gain when using sub-optimal algorithms.

In [63] the potential gain for several optimal rate adaptive system approaches is explored. The paper answers the question on how the performance gains of applying the individual adaptation schemes relate to each other. The average throughput and maximum delay results for a certain cell parameterization[1] while increasing the cell's radius are shown in Figure 2.18: poten-

[1]The simulation assumes a system bandwidth of 16.25 MHz divided into 48 sub-carriers, 4 different modulation types are available (BPSK, QPSK, 16-, and 64-QAM, target symbol error rate is 10^{-2}, the transmit power is set to 10 mW, one up-link and down-link phase has a duration of 1 ms in TDD mode. 8 terminals are located in the cell.

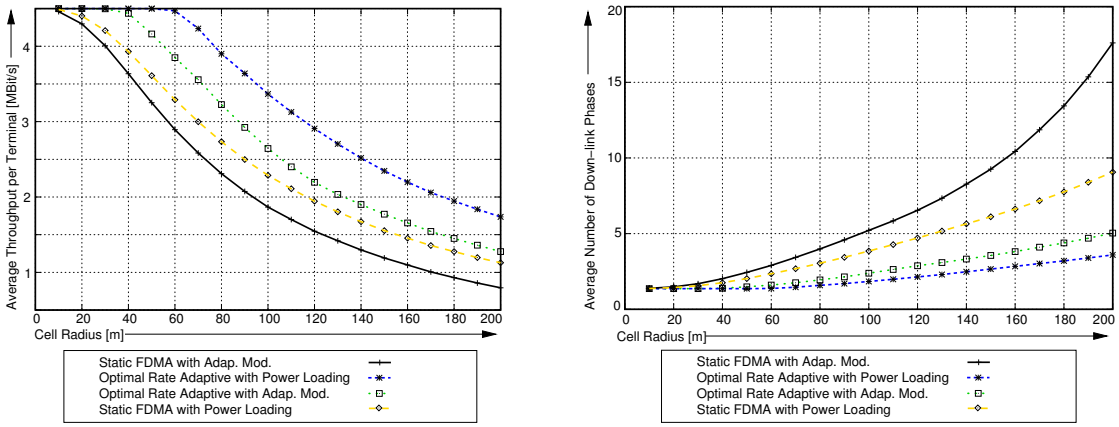


Figure 2.18: average throughput and maximum packet delay performance comparison between a static and several optimal rate-adaptive dynamic OFDMA schemes

tially, all dynamic schemes significantly outperform the static scheme. In terms of the average throughput the gain of a fully dynamic system is up to 100% compared to the static scheme. The gain in average packet delay (in terms of transmitting an IP packet of size 1500 Bytes) is even larger [64]. It is important to note that – in contrast to the point-to-point communication case – the power adaption does yield a significant performance increase. It is much larger in the case of dynamic sub-carrier assignments than in the case of static sub-carrier assignments, especially regarding the average throughput.

The question of how much of this optimal performance can be reached by sub-optimal algorithms has been answered for various different sub-optimal approaches individually. For example, Kim *et al.* [44] show that the *relaxation strategy* performs equally well when compared to the potential gain of the optimal solution. In [45], Yin *et al.* show that their *two-step approach* reaches about 90% of the optimum. Rhee *et al.* [46], however, find for a fairly simple *heuristic* a much smaller performance difference of about 2%. Concerning the BABS/ACG heuristics presented above, simulations in [53] show that the power requirements of the combination BABS/ACG are only slightly higher than the power requirements of Wong's relaxation approach [48], while CPU run times are smaller by a factor of 100. Due to inconsistencies between all these studies, however, it is not completely clear, how much the performance results of different sub-optimal approaches relate to each other. Still, it must be reasoned that even with sub-optimal mechanisms, large gains in performance can be achieved.

Apart from considering the theoretical performance gains of a dynamic OFDMA system over static systems, other studies consider the implementation aspects of such dynamic OFDMA systems. In [65] Rohling *et al.* consider different multiple access schemes for broadband OFDM systems. They find that dynamic OFDMA clearly outperforms TDMA and CDMA variants. However, FDMA also requires the largest overhead due to signaling, as described below.

2.4.5 Overhead

The performance gain due to dynamically adapting a wireless OFDM transmission system to the channel quality described above comes at the cost of additional signaling data that needs to be exchanged between the transmitter and the receiver, namely channel state information and control signaling data.

Channel state information The transmitter (particularly the scheduler at the transmitting side) requires knowledge regarding the instantaneous sub-carrier quality values (usually referred to as channel state information (CSI)) in order to use it as input for the dynamic adaptation algorithms. For the mechanisms presented in this section, the SNIR values per sub-carrier $\gamma_{j,s}^{(t)}$ are of major interest. In time division duplex (TDD), if the channel can be assumed to be reciprocal and continuous transmissions are going on in both directions, the data can be acquired directly by both transmitting entities. In all other cases (and particularly in any frequency division duplex (FDD) system), however, quality reports holding the sub-carrier CSIs have to be transmitted from the receiver to the transmitter, which introduces overhead and a delay ΔT_{csi} between the measurement of the channel quality data and its usage.

Control signaling The receiver requires information about the resource allocation for the following frame. As a consequence, system resources in terms of bandwidth and/or time need to be provided to form some kind of control channel that delivers the control signaling data. Depending on the type of dynamic mechanism in use, the amount of signaling information is more or less significant compared to the amount of user data to be transmitted. In the case of applying bit-loading schemes to point-to-point transmissions (*cf.* Section 2.4.1), the receiver has to be informed about the chosen modulation types per sub-carrier for the next allocation cycle to enable it to decode the sub-carrier symbols correctly. In the case of fully dynamic point-to-multi point transmissions that enable dynamic sub-carrier assignments (as in Section 2.4.2), the signaling information contains the full assignment set that determines which terminal is assigned which sub-carrier with which modulation type.

Overhead reduction The overhead, especially the control signaling, that is necessary to enable dynamic mechanisms in OFDMA systems can become quite costly. It has, however, been shown that it does not consume the complete performance gain achieved by the dynamic assignments. In fact, the signaling overhead is found to be rather small compared to the potential performance gains [66]. Still, there are attempts to reduce the signaling overhead by different means. One such opportunity is to build blocks of sub-carriers (so called *resource blocks*), which are treated as the smallest assignable unit within the system [56, 67, 68]. These studies show that block building is a good choice as long as the resource block bandwidth Δf_n is smaller than the channel's coherence bandwidth. Alternatively, the signaling overhead can be incorporated into the optimization problem. This has been proposed in [69]. Here, quite similar assignments from one down-link phase to the next one can be generated due to the correlation in time. Thus, the overall overhead is decreased. This option, however, can lead to higher computational resource requirements at the base station in order to generate these distinguished assignments.

2.5 Summary

Digital transmissions in wireless communication systems are subject to several transmission quality degrading effects. In modern cellular systems that target at very high data rates, mostly inter-symbol interference and inter-channel interference are the limiting factors. OFDM systems shield their transmissions against ISI by transmitting data at low data rates on multiple parallel sub-carriers. Additionally, by guaranteeing orthogonality among the sub-carrier frequencies, OFDM systems inherently avoid ICI. If the data that is simultaneously transmitted via the parallel carriers is sent by or destined for different entities that are concurrently using the available system spectrum, the system is referred to as OFDM-FDMA (or short: OFDMA) system. The application of dynamic mechanisms that adapt OFDM transmissions to the current channel states achieves significant performance gains (in terms of lower bit-error rate, lower power consumption and/or higher throughput) at the cost of increased computational system complexity and signaling overhead.

Dynamic OFDMA systems have been discussed for more than a decade now. The approaches presented in this chapter provide an insight into the basic ideas and some selected further approaches. Obviously, it does not cover the whole story, as innumerable original dynamic OFDMA approaches have been presented. The interested reader is referred to the numerous available survey and tutorial papers [70–73]. Still, the previous sections provide all the background information that is necessary to comprehend the ideas and computations of this thesis. Particularly, the cross-layer optimization approaches presented above build the basis for the research work that is presented in this thesis, as they are referenced and further developed throughout the following chapters.

Chapter 3

Scope of the Thesis

THE transmission technology OFDM is the per se transmission standard for modern digital communication systems. Due to its stability against inter-symbol interference and inter-channel interference it is especially appreciated as technical basis for wireless communication systems. Additionally, the opportunity to adapt its transmission parameters to instantaneous channel conditions on a very fine grained scale makes OFDM, and particularly OFDMA attractive for such systems. Broadcasting services, such as DAB [17] and DVB [18], as well as wireless LAN/MAN standards IEEE802.11a/g [19], ETSI Hiperlan2 [20], and WiMAX [21] successfully employ dynamic OFDM(A) techniques for several years now. However, even though dynamic OFDMA has been identified as a promising technique for cellular systems as well, and OFDMA based cellular standards such as 3GPP's Long Term Evolution (LTE) standard [22] have been around for several years now, there is no commercially working dynamic OFDMA based cellular system to date. One reason for that lies in the fact that it is not straight forward, but involves some great challenges to transfer well working dynamic mechanisms from a broadcasting or LAN/MAN system design to a cellular one.

This thesis is concerned with facing these particular challenges that arise when adopting well known dynamic OFDMA mechanisms in cellular systems. The system under consideration is a dynamic OFDMA based cellular system. Assuming downlink transmissions, in each cell a single transmitter (the base station) communicates with multiple receivers (the terminals) over a dedicated part of the wireless channel, as shown in Figure 3.1. Packets destined for different terminals and belonging to different quality of service classes arrive at the base station. Time is slotted into frames. For each frame, the base station has the task of scheduling the packets to be transmitted, and distributing the available wireless channel resources in terms of transmission power and frequency spectrum among the receivers accordingly. Thereby, the overall available resources are partitioned into two logically separated channels: a data channel for user data transmission and a control channel for signaling and meta data transmission.

3.1 Challenges

Major differences between cellular OFDMA systems and the systems for which dynamic OFDMA mechanisms have initially been deployed lie in the *packet nature of the traffic*, the *quality and*

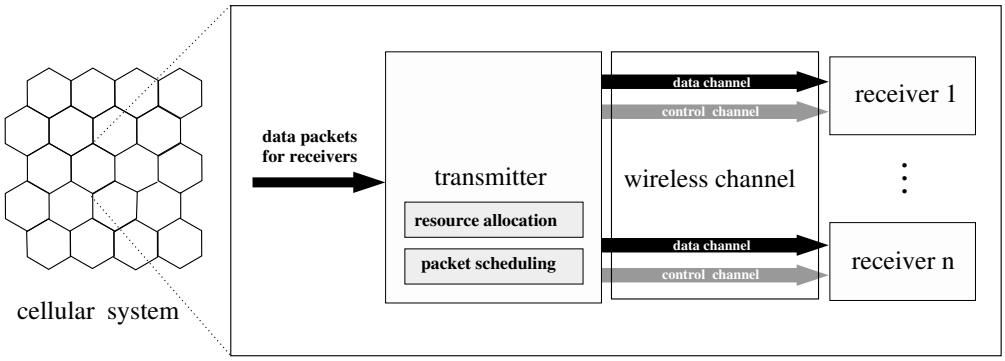


Figure 3.1: the downlink (DL) of a cellular system with a single transmitter (the base station) and multiple receivers (the terminals) per cell that are communicating via a wireless link

quantity of the required control signaling data, and the particular *interference situation* of cellular systems. The research challenges faced in this thesis arise from these three differences.

Challenge 1: the packet nature of traffic The data destined for users of modern cellular systems is delivered in packets, which might occasionally arrive back-to-back, but most of the time are delivered in bursts of packets resulting in periods of very high data rates on the one hand, and periods of zero data transmission on the other hand. At a cellular base station, the traffic burstiness is encountered by means of transmission buffers. It is expected that the potential throughput gains due to deploying dynamic mechanisms are smaller in packet-oriented networks than in streaming systems, where full buffers can be assumed at all times. Existing dynamic OFDMA optimization models do not consider the packet buffering issue, but assign the resources based on average rate requirements of the individual users, and/or certain fairness conditions. Consequently, so far it is not clear, how significant the limiting impacts of discontinuously arriving data are, and whether special packet-aware optimization strategies can compensate these impacts.

This thesis faces this challenge by developing a new dynamic OFDMA resource allocation optimization approach that integrates the resource allocation and packet scheduling decisions per frame. This way, optimal resource/packet matching per frame can be achieved. Furthermore, a new low-complexity packet-aware resource allocation technique is derived from this model. The optimal, as well as the low-complexity heuristic performance are compared to former non-packet-aware, as well as selected packet-aware scheduling approaches.

Challenge 2: the quality and quantity of required control signaling data Neither point-to-point, nor broadcasting systems require control signaling that indicates the terminal for which the data of the current transmission is destined. Today's most widespread wireless LAN and MAN systems realize multi access among users by splitting the available resources in the time domain (in a TDMA fashion), requiring a relatively low amount of control signaling data. In contrast, the control signaling effort in a cellular system featuring many users is relatively high, *i. e.,* it consumes a significant part of the available system resources. The necessary control

signaling data contains information about which terminal is assigned which part of the spectrum and/or which modulation and coding types are used. The overhead, thus, strongly depends on the grade of dynamics provided by the system (*i. e.,* the kind of resources that are dynamically assigned, the resource assignment granularity, and the number of supported modulation types/coding rates). It is mandatory that each terminal present in the system is able to receive and decode the control information, even if it is exposed to strong interference stemming from the surrounding cells. For these reasons, control channel reliability is a key issue in cellular dynamic OFDMA system design, Note that an error on the control channel has a much more significant impact on the data transmission performance than an error on the data channel. However, in current dynamic OFDMA optimization the impact of control signaling is not considered an important factor in the overall system optimization process.

This thesis faces this challenge by first investigating multiple configurations of a standard control channel approach concerning its vulnerability against interference in a cellular scenario, which statically reuses the same control channel resources in each cell – *i. e.,* assuming the worst interference case. The necessary amount of redundancy to reliably transmit the data is determined. Then, an optimization model that trades off user data channel and control channel resources with the goal of optimizing the overall system throughput performance for a dynamically configurable control channel is developed. Two new low-complexity control channel resource allocation techniques are derived from this model, based on dynamic power distribution and dynamic coding, respectively. A performance study shows the benefits of dynamically assigning data *and* control channel resources.

Challenge 3: the particular interference situation of cellular systems Usually, interference mitigation in cellular systems is handled based on standard interference coordination means, such as dividing the overall system bandwidth in frequency bands and distributing the bands among the cells following certain frequency reuse patterns. Interference mitigation in cellular systems is mandatory, as the cellular architecture yields a high number of similar transmission entities (terminals and more importantly base stations) generally working in the same part of the spectrum. As this situation is completely different from the situation in point-to-point or broadcasting systems, the original dynamic OFDM optimization models do not consider the impact of interference at all, and are thus applicable to single isolated cells only. Obviously, by using the frequency reuse techniques mentioned above an artificial isolation of the individual cells can be achieved. Standard frequency reuse schemes, however, are static and mostly adjusted before the system is put into operation, and, thus, known to be inefficient. Motivated by the fact that efficient and reliable control information delivery is possible reusing the same control channel resources in each cell, the question arises whether it is possible to also dynamically reuse the data channel resources in each cell (referred to as frequency reuse 1 scenario). Within such a scenario it is a challenge to integrate the interference coordination task and the resource allocation task at each base station. So far, however, it is not known if any performance gain can be expected, if neighboring cells jointly optimize their usage of transmission resources.

This thesis faces this challenge by first testing existing dynamic OFDM resource allocation approaches in combination with separate standard interference mitigation techniques. In order to compare their performance to an integrated interference cancellation and resource allocation

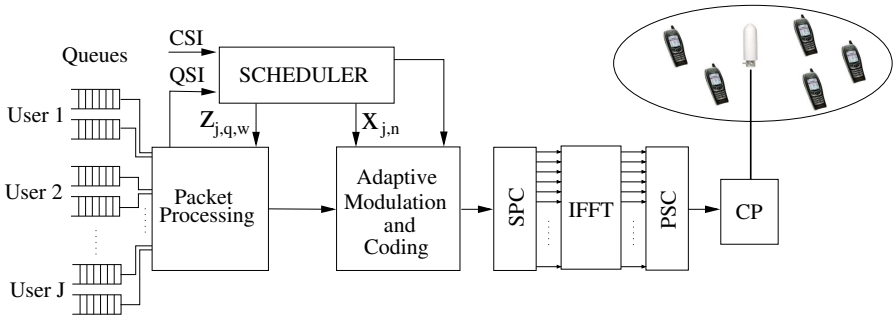


Figure 3.2: the detailed system view featuring several terminals and an OFDMA base station containing a scheduler, a packet processing and an adaptive modulation and coding unit, serial-to-parallel converter (SPC) and parallel-to-serial converter (PSC), as well as a cyclic prefix unit

approach, an according optimization problem is formulated. As the potential gain is shown to be significant, an interference coupling based, low complexity inter base station resource allocation heuristic is developed and shown to provide large gains especially to cell edge users.

By addressing the three challenges introduced above, this thesis contributes efficient means to deploy dynamic OFDMA techniques in cellular systems. While in principle all presented mechanisms are applicable to the up- and down-link direction of such systems, the up-link direction additionally faces significant challenges that are not within the scope of this thesis, such as synchronization issues, CSI signaling, and problems related to the peak-to-average ratio of the up-link transmission signal. By facing the above challenges and providing solutions to the connected problems, however, all significant cellular down-link related dynamic OFDMA research issues are considered in this thesis.

3.2 System Model

Throughout this thesis, the downlink (DL) of a Long Term Evolution (LTE) like cellular system according to [74] with site-to-site distance d_{s2s} is considered. In single cell scenarios, the cell radius r_{cell} is given. Within each cell, a base station coordinates all data transmissions (shown in Figure 3.2). J terminals are constantly moving with speed v.

3.2.1 Wireless channel model

In this section, the wireless channel models, which have been used throughout the thesis' simulations is introduced (a general introduction to wireless channel characteristics is provided in Section 2.2.4). In general, the impact of the wireless channel on the transmitted data is expressed as one signal-to-noise plus interference ratio value $\gamma^{(t)}_{c,j,s}$ per cell c, terminal j, sub-carrier s and time t instance:

$$\gamma^{(t)}_{c,j,s} = \frac{h^{(t)}_{c,j,s}\, p^{(t)}_{c,s}}{\sum\limits_{c_i \neq c} h^{(t)}_{c_i,j,s}\, p^{(t)}_{c_i,s} + \sigma^2_{\eta,s}}\,. \tag{3.1}$$

Here, $h^{(t)}_{c,j,s}$ is the attenuation experienced by terminal j on sub-carrier s at time t towards the base station of cell c, $p^{(t)}_{c,s}$ is the power radiated by the base station of cell c on sub-carrier s at time t, and $\sigma^2_{\eta,s}$ is the noise power present on sub-carrier s.

As described in Section 2.2.1, the attenuation $h^{(t)}_{c,j,s}$ is composed out of the three components *path loss*, *shadowing*, and *fading*. In the following, a description of how these effects are modeled in the simulations relating to this thesis is given. Note that also the co-channel interference term $\sum_{c_i \neq c} h^{(t)}_{c_i,j,s}\, p^{(t)}_{c_i,s}$ depends on these models, since the attenuation between the receiver terminal and all interfering base stations (in cells c_i) is determining the experienced CCI. In contrast, the noise is not explicitly modeled but assumed to be additive white gaussian noise (thermal noise) with a spectral density of -174dBm/Hz. Man made noise is not considered. The attenuation data used to calculate the SNIR at simulation run time is obtained in two generally different ways:

- empirical modeling, or
- measurement data.

While the empirical models are implemented as part of the simulator that delivers the values at simulation run-time, the measurement data is collected in *trace files* and read in by the simulator as necessary.

Empirical modeling

The empirical models for path loss and shadowing in use are the empirical models that have been suggested for LTE macro-cell system simulations by the 3GPP in [22] (Table A.2.1.1-3). Fading is modeled as Rayleigh fading.

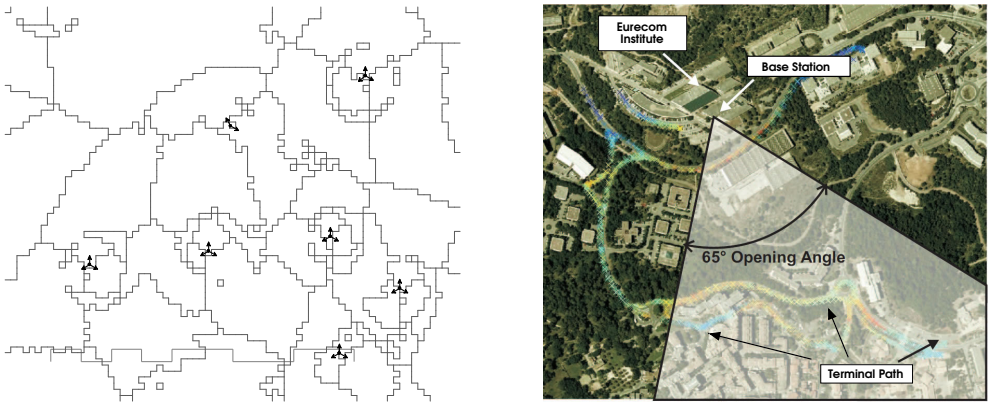
Path loss and shadowing The according path loss model delivers path loss attenuation values in dB depending on the distance between the transmitter and the receiver d:

$$h_{PL}[\text{dB}] = 128.1 + 37.6 \log_{10}(d/\text{kmg}) \tag{3.2}$$

where d is measured in kilometers. A penetration loss value of 20dB is added in accordance to simulation case 1 of Table A.2.1.1-1 of [22]. Also in accordance to [22] (Table A.2.1.1-3), shadowing is modeled as a lognormal random variable (*cf.* Equation (2.4)) with a standard deviation σ_{sh} of 8dB.

Fading The Rayleigh fading model in use follows the sum-of-sinusoids based model that has first been introduced by Clarke [75] and since then has been revised (*e. g.,* [76]), and served as a reference model in numerous works.

In general, the model calculates the signal attenuation due to fading h_{FA} as a function of the relative speed between transmitter and receiver v, the angle between the receiver's direction and the perpendicular to the transmitter φ, and the root mean square delay τ_{rms} among the fading paths p, where P paths are considered. It is assumed that the fading path delays τ_p follow an



(a) A cutout of the Lisbon's UMTS network.

(b) Eurecom's measurement campaign.

Figure 3.3: realistic channel value data acquisition environments

exponential distribution around mean τ_{rms}, and that the individual angles of the fading paths α_p are mutually independent and uniformly distributed over $[-\pi, \pi)$. Then the fading attenuation that impacts the transmitted signal can be computed as follows:

$$h_{FA} = |\hat{G}(t,f)| = \sqrt{G_{re}^2(t,f) + G_{im}^2(t,f)}\ , \tag{3.3}$$

where

$$G_{re}(t,f) \quad = \quad \sum_{p=1}^{P} \varrho_p \cos[2\pi f t \cos(\varphi_- \alpha_p) - \phi_p]\ \text{, and} \tag{3.4}$$

$$G_{im}(t,f) \quad = \quad \sum_{p=1}^{P} \varrho_p \sin[2\pi f t \cos(\varphi_- \alpha_p) - \phi_p]\ . \tag{3.5}$$

The attenuating impact ϱ_p on the amplitude of the wave traveling on path p is approximately a Gaussian random variable and has an expected value of

$$E[\varrho_p^2] = \frac{1}{P} \tag{3.6}$$

Phase shift ϕ_p depends on path delay τ_p and frequency f

$$\phi_p = 2\pi\tau_p f\ . \tag{3.7}$$

According fading values h_{FA} are computed for each sub-carrier in each frame, where the individual sub-carrier frequencies $f = f_c$ and the according frame start times are used as input variables. An exemplary fading profile produced with this model is shown in Figure 2.6.

Trace file data

Path loss and shadowing In the trace file data case, path loss data that has been generated within the EU-project Momentum [77] are used (see Figure 3.3a). The highly detailed data

comprises network configuration information, such as propagation features, environmental, service and traffic information for several real existing urban UMTS networks (Berlin, Lisbon, The Hague). The data are derived from isotropic predictions that are based on the individual cell-site data, such as each cell's 3D physics, the terminal's 3D position, the antenna's azimuth angle, tilting, the maximum output power, etc. Using the collected data, the realistic cell specific data can be computed for an arbitrary number of terminals transmitting and receiving data on either of the UMTS frequency bands.

The path loss data that is used in this thesis has been provided by the authors of [77] as a $C_{\mathrm{mom}} \times J_{\mathrm{mom}}$ matrix of path loss values, where C_{mom} is the number of observed cells and J_{mom} is the number of observed terminals. Hence, one path loss value is available per terminal/base station combination. Note that the provided amount of data is very large such that only a subset $C \times J$ of it is used in a single simulation run. In different simulation runs, however, different sub-sets are randomly and independently chosen, yielding a higher degree of confidence for the presented simulation results.

Fading The trace file OFDMA fading profiles have been measured using the Eurecom MIMO Openair Sounder (EMOS), which is based on Eurecom's Open Air Interface platform [78]. The platform consists of a base station that continuously sends a signaling frame, and one or more terminals that receive the frames to estimate the channel. In Section 3.4 details on the utilized hardware are provided. In a measurement campaign that has been conducted by Kaltenberger *et al.* [79] the correlation and capacity of measured multi-user OFDMA multiple input multiple output (MIMO) channels has been explored. The measurements were conducted outdoors in the vicinity of the Eurecom institute. The scenario is characterized by a semi-urban hilly terrain, composed of small buildings and vegetation with a predominantly present LOS. Figure 3.3b shows a map of the environment.

In the context of these measurements, OFDMA fading profiles of the participating terminals have been stored. The authors of [79] have provided the data in terms of fading matrices (in time and frequency) per terminal. Out of the available number of fading profiles, in the simulation, the individual profile per terminal has been randomly chosen. An analysis of the measured fading profiles is provided in the appendix.

3.2.2 Mobility model

In order to model the terminals' movements, the Manhattan Grid Mobility Model (MGMM) has been chosen. Generally, the MGMM is used to model moving objects in an urban area consisting of a grid of streets. It can be used to investigate the route of buses running through metropolises, as well as for people on shopping trips through department stores, since in both cases all movements are done on a grid topology. Particularly, for the MGMM used in this thesis, the following rules apply:

- at simulation start, the terminals are uniformly distributed over the simulated area
- all nodes move forward only with a constant speed v
- at each cross-roads, the node decides on going right, left or straight ahead with the same probability

- if the cross-roads lies at the border of the simulated area, the directions leading out of the simulated area cannot be chosen

Note that modeling the terminal movements as described above yields a transient terminal distribution. The mobility model is, thus, not valid for simulation runs with long durations. For the relatively short simulations of this thesis that in each scenario utilize the same set of terminal distributions to reproduce identical initial conditions, however, the model can be used. In the case of trace-file based simulation, terminal mobility does not follow the model described above, but is implicitly modeled within the traces.

3.2.3 Physical layer model

The system under consideration uses orthogonal frequency division multiplexing as transmission scheme for downlink data transmission. It features a total bandwidth of B at center frequency f_{c}. The given bandwidth is split into S sub-carriers, each featuring a sub-carrier bandwidth (also referred to as sub-carrier frequency spacing) of

$$\Delta f = \frac{B}{S} \, . \tag{3.8}$$

In order to guarantee orthogonality between the sub-carriers, the symbol length T_{s} is identical for all sub-carriers and is related to the sub-carrier bandwidth as follows:

$$T_{\mathrm{s}} = \frac{1}{\Delta f} = \frac{S}{B} \, . \tag{3.9}$$

A maximum transmit power $p_{\max}$ is split among the sub-carriers. The modulation type and coding rate per sub-carrier are selected by the base station's adaptive coding and modulation unit. Time is slotted into transmission time intervals (TTIs) of duration T_{TTI}, where T_{TTI} is assumed to be smaller than the coherence time of the wireless channel (*i. e.,* in the order of milliseconds). Hence, the smallest *addressable* resource unit has a size of one OFDM symbol length T_{s} in the time domain and one sub-carrier bandwidth Δf in the frequency domain, and is referred to as resource element (RE) (see Figure 3.6). Prior to the transmission of the time domain OFDM symbol, the cyclic prefix (see Section 2.3.1) of length T_{g} is added.

3.2.4 Link layer model

The task of the link layer model in simulations is to translate the channel quality experienced by a receiver into receiver performance in terms of capacity, throughput or error values.

Data channel model

The link layer model applied to the data channel simulations of this thesis follows the link level performance model suggested by the 3GPP for LTE link level simulations [80]. It assumes that the throughput of a modem with adaptive coding and modulation can be approximated by an attenuated and truncated form of the Shannon bound

$$\vartheta_{\mathrm{shannon}}[\mathrm{bps}] = B \log_2(1+\gamma) \, . \tag{3.10}$$

Parameter	Symbol	Value	Notes
attenuation factor	α_{tc}	0.6	represents implementation losses
maximum throughput in bps/Hz	β_{tc}	4.4	based on 64QAM, code rate 4/5
minimum SNIR in dB	$\gamma_{\min}$	-10	Based on QPSK, code rate 1/8

Table 3.1: LTE simulation baseline link level downlink performance parameters [80]

Shannon bound (3.10) represents the maximum throughput that can theoretically be achieved over an AWGN channel for a given SNIR [81]. It is known to be non-achievable by practical wireless transmission systems. Still, it is a very common means to evaluate the theoretical upper bound of transmission systems that are under development. A more realistic model for throughput estimation is given by the *truncated Shannon* model given in Equation (3.11). It approximates the throughput ϑ in bits-per-second (bps) over a channel with a given SNIR γ, when using link adaptation:

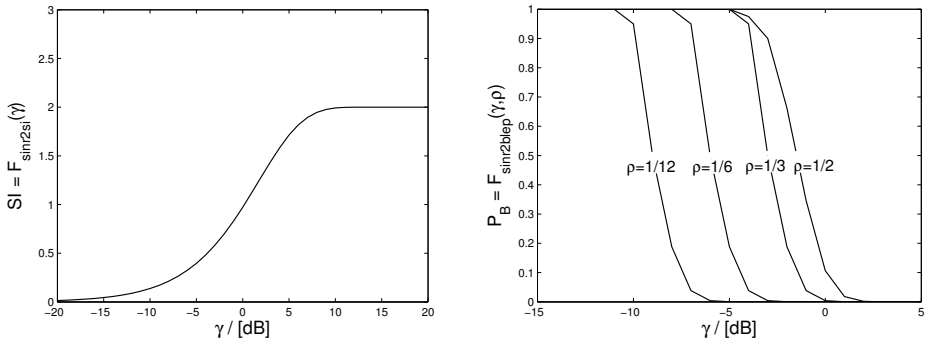
$$\vartheta_{\mathrm{tc}}[\mathrm{bps}] = \begin{cases} 0 & \text{for } \gamma < \gamma_{\min} \\ \alpha_{\mathrm{tc}} B \log_2(1+\gamma) & \text{for } \gamma_{\min} < \gamma < \gamma_{\max} \\ \beta_{\mathrm{tc}} B & \text{for } \gamma > \gamma_{\max} \end{cases}, \tag{3.11}$$

where B is the channel bandwidth, α_{tc} is the attenuation factor, representing implementation losses, β_{tc} is the maximum throughput of the adaptive coding and modulation codeset in bps/Hz, and $\gamma_{\min}$ and $\gamma_{\max}$ in dB are the minimum SNIR of the codeset, and the SNIR at which the maximum throughput is reached respectively. Parameters α_{tc}, β_{tc}, $\gamma_{\min}$, and $\gamma_{\max}$ can be chosen to represent different modem implementations and link conditions. The parameters proposed in Table 3.1 represent a baseline case, which assumes:

- link adaptation (see Table 3.1 for details of highest and lowest rate codes),
- turbo coding,
- channel prediction, and
- hybrid automatic repeat request (HARQ).

Control channel model

In general, the data channel model introduced in the last section could also be used for the control channel provided that a different parameter set is selected (since the control channel doesn't use turbo codes and has a smaller set of available modulation/coding combinations). However, as the block error probability (BLEP) subject to the choice of a particular coding scheme is of major interest for the following control channel observations, a slightly more detailed model is necessary. In order to estimate the BLEP as a function of the link quality, the mutual information effective SNIR metric (MIESM) is used, which has been shown to achieve very high BLEP prediction accuracy in OFDM systems in [82].



(a) $F_{\mathrm{sinr2si}}(\gamma)$ (b) $F_{\mathrm{sinr2blep}}(\gamma, \rho)$

Figure 3.4: link level performance functions for code rates $\rho = \{\frac{1}{12}, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}\}$

The mutual information effective SNIR metric involves two steps of transforming the experienced SNIR per control channel sub-carrier. First, the *effective* SNIR per terminal is calculated

$$\gamma_{\mathrm{eff},j}^{(t)} = F_{\mathrm{sinr2si}}^{-1}\left(\sum_{s=1}^{S} x_{c,j,s}^{(t)} F_{\mathrm{sinr2si}}(\gamma_{j,s}^{(t)})\right) , \tag{3.12}$$

where F_{sinr2si} (*cf.* Figure 3.4a) is a function that maps SNIR values to the related symbol-information value, and $x_{c,j,s}^{(t)}$ is the binary assignment variable that determines the terminal/sub-carrier allocations (*cf.* Equation (2.22)) in cell c. In the second step, the block error probability for this terminal is determined by using $F_{\mathrm{sinr2blep}}(\gamma_{\mathrm{eff},j}^{(t)}, \rho)$ (*cf.* Figure 3.4b), where ρ is the per-terminal control channel code-rate. Note that if there is no set of values for a certain code-rate ρ, the set of the closest available code-rate is chosen. During simulation run-time, the error probability is translated into an error rate by means of random number drawing. In case of a control channel error, all related user data on the data channel is lost.

Traffic models

Three qualitatively different traffic models are assumed for the investigations of this thesis: a *full buffer* model, an *Internet traffic* model, as well as a *voice traffic* model.

Full buffer model In general, the full buffer model assumes infinite transmission buffers that are filled up at any time for each of the participating entities. Particularly, in a cellular downlink scenario this translates into a situation, where the base station has data available for each terminal in each TTI. Consequently, whenever the base station assigns transmission resources to a terminal, these resources can be fully utilized. This full buffer model is commonly used for average system throughput analysis, but, obviously, is not applicable to more specific QoS aware (in terms of delay, jitter, PESQ,...) packet scheduling approach investigations. For the latter, the following packet specific traffic models are used.

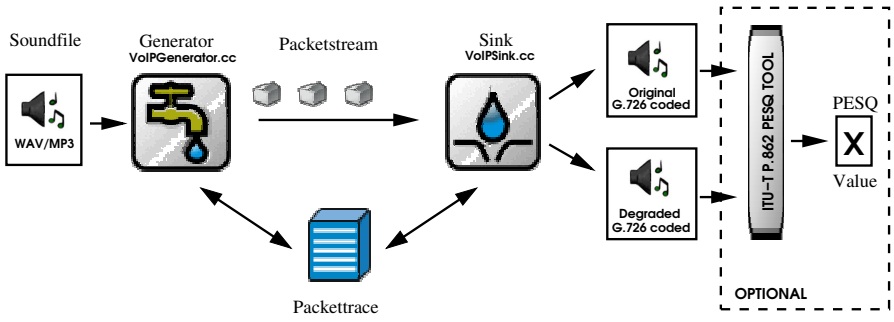


Figure 3.5: For VoIP traffic simulation and evaluation a particular VoIP packet generator, and a VoIP sink are utilized. Optionally, the transmission results can be used as input for ITU-T's PESQ tool to obtain the according MOS value.

Internet traffic model General Internet traffic arriving at the base stations per terminal is modeled using the Poisson Pareto burst process (PPBP) model suggested in [83]. The PPBP is a simple but accurate model for Internet traffic whose general idea is based on the observation that a significant part of the self-similar Internet traffic in broadband networks is supplied by a range of bursty traffic sources. In [84] it was shown that one possible source of this burstiness was in the aggregation of independent on-off sources with heavy tailed 'on time' and/or 'off time' distributions. The PPBP consists of formulas that provide an according traffic pattern.

In particular, the arrival of packet bursts at the base station is a Poison process with rate λ, so the intervals between adjacent burst arrival times are negative exponentially distributed with mean $1/\lambda$, and the mean number of new bursts arriving during one down-link frame of length T_{f} is Poisson distributed with mean λT_{f}. Accordingly, the number of concurrently active bursts is Poisson-distributed, with mean λ times the mean burst duration [85]. Each arriving burst is destined for one of the available terminals j with equal probability. Depending on the burst duration d_{burst} an according number of equally-sized packets is added to j's queue. In the PPBP, the burst durations d_{burst} are independent and identically distributed Pareto random variables:

$$\Pr(d_{\mathrm{burst}} > x) = \begin{cases} (\frac{x}{\delta})^{-\xi}, & \text{if } x \geq \delta \\ 1, & \text{otherwise,} \end{cases} \tag{3.13}$$

with burst duration $\delta > 0$. For $1 < \xi < 2$, it holds that $\mathrm{E}(d_{\mathrm{burst}}) = \frac{\delta\xi}{\xi-1}$ and the variance of d_{burst} is infinite. Throughout the thesis for the burst arrival process $\lambda = 0.005$, and for the burst duration $\delta = 1$ and $\xi = 1.5$ are chosen. This parameterization allows for high system utilization, such that the system is working to full capacity in more than 50% of the simulated cases.

Voice traffic model VoIP traffic is modeled separately from the general Internet traffic for three reasons: (1) the transmission of voice traffic is still the major application in cellular networks and should thus be investigated individually, (2) VoIP with its particular QoS requirements (throughput, delay, jitter) is used as representative for differentiated QoS traffic, which might be prioritized by the scheduler, and (3) by using real voice wave files and ITU-T's perceptual eval-

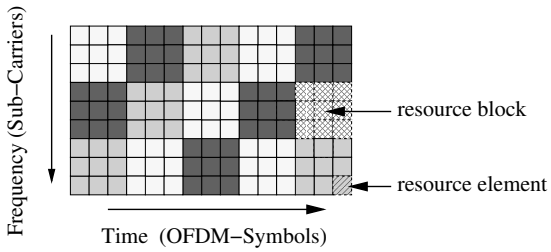


Figure 3.6: example RBs consisting of $3x3 = 9$ REs

uation of speech quality (PESQ) tool [86], the obtained mean opinion score (MOS) [87] can be used as an additional meaningful metric to evaluate system performance.

In general, the appearance of VoIP packets within the simulations of this thesis is not based on a mathematical model, but following the approach of [88], it relies on real VoIP stream data. The according VoIP traffic generator consists of a packet generator and a sink (see Figure 3.5). Roughly speaking, the packet generator first creates a stream of 64 samples comprising VoIP packets out of an arbitrary sound file using an appropriate VoIP codec. Throughout this thesis VoIP standard codec G.726 [89] is used (as suggested in [88]). Depending on the individual maximum sample amplitude, for each packet the packet generator decides on whether a VoIP packet or a (shorter) silence packet should be sent. The packet creation information is stored in a packet trace list. Each packet that arrives at the sink is first checked for errors and then its status and its arrival time (*i. e.,* delay information) are logged in the packet trace. Once a VoIP session is finished, the received wave file is created using the packet trace and the original sound samples. Using both, the original and the received (impacted) wave file, the MOS value that quantifies the quality of the particular VoIP session is derived using ITU-T's PESQ tool.

Within the simulations of this thesis, there is a single VoIP generator and sink pair per active terminal. The VoIP calls start randomly within the simulation time and the call lengths are normally distributed with a mean of $5s$ and a standard deviation of $2s$. The relatively low mean time allows for several connections per terminal and simulation run. Per simulation run, each generator constructs VoIP packets out of the same wave file. Hence, ideally, each terminal should experience the same sound quality. The wave files in use contain the sound information of one side (*i. e.,* solely the down-link data) of a regular telephone call, characterized by talk spurts and silence periods.

3.2.5 Medium access layer model

Data multiplexing in the downlink is based on orthogonal frequency division multiple access, where the smallest *allocatable* resource unit is a *resource block* n (*cf.* Figure 3.6). A resource block consists of a carefully defined number of adjacent REs in the frequency domain (sub-carriers) and spans all REs available for user data transmission of a TTI [22]. The resource block bandwidth (number of REs in the frequency domain) is chosen, such that channel quality differences among the sub-carriers belonging to a single resource block is negligible small (*i. e.,* the resource block is assumed to experience mostly flat fading). If N_{subs} is the number of

adjacent sub-carriers (REs in the frequency domain) belonging to a single resource block n, and S is the number of sub-carriers in the system, the number of available resource blocks to transmit data per TTI is

$$N = \frac{S}{N_{\text{subs}}} \,. \tag{3.14}$$

Scheduling decisions

At the beginning of each TTI, a base station scheduler (as shown Figure 3.2) assigns resource blocks to the terminals in each cell c. The distribution of the resource blocks among the terminals has a significant impact in the system performance. In this thesis, different sophisticated dynamic OFDMA scheduling strategies are explored, whereas a simple round robin (RR) scheduler serves as a lower bound reference.

It is also the scheduler's task to distribute the maximum available transmission power $p_{\max}$ among the resource blocks. Within each resource block n all sub-carriers s obtain the same power share. Moreover, the scheduler applies adaptive coding and modulation with resource block granularity based on the available channel state information values. The achievable throughput per resource block is calculated according to Equation (3.11) of the link layer model introduced in Section 3.2.4.

CSI availability

At the beginning of each TTI a snapshot of the current channel state information values per terminal and base station is taken. Throughout this thesis it is assumed that the CSI measurement is not subject to any measurement errors. The snapshot values are used as input for the scheduler. In terms of CSI availability, two generally different approaches are considered:

1. CSI is available without any delay, *i. e.,* the scheduler is assumed to be able to adapt to channel changes immediately, or

2. the scheduler works with delayed and thus maybe outdated channel measurements.

While the first one accounts for an ideal channel knowledge scenario, the second simulates the case of outdated channel information. Note that the latter is a worst case estimation, as the option of using promising channel prediction techniques (*e. g.,* [90–92]) is abandoned. Note that the advantage of a highly accurate adaptation to the channel state in the optimal world might turn into a disadvantage in the real world. Therefore, the first assumption is used whenever general statements about the potential of a mechanism is needed. If a mechanism is shown to have a potential for performance improvement under the idealistic assumptions, the impact of CSI processing delay ΔT_{csi} on the performance of the dynamic sub-carrier and power allocation is studied utilizing the second approach. In that case different time spans between the measurement of CSI and its actual usage for channel adaptation are considered. Hence, instead of exactly knowing which throughput is achievable by choosing certain terminal/resource combinations (as in the ideal case), the scheduler has to base its scheduling decisions for resource block n on each terminal j's *expected throughput* values

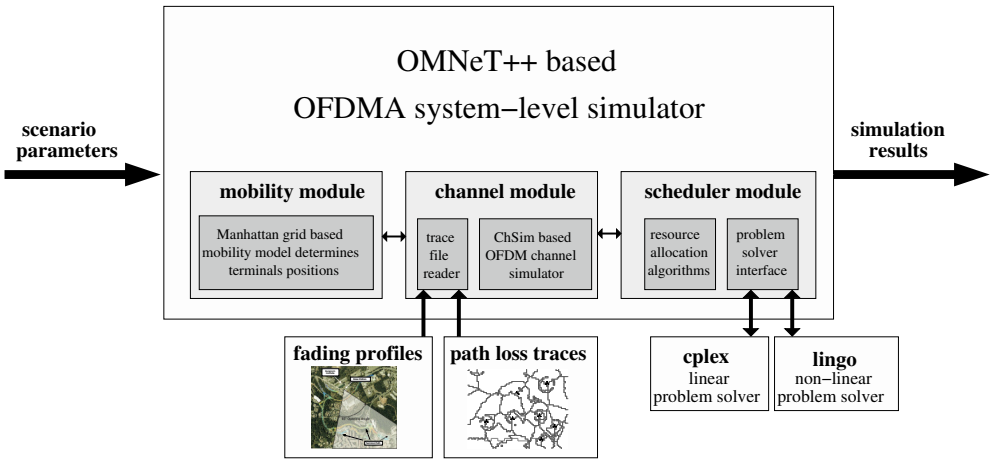


Figure 3.7: the thesis' methodology and simulation framework

Parameter	**Symbol**	**Value**
center frequency	f_{c}	2GHz
system bandwidth	B	5MHz
resource block frequency spacing	Δf_n	200 kHz
number of resource blocks	N	25
sub-carriers per resource block	N_{subs}	12
OFDM symbols per resource block	N_{symbs}	7
TTI duration	T_{TTI}	1ms
OFDM symbol length	T_{s}	60μs
OFDM guard length	T_{g}	11.4μs
maximum transmission power per cell	$p_{\max}$	43 dBm

Table 3.2: OFDMA system simulation parameters

$$\hat{\vartheta}_{j,n}^{(t)} = F\left(\hat{\gamma}_{\check{c},j,n}^{(t)}\right) , \tag{3.15}$$

where $\hat{\gamma}_{\check{c},j,n}^{(t)}$ is terminal j's *expected SNIR*:

$$\hat{\gamma}_{\check{c},j,n}^{(t)} = \frac{p_{\check{c},n}^{(t)}\, h_{\check{c},j,n}^{(t-\Delta T_{\mathrm{csi}})}}{\sum\limits_{\forall c \neq \check{c}} p_{c,n}^{(t)}\, h_{c,j,n}^{(t-\Delta T_{\mathrm{csi}})} + \sigma_\eta^2} . \tag{3.16}$$

Recall that $F(\dots)$ is a function that determines the throughput depending on the SNIR.

Parameter	Symbol	Value
number of cells	C	1
number of terminals	J	10
antenna type		omni directional
cell geometry		hexagonal
path loss / shadowing		empirical model
fading		Rayleigh/Clarke model
link-to-system function	$F(\dots)$	truncated Shannon
cell-radius	r_{cell}	500m
terminal speed	v	$\{0; 10\}$ m/s
mobility model		Manhattan grid
CSI processing delay	ΔT_{csi}	$\{0 \dots 5\}$ TTIs

Table 3.3: basic single cell reference model parameters

3.3 Methodology and Simulation Setup

The performance of the system introduced in Section 3.2 is evaluated by means of simulation. The C++ based timed discrete event simulator OMNeT++ [93] is used to implement the base station resource allocation mechanisms and packet scheduling functionality (*scheduler module*), the terminal movement and data processing (*mobility module*), and the wireless channel behavior (*channel module*)[1]. Alternatively, the wireless channel behavior is determined by trace files that are loaded during the simulation run-time. Throughout the thesis, linear, as well as non-linear resource allocation optimization problems are formulated. To solve these problems, special optimization problem solving software is integrated into the simulation framework. For the linear and non-linear program optimization approaches (*e. g.,* the constraint max sum rate problems (4.12), or (6.6)), the simulator creates one global scheduling problem instance per TTI, or one local instance *per cell and TTI*. The linear scheduling optimization problem instances are piped into ILOG's CPLEX linear problem solver [94]. Non-linear optimization problem instances are piped into LINDO's LINGO non-linear optimization problem solver [95]. After solving the individual instances, the solving software pipes the optimal scheduling decisions back into the simulator that uses it for further processing and throughput calculations. The overall simulation framework is depicted in Figure 3.7.

The general simulation parameterization largely follows simulation case 1 for LTE system simulation as presented in Tables A.2.1.1-1 and A.2.1.1-2 of [22]. The parameters are summarized in Table 3.2.

[1] for the according empirical channel models in use and information about the acquisition of the trace file channel data see Section 3.2.1

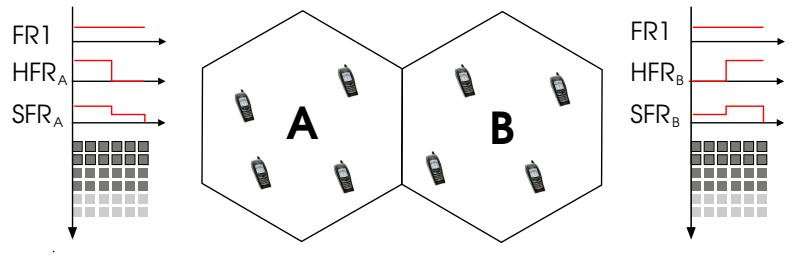


Figure 3.8: reference scenario for the basic multi-cell/CCI investigations

Parameter	**Symbol**	
number of cells	C	2
number of terminals	J	8
antenna type		omni directional
cell geometry		hexagonal
path loss / shadowing		empirical model
fading		Rayleigh/Clarke model
link-to-system function	$F(\dots)$	Shannon
inter-site distance	d_{s2s}	500m
max. object speed	v	3 m/s
req. throughput per terminal	$\vartheta_{\mathrm{req},j}$	$\{1 \dots 5\}$ Mbps

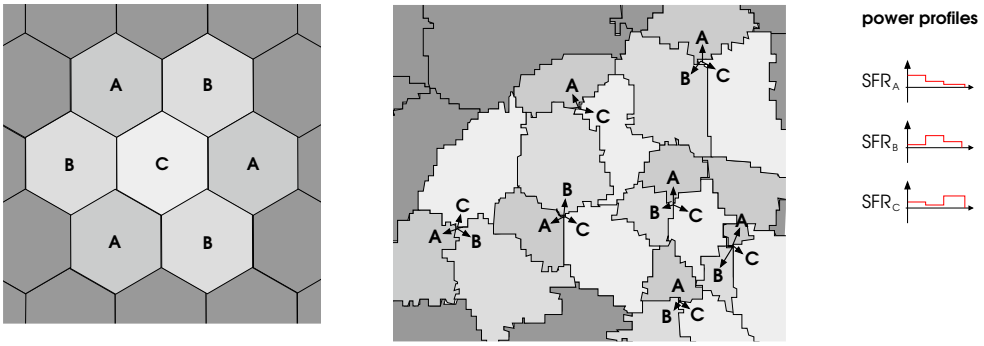
Table 3.4: basic multi cell reference model parameters

3.3.1 Single-cell reference scenario

For the inter-terminal fairness and packet scheduling explorations of Chapter 4, a single cell reference model featuring one base station (containing one scheduler) and J connected terminals in one hexagonal cell is considered (*cf.* Figure 3.2). The cell radius is set to 500 m. Terminals are uniformly distributed over the cell area. The truncated Shannon link level performance model (*cf.* Section 3.11) is used to translate channel quality into throughput values (*i. e.,* $F(\dots)$ is the truncated Shannon function (3.11)). The single cell reference parameterization can be found in Table 3.3.

3.3.2 Basic multi-cell reference scenario

The basic multi-cell reference scenario is shown in Figure 3.8. In order to achieve the theoretical optimum and refrain from clipping effects that might appear when using the truncated Shannon link level performance model, it adopts the Shannon capacity formula (3.10) as link-to-system



(a) hexagonal 7-cell layout (b) cutout of Lisbon's UMTS network, profiled by the momentum project (*cf.* Section 3.2.1) (c)

Figure 3.9: extended and trace based multi-cell reference scenarios with SFR power profiling

function $F()$ for the basic multi-cell investigations of Section 6.3. Two hexagonal cells are considered. The reason for this limitation lies in the fact that for significantly larger scenarios the non-linear *global constrained max sum rate* problem (6.6) is not solvable within reasonable time bounds using standard hard- and software equipment. Eight stationary terminals, four in each cell, are uniformly distributed over the cell area. Frequency- and time-selective fading is present due to objects moving at speed v. All basic multi-cell reference model parameters are summarized in Table 3.4.

3.3.3 Extended Multi-Cell Reference Scenarios

The less complex linear optimization and heuristic CCI aware resource allocation approaches of Chapter 6 are studied in two qualitatively different extended multi-cell reference scenarios: an idealized hexagonal, and a real-world trace based scenario.

Hexagonal scenario The extended hexagonal scenario is shown in Fig. 3.9a. 70 terminals are uniformly distributed over the system area that consists of seven cells at an inter-site distance of 500 m. It adopts the truncated Shannon link level performance model. Additional parameters are summarized in Table 3.5.

Trace based scenario In a second step, the hexagonal geometry is replaced by a realistic network layout and an even larger number of terminals in the second scenario (Fig. 3.9b). Moreover, while in the first scenario the wireless channel behavior is modeled, wireless channel measurement data and fading profile traces are used in the second scenario. The reason for doing the evaluation in the two-tier fashion is to segregate the effects of an increased network size from the effects that come with a realistic network layout. The particular parameterization values are listed in Table 3.5. An analysis of the measured fading profiles is provided in the appendix.

Parameter	Symbol	hexagonal	trace-based
number of cells	C	7,9	20
number of terminals	J	70,45	200
antenna type		omni directional	120° sectors
cell geometry		hexagonal	real-world
path loss / shadowing		empirical model	Momentum data
fading		Rayleigh/Clarke model	OpenAirInteface data
link-to-system function	$F(\dots)$	truncated Shannon	
inter-site distance	d_{s2s}	500m	irregular
terminal speed	v	$\{0; 10\}$ m/s	varying 0-10 m/s
mobility model		Manhattan grid	real pattern
max per terminal throughput	$\vartheta_{\mathrm{req},j}$	1 Mbps	
CSI processing delay	ΔT_{csi}	$\{0 \dots 5\}$ TTIs	

Table 3.5: extended hexagonal and trace based multi-cell reference model parameters

3.3.4 Performance metrics

The performance of a simulated system is measured in:

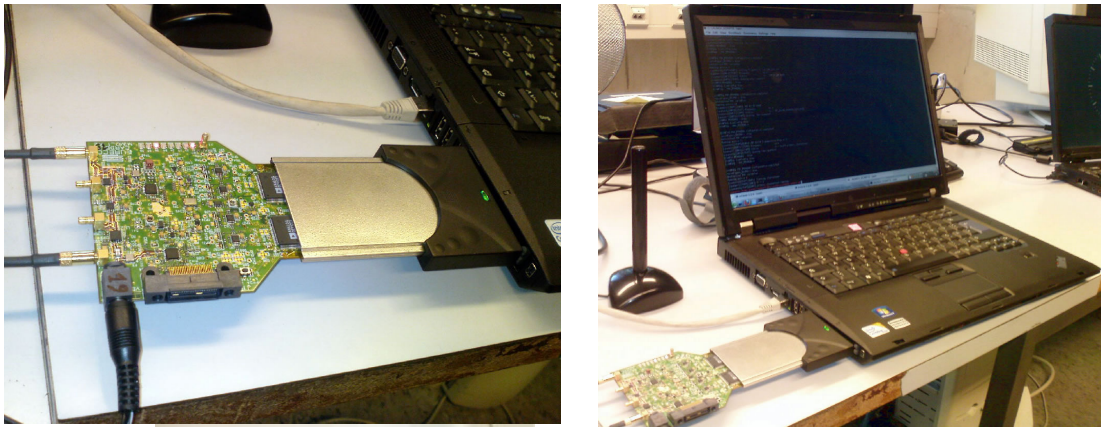
- *spectral efficiency*: the mean system throughput in bits/s related to the constrained system bandwidth in 1/s;
- *fairness*: the span between the mean throughput of the weakest terminal and the strongest terminal in bits/s; and (in the case of voice over IP traffic)
- *speech quality*: the perceptual evaluation of speech quality (PESQ) metric.

3.4 OFDMA Hardware

To acquire the traces used in the trace-based simulation scenario, Eurecom's CardBus MIMO I software defined radio (SDR) platform, which is part of the OpenAir Interface project, has been used. Eurecom equips these radios with a complete OFDMA stack and an integrated OFDM channel sounder. The following information has been extracted out of [96].

Eurecom's CardBus MIMO I CardBus MIMO is a dual-RF (dual-frequency or dual-antenna) system with both data acquisition and DSP functionality in a very small form-factor (PCMCIA/CardBus). A complete OpenAir Interface modem (5 MHz TDD two-way 2x2 MIMO-OFDM) runs on the card where the PC hosts the PHY and MAC-layers in the kernel space of a real time enabled Linux kernel.

EMOS channel sounder The Eurecom MIMO Openair Sounder is a part of the openairinterface platform which allows real-time MIMO channel measurements synchronously over multi-



(a) Eurecom's CardBus MIMO card

(b) laptop with CardBus MIMO card and antenna

Figure 3.10: Eurecom's openairinterface platform: EMOS trace-file acquisition setup

ple users moving at vehicular speed. EMOS consists of a base station that continously sends out a signalling frame, and one or more users that receive the frames to estimate the channel.

To generate the traces used in the trace based reference scenario (*cf.* Section 3.3.3), for the base station a workstation with four PLATON data acquisition cards is employed. The PLATON boards are used along with a Powerwave 3G broadband antenna (part no. 7760.00) composed of four elements which are arranged in two cross-polarized pairs. The terminals consist of a laptop computer with the CardBus MIMO I data acquisition card (see above) and two clip-on 3G Panorama Antennas (part no. TCLIP-DE3G).

Sounding signal

The EMOS is using an OFDM modulated sounding sequence. The duration of one transmit frame is 2.667 ms and it consists of a synchronization symbol (SCH), a broadcast data channel (BCH) comprising 7 OFDM symbols, a guard interval, and 48 pilot symbols used for channel estimation. The pilot symbols are taken from a pseudo-random QPSK sequence defined in the frequency domain. The subcarriers of the pilot symbols are multiplexed over the four transmit antennas to ensure orthogonality in the spatial domain. The BCH contains the frame number of the transmitted frame that is used for synchronization among the terminals.

Channel estimation procedure

Each terminal first synchronizes to the base station using the SCH. It then tries to decode the data in the BCH. If the BCH can be decoded successfully, the channel estimation procedure is started. The channel estimation procedure consists of two steps. Firstly, the pilot symbols are derotated with respect to the first pilot symbol to reduce the phase-shift noise generated by the dual-RF CardBus/PCMCIA card. Secondly, the pilot symbols are averaged to increase the measurement SNR. The estimated MIMO channel is finally stored to disk.

Multi-user measurement procedure

In order to conduct multi-user measurements, all the terminals need to be frame-synchronized to the base station. This is achieved by storing the frame number encoded in the BCH along with the measured channel at the terminals. This way, the measured channels can be aligned for later evaluation. The frame number is also used to synchronize the data acquisition between terminals. One measurement run (file) starts every 22.500 frames (60 sec) and is exactly 18.750 frames (50 sec) long.

Chapter 4

Packet Traffic aware Resource Allocation

THE DATA destined for users of modern cellular systems is delivered in packets, which might occasionally arrive back-to-back, but most of the time are delivered in bursts of packets featuring periods of very high data rates on the one hand, and periods of zero data transmission on the other hand. Obviously, this instance yields to situations, where in a certain cell there is no data to be transmitted to the best situated user, and, thus, multi-user diversity cannot fully be exploited. Consequently, it is expected that the potential of throughput gains due to deploying dynamic mechanisms is smaller in packet-oriented networks than in streaming systems, where full buffers can be assumed at all times. Moreover, large amounts of transmission resources might be wasted, if a subset of sub-channels is assigned to a terminal, for which only one small packet is enqueued. Consequently, this chapter is devoted to answering the question:

> *Are special packet-aware resource allocation techniques necessary in packet-oriented OFDMA systems?*

So far, mathematical models to solve the down-link resource assignment optimization problem do not consider packet buffering, *i. e.,* they do not match the resources assigned to a single terminal to the data to be transmitted to it. There are heuristics that influence standard optimization problems by excluding terminals with empty buffers and adding weights according to the buffer fill levels. It is, however, not clear, how close the performance of the latter approach optimality, and whether specialized packet-aware assignment strategies would be much more complex and perform significantly better in terms of

1. system throughput,
2. individual per terminal throughput, and
3. individual per terminal experienced QoS (*e. g.,* in terms of MOS values).

This chapter is concerned with developing packet traffic aware resource allocation mechanisms that aim in optimizing the system performance in terms of these three metrics. Note that there is

a conflict in achieving the first and the latter two metrics, as optimizing the system throughput in most cases yields to scenarios with little *inter terminal fairness*, *i. e.,* to situations, where good situated receive large parts, whereas bad situated are assigned small parts of the resources (*cf.* Section 2.4.2). Achieving fairness among terminals, however, is a major design issue in cellular network design. Thus, in this chapter, first, existing fairness and rate control strategies are reviewed and evaluated in an exemplary OFDMA based cellular scenario, in which continuous data streams per terminal are presented. Then, in a second step, the strategies are adopted in a packet-oriented cellular system setup and its shortcomings are determined. In this context, packet-traffic aware mathematical optimization models are derived from the basic models introduced in Chapter 2. The optimal results are compared to the performance results of existing packet scheduling heuristics, and of a new combined resource and packet scheduling algorithm, which is derived from the packet-aware optimization problem.

4.1 Fairness

The issue of providing fairness among terminals has already been identified as a major challenge in OFDMA point-to-multi-point transmissions in Section 2.4.2. Two resource scheduling alternatives to provide fairness have been introduced: the *rate adaptive* and the *margin adaptive* optimization strategy. Recall that while the rate adaptive approach (2.24) generally tries to maximize the throughput constraint to a certain power budget, the margin adaptive approach (2.25) aims at minimizing the power that is needed to serve a requested data rate. In the context of cellular networks, the rate adaptive optimization approach is the more interesting approach. The reason for that lies in the fact that in a typical cellular scenario, there is a limited power budget per base station (*e. g.,* 43dBm in a 5MHz LTE system [80]), while there is an ever increasing need for high data rates. The rate adaptive approach is also referred to as *maximize the minimum* (or short *max-min*) throughput approach, as it aims in providing all users with the same data rate ϵ. Obviously, the weakest terminal determines this rate. This circumstance yields situations, in which the overall system throughput is highly impacted to the negative by the weakest terminal.

A well received alternative approach to provide system fairness in such a scenario has initially been described by Kelly *et al.* [97, 98], and later extended by several other authors [71, 99, 100]: users are associated with *utility functions* that tell about the usefulness of a resource q to a user j in a network with a limited set or resources Q. Assume that $\mathbf{C}_q$ is the finite capacity of resource $q \in Q$, vector $\mathbf{C} = (\mathbf{C}_q, q \in Q)$, and allocation $a_{j,q} \in \{0, 1\}$ is a binary assignment of resource q to user j with matrix $\mathbf{A} = \{a_{j,q}\}$. If a rate r_j is allocated to user j, than this has utility $U_j(r_j)$ to user j, and if furthermore $U_j(r_j)$ is an increasing, strictly concave and continuously differentiable function of r_j, and the utilities are additive, such that the aggregate utility of rates $\mathbf{r} = (r_j, j \in J)$ is $\sum_{j \in J} U_j(r_j)$, the system optimal rates solve the following problem:

$$\max_{\mathbf{A}} \sum_j U_j(r_j) \tag{4.1a}$$

$$\text{s. t.} \quad \mathbf{Ar} \leq \mathbf{C} \tag{4.1b}$$

$$\text{over} \quad \mathbf{r} \geq \mathbf{0} \tag{4.1c}$$

This optimization problem is mathematically fairly tractable with a strictly concave objective function and a convex feasible region. In general, however, the problem involves utilities that are unlikely to be known by the network [98]. For a comprehensive discussion on how to obtain and use utility functions see [101]. Instead of defining and distributing utility functions for all kinds of applications and usage scenarios in the network, Kelly [97] suggests a game theoretic fairness definition with respect to user utility: the so called *proportional fairness*. Referring to Equation (4.1), a vector of rates $\mathbf{r}$ is *proportionally fair*, if it is feasible, that is $\mathbf{r} \geq \mathbf{0}$ and $\mathbf{Ar} \leq \mathbf{C}$, and if for any other feasible vector $\mathbf{r}^*$, the aggregate of proportional changes is zero or negative:

$$\sum_{\forall j} \frac{r_j^* - r_j}{r_j} \leq 0\,. \tag{4.2}$$

Besides the *max min* fairness (2.24), and *individual user utility* fairness (4.1) approaches, Kelly's proportional fairness definition has widely been accepted as one of the three most important fairness definitions in the context of cellular networks. Due to its robustness against outliers and its easy applicability (no prior knowledge of utility functions necessary), it has the most practical relevance among the three.

4.1.1 Proportional fair multi-carrier scheduling

Kim *et al.* [102] have reformulated Kelly's definition and applied it to a scheduler of a multi-carrier network: a scheduling P^* is proportional fair (PF) if and only if for any feasible scheduling P we have:

$$\sum_{\forall j} \frac{\bar{\vartheta}_{j,k}^{(P^*)} - \bar{\vartheta}_{j,k}^{(P)}}{\bar{\vartheta}_{j,k}^{(P)}} \leq 0\,. \tag{4.3}$$

where $\bar{\vartheta}_{j,k}^{(P)}$ is the average throughput of user j over the last k scheduling epochs, if scheduling P is selected. In other words, any positive change of a user in the allocation results in a negative average change for the system. It is known from [97] that this condition is met, if the scheduling maximizes the sum of logarithmic average user throughputs, which can be formally expressed as

$$P^* = \arg\max_P \sum_{\forall j} \log \bar{\vartheta}_{j,k}^{(P)}\,. \tag{4.4}$$

In [102] Kim *et al.* prove that in the context of a *multicarrier system* with S carriers, this is congruent with the following claim: a scheduler is proportionally fair in a multicarrier system, if and only if at scheduling epoch t it satisfies:

$$P^* = \arg\max_P \prod_{\forall j} \left(1 + \frac{\sum_{\forall s} a_{j,s}^{(P)} F\left(\gamma_{j,s}^{(t)}\right)}{(k-1)\bar{\vartheta}_{j,k}^{(t-1)}} \right)\,, \tag{4.5}$$

where $a_{j,s}^{(P)}$ is the binary allocation variable that is one if scheduling P assigns sub-carrier s to user j at scheduling epoch t, and zero else, $F(...)$ is a function that delivers the instantaneous throughput achievable by user j on sub-carrier s, if it experiences SNIR $\gamma_{j,s}^{(t)}$ on s, and $\bar{\vartheta}_{j,k}^{(t-1)}$ is the average throughput of user j over the last k scheduling epochs, computed at the previous scheduling epoch.

4.1.2 Optimization model

Slightly modifying Kim's multi-carrier proportional fair optimization formulation (4.5) yields a binary integer problem formulation of the kind that delivers terminal/sub-carrier assignments as optimization result:

$$\max_{\mathbf{X}^{(t)}} \prod_{\forall j} \left(1 + \frac{\sum\limits_{\forall s} x_{j,s}^{(t)} F\left(\gamma_{j,s}^{(t)}\right)}{(k-1)\bar{\vartheta}_{j,k}^{(t-1)}} \right) \tag{4.6a}$$

$$\text{s.t.} \sum_{\forall j} x_{j,s}^{(t)} \leq 1 \qquad \forall s \tag{4.6b}$$

where constraint (4.6b) is the *disjunctive sets* constraint already known from the *max sum rate* optimization problem formulation (2.22). Solving *multi carrier proportional fair* problem (4.6) yields terminal/sub-carrier assignments that comply with condition (4.3). It is, however, a highly complex approach, as the number of necessary comparisons for an exhaustive serach to find the optimal allocation is J^S. Moreover, as problem (4.6) contains a product over all user allocations, it is a *non-linear* problem, and, thus, not solvable with standard linear optimization problem solving techniques.

To counter the problem of high complexity, Kim *et al.* [102] propose two less complex alternatives. The first one selects only one user at each scheduling epoch, omitting the exploitation of frequency diversity over the sub-carriers. The according scheduling rule is written as

$$P^* = \arg\max_{\forall j} \left(\frac{\sum\limits_{\forall s} a_{j,s}^{(P)} F\left(\gamma_{j,s}^{(t)}\right)}{(k-1)\bar{\vartheta}_{j,k}^{(t-1)}} \right) . \tag{4.7}$$

From the single user proportional fair scheduling formulation (4.7), a linear binary integer problem to achieve optimal terminal/sub-carrier assignments can be derived:

$$\max_{\mathbf{X}^{(t)}} \sum_{\forall j} \left(\frac{\sum\limits_{\forall s} x_{j,s}^{(t)} F\left(\gamma_{j,s}^{(t)}\right)}{(k-1)\bar{\vartheta}_{j,k}^{(t-1)}} \right) \tag{4.8a}$$

$$\text{s.t.} \quad \sum_{\forall j} x_{j,s}^{(t)} \leq 1 \qquad \forall s \tag{4.8b}$$

$$\sum_{\forall s} x_{j,s}^{(t)} = S \vee 0 \qquad \forall j \tag{4.8c}$$

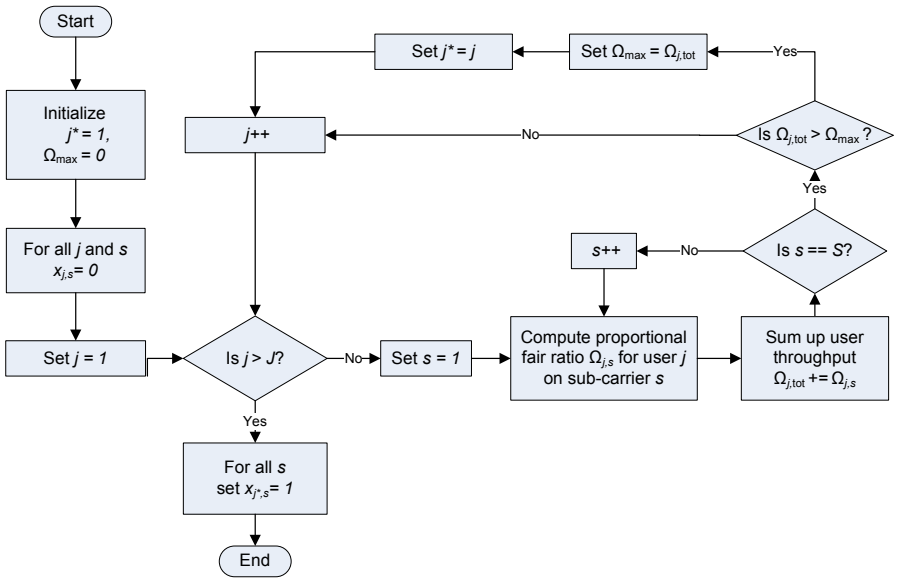


Figure 4.1: algorithm implementing Kim's *single user proportional fair* approach

Problem (4.8) can easily be solved by standard linear integer program solvers. Kim *et al.*'s second approach takes frequency diversity into account by considering *carrier-wise* proportional fairness:

$$p_s^* = \arg\max_{\forall j} \frac{F\left(\gamma_{j,s}^{(t)}\right)}{\bar{\vartheta}_{j,k}^{(t-1)}} \qquad \forall s. \tag{4.9}$$

Intuitively spoken, this approach handles each sub-carrier as isolated system for which proportional fair criterion (4.3) must be satisfied. The *carrier-wise* proportional fairness approach has widely been accepted [100, 103, 104] as the standard way to achieve proportional fairness in multi carrier systems. To obtain the system-wide optimal terminal/sub-carrier assignments, the sum over all per sub-carrier decisions needs to be optimized:

$$\max_{\mathbf{X}^{(t)}} \sum_{\forall j} \sum_{\forall s} \frac{x_{j,s}^{(t)} F\left(\gamma_{j,s}^{(t)}\right)}{\bar{\vartheta}_{j,k}^{(t-1)}} \tag{4.10a}$$

$$\text{s. t.} \sum_{\forall j} x_{j,s}^{(t)} \leq 1 \qquad \forall s \tag{4.10b}$$

As the carrier-wise proportional fairness problem (4.10) is linear, it can be solved with standard linear integer problem solvers. But both low complexity problems, the *single user* (4.8) and the *carrier-wise* proportional fairness approach (4.10) can be solved by means of heuristic algorithms as well. In order to decide on the terminal/sub-carrier assignments $x_{j,s}^{(t)}$ at time t, these algorithms need to compute the proportional fairness ratios

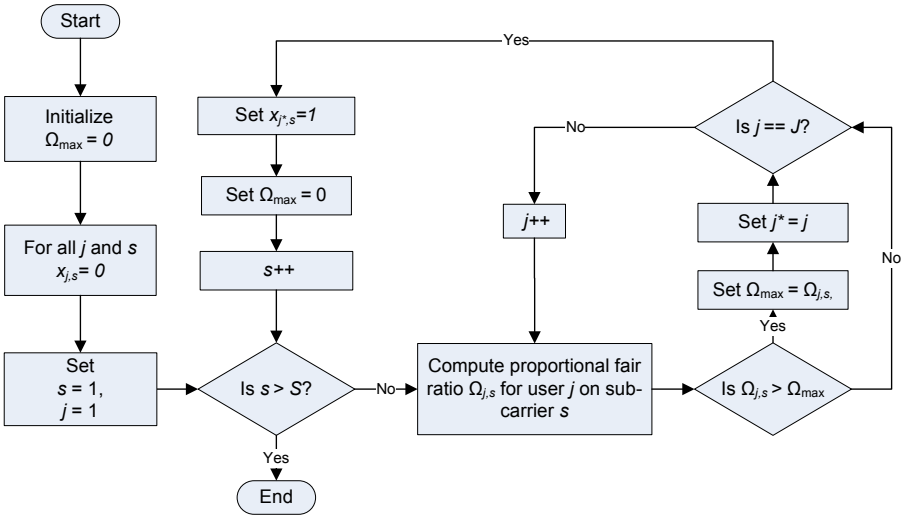
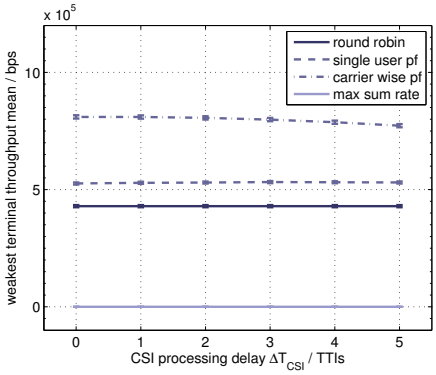


Figure 4.2: algorithm implementing Kim's *carrier-wise proportional fair* approach

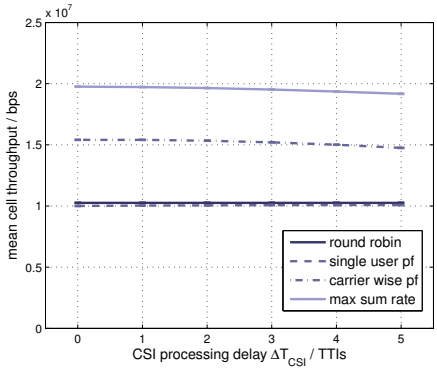
$$\Omega_{j,s}^{(t)} = \frac{F\left(\gamma_{j,s}^{(t)}\right)}{\bar{\vartheta}_{j,k}^{(t-1)}} \tag{4.11}$$

of all possible sub-carrier s terminal j combinations. Accordingly, two nested loops, over all sub-carriers and terminals respectively, need to be run through per scheduling epoch. Recall that k is the PF window size, *i. e.,* the number of TTIs that are evaluated to compute the average achieved throughput $\bar{\vartheta}_{j,k}^{(t-1)}$. Obviously, k needs to be carefully chosen, as it has a significant impact on the PF scheduling behavior: while small sizes favor the stronger and harm the weaker terminals, large sizes account for an increased memory consumption and computational complexity. Throughout this thesis, the window size is set to $k = 50$TTIs.

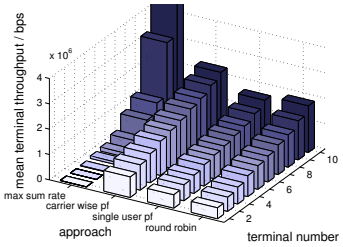
The algorithms' central and most complex task is the comparison of the PF ratios $\Omega_{j,s}^{(t)}$. A flow-chart for an example algorithm implementing Kim's single user proportional fairness approach is shown in Figure 4.1. Here, the outer loop goes over all terminals $j \leq J$, and the inner loop goes over the sub-carriers s. Within the inner loop, for each terminal j, the cumulated PF ratio $\Omega_{j,\mathrm{tot}}^{(t)}$ is computed. In the outer loop, the comparison to the current maximum cumulated ratio is done, and hence, only J comparisons need to be done per scheduling epoch. In contrast, the comparison is done individually per sub-carrier s and terminal j in the inner loop of the carrier-wise proportional fair algorithm (*cf.* Figure 4.2). Accordingly, in this approach $J \cdot S$ comparisons are performed per scheduling epoch. As previously stated, the surplus in complexity comes along with the opportunity to exploit frequency diversity. So far in literature, only the performance advantage of the original *multi carrier proportional fair* problem (4.6) of complexity J^S over the low complexity *carrier-wise* approach (4.10) has been shown [102]. The more interesting matter for the scope of this thesis, however, remains, namely the question, if the performance difference between the two low complexity approaches is significant. To find the answer, the following computational tests are conducted.



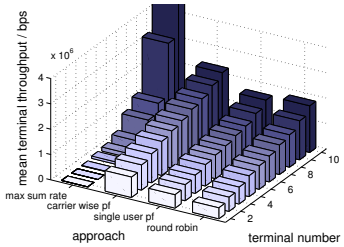
(a) mean throughput (bps) of weakest terminal



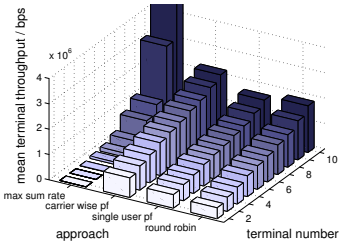
(b) mean cell throughput (bps)



(c) per terminal tp, ΔT_{csi}= 0 TTIs



(d) per terminal tp, ΔT_{csi}= 2 TTIs



(e) per terminal tp, ΔT_{csi}= 5 TTIs

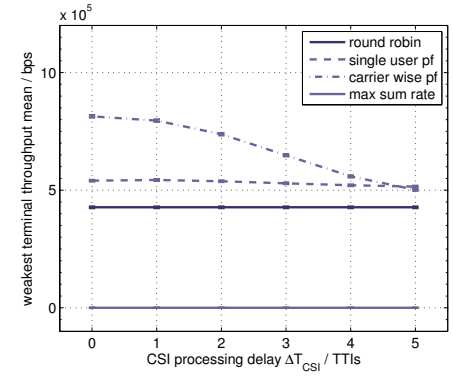
Figure 4.3: single user vs. carrier-wise proportional fair policy, *stationary terminals*

4.1.3 Performance analysis

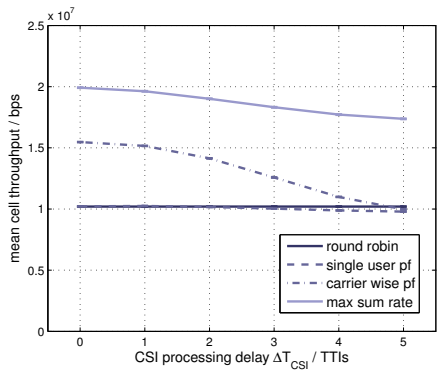
In this section, the performance of the lowest complexity *single user* proportional fair approach (4.8) is compared to the *carrier-wise* proportional fair approach (4.10). The system model from Section 3.2 is applied, the parameterization is set according to the single cell scenario given in Section 3.3.1. In order to relate their system performances to a lower and upper bound, two more scheduling approaches are simulated, the *max sum rate* approach (2.22) (that delivers maximal system throughput at the cost of poor fairness) as the upper, and a simple round robin approach as lower bound.

Single user proportional fair resource allocation

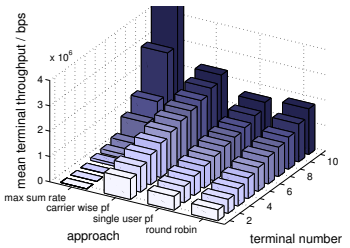
The throughput results for the case of stationary terminals are summarized in Figure 4.3, the results for moderately moving terminals are shown in Figure 4.4. The general impact of applying the proportional fair policy is well shown Figure 4.4a: the mean throughput of the individual weakest terminal is increased. In the case of single user PF scheduling a throughput gain of approx. 20% is achieved compared to the standard round robin results. At the same time, the mean cell throughput is slightly decreased. This is due to the fact that when PF scheduling is applied, well situated terminals that easily achieve high throughput values (and, thus, account



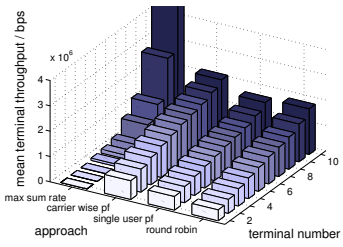
(a) mean throughput (bps) of weakest terminal



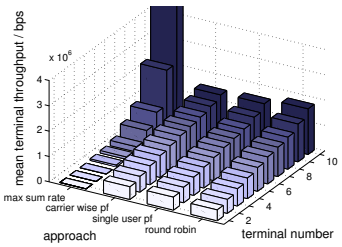
(b) mean cell throughput (bps)



(c) per terminal tp, ΔT_{csi}= 0 TTIs



(d) per terminal tp, ΔT_{csi}= 2 TTIs



(e) per terminal tp, ΔT_{csi}= 5 TTIs

Figure 4.4: single user vs. carrier-wise proportional fair policy, *moving terminals,* $v = 10\frac{m}{s}$

for a high cell throughput) are scheduled less often than in the RR case. This effect is shown in Figure 4.3c, which shows the mean throughput per terminal values for the ideal channel knowledge case. By considering not only the individual terminal throughput, but also the average channel quality (over all sub-carriers) per terminal when making the scheduling decision, the single user PF approach, however, exploits the system's multi-user diversity (*cf.* Section 2.4.2), and, hence, the overall decrease in system performance is low. Since the terminal's average channel quality changes on a relatively large time scale, the single user PF performance is hardly affected by outdated channel state information. A slight decrease in mean weakest terminal and cell throughput performance is present only in the case of moving terminals (Figure 4.4). Considering the fact that in most systems a processing delay of $\Delta T_{\text{csi}} = 2 \ldots 3$ TTIs is realistic, and since no channel prediction techniques are applied, this decrease is, however, neglectible.

Carrier-wise proportional fair resource allocation

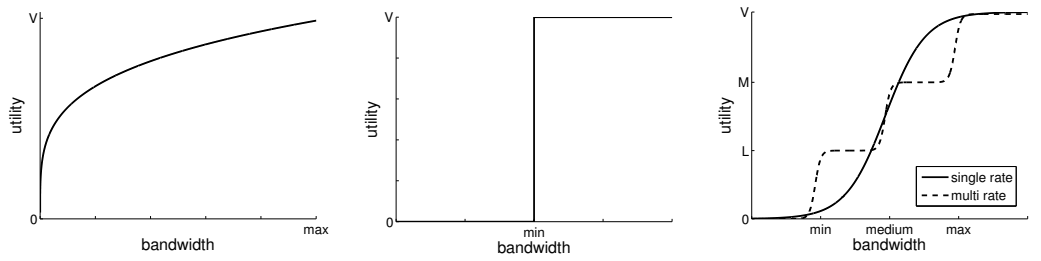
In addition to multi-user diversity, the carrier-wise PF approach exploits the channel's frequency diversity (*i. e.,* the circumstance that for a single terminal the sub-carriers are differently attenuated). By exploiting both effects at the same time, *i. e.,* by determining *per sub-carrier*, which terminal will use it in the upcoming TTI, large gains in throughput on the individual sub-carriers, and, hence for the overall system, can be achieved. As previously stated (*cf.* Section 2.4.2), the

largest system gains are possible if each sub-carrier is assigned to the terminal experiencing the best channel quality on it (the *max sum rate* approach). As can be seen in Figure 4.3b, mean cell throughput gains up to 100% compared to the RR case are possible in the chosen reference scenario. The lack of fairness that comes with this system gain, however, is shown in Figure 4.3b: the weakest terminal does not get any resources at all, and hence has zero throughput. In fact, it is not only the weakest terminal that has zero throughput, if resources are scheduled according to the max sum rate approach, as can be seen on Figure 4.3c. The largest part of the available resources is assigned to the best 2-3 terminals, while all other terminals significantly suffer from the chosen policy. In contrast, when applying the carrier-wise PF policy, all terminals gain from exploiting multi-user and frequency diversity. The percentual gain for the weaker terminals is is larger (up to approx. 100%) than for the strong terminals (approx. 20%). In average, for the overall cell, a throughput gain of approx. 50% is possible compared to the round robin case. The carrier-wise PF policy, thus, halves the system gains that are possible due to multi-user and frequency diversity, but achieves the highest degree of fairness among the considered cases.

Note, however, that both frequency diversity exploiting approaches suffer from outdated channel state information. While in the stationary terminals scenario the CSI age has a rather low impact on the performance, the impact is significant in the case of moderately moving terminals (Figure 4.4). If no channel prediction technique is applied, the gains stemming from exploiting the frequency diversity in the carrier-wise PF case are completely vanished if the available CSI has an age of 5TTIs. In other words, the performance of the carrier-wise PF approach goes down to the performance of the single user PF approach, and accordingly the higher complexity of the carrier-wise approach does not pay off in this case. Still, considering again the fact that in most systems a processing delay of $\Delta T_{\text{csi}} = 2 \ldots 3$ is realistic, it can be stated that carrier-wise allocation has the potential to significantly increase the system fairness and throughput performance in the case of stationary of moderately moving terminals.

Conclusions

Providing fairness among the terminals is a major challenge when allocating the resources in cellular networks. Among the three well known fairness policies *max min*, *individual user utility*, and *proportional fairness*, the latter has been identified to be of most practical relevance, as it provides a well-balanced relationship between user fairness and system performance. Two different algorithmic PF alternatives have been compared: *single user* and *carrier-wise* proportional fairness. While the latter achieves significant gains in weakest terminal *and* system throughput performance, the first achieves only some gains in weakest terminal performance – this, however, with higher stability against outdated CSI and at lower complexity. As this thesis aims in determining the potential throughput gains of dynamic mechanisms, the carrier-wise approach is of higher interest. From this point, the thesis thus concentrates on the carrier-wise PF. In the remainder of the thesis it is simply referred to as proportional fair scheduling approach.



(a) elastic traffic (FTP,HTTP) (b) hard real-time traffic (VoIP) (c) soft real-time traffic (MPEG)

Figure 4.5: application specific utility curves

4.2 Rate Constraints

Purely maximizing the proportional fairness yields a disadvantage compared to utility-based optimization: the individual rate requirements of the individual terminals are not considered.

4.2.1 Constrained max sum rate optimization

Maximizing throughput without considering the individual terminal's needs might lead to a situation in which terminals are allocated resources that they do not need at a certain point in time. To circumvent this situation, in the following utility-based optimization – as introduced in Equation (4.1) – is considered again. Recall that in general utility-based optimization aims in maximizing the sum of the individual terminal utility values. The utility function to be used by a certain terminal j is determined by the type of its application. An example utility curve for voice over IP traffic is shown in Figure 4.5b. Since VoIP applications are extremely sensitive to packet delay and loss caused by bandwidth insufficiency, its utility function falls into the category of hard real-time kind [105]. If the allocated bandwidth for a VoIP user is less than min, its user utility will drop to zero. On the other hand, utility does not increase for larger bandwidth shares. Due to this characteristic, VoIP traffic significantly differs from so called *elastic traffic* resulting from the file transfer protocol (FTP), the hyper text transfer protocol (HTTP), or the simple mail transfer protocol (SMTP) [106]. Elastic traffic always gains from increasing the bandwidth. The relative gain, however, decreases with the user's momentary bandwidth share, such that once the user holds a certain bandwith share max, increasing the bandwidth yields only small utility improvements (Figure 4.5a).

The utility characteristics of soft real-time applications are somewhat a mixture of the elastic and hard real-time characteristics. Due to buffering and advanced jitter control technologies, soft real-time applications, such as streaming video, can tolerate occasional delay-bound violations and packet drops, the so called *delay-adaptive* capability. The minimal encoding rate min is independent of network congestion. Below this threshold, video user utility drops to zero. Bandwidth shares that exceed the encoding rate max, on the other hand, do not add utility. Figure 4.5c shows example utility curves of single and multi-rate video codecs. A multi-rate codec that is able to adapt to the current bandwidth availability by switching among a low L, a medium

M, and the target rate V features a utility curve that resembles a "softened" step function: as the coders offer video streams only at certain bit rates, any bit rate between the offered data rates does not provide additional utility to a terminal receiving the stream. If the next higher rate can't be achieved, thus, no additional bandwidth should be allocated for this terminal.

Utility-based optimization is a promising approach to achieve fairness and integrate congestion control into the resource allocation process [106]. It is, however, highly complex. Moreover, it involves the presence of application-specific utility functions. In the following, a more general and simplified utility-based optimization model is introduced, the *constrained max sum rate (CMSR)* strategy.

4.2.2 Optimization model

It is assumed that for each terminal j there is a maximum throughput $\vartheta_{\mathrm{req},j}$, and that any throughput beyond $\vartheta_{\mathrm{req},}$ is useless. In other words, the goal is to maximize the system throughput while assuring that none of the users gets more than a certain maximum rate:

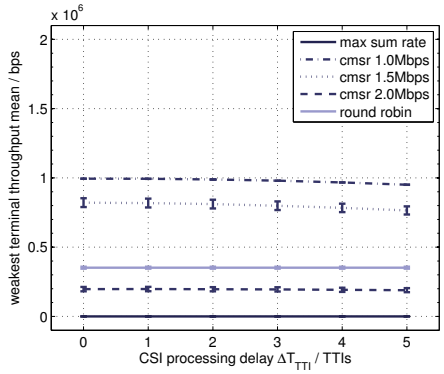
$$\max_{\mathbf{X}^{(t)}} \quad \sum_j \sum_s x_{j,s}^{(t)} F\left(\gamma_{j,s}^{(t)}\right) \tag{4.12a}$$

$$\text{s. t.} \quad \sum_s x_{j,s}^{(t)} F\left(\gamma_{j,s}^{(t)}\right) \leq \vartheta_{\mathrm{req},j} \qquad \forall\, j \in J \tag{4.12b}$$

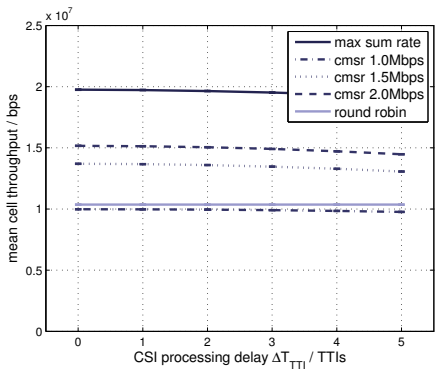
This approach corresponds to a simple piecewise linear utility function. As it is the case with the three utility functions above (elastic, hard, and soft real-time), it assures that bandwidth is not wasted on users that only slightly or not at all would gain from additional bandwidth allocations. It does not provide guarantees to meet minimum bounds. As the overall goal is to maximize the overall system throughput, terminals will only suffer from bandwidth under provisioning, if it is not possible to meet all user requirements. Compared to the max-min optimization approach (2.24), allocating the resource according to the *constrained max sum rate optimization* approach has the advantage that a momentarily bad situated terminal does not impact the overall performance. It matches a well known provider policy that does not guarantee a certain rate, but provides the best possible performance up to a maximum rate $\vartheta_{\mathrm{req},}$.

4.2.3 Performance analysis

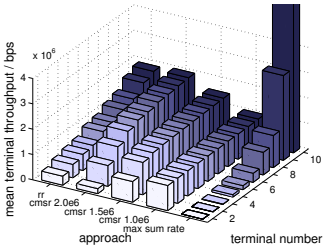
As it has previously been the case for the proportional fair results in Section 4.1.3, in this section the constrained max sum rate system performance is related to the upper and lower bound given by the *max sum rate* approach (2.22) (that delivers maximal system throughput at the cost of poor fairness – upper bound), and the simple round robin approach (lower bound). The system model from Section 3.2 is applied, the parameterization is set according to the single cell scenario given in Section 3.3.1. The throughput results for the case of stationary terminals are summarized in Figure 4.6, the results for moderately moving terminals are shown in Figure 4.7. Both figures show the results for three different maximum rate requirement $\vartheta_{\mathrm{req},j}$ settings: 1Mbps, 1.5Mbps, and 2Mbps.



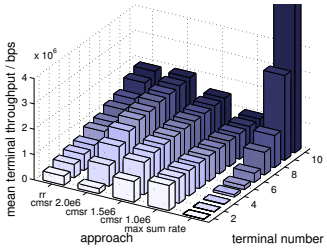
(a) mean throughput (bps) of weakest terminal



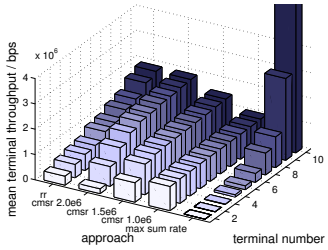
(b) mean cell throughput (bps)



(c) per terminal tp, ΔT_{csi}= 0 TTIs



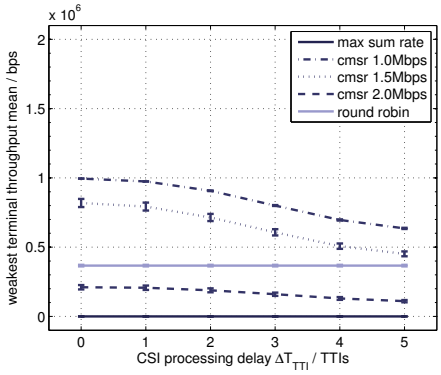
(d) per terminal tp, ΔT_{csi}= 2 TTIs



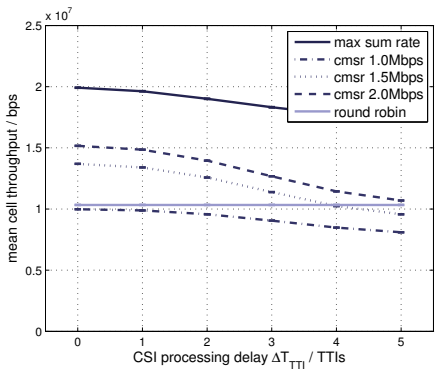
(e) per terminal tp, ΔT_{csi}= 5 TTIs

Figure 4.6: constrained max sum rate policy for various max. rates, *stationary terminals*

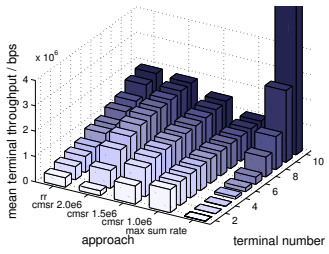
The general impact of applying the constrained max sum rate policy is well shown in Figures 4.6c through 4.6e, the individual terminal throughput for different CSI processing delay values ΔT_{csi}: if enough resources are available (as it is the case for $\vartheta_{\text{req},j} = 1$Mbps), all terminals receive the same share in terms of individual terminal throughput; the application of the CMSR, hence, yields *same rate fairness*, if the rate requirements are the same for all active terminals. If not enough resources are available ($\vartheta_{\text{req},j} > 1Mbps$), the throughput of the weaker terminals quickly drops down to zero. This property inherits the CMSR policy from the max sum rate approach. Compared to the latter, however, the majority of terminals has a significant higher individual throughput. In comparison with round robin, the difference between *same Rate fairness* and *same (bandwidth) share fairness* (as provided by the round robin policy) becomes evident: the typical step function (see *e. g.,* Figure 4.6c) that shows higher throughput for better situated terminals is replaced by an equal terminal throughput distribution (that results from an unequal – step-like – bandwidth share distribution, where bandwidth share is the larger the weaker the terminal is). Still CMSR is superior in both, the system, as well as the weakest terminal throughput, if the max rate $\vartheta_{\text{req},j}$ is chosen accordingly (see Figures 4.6a and 4.6b, $\vartheta_{\text{req},j} = 1.5$Mbps). In the case of a low max rate ($\vartheta_{\text{req},j} = 1$Mbps), however, the system throughput falls below the round robin performance, for the sake of weakest terminal performance. For a high max rate ($\vartheta_{\text{req},j} = 2$Mbps) the opposite is the case. Note that the CSI processing ΔT_{csi} delay has hardly any impact on the performance of any approach in the stationary terminal case.



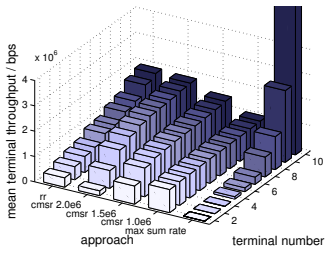
(a) mean throughput (bps) of weakest terminal



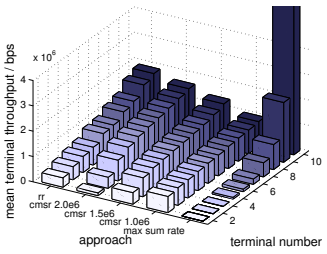
(b) mean cell throughput (bps)



(c) per terminal tp, ΔT_{csi}= 0 TTIs



(d) per terminal tp, ΔT_{csi}= 2 TTIs



(e) per terminal tp, ΔT_{csi}= 5 TTIs

Figure 4.7: constrained max sum rate policy for various max. rates, *moving terminals,* $v = 10\frac{m}{s}$

If the terminals are moving, the CMSR performance decreases with an increasing ΔT_{csi}. Recalling that the results shown in Figure 4.7 are the worst case estimation (considering no channel prediction techniques at all), it must be reasoned that at moderate CSI delay $\Delta T_{\text{csi}} =$ 2ms and speed $v = 10\frac{\text{m}}{\text{s}}$, the performance of the CMSR approach has significant gains compared to the round robin case. Note that, depending on the max rate requirement, even under these conditions weakest terminal throughput gains of 100% and mean system throughput gains of 40% are possible.

Conclusions

The constrained max sum rate scheduling approach corresponds to a simple piecewise linear utility function based approach: constraining the average rate per terminal avoids wasting capacity on users that do not gain from bandwidth increments beyond a certain share – as it is the case with the *raw max sum rate* approach. The resources can instead be used to improve the performance of weaker users, yielding *same rate fairness*. In the case of bandwidth underprovisioning (*i. e.,* the provided bandwidth is lower than the sum of required terminale rates), however, the weak terminals suffer first. Consequently, in order not to waste resources and protect the weakest terminals, a combination of a fair and individual rate requirements aware scheduling approach is necessary.

4.3 Packet Awareness

In the previous sections, optimization has been done in systems, which emanate from persistent down-link data transmission, *i. e.,* for each terminal endless data reserves are assumed that can be broken up into data pieces of arbitrary size. There is no notion of packetization in these systems. The major difference between such systems and packet-oriented cellular networks lies in the partitioning of the wireless link capacity. While in the first case the link capacity is shared between all active terminals in a deterministic manner, the resource is shared on a per-packet basis in the packet-oriented case. Thus, for each packet a decision on when to deliver it to its destination has to be made. In each TTI several packets can be transmitted in parallel. Hence, a consequent problem formulation for optimizing data transmissions in a packet-oriented cellular OFDMA based system is the question:

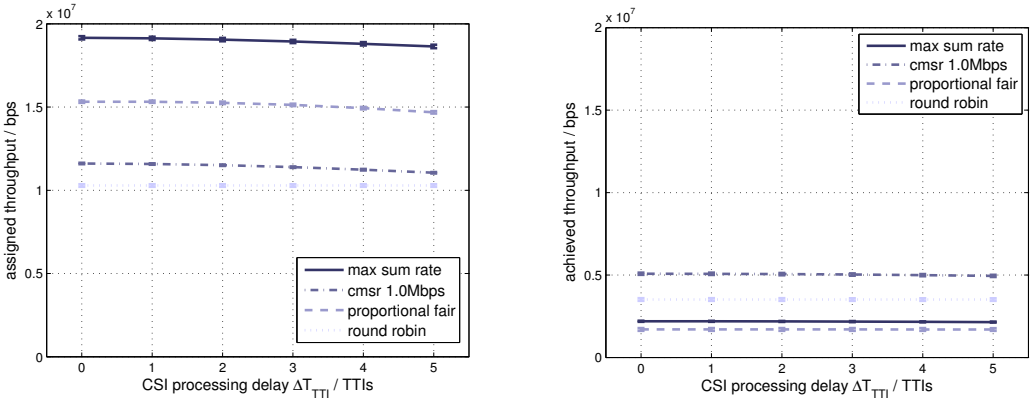
> *When to transmit a buffered packet on which sub-carriers in order to obtain optimal performance?*

Packet scheduling At each base station, the scheduler (*cf.* Figure 3.2) is responsible for selecting the packets to be transmitted during the next TTI. It is also responsible for the dynamic resource allocation task. Note that none of the optimization models presented so far in this thesis deals with balancing the resource scheduling decisions with the packet scheduling decisions to be made in a packet-oriented system. As a result, the danger of having a discrepancy between the packets that need to be sent and the amount of available resources at a certain terminal are quite high, when those persistent data optimization models are applied to packet-oriented networks [107]. In general, the interaction of resource and packet scheduling in a dynamic packet-oriented system can follow two different strategies: *sequential* and *integrated* resource allocation/packet scheduling.

4.3.1 Sequential resource and packet scheduling

The sub-carrier allocation decisions are based on channel state information, but chosen independently of the packet scheduling process. A subsequently executed algorithm schedules the packets according to the outcome of the resource allocation process. Figure 4.8 demonstrates the problems that arise, if sequential resource and packet scheduling is done in a packet-oriented network without considering the instantaneous terminal rate requirements (*i. e.,* their buffer fill levels).

Simulation parameterization The above results have been obtained by means of simulation using the system model introduced in Section 3.2 and following the methodology, the system parameterization and assumptions according the single-cell scenario of Section 3.3. Ten terminals are active in the single cell. Each terminal receives data of two different classes: best effort web traffic, and VoIP packets requiring a certain level of QoS following the models introduced in Section 3.2.4. While the G.726 encoded VoIP packets per terminal arrive mostly individually with inter arrival times of approx. 8ms (subject to delay and jitter), the web traffic arrives



(a) assigned cell throughput (bps)

(b) achieved cell throughput (bps)

Figure 4.8: assigned vs. achieved cell throughput in a packet-oriented network

in bursts. Consequently, in each TTI the transmit buffers at the base station for some of the terminals are filled, others are empty. The terminals are assumed to be stationary.

The discrepancy between *assigned* and *achieved* throughput The resource scheduling strategies presented in the previous sections do not consider the instantaneous terminal requirements, but base their decisions solely on the individual terminal's channel state (in the max sum rate case), on the channel state and average rate requirements (as in the constrained max sum rate case), or on the channel states and former average throughput results (proportional fair based resource scheduling). Based on this information, the different strategies *assign* distinct parts of the overall available spectrum to the terminals, which translates into an expected *assigned throughput*, if the according channel states, transmit power values and modulation and coding types are related to each other. In the case of sequential resource and packet scheduling, the packet scheduling is done in a second step based on this assigned throughput. The results shown in Figure 4.8 are obtained by applying a simple packet scheduling strategy based on the resource allocation decisions made by the MSR, CMSR, and PF approaches of the previous sections: for each user that is assigned resources in the next TTI first the VoIP packets are scheduled (if any are in the transmission buffer), and then as many web traffic packets as match the remaining assigned resources are scheduled. If the next packet to be scheduled is too large for the remaining resources, it is fragmented.

The *achieved throughput* displayed in Figure 4.8b shows the amount of data that actually arrives at the terminals. Obviously, there is a significant difference between the expected throughput that the scheduler assigns by making its resource scheduling decisions and the actually achieved throughput. Note that the largest differences are present for the max sum rate case, where large parts of the available spectrum are concentrated on a single (the strongest) terminal. In scenarios, where this particular terminal experiences mostly empty buffers, the assigned resources are almost completely wasted. Moreover, the assigned throughput can only be achieved in scenarios, where at each point in time the momentary strongest terminal has packets in its

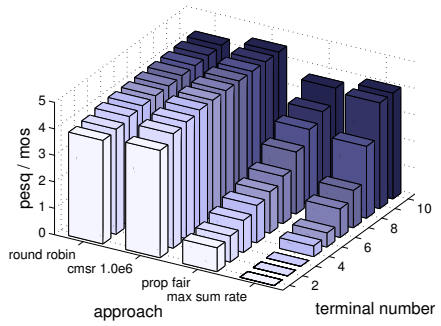


Figure 4.9: achieved PESQ values

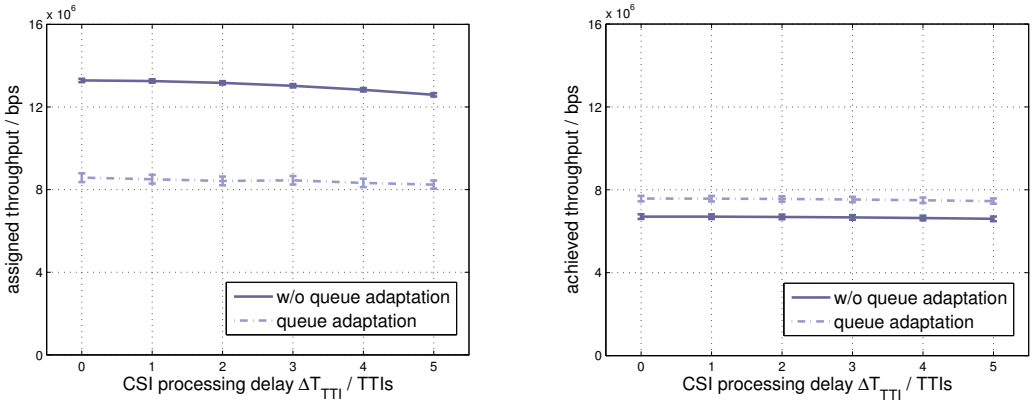
MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Table 4.1: ITU-T's Mean opinion score (MOS).

base station transmission buffer. In a weaker form, the same reasoning holds for the differences in assigned and achieved throughput of the other approaches.

An additional problem that arises if the actual buffering situation per terminal is not considered during the resource allocation process is evident in Figure 4.9. The figure shows the perceptual evaluation of speech quality (PESQ) performance in terms of the individual terminal's mean opinion score (MOS) values. While ITU-T's original MOS scale generally ranges up to 5.0 (*cf.* Table 4.1), the maximal achievable PESQ-MOS value as defined by the ITU recommendation P.862 [87] is 4.5. This due to the fact that a PESQ-MOS value is assumed to reflect the average sound quality experience of all potential listeners. Statistics show that the best average result one can expect from a listening test is ca. 4.5 (it seems like human beings are always cautious to score a 5, meaning "excellent", even if there is no degradation at all). Depending on the voice source and the codec in use, the maximal achievable PESQ values for a certain voice transmission is even lower. For the telephone call sound file that is used to model all VoIP traffic in this thesis, a maximum PESQ value of approx. 4.2 is achievable under ideal conditions. Comparing this optimum to the achieved PESQ values in Figure 4.9 reveals a fairness problem: as the individual VoIP packet QoS requirements cannot not be respected, if the actual buffering situation at the base station is not considered, the PESQ performance of most users is poor, or even bad in the case of PF scheduling. Good to excellent PESQ performance for the majority of terminals is only achieved for the static round robin and the 1Mbps CMSR case, which in the previous section have been shown to provide a high degree of inter-terminal fairness at the cost of system performance in terms of mean cell throughput (*cf.* Figure 4.6).

Considering the buffering situation It is not surprising that dynamic techniques, which do not consider the buffering situation at the base station yield bad performance in packet-oriented systems. Particularly, it is an obvious mistake to schedule transmission resources for terminals having empty buffers. To encounter this problem, simple mechanisms that exclude such terminals from the resource scheduling process have been suggested, and more sophisticated buffer aware algorithms have been proposed. In [108], the authors present an algorithm that assigns the sub-carriers based on buffer state information and measured traffic statistics in addition to channel state feedback. They consider the total number of bits in the queue, as well as the delay of



(a) assigned cell throughput (bps)

(b) achieved cell throughput (bps)

Figure 4.10: assigned vs. achieved cell throughput in a packet-oriented network

the first packet in each FIFO queue as the most important buffer information. It is shown that the algorithm performs significantly better than simple max rate optimization and the rate adaptive strategy by Rhee *et al.* [46] in terms of system throughput and inter-terminal fairness. The authors of [109] introduce a heuristic that optimizes the packet scheduling process among several terminals receiving video data packets in a cellular OFDMA system. In a later work [110], the according resource allocation process has been formulated as an optimization problem, which is related to the *rate adaptive* optimization problem (2.24). The difference between the original rate adaptive formulation and the buffer-aware version is the application of a factor $\alpha_j^{(t)}$ that relates the queue length $q_j^{(t)}$ of each terminal j to the queue length of the terminal featuring the smallest queue length $q_{\tilde{j}}^{(t)}$:

$$\alpha_j^{(t)} = \frac{q_j^{(t)}}{q_{\tilde{j}}^{(t)}} \quad . \tag{4.13}$$

Factor $\alpha_j^{(t)}$ is used as a weight in the rate adaptive resource constraint as follows:

$$\begin{aligned} &\max_{\mathbf{X}^{(t)}} \quad \epsilon \\ &\text{s. t.} \quad \sum_s F\left(\gamma_{j,s}^{(t)}\right) x_{j,s}^{(t)} \geq \alpha_j^{(t)} \epsilon \quad \forall j \ , \end{aligned} \tag{4.14}$$

As it is the case with the original rate adaptive problem formulation (2.24), the disjunctive sets constraint (2.23) is also part of the overall buffer-aware rate adaptive problem formulation.

Figure 4.10 shows the impact on the overall system throughput, if buffer factor $\alpha_j^{(t)}$ is utilized. First, regarding the assigned throughput shown in Figure 4.10a, the results show a significant decrease (almost one third) in the case of queue adaptation (*i. e.,* using the modified formulation (4.14)) compared to the case without queue adaptation (*i. e.,* following the original formulation (2.24)). Again, this decrease is mainly due to the fact that at certain points in time

the buffers of the stronger terminals are empty, and thus, the multi-user diversity gain cannot be exploited. On the other hand, referring now to the achieved throughput in Figure 4.10b, it is shown that the achieved throughput is almost equal to the assigned throughput. The small performance difference among the two is due to the resource wastage that stems from leaving parts of the assigned sub-carrier resources unused (*e. g.,* if a complete resource block is assigned to a terminal, which only receives a relatively small VoIP packet or an even smaller silence packet). The difference between assigned and achieved throughput is rather large in the case without queue adaptation. Moreover, it is smaller than the achieved throughput in the case of applying queue adaptation. The gain is around 10%, which is not very impressive, but a consequence of the max-min principle that favours the weakest terminal, and, thus, does not provide for significant gains due to multi-user diversity exploitation in the case of optimal assignments. Two questions arise from the facts above:

1. How to adopt the buffer awareness factor $\alpha_j^{(t)}$ to the more sophisticated resource scheduling approaches of the last sections?

2. If the application is possible, how close is its performance to the optimum?

While it more or less takes some trial-and-error attempts to answer the first question per resource scheduling approach, answering the second question is more tricky. In order to determine the optimal combined resource and packet scheduling performance in a packet-oriented system, the two decisions need to be integrated in a single optimization problem. Thus, before integrating buffer-awareness factor $\alpha_j^{(t)}$ into the resource scheduling approaches of the previous sections, in the following an integrated resource and packet scheduling optimization model is introduced.

4.3.2 Integrated resource allocation and packet scheduling

Integrating the sub-carrier allocation and packets scheduling decisions in a single optimization problem that is solved per TTI directly yields decisions on *which packets* are delivered using *which resources*. In other words, there is no need for a subsequently executed packet scheduling algorithm, as the output of the optimization problem delivers the packet scheduling decisions along with the resource scheduling decisions.

Some work has previously been done regarding that topic: In [111], Diao *et al.* propose a modified integration of generalized processor sharing scheduling with dynamic OFDMA resource allocation in order to account for the packet nature of data flows. While this approach matches packets of variable size to sub-carriers, the model cannot cope with adaptive modulation. It is, thus, not possible to determine optimal throughput values with it. Instead, the primary objective is to decrease bit error rates for a given, fixed total transmit power. Other work by Zhang *et al.* [112] considers packet-based optimization combined with dynamic resource allocation. Their model, however, only works for fixed-size packets. There is a lack of integrated resource and packet scheduling optimization problem formulations that can be used to determine optimal system throughput, terminal fairness and QoS performance. Consequently, in the following, an according model is developed. First, the packet scheduling decision is modeled as

a binary optimization variable. Then, this variable is integrated into existing resource allocation strategies.

The packet scheduling decision variable The integrated approach mainly differs from the sequential one in its goal to optimize packet delivery instead of maximizing rates. In order to integrate the packet scheduling decisions into the optimization problem formulation, the packet scheduling decision variable $z_{j,q,w}^{(t)}$ is introduced, where at time t:

$$z_{j,q,w}^{(t)} = \begin{cases} 0 \text{ if packet on waiting position } w \text{ in terminal } j\text{'s queue } q \text{ is not scheduled,} \\ 1 \text{ if packet on waiting position } w \text{ in terminal } j\text{'s queue } q \text{ is scheduled.} \end{cases}$$

The set of all packet scheduling decision variables $z_{j,q,w}^{(t)}$ is used to form the 3D packet scheduling decision matrix $\mathbf{Z}^{(t)}$. In contrast to the rate-based models, in the packet-aware optimization case the system throughput ϑ is obtained by taking the sum over the product of all scheduled packets and their packet sizes $\varsigma_{j,q,w}^{(t)}$:

$$\vartheta = \sum_{j,q,w} \left[z_{j,q,w}^{(t)} \varsigma_{j,q,w}^{(t)} \right] . \tag{4.15}$$

In order to obtain the integrated optimization problem formulations from the former rate based ones, the throughput term needs to be substituted accordingly. In the following, this is done for the already introduced strategies max-sum rate (2.22), proportional fair (4.5), and constrained max-sum rate optimization (4.12).

4.3.3 Packet aware max-sum optimization

Starting with the simplest and least complex dynamic sub-carrier assignment problem, two packet aware optimization approaches are derived from sum-rate maximization problem (2.22):

1. the sequential *packet aware sum-rate maximization* employing buffer factor $\alpha_j^{(t)}$, and
2. the integrated *sum-packet maximization* using the packet decision variable $z_{j,q,w}^{(t)}$.

The performance of both approaches is then compared to each other. The goal is to answer the question on how close the performance of the less complex sequential strategy gets to the optimum performance of the integrated approach.

Packet aware max sum-rate Recall that the sequential approach consists of two parts: the packet-aware rate-based resource allocation and the heuristic packet scheduling part. Again, a simple scheduling heuristic is utilized. It schedules the packets for each user that is assigned resources in the next TTI as follows: first the VoIP packets are scheduled (if any are in the transmission buffer), and then as many web traffic packets as match the remaining assigned resources are scheduled. Packets can be fragmented, if the remaining resources do not suffice to deliver

a complete packet. For the resource allocation part, sum-rate maximization approach (2.22) is extended to include buffer factor $\alpha_j^{(t)}$:

$$\max_{\mathbf{X}^{(t)}} \sum_{j,s} \alpha_j^{(t)} F\left(\gamma_{j,s}^{(t)}\right) x_{j,s}^{(t)}. \tag{4.16a}$$

$$\text{s. t.} \qquad \sum_{j} x_{j,s}^{(t)} \leq 1 \quad \forall s. \tag{4.16b}$$

Recall that $F\left(\gamma_{j,s}^{(t)}\right)$ delivers the number of information bits available for transmission on sub-carrier s for terminal j. Constraint (4.16b) is the exclusive sub-carrier utilization constraint first presented in Section 2.4.2. To obtain the packet scheduling decisions, the scheduling heuristic is applied on the outcome of the resource allocation decisions $\mathbf{X}^{(t)}$.

Max sum-packet In order to obtain the integrated max sum-packet problem formulation, sum-rate maximization problem (2.22) needs to be reformulated to include the packet scheduling decision variable $z_{j,q,w}^{(t)}$:

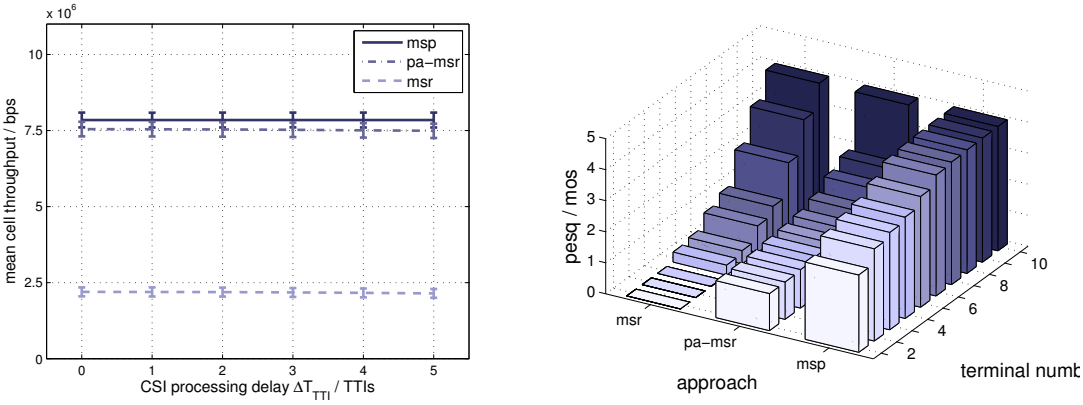
$$\max_{\mathbf{X}^{(t)},\mathbf{Z}^{(t)}} \quad \sum_{j,q,w} \left[z_{j,q,w}^{(t)} \varsigma_{j,q,w}^{(t)}\right] \tag{4.17a}$$

$$\text{s. t.} \qquad \sum_{s} \left[x_{j,s}^{(t)} F\left(\gamma_{j,s}^{(t)}\right)\right] \geq \sum_{q,w} \left[z_{j,q,w}^{(t)} \varsigma_{j,q,w}^{(t)}\right] \ \forall j, \tag{4.17b}$$

$$z_{j,q,w}^{(t)} \geq z_{j,q,w+1}^{(t)} \quad \forall j, q, (w < W) \tag{4.17c}$$

$$\sum_{j} x_{j,s}^{(t)} \leq 1 \quad \forall s. \tag{4.17d}$$

where $\varsigma_{j,q,w}^{(t)}$ is the size of terminal j's packet in queue q at position w and W is the number of available places in the queue. Optimization goal (4.17a) maximizes the sum of scheduled packet sizes, *i. e.,* the cell throughput. Note that this time the packet scheduling decision variable $z_{j,q,w}^{(t)}$ is part of the optimization goal, whereas the SNIR to bit mapping function $F\left(\gamma_{j,s}^{(t)}\right)$ appears only in the constraints. In constraint (4.17b) the sum over the product of the terminal/sub-carrier assignments and the bits per sub-carrier value per terminal (*i. e.,* the overall allocated sub-carrier resources for that terminal) is compared to the sum of its scheduled packet sizes. Thus, the constraint assures that adequate sub-carrier capacity is allocated to each terminal that is receiving packets. Constraint (4.17c) is the FIFO constraint that guarantees that packets are dequeued in FIFO order. Solving the problem leads to the optimal configurations of the binary decision matrices $\mathbf{X}^{(t)}$ and $\mathbf{Z}^{(t)}$ with respect to the maximum cell throughput. Note that maximizing the sum over scheduled packets inherently avoids solutions that assign resources to terminals with empty buffers.



(a) achieved cell throughput (bps)

(b) PESQ values (MOS)

Figure 4.11: achieved cell throughput and PESQ values for max sum rate resource allocation

Performance comparison

Figure 4.11 shows the performance results for the sequential packet aware max sum-rate (*pa-msr*) and the integrated max sum-packet (*msp*) approaches. For comparison reasons, the results of the simple max sum rate (*msr*) approach (2.22) are presented as well.

As expected, there are significant differences between the achieved throughput of the max sum-packet and the (non packet-aware) max sum-rate approach shown in Figure 4.11a: not adapting to the buffer fill levels yields a loss of more then two thirds of the system throughput. More interestingly, the packet-aware max sum-rate approach gets pretty close to the optimal throughput performance: the gap is only a few percent. Regarding the PESQ performance, however, significant differences are present. The very bad PESQ performance of most of the terminals in the simple max sum-rate case result from its fairness issues discussed in Section 2.4.2 (good situated terminals are favored while terminals in bad positions might not receive anything at all). These issues are partly overcome in the case of the packet aware max sum rate approach, as at some point in time there are no packets to be delivered to the dominating terminals, and, thus, the weaker terminals get their share. However, as this gain moves the weaker terminals from the *bad* to the *poor* MOS status, it does not really help.

In contrast, the integrated max sum-packet approach performs surprisingly well. This difference in PESQ performance can be traced back to the fact that the relatively small VoIP packets do not have a great impact on the buffer factor $\alpha_j^{(t)}$ in the sequential optimization case, but most of the times fit a single resource block assignment. Since in such a case a weak terminal requires the same resources as a strong terminal, both terminals have the same priority in the integrated *max sum-packet* case and, thus, the weak terminal is in average scheduled much more often in this case than in the *max sum-rate* case.

4.3.4 Packet aware proportional fair optimization

To encounter the fairness issues of the max-sum rate approaches, proportional fair resource allocation has been introduced in Section 4.1. Again, a sequential and an integrated packet-aware proportional fairness approach are derived from the original rate-based formulation (4.10).

Packet-aware proportional fairness To obtain the sequential packet-aware proportional fairness optimization approach, the original proportional fair optimization goal (4.10a) is extended by including buffer factor $\alpha_j^{(t)}$:

$$\max_{\mathbf{X}^{(t)}} \sum_{j,s} \alpha_j^{(t)} \frac{x_{j,s}^{(t)} F\left(\gamma_{j,s}^{(t)}\right)}{\bar{\vartheta}_{j,k}^{(t-1)}} . \tag{4.18}$$

Recall that $\bar{\vartheta}_{j,k}^{(t)}$ is terminal j's average achieved throughput over the last k frames. In combination with Constraint (4.16b), packet-aware proportional fair resource allocation is provided; and by executing the simple scheduling algorithm known from the sequential packet aware max sum-rate approach, an according proportional fair packet delivery is implemented.

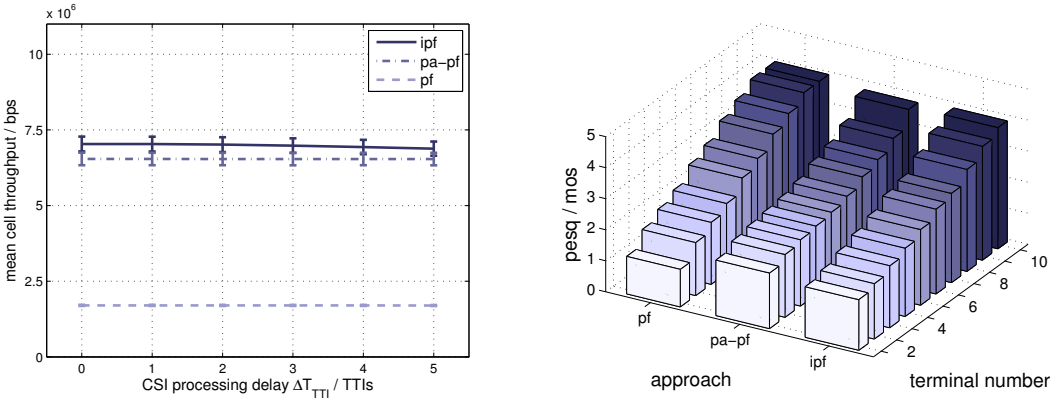
Integrated proportional fairness In order to achieve the same kind of fairness with integrated optimization, the integrated max-sum packet Objective (4.17a) can be modified to include each terminal's throughput $\bar{\vartheta}_{j,k}^{(t)}$ over the last k TTIs:

$$\max_{\mathbf{X}^{(t)},\mathbf{Z}^{(t)}} \quad \sum_{j,w} \left[z_{j,q,w}^{(t)} \frac{\varsigma_{j,q,w}^{(t)}}{\bar{\vartheta}_{j,k}^{(t-1)}} \right] . \tag{4.19}$$

Here, terminal j's ability to transmit packets during the upcoming TTI is related to its average achieved throughput $\bar{\vartheta}_{j,k}^{(t)}$. Adding the integrated problem Constraints (4.17b)-(4.17d) yields the complete integrated proportional fairness optimization problem formulation.

Performance comparison

Figure 4.12 shows the performance results of the sequential packet-aware proportional fairness (*pa-pf*), the integrated proportional fairness (*ipf*), as well as the standard (non packet-aware) proportional fair (*pf*) approach. As it is the case with the max-sum optimization, not adapting to the buffer situation yields a significant throughput break-in. Here, it is even more severe – up to three fourths of the overall system throughput are lost. Again, the system throughput performance of the sequential resource allocation and packet scheduling approach is very close to the optimal integrated performance. In terms of PESQ performance, the average MOS values of the users are almost the same for all three considered approaches. This is due to the fact that by considering the achieved average throughput performance, the proportional fair policy guarantees that each terminal is scheduled once in a while. As the VoIP packets are relatively small in size, it is almost sure that the packet can be delivered at once, if the terminal is scheduled for data delivery. Since, however, the stronger terminals are scheduled much more often than the



(a) achieved cell throughput (bps)

(b) pesq values (mos)

Figure 4.12: achieved cell throughput and PESQ values for proportional fair resource allocation

weak terminals, the VoIP packets of the latter experience partly, significant delays, such that the MOS performance of the weaker half of terminals even in the optimal integrated optimization case ranges only between *poor* and *fair*, while for the stronger half it ranges between *fair* and *good*. None of the terminals experiences *bad* PESQ performance though.

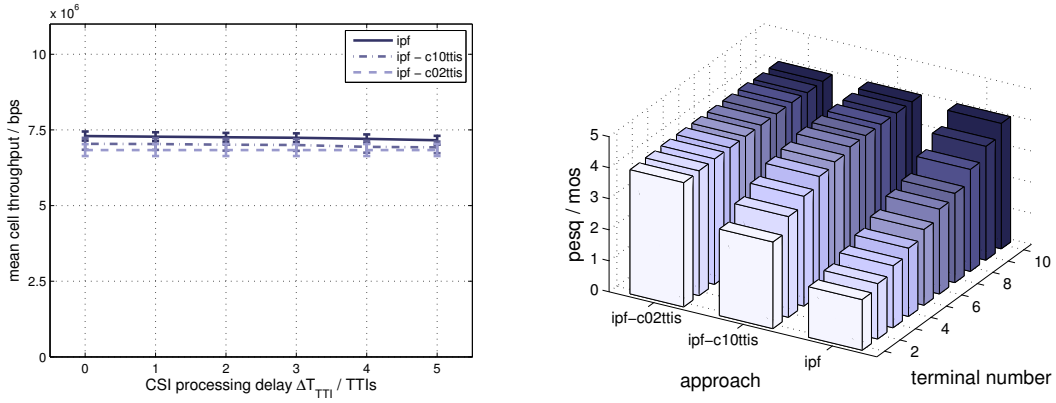
4.3.5 Constrained packet aware optimization

The concept of constraint optimization has been shown for the rate based max-sum rate approach in Section 4.2. The general idea of adding special purpose constraints to shift the optimization results into a certain direction can obviously also be applied to the integrated resource allocation and packet scheduling approaches (4.17) and (4.19). In contrast to the rate based optimization models, constraints on individual packet granularity are possible in the case of integrated resource allocation and packet scheduling optimization.

Throughput constraint As in the rate-based cases, constraining the throughput is possible by bounding the assigned resources by a *required throughput* $\vartheta_{\mathrm{req},j}$:

$$\sum_s x_{j,s}^{(t)} F\left(\gamma_{j,s}^{(t)}\right) \leq \vartheta_{\mathrm{req},j} \quad \forall\, j \in J \tag{4.20}$$

Recall that this approach corresponds to a simple piecewise linear utility function. It assures that bandwidth is not wasted on users that only slightly or not at all would gain from additional bandwidth allocations. While this constraint is of great use in rate-based optimization, it's usage in packet optimization scenarios is rather limited, as terminals with a buffer that is far more filled than the buffers of the other terminals should be served with priority, and not constrained to an average rate. Generally, in packet-based optimization, constraints that work on packet granularity level instead of average rates are of greater use. An according example is the delay constraint.



(a) achieved cell throughput (bps)

(b) pesq values (mos)

Figure 4.13: achieved cell throughput and PESQ values for constrained optimal proportional fair resource allocation

Delay constraint The general idea of the delay constraint is to include the delivery deadline requirements of so called timely delivery (TD) data packets into the optimization problem formulation. In order to guarantee TD packet timeliness, the following QoS constraint has to be added to the integrated optimization models (4.17) or (4.19):

$$\left(1 - z^{(t)}_{j,\mathrm{TD},w}\right) \cdot \left(\tau^{(t)}_{j,\mathrm{TD},w} + T_{\mathrm{TTI}}\right) \leq \mathcal{T}_{\mathrm{TD}} \quad \forall j, w \tag{4.21}$$

where $\tau^{(t)}_{j,\mathrm{TD},w}$ is the momentary delay of j's TD packet at position w and $\mathcal{T}_{\mathrm{TD}}$ is the maximum admitted TD delay.

Performance comparison

Adding delay constraint (4.21) to the original integrated proportional fair problem formulation (*ipf*) yields the desired effect of improved PESQ performance. If the maximum admitted delay is set to $\mathcal{T}_{\mathrm{TD}} = 2$ TTIs (*ipf-c02ttis*), *excellent* PESQ performance is experienced by all users, at the cost of a slight decrease in the overall throughput performance. An even smaller decrease is possible, if the maximum admitted delay is chosen less tightly, *e. g.,* $\mathcal{T}_{\mathrm{TD}} = 10$ TTIs (*ipf-c10ttis*). In this case, however, the PESQ performance of the weaker terminals is slightly worse.

4.3.6 Heuristics

Solving the optimization problems introduced in Section 4.3.2 delivers the optimal binary subcarrier and packet scheduling assignments according to the Objectives (4.17a) or (4.19). Solving these problems, however requires a large amount of time and computational power, mainly due to the integer constraint on the assignment variables. As a low complexity alternative that achieves sub-optimal assignment results, heuristic algorithms can be used (*cf.* Section 2.4.3). In general with heuristic approaches, there is a trade-off between the algorithmic complexity and the gap

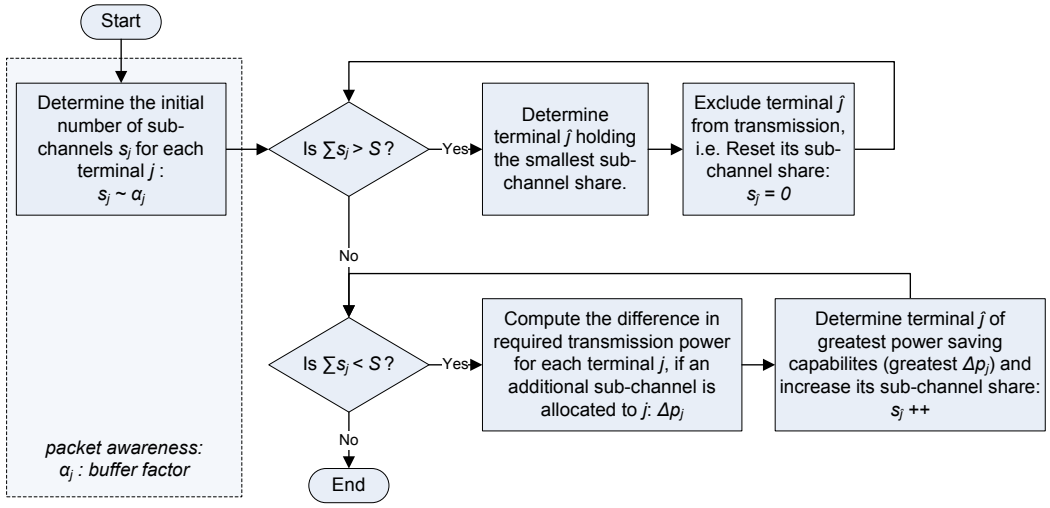


Figure 4.14: packet aware Bandwidth Assignment Based on SNIR (PA-BABS) algorithm

to optimal performance. In the following, two heuristic approaches are presented to solve the sequential and the optimal integrated problems respectively.

Sequential heuristics

The sequential heuristics solve the sequential resource assignment and packet scheduling approach of Section 4.3.1.

Packet-aware BABS algorithm A first packet-aware sequential heuristic is a modified version of the combined BABS/ACG algorithm presented in Section 2.4 and the simple packet scheduling algorithm introduced in the last section. The modification that particularly makes the BABS algorithm packet-aware is shown in Figure 4.14: the number that a certain terminal is assigned in the following TTI does now depend the buffer factor $\alpha_j^{(t)}$, instead of average rate requirements. Particularly, the inital number of assigned sub-carriers is determined by relating the own buffer factor to the sum of buffer factors:

$$s_{\tilde{j}} = S \cdot \frac{\alpha_{\tilde{j}}^{(t)}}{\sum_j \alpha_j^{(t)}} \tag{4.22}$$

The ACG and the simple packet scheduling algorithm remain unchanged.

Packet-aware proportional fair algorithm As proportional fair sequential heuristic alternative, Kim's carrier-wise proportional fair scheduling approach is modified to be packet aware. The modification is shown in the flowchart in Figure 4.15: the proportional fair ratio $\Omega_{j,s}^{(t)}$ is scaled by the buffer factor. Note that scaling the ratios excludes terminals with empty buffers from the scheduling process.

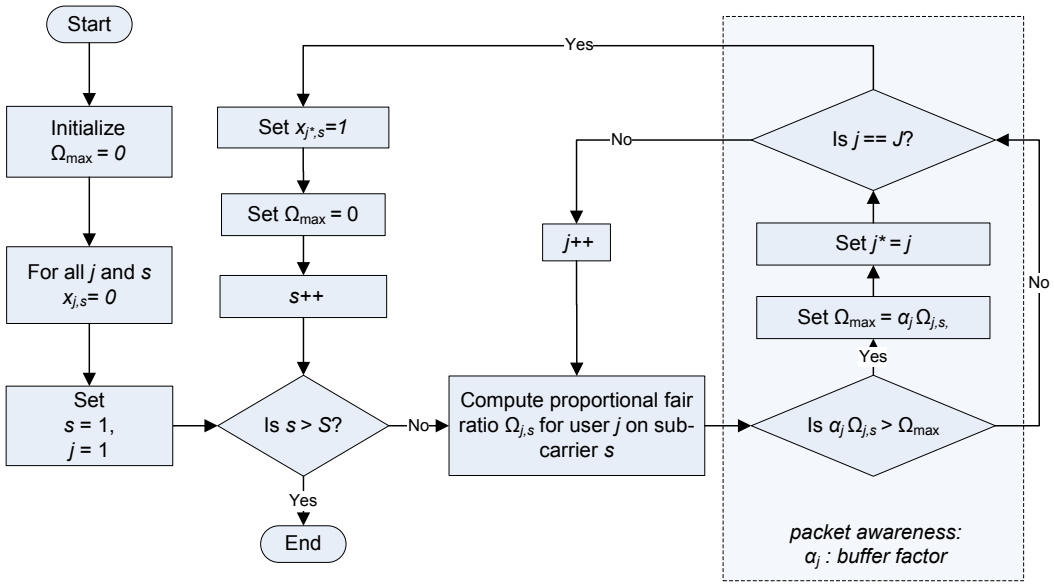


Figure 4.15: packet aware version of Kim's carrier-wise proportional fair (PA-PF) algorithm

Integrated heuristic The integrated resource allocation and packet assignment heuristic is derived from the integrated optimization approach introduced in Section 4.3.2. Generally speaking, it selects a terminal according to an arbitrary priority metric and then assigns as many sub-carriers as necessary in order to transmit the selected terminal's highest priority packet (see flowchart in Figure 4.16). The algorithm is, thus, referred to as sub-carrier/packet matching (SPM) algorithm. The pseudo-code of one particular implementation that uses the max-sum approach (2.22) as priority function is shown in Pseudo-Code Box 1. In particular, the algorithm first selects the terminal/sub-carrier pair $<\tilde{j}, \tilde{s}>$ of highest priority - here, in the max-sum case, it is the highest achievable transmission capacity $\chi_{j,s}^{(t)}$. Then, in Steps 5-9, it looks for terminal $\tilde{j}$'s highest priority packet by going through the QoS classes and queue positions, starting at the latest remembered highest priority packet QoS queue $q_{\tilde{j}}$ and queue position $w_{\tilde{j}}$ of the selected terminal. If there is a packet to be delivered for the selected terminal, it is scheduled for transmission in Step 10. If this is not the case, the terminal is excluded from the scheduling process for this TTI and another terminal is selected. In Steps 11-19, as many sub-carriers out of the set of available sub-carriers $\mathcal{S}$ as required to deliver the selected packet are assigned to the selected terminal, where the sub-carriers with the highest capacity for the selected terminal are chosen in Step 19. In order to determine whether the cumulated sub-carrier capacity $\mathcal{X}_{\tilde{j}}^{(t)}$ suffices to transmit the scheduled packets, it is compared to the cumulated packet size $\mathcal{B}_{\tilde{j}}^{(t)}$ in Step 15. Once enough sub-carrier capacity has been assigned to the selected terminal, the algorithm continues with finding the next highest priority terminal/sub-carrier pair, as long as there are sub-carriers available. If there are not enough sub-carriers in order to complete a scheduled packet, it is fragmented.

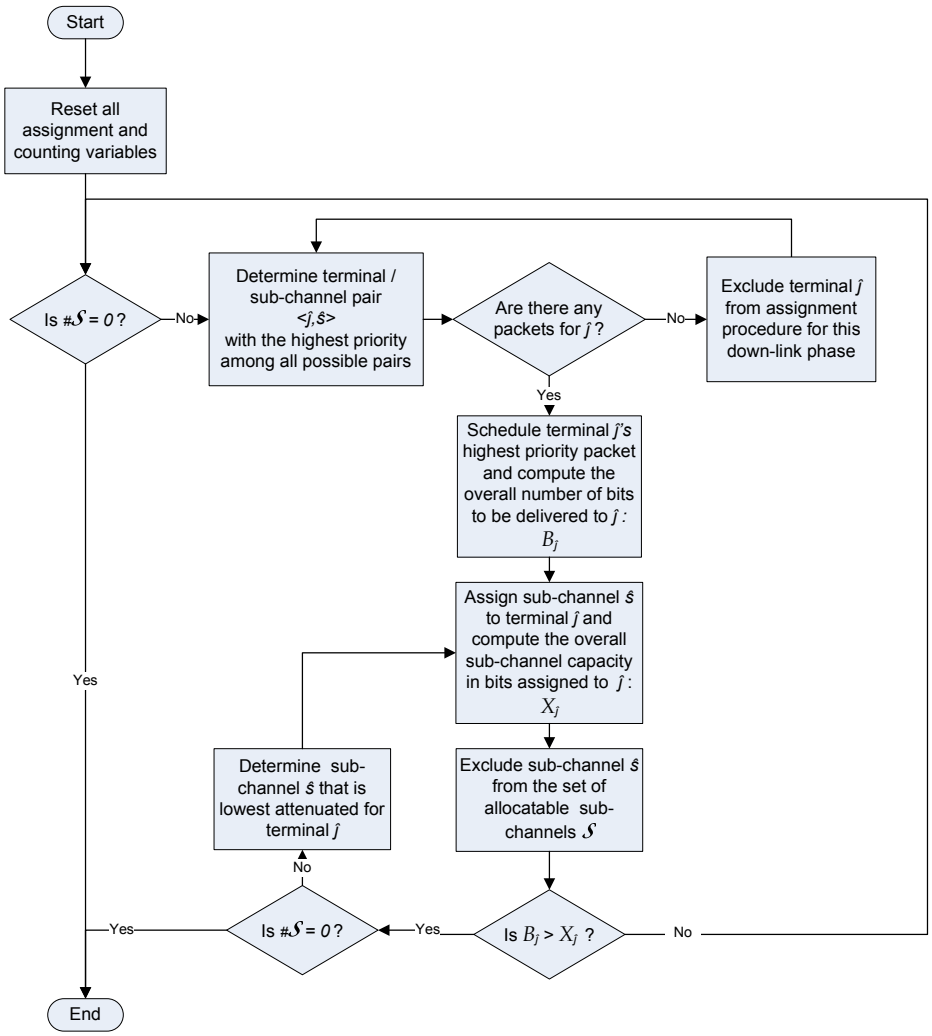


Figure 4.16: the sub-carrier/packet matching (SPM) algorithm

Instead of following the max-sum policy, the terminal priority can be chosen according to arbitrary priority approaches. An alternative proportional fair terminal selection policy can easily be integrated by substituting the equation in Step 3 with the following expression:

$$<\tilde{j}, \tilde{s}> = \underset{<j,s> \in \mathcal{J} \mathrm{x} \mathcal{S}}{\arg\max} \left(\frac{\chi(\gamma_{j,s}^{(t)})}{\bar{\vartheta}_{j,k}^{(t-1)}} \right) \quad . \tag{4.23}$$

Note that – in contrast to the optimal MILP decisions – the order with which the QoS classes are arranged in $\mathcal{Q}$ has a major impact on the heuristic scheduling results. As lower indices in $\mathcal{C}$ are preferred, the QoS classes have to be added to $\mathcal{C}$ in decreasing priority order. To further prioritize a certain QoS class, the objective definition in Step 3 can be modified accordingly.

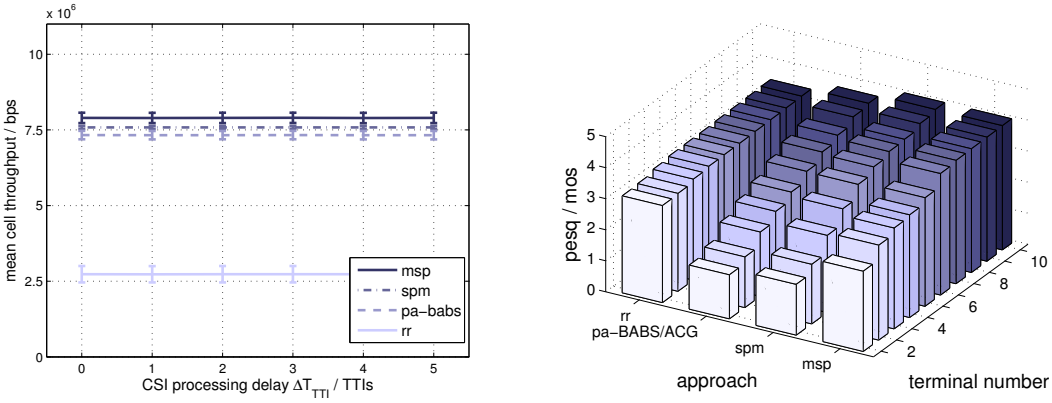
```
1  Initialize: ∀j, s, q, w : z^(t)_{j,q,w} = 0, x^(t)_{j,s} = 0, X^(t)_j = B^(t)_j = 0; q_j = 0, w_j = 0

2  while (#S > 0) do
3      <j̃, s̃> = arg max_{<j,s> ∈ JxS} (χ^(t)_{j,s})
4      q̃ = q_j̃; w̃ = w_j̃

5      while (ς^(t)_{j̃,q̃,w̃} == 0) do
6          q̃ ++; w̃ = 0
7          if (q̃ == Q) then
8              J := J \ {j̃}
9              goto Step 3;
           end
       end
10     z^(t)_{j̃,q̃,w̃} = 1
11     while (#S > 0) do
12         x^(t)_{j̃,s̃} = 1
13         X^(t)_j̃ = X^(t)_j̃ + χ^(t)_{j̃,s̃}
14         S := S \ {s̃}
15         if (ς^(t)_{j̃,q̃,w̃} ≤ X^(t)_j̃ − B^(t)_j̃) then
16             B^(t)_j̃ = B^(t)_j̃ + ς^(t)_{j̃,q̃,w̃}
17             w̃ ++
18             break;
           end
19         s̃ = arg max_{s∈S}(χ^(t)_{j̃,s})
       end
   end
```

Pseudo-Code Box 1: the sub-carrier/packet matching (SPM) algorithm

Performance comparions

In the following, the performance of the sequential and the integrated heuristic approaches is compared to the optimal solutions of the according integrated optimization approach as upper, and to the results of the simple round robin (RR) scheduling strategy as the lower bound.

Max-rate heuristics The performance results of the max-rate heuristics are shown in Figure 4.17. It is evident that the overall throughput performance of both heuristic approaches sequential packet-aware BABS (*pa-babs*), and max-rate SPM (*spm*) is quite close to the performance that is obtained, if the max-sum packet optimization problem (4.17) is optimally solved (*msp*). The throughput performance of the integrated SPM approach is slightly better than the one of the sequential approach, but both dynamic approaches perform significantly better than



(a) achieved cell throughput (bps)

(b) pesq values (mos)

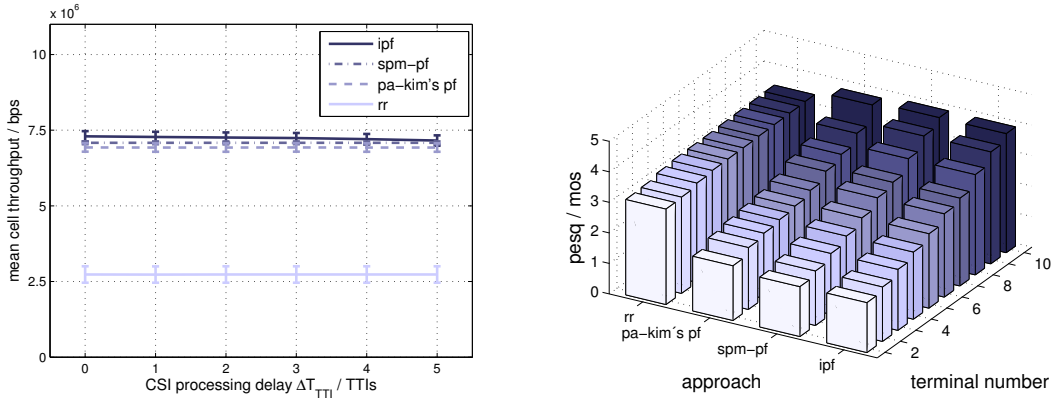
Figure 4.17: achieved cell throughput and PESQ values for heuristic SPM resource allocation

the non-adaptive round robin approach (*rr*): the gain is about 200%. In terms of PESQ performance some more obvious differences are present. Here, the simple RR approach performs quite good, as it regulalry reserves resource shares for all terminals. This is outperformed only by the optimal intgrated case, which, by perfectly matching the small VoIP packets to remaining sub-carrier resources achieves a PESQ performance beween *good* and *excellent* for almost all of the terminals. The integrated SPM heuristic yields excellent PESQ performance for the stronger half of the terminals. For the weaker half, the PESQ performance almost linearly decreasing. The same holds for the sequential BABS approach, except for the fact that only one third of the terminals experience *excellent* PESQ performance. Note again, that the quite good PESQ results of the integrated approaches are obtained by solely optimizing the max-packet delivery. By adding delay constraint (4.21) these PESQ results can even be improved.

Proportional fair heuristics Both heuristic proportional fair scheduling approaches, the sequential packet aware combination of Kim's PF and the simple packet scheduling approach (*pa-kim's pf*), as well as the integrated proportional fair SPM approach (*spm-pf*), perform close the optimum integrated solution (*ipf*), in terms of system throughput, as well as PESQ performance. As previously discussed, the PF policy causes the almost linear decline of the PESQ performance among the terminals for the sake of proportional fair resource distribution among the terminals. As is it the case with the max-rate heuristics, the PESQ performance can be improved by considering delay constraint (4.21).

4.4 Conclusions

The data destined for users of modern cellular systems is delivered in packets, which might occasionally arrive back-to-back, but most of the time are delivered in bursts of packets. This kind of data delivery is generally different from the way data is transmitted in the systems, for



(a) achieved cell throughput (bps)

(b) pesq values (mos)

Figure 4.18: achieved cell throughput and PESQ values for heuristic PF resource allocation

which the dynamic OFDMA resource allocation techniques have initially been developed.

In this chapter, the consequent research question '*Are special packet-aware resource allocation techniques necessary in packet-oriented OFDMA systems?*' has been considered. It has been shown that special packet-aware resource allocation techniques are not urgently necessary in packet oriented OFDMA systems. Standard resource allocation techniques used in systems with continuous data flows cannot directly be applied to those systems, but already the combination of according schemes with simple heuristics yields good results, such that the far more complex integrated resource allocation / packet scheduling resource optimization techniques do not perform significantly better. Still, it is evident that by applying specialized packet resource scheduling algorithms more attention can be payed to the QoS requirements of the individual packet streams, resulting in slightly improved quality of *e. g.,* voice transmissions. In general, however, the sequential application of slightly modified standard rate-aware dynamic mechanisms and simple packet scheduling algorithms works well in packet-oriented cellular OFDMA systems, such that for the following investigations of this thesis the emphasis is on the resource allocation task, while the packet scheduling aspect is left out.

Chapter 5

Optimized Control/Data Channel Capacity Split

IN the previous chapter it has been demonstrated, how dynamic orthogonal frequency division multiple access mechanisms can be used within the scheduling process to improve the system's QoS performance. It is, however, well known that the gain of applying adaptive mechanisms in OFDMA networks comes at the cost of additional signaling data, namely the resource allocation information that needs to be delivered to the receiver (*cf.* Figure 5.1). As a consequence, system resources in terms of bandwidth and/or time need to be provided to form some kind of control channel. Control channel reliability is a key issue in dynamic OFDMA system design, as an error on the control channel has a much more significant impact on the data transmission performance than an error on the data channel. Consequently, this chapter is devoted to answering the question:

> *Does reliable control data transmission require a disproportional split of the available transmission resources in favor of the control channel redundancy?*

As the control channel reliability is mainly impacted by the co-channel interference stemming from base stations of neighboring cells, observing a single cell – as has been done in the previous section – is not enough. Instead, the following investigations are performed in a multi-cell scenario. A limited set of different code rates is applied in order to answer the question of how much redundancy is needed to reliably deliver the resource allocation information (in terms of meeting a target block error rate) despite the impact of co-channel interference. The first goal is to determine the code rate that consumes the smallest amount of system resources among all considered code rates that allow for reliable control data transmission for a certain scenario. There is a trade-off when increasing the redundancy: on one hand the increase results in lower code rates and thus in an improved control data protection, whereas on the other hand a larger resource share is necessary and consequently the impact of CCI is increased. In a second step, power control and adaptive coding are examined as a possibility to improve the control channel's error characteristic. In this context, according optimization problems are formulated and sub-optimal heuristics for the dynamic power and adaptive coding approach are presented.

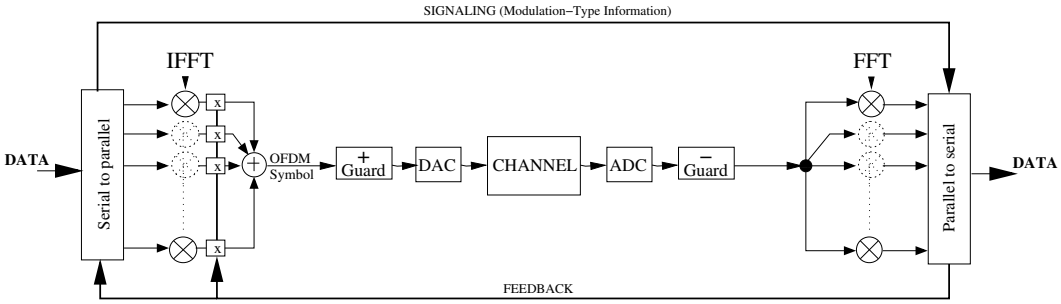


Figure 5.1: OFDMA system model with control data signaling feedback

5.1 Inband Control Channel

In general, throughout this chapter, a control channel model that includes significantly unreliable behavior is considered. The unreliability is due to the wireless channel's variability in time, frequency and space (*cf.* Section 2.2) that excludes fully reliable wireless transmissions.

Four different factors decide on the amount of system resources that are necessary to carry the signaling information to the receiving terminal:

1. the system setup, *i. e.,* the number of available sub-carriers/resource blocks, modulation types, coding rates, etc.,
2. the wireless channel's dynamic, as the necessary signaling frequency is proportional to the channel state changes,
3. the granularity of the dynamic assignments, and
4. the redundancy that protects the resource allocation information.

While in given network scenarios the first two factors are usually fixed, the latter two can be modified. Choosing a finer assignment granularity yields a better exploration of the system's frequency diversity, and thus a better system performance in terms of throughput. Selecting lower code-rates on the control channel accounts for more reliable control data delivery. This also has an impact on the user throughput, as user data is lost if the receiver holds erroneous allocation information. On the other hand both modifications increase the control channel's resource requirements and, thus, lower the system resources available for user data delivery. Hence, in order to optimize the system's performance in terms of user throughput, an optimal splitting of the overall amount of system resources into control channel and user data channel resources needs to be determined.

5.1.1 General structure

The general structure of the considered control channel follows the structure of 3GPP's physical downlink control channel (PDCCH) – the control channel that has been standardized in the

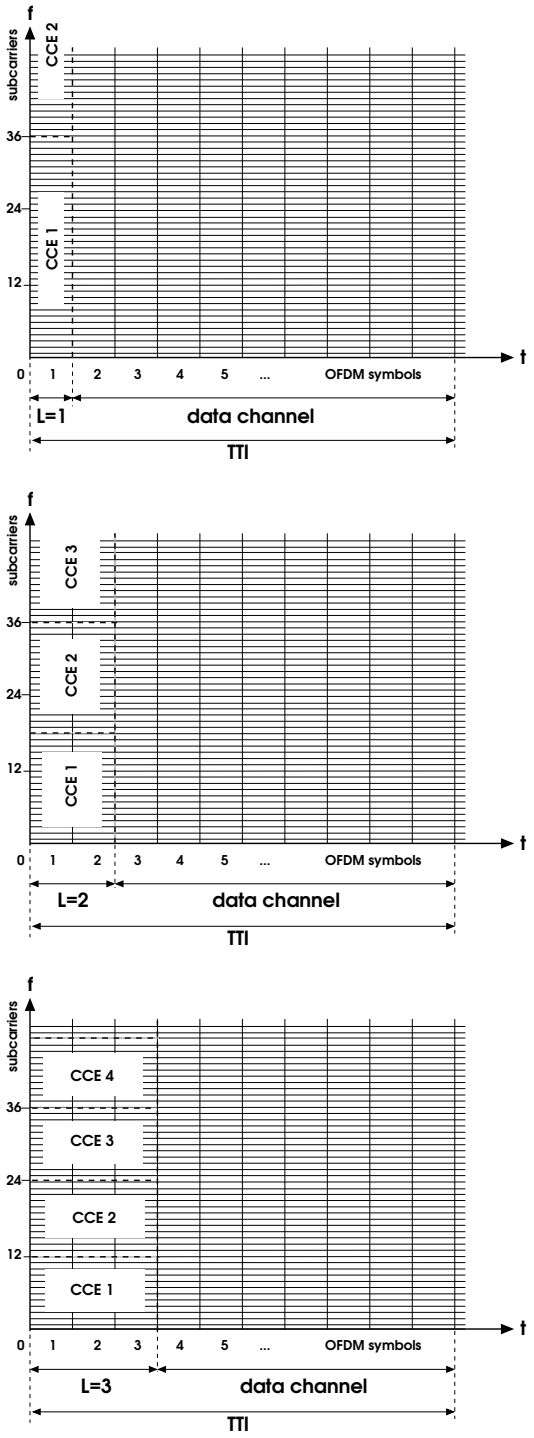


Figure 5.2: each control channel element consists of χ =36 REs, where one resource element spans one sub-carrier in frequency and one OFDM symbol in time. Depending on the control channel sizes $L \in \{1, 2, 3\}$, one CCE spans 12, 18, or 36 sub-carriers, determined by the function S_L

	Item	Size
1	C-RNTI (terminal id coded in CRC)	16 bits
2	resource indicator resource block selection)	12 bits
3	TBS and modulation type	8 bits
4	power control	4 bits
	in total	40 bits

Table 5.1: exemplary assignment item sizes according to the LTE standard [113]

context of the Long Term Evolution system standard [113]. Accordingly, the control information is multiplexed with the physical down-link data channel, occupying the first L OFDM symbols of each TTI, as shown in Figure 5.2. Hence, in each TTI an overall number of $L \times S$ REs is available for control signaling, where one RE spans one OFDM-symbol in the time domain and one sub-carrier in frequency. In order to lower the number of allocatable units, the REs are grouped into control channel elements (CCEs) of size χ REs. Note that one CCE spans all L OFDM symbols occupied by the control channel. Thus, the necessary number of sub-carriers to form a single CCE is a function of L:

$$S_L = \frac{\chi}{L} \ . \tag{5.1}$$

Also note that – in contrast to the way it is shown in Figure 5.2 – the S_L sub-carriers of a single CCE are non-adjacent, but spread across the frequency band in order to balance CCE channel quality (as described in [114]). For the ease of presentation, in the following the depicted way of CCE aggregation will be referred to as the *logical CCE domain*, whereas for all calculations and in the simulation, the spread *physical domain* CCE aggregation is utilized. The mapping between the logical and the physical domain is done with a logical-to-physical spreading function that differs among neighboring cells.

Signaling data multiplexing For each terminal that the base station scheduler has scheduled for data delivery in the upcoming TTI, one *downlink assignment* must be delivered in advance. It contains the terminal's id for identification purposes protected with a cyclic redundancy check (CRC) value (in the LTE context, the combination of terminal id and CRC is referred to as cell radio network temporary identity (C-RNTI)). Moreover, the assignment includes a resource indicator identifying the terminal's resource block assignments, the transport block size (TBS), as well as the id of the modulation type that is applied on the indicated resource blocks.

The resource indicator is a bitmap whose size equals the number of resource blocks in the system. A one in the bitmap means that the according resource block is assigned to the terminal that is the assignment's addressee (the terminal whose id is coded in the C-RNTI). Alternatively, in order to save control channel resources a smaller bitmap might be chosen that stands for only the upper or lower half of the resource block space. Then, if resource blocks of both halves are assigned to a single terminal, it needs to receive two assignments. Moreover, the resource blocks

of a single assignment are all modulated with the same modulation type. If the modulation types differ among the resource blocks, multiple assignments need to be sent. Note that the code-rate in use does not need to be signaled, as it can be derived from the other TBS value in combination with the modulation type and resource block assignments. In addition to the items above, hybrid automatic repeat request information might be included in the assignments.

The total size of a single downlink assignment Ψ_{dl} depends on the MAC's degree of flexibility. An exemplary assignment configuration is shown in Table 5.1. In the following, the according assignment size of 40 bits is assumed. In order to allow different code-rates on the control channel, a varying amount of CCEs ϕ_{cce} can be aggregated per assignment. The assignment code-rate ρ then is

$$\rho = \frac{\Psi_{\mathrm{dl}}}{2\phi_{\mathrm{cce}}\chi} \ . \tag{5.2}$$

Note that the overall number of assigned control REs ($\phi_{\mathrm{cce}}\chi$) is multiplied by 2 because quadrature phase shift keying (*cf.* Section 2.1) is consistently used as control channel modulation type. At the receiver side, blind detection mechanisms are used to allow different code rates (*i. e.,* CCE aggregations per assignment). As these mechanisms require a lot of computational power, the set of available code rates is limited.

Control channel load In general, the base station scheduler assumed for control channel reliability investigations is the multi-carrier proportional fair OFDMA scheduler introduced in Section 4.1.1. However, as the load on the control channel depends on the number of scheduled terminals, different scheduling strategies need to be used to simulate different control channel load settings in the same scenario. Thus, a functionality to control each scheduler's maximum resource share κ per terminal j and TTI has been included. For example, a choice of $\kappa = 1/3$ restricts a scheduler to provide a maximum of one third of all *data-channel resources* in this TTI to a single terminal. As a consequence, at least three terminals are scheduled, if available in the cell. Note that full buffers are assumed for all terminals in the system, such that there is a perpetual need for each terminal to be scheduled.

5.2 Basic Control Channel Investigations

First, the reliability of the basic control channel without dynamic OFDMA enhancements (as *e. g.,* standardized by the 3GPP in [113]) is investigated. In this basic setting, the maximum admitted transmission power $p_{\max}$ per cell c is equally split and distributed among all S subcarriers:

$$p_{c,s}^{(t)} = \frac{p_{\max}}{S}. \tag{5.3}$$

The control channel size L, as well as the number of CCEs per assignment ϕ_{cce} (and accordingly the per user code-rate ρ) are assumed to be flexible, but fixed in each simulation scenario.

Parameter	Symbol	Value
downlink assignment size in bits	Ψ_{dl}	40
control channel size in OFDM symbols	L	1,2,3
CCEs per assignment	ϕ_{cce}	1,2,4,8
CCE size in REs	χ	36
maximum resource share per TTI per user	κ	$\frac{1}{5}, \frac{1}{3}, 1$
maximum transmission power per cell	$p_{\max}$	46 dBm
power increment and decrement values	$\Delta p_{\mathrm{incr}} = \Delta p_{\mathrm{decr}}$	2 dBm
target block error rate	$\hat{R}_{\mathrm{bler}}$	0.01
CSI processing delay	ΔT_{csi}	1

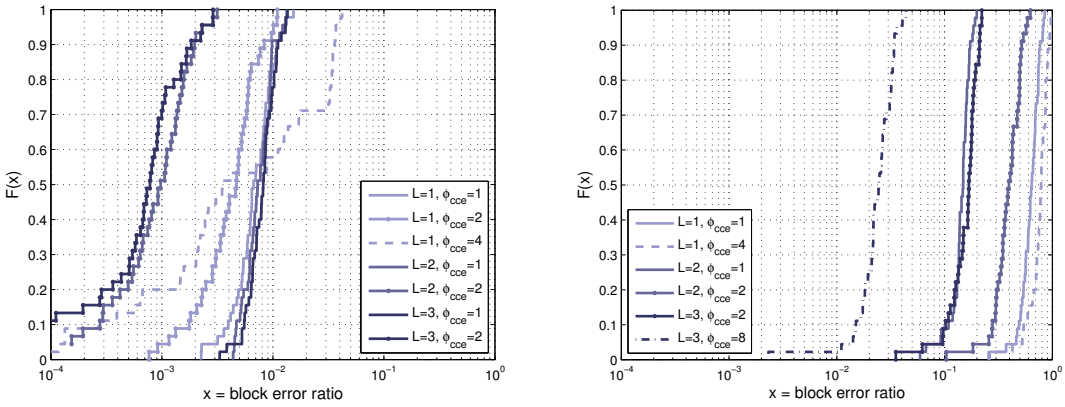
Table 5.2: control channel reliability simulation parameters

Note that only those terminals that have been selected for data delivery during the next TTI are assigned control channel resources. The number of scheduled terminals in cell c is referred to as J_c. The terminals are arranged in proper order according to their priority, which has previously been determined by the scheduler. Starting with the highest priority terminal ($j = 1$), the terminal/sub-carrier assignments in the logical domain of cell c are selected as follows:

$$\forall\, c : \begin{cases} x^{(t)}_{c,1,1} = \cdots = x^{(t)}_{c,1,\phi_{\mathrm{cce}} S_L} = 1, \\ x^{(t)}_{c,2,\phi_{\mathrm{cce}} S_L + 1} = \cdots = x^{(t)}_{c,2,2\phi_{\mathrm{cce}} S_L} = 1, \\ \vdots \\ x^{(t)}_{c,J_c,(J_c-1)\phi_{\mathrm{cce}} S_L + 1} = \cdots = x^{(t)}_{c,J_c,J_c \phi_{\mathrm{cce}} S_L} = 1, \end{cases} \tag{5.4}$$

all other $x^{(t)}_{c,j,s} = 0$. Note that the overall number of available CCEs per TTI is limited to S/S_L. In case there are too few CCEs available to create all necessary assignments, low priority terminals are excluded from control channel usage during this TTI.

Redundancy estimation In order to explore the reliability of the basic control channel configuration, a limited set of different code rates is applied, *i. e.,* the number of CCEs per assignment is varied. The main question is how much redundancy is necessary to reliably deliver the resource allocation information (in terms of meeting a target block error rate) despite the impact of co-channel interference. In other words, the code rate that consumes the smallest amount of system resources among all considered code rates that allow for reliable control data transmission for a certain scenario is to be determined.



(a) low control channel load scenario ($\kappa = 1.0$)

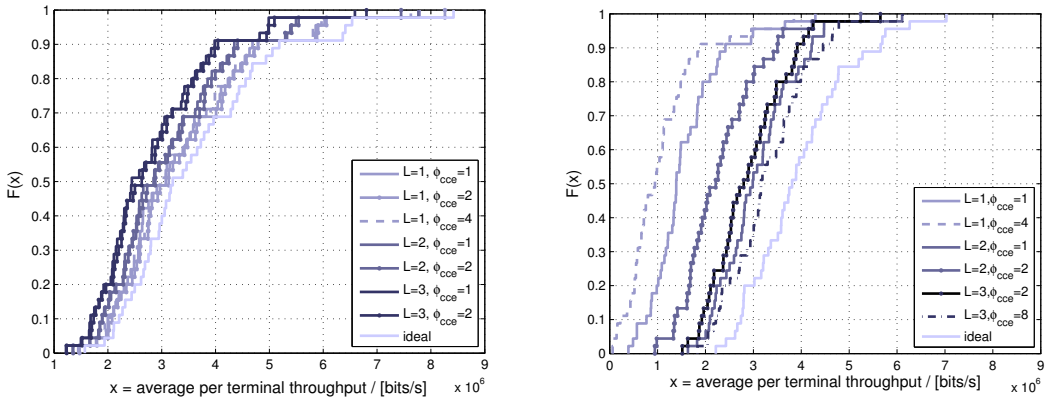
(b) high control channel load scenario ($\kappa = 0.2$)

Figure 5.3: Per terminal block error rate CDFs for selected low control channel load ($\kappa = 1.0$, in average 0.9 scheduled terminals per cell and TTI – left) and high control channel load ($\kappa = 1/5$, in average 4.5 scheduled users per cell and TTI – right) basic control channel configuration scenarios.

5.2.1 Results

The following control channel reliability investigation results have been obtained assuming the channel and system models introduced in Section 3.2. A high, a medium, and a low control channel load scenario (see Table 5.2) are studied. Among the different simulation scenarios, the number of CCEs per assignment ϕ_{cce}, as well as the control channel size L are varied. The system parameterization is selected according to an extended hexagonal multi-cell scenario featuring 9 cells and 45 terminals according to Table 3.5. The control channel specific link-level-performance model is presented in Section 3.2 as well. All additional simulation parameters are listed in Table 5.2.

Figure 5.3 shows the results for the selected scenarios featuring static power allocation. Low control channel load results with an average number of 0.9 scheduled users per cell and TTI are shown left, high load results with an average of 4.5 scheduled users per cell and TTI are shown right. From the result graphs of the low load scenario it can be stated, that 90% of all users have a block error rate that corresponds to the chosen per-user system target error probability $\hat{R}_{\text{bler}} = 0.01$, if one CCE per terminal is utilized ($\phi_{\text{cce}} = 1$, corresponding to a per-user code rate of approximately $\rho = 1/2$). While the results for different control channel sizes L are similar for $\phi_{\text{cce}} = 1$, they vary for $\phi_{\text{cce}} > 1$. Here, the block error rate is the smaller, the greater L is (compare the graphs for $\phi_{\text{cce}} = 2$). This is mainly due to the fact that the impact of CCI grows with the fraction of used sub-carriers, which is proportional to ϕ_{cce}, but anti-proportional to L. Note that for $L = 1$ the BLER gets worse when the redundancy is increased from $\phi_{\text{cce}} = 2$ to $\phi_{\text{cce}} = 4$ CCEs per assignment. Here, the coding gain does not balance the increased CCI.



(a) low control channel load scenario ($\kappa = 1.0$)

(b) high control channel load scenario ($\kappa = 0.2$)

Figure 5.4: Per terminal data throughput CDFs for selected low control channel load ($\kappa = 1.0$, in average 0.9 scheduled users per cell and TTI – left) and high control channel load ($\kappa = 1/5$, in average 4.5 scheduled users per cell and TTI – right) in the basic control channel configuration case. The curves that are labeled 'ideal' show the throughput values for according systems that rely on an ideal (error-free and zero-resource consuming) control channel.

This trend is even more visible in the graphs of the high load case. Here, the coding gain when switching from $\phi_{\text{cce}} = 1$ to $\phi_{\text{cce}} = 2$ in the $L = 2$ case – which accounts for better BLER performance at low load – leads to worse performance, whereas switching from $\phi_{\text{cce}} = 2$ to $\phi_{\text{cce}} = 8$ if $L = 3$ leads to the only scenario where 90% of the users have a BLER performance that is at least in the order of magnitude of the target BLER $\hat{R}_{\text{bler}}$.

Thus, it must be reasoned that the control channel size L, as well as the number of CCEs per assignment ϕ_{cce} must be carefully chosen with respect to the expected control channel load (that mainly depends on the scheduling strategy). Stronger code rates might lead to a worse BLER performance. Taking additionally the average terminal throughput values into account, which reflect the reliability versus efficiency trade-off, a rule of thumb can be derived: use a small amount of redundancy in combination with small control channel sizes and a large amount of redundancy with large ones.

5.3 Dynamic Power Assignment Approach

The general possibility to distribute the power in a terminal specific way exists for a system according to our system model, since it features a dedicated control-channel structure (the downlink assignments are sent individually to each user using dedicated per-user resources (*cf.* Section 3.2.4)). Accordingly, if the power is to be assigned dynamically, the transmission power on

sub-carrier s depends on the terminal it is assigned to:

$$p_{c,s}^{(t)} = \sum_{j=1}^{J_c} x_{c,j,s}^{(t)} p_j^{(t)} \ , \tag{5.5}$$

where $p_j^{(t)}$ is terminal j's transmission power per sub-carrier.

5.3.1 Optimization model

In contrast to the data channel optimization approaches introduced in Section 2.4, in the control channel optimization case, the goal lies neither in maximizing the control channel capacity nor in saving energy. Instead, the channel's reliability, *i. e.,* its resistance against errors is of major interest. In other words, the goal is to have as many terminals as possible meet the target block error rate $\hat{R}_{\text{bler}}$ in each TTI (Equation (5.6a)), while the maximum power constraint (Equation 5.6b) has to be met:

$$\max_{\mathbf{p}^{(t)}} \sum_{j=1}^{J_c} \delta_j^{(t)} \tag{5.6a}$$

$$\text{s. t.} \sum_{j=1}^{J_c} \sum_{s=1}^{S} x_{c,j,s}^{(t)} p_j^{(t)} \leq p_{\max} \quad \forall c \ , \tag{5.6b}$$

where $x_{c,j,s}^{(t)}$ is the terminal/sub-carrier assignment at time t, $p_j^{(t)}$ is terminal j's transmission power per sub-carrier, and $\delta_j^{(t)}$ is a binary auxiliary variable with

$$\delta_j^{(t)} = \begin{cases} 0, & \text{if } F_{\text{snir2bler}}(\tilde{\gamma}_{\text{eff},j}^{(t)}, \rho) > \hat{R}_{\text{bler}} \\ 1, & \text{if } F_{\text{snir2bler}}(\tilde{\gamma}_{\text{eff},j}^{(t)}, \rho) \leq \hat{R}_{\text{bler}} \ . \end{cases} \tag{5.7}$$

Note that, in contrast to the power optimization problems for data channel capacity maximization (as presented in Section 2.4) that maximize the capacity by assigning more power to good than to bad quality sub-carriers according to the water-filling theorem, here the power is distributed among the terminals, such that badly located terminals receive more power than terminals in a good position. This way a harmonization among the terminals is realized. Just as it is the case with the optimization problems related to data channel optimization, solving problem (5.6) requires immense computational power and base station signaling information exchange. In the next section, a sub-optimal low complexity heuristic that locally determines the terminal power values base station-wise in order to increase the overall system performance is presented.

5.3.2 Heuristic

The general idea of the dynamic power allocation algorithm for the control channel is to distribute the available power, such that as many terminals as possible are estimated to meet the target BLER $\hat{R}_{\text{bler}}$. It is executed at *each base station c individually*.

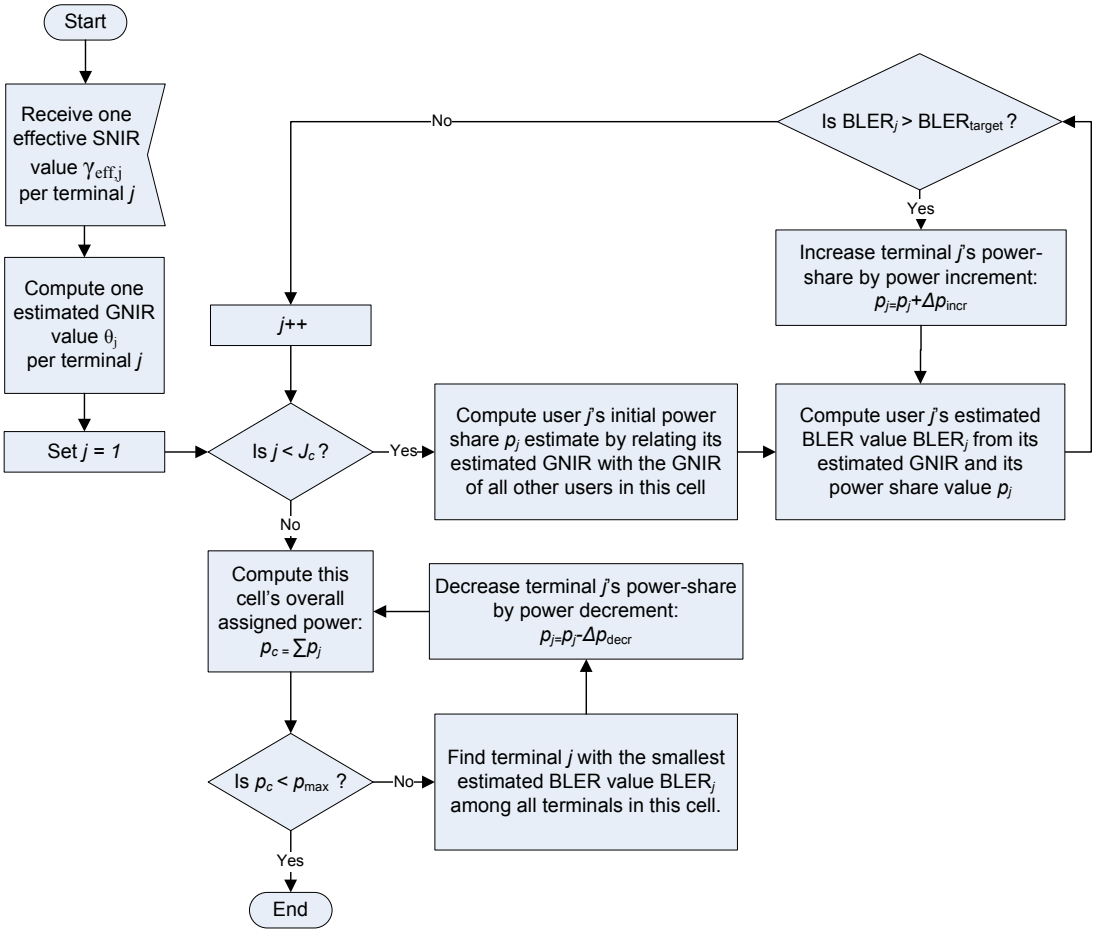


Figure 5.5: control channel dynamic power allocation algorithm flowchart

In a first step, an initial power distribution is calculated according to the experienced effective channel quality per terminal. Note that this initial distribution does not guarantee any terminal to meet $\hat{R}_{\mathrm{bler}}$, but serves as a good starting point for the power distribution. From that starting point, the power is adapted per terminal in small power increment Δp_{incr} and decrement Δp_{decr} steps, such that $\hat{R}_{\mathrm{bler}}$ is met for as many terminals as possible within the given power budget $p_{\max}$. The overall algorithm flow chart is depicted in Figure 5.5. Pseudo-Code Box 2 shows the details of the dynamic power control algorithm. Initially, each terminal j is assumed to be able to measure its momentary effective SNIR value $\gamma_{\mathrm{eff},j}^{(t)}$ and to signal it to its base station, where it arrives being ΔT_{csi} TTIs old. Third, the base station computes the corresponding gain-to-noise and interference ratio (GNIR) value $\bar{\theta}_j^{(t-\Delta T_{\mathrm{csi}})}$:

$$\bar{\theta}_j^{(t-\Delta T_{\mathrm{csi}})} = \frac{\gamma_{\mathrm{eff},j}^{(t-\Delta T_{\mathrm{csi}})}}{p_j^{(t-\Delta T_{\mathrm{csi}})}} \,. \tag{5.8}$$

for $(j = 1; j \leq J_c; j++)$ **do**

  $p_j^{(t)} = (\sum_{\substack{i=1 \\ i \neq j}}^{J_c} \bar{\theta}_i^{(t-\Delta T_{\text{csi}})} / \bar{\theta}_j^{(t-\Delta T_{\text{csi}})})(p_{\max} / S_L \phi_{\text{cce}})$

  while $\left(F_{\text{sinr2blep}}(\tilde{\gamma}_{\text{exp},j}^{(t)}, \rho) > \hat{P}_B\right)$ **do**

    $p_j^{(t)} = p_j^{(t)} + \Delta p_{\text{incr}}$

  end

end

$p_c^{(t)} = \sum_{j=1}^{J_c} \sum_{s=1}^{S} x_{c,j,s}^{(t)} p_j^{(t)}$

while $(p_c^{(t)} > p_{\max})$ **do**

  $j = \arg\min_{\forall j} F_{\text{sinr2blep}}(\tilde{\gamma}_{\text{exp},j}^{(t)}, \rho)$

  $p_j^{(t)} = p_j^{(t)} - \Delta p_{\text{decr}}$

  $p_c^{(t)} = \sum_{j=1}^{J_c} \sum_{s=1}^{S} x_{c,j,s}^{(t)} p_j^{(t)}$

end

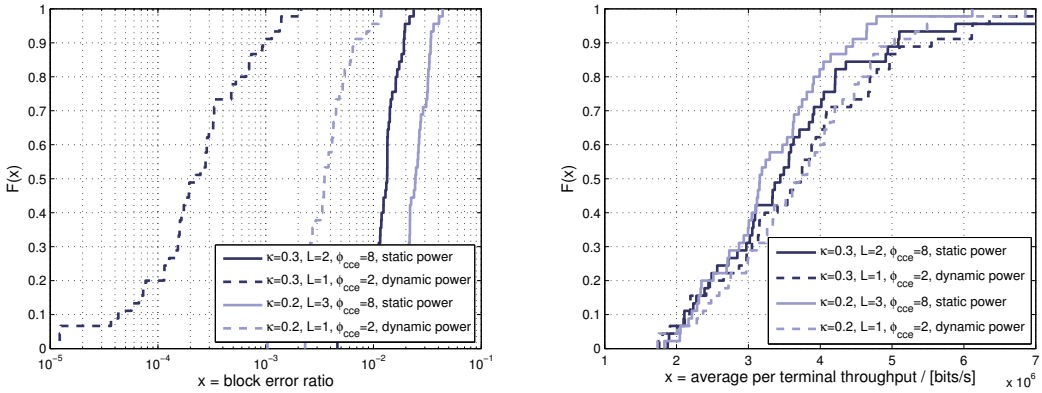
Pseudo-Code Box 2: control channel dynamic power allocation algorithm

Note that it is necessary for the base station to remember the power value $p_j^{(t-\Delta T_{\text{csi}})}$ of user j in order to determine the GNIR in due time. The set of all measured GNIR values per base station is used as input for the dynamic power control algorithm, which with these values tries to deliver a result as close to (5.6) as possible, while independently being executed in each cell c. In the following, the related pseudo-code in box 2 is explained in detail:

In Step 1 and 2, for each terminal j the per sub-carrier power value $p_j^{(t-\Delta T_{\text{csi}})}$ is estimated by putting its latest available average GNIR value in relation to the values of all other scheduled terminals in this cell and the overall power $p_{\max}$. Then, in Step 3, this estimated power value is used to compute the expected average SNIR for each terminal

$$\tilde{\gamma}_{\text{exp},j}^{(t)} = p_j^{(t)} \bar{\theta}_j^{(t-\Delta T_{\text{csi}})} \,, \tag{5.9}$$

and derive the user's expected BLER using function $F_{\text{sinr2blep}}$. Recall that ρ in Step 3 is the code rate that is derived from the fixed number of CCEs per user ϕ_{cce}. If the expected BLER as a function of the expected average SNIR is greater than the target BLER $\hat{R}_{\text{bler}}$, the terminals's per sub-carrier power value is incremented using the power-increment value Δp_{incr} until the requirement is met. In Step 5, the overall radiated power in the cell $p_c^{(t)}$ is computed. As this value needs to be smaller than the maximum admitted value $p_{\max}$, in the following Steps 6-9 power decrements are deducted from the per sub-carrier power values of those terminals that hold the lowest expected BLER values – until the requirement is met.



(a) block error rate CDF

(b) per terminal throughput CDF

Figure 5.6: The gain due to dynamic power allocation for the best performing combinations of control channel size L and number of CCEs per assignment ϕ_{cce} in the medium ($\kappa = 1/3$), and high ($\kappa = 1/5$) load scenario.

5.3.3 Results

The performance gain that stems from applying the dynamic power allocation algorithm is shown in Figure 5.6. Here, for the medium and high control channel load scenario those configurations are chosen that deliver the highest average user data throughput (*cf.* Figure 5.4). The average BLER curves show that by applying adaptive power allocation as means to control CCI, the system's BLER performance can be significantly improved:

All terminals in the medium load scenario, and 95% of the terminals in the high load scenario feature an average BLER beneath 10^{-2}. Moreover, note that dynamic power loading significantly lowers the required number of CCEs per assignment ϕ_{cce}. As a consequence, the control channel size L can be reduced and more resources provided to the data channel. Consequently, the average per user throughput is increased.

5.4 Dynamic Coding Approach

In the dynamic coding case, the control channel size L and the number of CCEs per assignment ϕ_{cce} (and accordingly the per terminal code rate ρ_j) are individually determined by each base-station per TTI. In order to enable the terminals to handle the varying control channel size, two bits per TTI are used to signal the control channel size L. Blind detection mechanisms allow for different combinations of code rates at the terminal side. However, since blind detection mechanisms require high computational power, the set of available code rates is limited by a limited set of available CCEs per assignment and terminal $\phi_{\text{cce},j} \in \{1, 2, 4, 8\}$ (in accordance with [114]). Again, note that only those terminals that have been selected for data delivery during the next TTI are assigned control channel resources.

5.4.1 Optimization model

Again, the primary optimization goal is to maximize the number of terminals that are estimated to meet the target BLER $\hat{R}_{\mathrm{bler}}$ (Equation (5.10a)). In addition, however, in order to maximize the data channel capacity, the primary goal has to be met with the smallest possible control channel size L. Hence, the control channel size has to be taken into account in the optimization goal as well:

$$\max_{\mathbf{X}^{(\mathbf{c})}} \sum_{j=1}^{J_c} \delta_j^{(t)} - \omega L \ , \tag{5.10a}$$

where $\mathbf{X}^{(c)}$ is the 3D terminal/sub-carrier per cell assignment matrix. Again, $\delta_j^{(t)}$ is the binary auxiliary variable as defined in Equation (5.7). Scaling factor ω decouples the primary and the secondary goal. By chosing ω smaller than the inverse of the greatest possible control channel size (in this case $\omega < 1/3$), it is guaranteed that the primary goal has more significance than the secondary goal. Moreover, it has to be assured that each sub-carrier is assigned to one terminal per cell only (Equation (5.10b)), that – since the CCE scheduling decision are done in the logical domain, *cf.* Section 5.1.1 – sub-carriers are assigned *en-block* only (Equation (5.10c)), and that – due to the blind detection contstraints — the number of CCEs per assignment ϕ_{cce} (*i. e.,* the per terminal code rate) is out of the set of available combinations (Equation (5.10d)):

$$\text{s. t.} \qquad \sum_{j=1}^{J_c} x_{c,j,s}^{(t)} \leq 1 \qquad \forall\ c, s \tag{5.10b}$$

$$if\ [x_{c,j,s}^{(t)} == 1 || (s \bmod S_L) == 1]$$
$$then\ x_{c,j,s-S_L+1}^{(t)} = \ldots = x_{c,j,s-1}^{(t)} = 1 \qquad \forall\ c, j, s \tag{5.10c}$$

$$\tfrac{1}{S_L} \sum_{s=1}^{S} x_{c,j,s}^{(t)} \in \{1, 2, 4, 8\} \qquad \forall\ c, j \tag{5.10d}$$

Remember that each CCE consists of the same number of REs, such that the number of sub-carriers per CCE S_L depends on the control channel size L. As in the dynamic power optimization case, optimization problem (5.10) can't be solved in reasonable time by a base station scheduler. In the following, a heuristic alternative that delivers sub-optimal dynamic coding assignments is presented.

5.4.2 Heuristic

The general idea of the dynamic coding algorithm for the control channel is to distribute the available CCEs, such that as many terminals as possible are estimated to meet the target BLER $\hat{R}_{\mathrm{bler}}$, while keeping the control channel size L as small as possible. The available transmission power per base station $p_{\max}$ is equally split among the sub-carriers, *i. e.,* $p_j^{(t)} = p_{\max}/S$. The algorithm is executed at each base station individually.

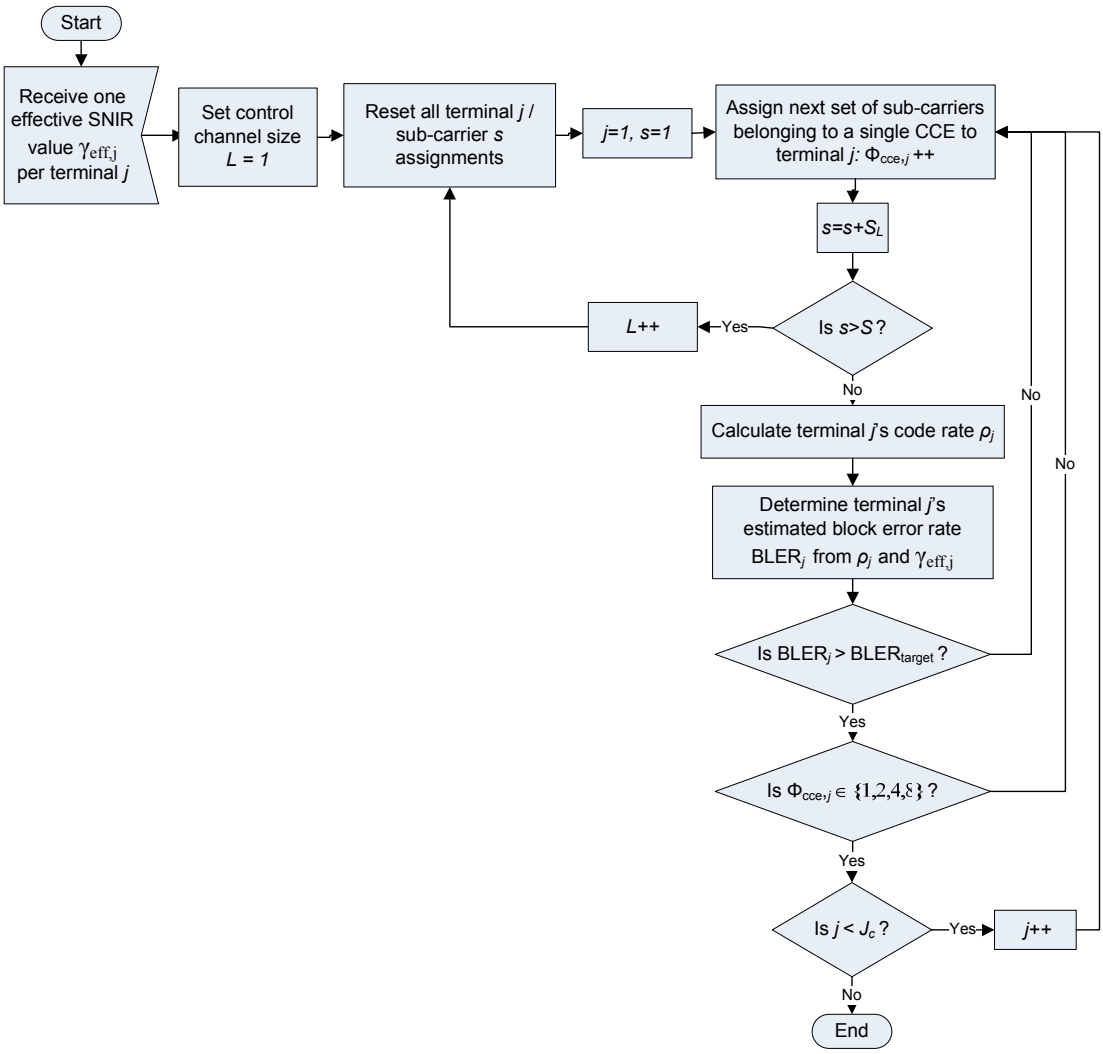


Figure 5.7: control channel dynamic coding algorithm flowchart

Pseudo-Code Box 3 shows the details of the dynamic coding algorithm. Initially, the smallest available control channel size $L = 1$ is selected. As a consequence, each CCE is composed of S_L sub-carriers. In Step 2 and 3, all user/sub-carrier assignments in this cell c are reset and the overall number of assigned sub-carriers s_{ass} is set to zero. Then, inside the loop that comprises Steps 5–13, for each terminal assignment as many CCEs as necessary to achieve an expected block error rate smaller than the target BLER $\hat{R}_{\mathrm{bler}}$ are assigned. Note that in each iteration the sub-carrier counter s is increased by the number of sub-carriers per CCE S_L (Step 5). Thus, in each iteration a complete CCE is assigned in Step 6. In Step 7 the momentary code rate $\rho_j^{(t)}$ of terminal j is calculated according to Equation (5.2).

Using code rate $\rho_j^{(t)}$ and the ΔT_{csi} TTIs old average effective SNIR value $\gamma_{\mathrm{eff},j}^{(t-\Delta T_{\mathrm{csi}})}$, which has previously been provided by terminal j, the expected BLER is determined by function

Initialize: $isDone = false, p_j^{(t)} = (p_{\max}/S) \quad \forall\, j$

1 **for** $(L = 1; L \leq 3; L++)$ **do**

2 $\quad \forall\, \{j, s\}$ in this cell c: $x_{c,j,s}^{(t)} = 0$

3 $\quad s_{\text{ass}} = 0$

4 $\quad$ **for** $(j = 1; j \leq J_c; j++)$ **do**

5 $\quad\quad$ **for** $(s = 1; s \leq (S - s_{\text{ass}}); s = s + S_L)$ **do**

6 $\quad\quad\quad x_{c,j,s+s_{\text{ass}}}^{(t)} = \cdots = x_{c,j,s+s_{\text{ass}}+S_L}^{(t)} = 1$

7 $\quad\quad\quad \rho_j^{(t)} = \Psi_{\text{dl}}/2L(s + S_L)$

8 $\quad\quad\quad$ **if** $\left(F_{\text{snir2bler}}(\gamma_{\text{eff},j}^{(t-\Delta T_{\text{csi}})}, \rho_j^{(t)})) \leq \hat{R}_{\text{bler}}\right)$ **then**

9 $\quad\quad\quad\quad$ **if** $(((s + S_L)/S_L) \in \{1, 2, 4, 8\})$ **then**

10 $\quad\quad\quad\quad\quad s_{\text{ass}} = s_{\text{ass}} + s + S_L$

11 $\quad\quad\quad\quad\quad$ **if** $(j == J_c)$ **then**

12 $\quad\quad\quad\quad\quad\quad isDone = true$

$\quad\quad\quad\quad\quad$ **end**

13 $\quad\quad\quad\quad\quad break$

$\quad\quad\quad\quad$ **end**

$\quad\quad\quad$ **end**

$\quad\quad$ **end**

14 $\quad$ **if** *(isDone)* **then**

15 $\quad\quad break$

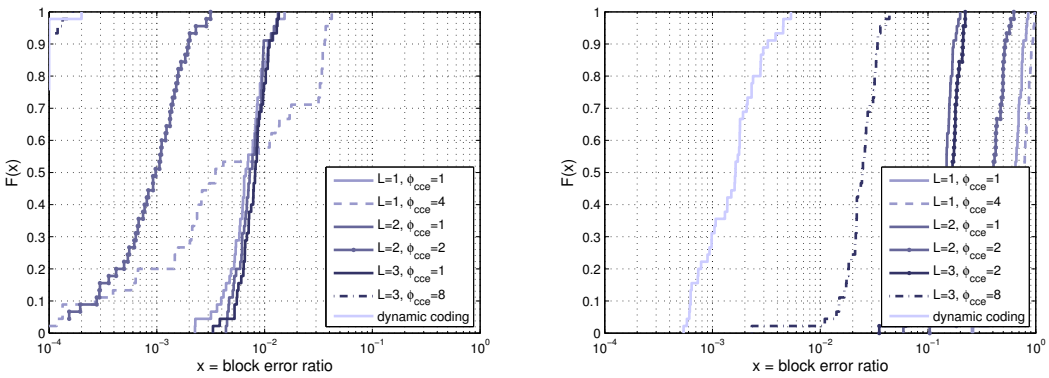
$\quad$ **end**

end

Pseudo-Code Box 3: control data dynamic coding algorithm

$F_{\text{snir2bler}}(\gamma_{\text{eff},j}^{(t-\Delta T_{\text{csi}})}, \rho_j^{(t)})$. It is compared to the target BLER $\hat{R}_{\text{bler}}$ in Step 8. If the expected BLER is larger than the target BLER value, another CCE is assigned by repeating Steps 5 through 7. If it is smaller, it is checked in Step 9, whether a valid number of CCEs has been assigned to j. Recall that there is a limited set of available code rate (CCEs per assignment) in order to keep the complexity of the blind detection mechanisms at the receiver side low. If an invalid number of CCEs has been assigned, additional CCEs are selected. Once a valid number of CCEs yields an expected BLER that is smaller than the target BLER $\hat{R}_{\text{bler}}$, terminal j's assignment is done. Then, in Step 10 the overall number of assigned sub-carriers s_{ass} is updated accordingly. If j is the terminal of least priority (*i. e.,* the last terminal to be scheduled), the process is marked as finished using the $isDone$ flag in Step 12, before the loop is broken in Step 13.

Arriving at Step 14, two different states need to be distinguished: either all terminals fulfill



(a) low control channel load scenario ($\kappa = 1.0$)

(b) high control channel load scenario ($\kappa = 0.2$)

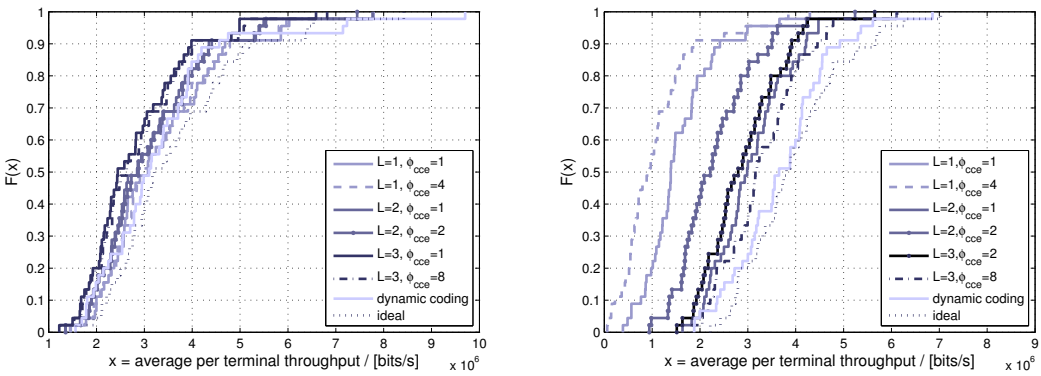
Figure 5.8: The gain in BLER due to dynamic coding. The CDFs of two different control channel load cases are illustrated: low control channel load ($\kappa = 1.0$, in average 0.9 scheduled users per cell and TTI) and high control channel load ($\kappa = 0.2$, in average 4.5 scheduled users per cell and TTI).

the requirement that their expected BLER is smaller than the target BLER- in this case the assignment process is finished and the main loop is broken (Step 15), or the momentary available number of CCEs is too small to meet all terminals' requirements. In the latter case, the complete process is repeated with an increased control channel size L. In the very rare case that even for $L = 3$ there is one or more users that do not get enough resources, those users are neglected.

5.4.3 Results

Figures 5.8 and 5.9 show the results of the dynamic coding approach, as well as the results of selected static coding scenarios. Low control channel load results with an average number of 0.9 scheduled terminals per cell and TTI are shown left, high load results with an average of 4.5 scheduled terminals per cell and TTI are shown at the right side of the respective Figure. The static coding parameterization that shows the best BLER performance in the respective load case have been chosen for comparison.

From the low load BLER graph in Figure 5.8a it can be seen that the dynamic coding BLER performance approximately corresponds to the performance of the strongest static coding case ($L = 3, \phi_{\text{cce}} = 8$): more than 90% of all terminals experience a block error rate below 10^{-4}. In the high load case (Figure 5.8b), the dynamic coding approach significantly outperforms the strongest coding case. This is mainly due to the fact that the CCI generating impact of additional redundancy is minimized, if the coding rate is adapted on a per terminal basis. In other words, the probability of two sub-carriers being used in neighboring cells is much lower than in the strongest static case, whereas the probability of having enough redundancy in order to decode the signaling information at the receiver is much higher than in the weakest coding case. The gain in BLER performance, as well as the smaller average control channel size L translate into an



(a) low control channel load scenario $\kappa = 1.0$

(b) high control channel load scenario $\kappa = 0.2$

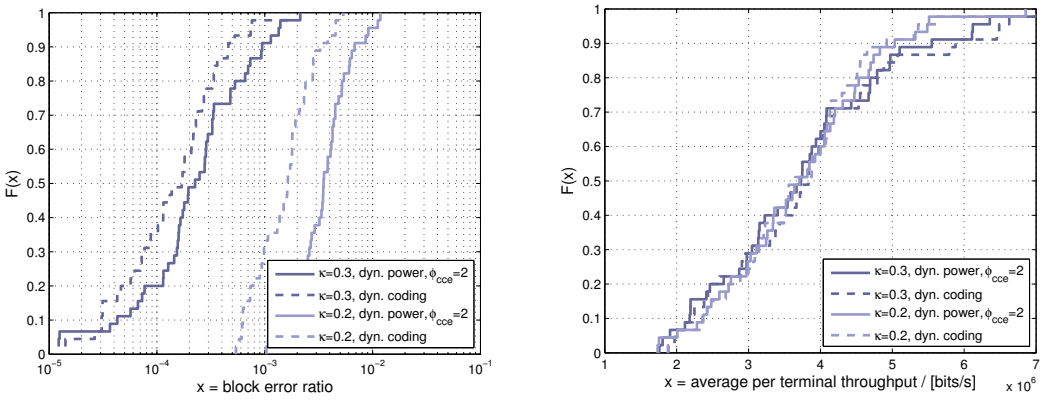
Figure 5.9: The gain in per terminal throughput due to dynamic coding. The curves that are labeled 'ideal' show the throughput values for according systems that rely on an ideal (error-free and zero-resource consuming) control channel.

average per user throughput that is close to the ideal case, as shown in Figure 5.9. Note, that for comparison purposes a system that relies on an ideal (error-free and zero-resource consuming) control channel (referred to as *ideal*) is additionally depicted. In the high control channel load scenario, the achievable per terminal throughput when adaptive coding is applied gets close to this ideal case.

Since in the low load scenario the weak static coding approaches with small control channel sizes (*e. g.,* $L = 1, \phi_{\text{cce}} = 1$) deliver throughput results that are as good, the throughput gain due to dynamic coding is not that significant. Explicit gains are present for the best 10% of the users only. Some other might even experience slightly worse performance. This is mainly due to the fact that in the low load scenario all shown combinations perform well in terms of BLER performance.

5.5 Dynamic Power vs. Dynamic Coding

In Figure 5.10 the performance of the dynamic power allocation and coding approaches presented in the last two sections are compared. For the dynamic power case, the coding configuration ($L = 1, \phi_{\text{cce}} = 2$) that delivers the best BLER performance in the medium and high load scenarios has been chosen. From the graphs it can be seen that even though the adaptive coding approach delivers slightly better BLER results, both approaches perform almost the same in terms of average per user data throughput. As a consequence, at this point there is no reason to prefer one of the two when it comes to implementing an according system. However, as mentioned before, the adaptive coding algorithm depends on blind detection at the receiver side, whereas the adopted dynamic power algorithm relies on information storage and processing at the transmitter side. Moreover, the two algorithms consist of algorithmic steps of significantly



(a) block error rate CDF

(b) per terminal throughput CDF

Figure 5.10: The difference in block error rate and per terminal data throughput between the dynamic coding and the dynamic power approach. The graphs show the results for the medium ($\kappa = 1/3$), and the high ($\kappa = 1/5$) load scenario.

differing computational complexity. Depending on the system parameterization, thus, one or the other approach might be the better choice. For an LTE-like system, that features powerful base stations and – in comparison to the base station – relatively weak terminals, the application of the dynamic power mechanism seems to be the better choice, as it does not increase the blind detection complexity.

5.6 Conclusions

The gain of applying adaptive mechanisms to the data channel of orthogonal frequency division multiple access networks comes at the cost of additional signaling data, namely the resource allocation information that needs to be delivered to the receiver. System resources in terms of bandwidth and/or time need to be provided to form some kind of control channel. Control channel reliability is a key issue in dynamic OFDMA system design, as an error on the control channel has a much more significant impact on the data transmission performance than an error on the data channel.

In this chapter, the consequent research question '*Does reliable control data transmission require a disproportional split of the available transmission resources in favor of the control channel redundancy?*' has been considered. For this purpose, OFDMA system performance in terms of per terminal throughput subject to control channel reliability has been investigated. It has been shown that a concept of a time-multiplexed physical control channel that uses the first L OFDM symbols per TTI can be used, even in a frequency reuse 1 environment. The coding, however, has to be quite strong, which means that a relatively high fraction of the system resources are used for the control channel, making the system less efficient. In order to increase

the system efficiency, dynamic power allocation and dynamic coding approaches that adapt the control channel resources to the individual terminal situations have been developed. According optimization models have been introduced and heuristic algorithms that deliver sub-optimal problem solutions at low computational complexity have been presented. It has been shown that the application of both approaches significantly increases the system performance in terms of per terminal throughput. In both cases, a throughput close to the ideal case featuring a reliable, non-resource consuming control channel is achieved. Moreover, as the difference in performance gain in negligibly small, the usage of either of the two mechanisms is suggested depending on the system configuration.

In summary, reliable control data transmission in a cellular OFDMA based system is possible in a frequency reuse 1 scenario, even if only a relatively small part of the transmission resources is reserved to implement the control channel.

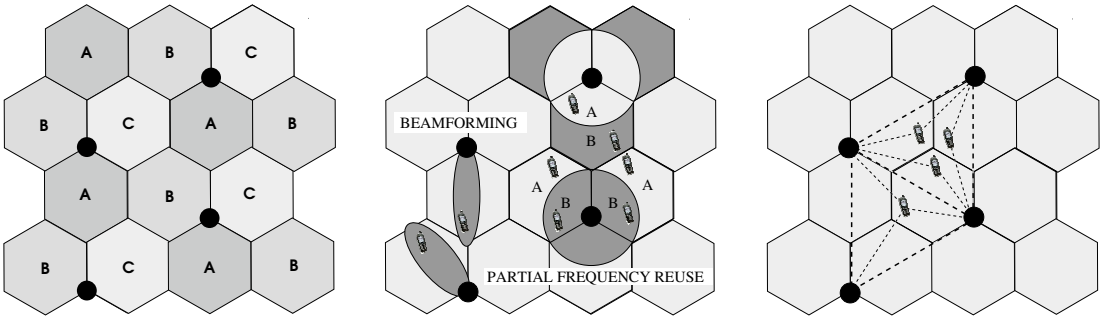
Chapter 6

Interference Mitigation for Frequency Reuse 1

A KEY ISSUE with the application of orthogonal frequency division multiple access in mobile cellular systems is co-channel interference. Especially terminals located at the cell border largely suffer from the power radiated by the base station of neighboring cells in their communication band. Consequently, some kind of co-channel interference coordination (CCIC) technique (also referred to as inter-cell interference coordination (ICIC) [115], or interference coordination (IFCO) [116] techniques) is required to reduce the interference and to enable also the worst situated terminals to receive data. A standard interference coordination technique is to divide the overall system bandwidth in frequency bands and to use distinct frequency bands in neighboring cells following certain frequency reuse patterns. This reuse pattern scheme, however, is static and mostly adjusted before the system is put into operation. It is, thus, known to be inefficient. An alternative approach for the more flexible modern systems is to allow the usage of the complete spectrum in each cell and to integrate the interference coordination and the resource allocation task at each base station. So far, however, the general feasibility of this approach has not been investigated. Consequently, this chapter is devoted to answering the question:

> *Can any performance gain be expected, if neighboring cells jointly optimize their usage of transmission resources?*

Particularly, the potential of dynamic power allocation and sub-carriers assignment mechanisms (*cf.* Section 2.4) to mitigate CCI in cellular OFDMA systems is explored. First, an overview of existing co-channel interference coordination techniques in cellular systems is provided. Among them, interference aware scheduling is presented as a medium complex, but efficient means to realized CCI mitigation. Particularly, soft frequency reuse (SFR) is introduced as computationally cheap yet efficient CCI mitigation technique. Then, a scheduler employing soft frequency reuse is compared to an optimally working fully dynamic resource allocation enabled scheduler. It is shown that there is a high potential to mitigate interference by SFR, but that there is still a significant gap to the optimal approach, where power and frequency shares are distributed optimally. Consequently, an interference-coupling based adaptive



(a) hard frequency reuse (FR) (b) interference aware scheduling (c) coordinated multi-point (CoMP)

Figure 6.1: interference coordination techniques in cellular networks

frequency reuse (AFR) optimization approach is developed as an opportunity to dynamically create SFR power profiles that approach optimality.

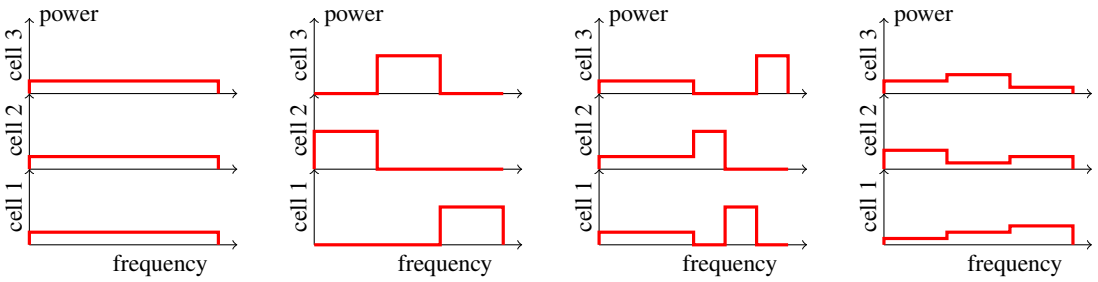
6.1 Co-channel Interference Coordination in Cellular Systems

Schemes that coordinate co-channel interference in cellular networks can generally be grouped in three categories: hard frequency reuse (HFR), interference aware scheduling, and coordinated multi-point transmission (CoMP) approaches (*cf.* Figure 6.1). While the first is a well known, easy to implement non-efficient approach, the latter two are highly efficient and highly complex, CoMP being the most efficient and complex among them. Moreover, the latter two rely on the availability of up-to-date channel state information.

6.1.1 Hard frequency reuse

In legacy cellular systems, the major alternative for mitigating CCI was hard frequency reuse (HFR), where neighboring cells do not use the same frequency sub-band. Frequency sub-bands are reused only outside a cluster, which typically has a size of 3 to 15 cells. The *frequency reuse factor* r_f gives information about the cluster's size. A reuse factor of *e. g.,* $r_f = 3$ splits the system bandwidth in 3 sub-bands (A,B,C), all of which are used by exactly one cell of each cluster of size 3 (*cf.* Figure 6.1a). While this is a very simple and easy to implement way to realize protection against interference, it obviously is not very efficient to use only a fraction of the overall bandwidth in each cell.

Nowadays, wideband CDMA-based third generation cellular systems use the overall available spectrum in each cell, adopting a frequency reuse 1 scheme. Such systems cope with CCI by means of signal spreading/despreading, channel encoding, and fast transmit power control. The application of these mitigation techniques is intrinsically tied to the inherent properties of CDMA, and cannot be applied to OFDMA systems such as LTE and WiMAX. Still, a reuse factor of one is preferable. It is also feasible in such systems, as the radio resource management and channel coding can cope with some interference. The resource assignment decisions



(a) uniform (b) hard reuse 3 (c) fractional reuse 3 (d) soft reuse 3

Figure 6.2: example HFR and SFR power profiles featuring 3 subbands

then need to take the current CCI situation per terminal into account. The following approaches make use of the very properties of OFDMA systems, such as allowing users to be addressed on certain parts of the frequency only, and multiple transmit antenna solutions that are possible due to much less complex signal processing at the receiver than in CDMA systems.

6.1.2 Interference aware scheduling

Assuming advanced CSI signaling opportunities, the impact of co-channel interference on the individual terminals can be estimated at the base stations. Using this knowledge, base stations can schedule transmissions such that the impact of interference is minimized. There are in general three dimensions in which the terminals can be scheduled: in space, time, and frequency. For a cell to be able to schedule their terminals in the spatial domain, it is mandatory for it to possess an array of adaptive transmit antennas. In conjunction with a specialized processor, the power over the array can be masked such that it is radiated in a lobe (a beam) pointing to the individually scheduled terminal, as shown in Figure 6.1b. The figure shows also an alternative approach: partial frequency reuse (PFR). With PFR, the base station divides the frequency band into low and high interference regions by means of power profiling. It then schedules highly interference impacted terminals (*i. e.,* cell-edge terminals) in the low interference regions. This partial frequency reuse approach requires FDMA as multiple access technique, but is implementable independently of the number of transmit antennas and is computationally less complex. In the following, both approaches are explained in more detail, and an overview of state-of-the-art approaches is provided.

Partial frequency reuse based scheduling

While HFR splits the system bandwidth into a number of distinct sub-bands according to a chosen reuse factor and lets neighboring cells transmit on different sub-bands, partial frequency reuse applies a frequency reuse factor of one to terminals located in the cell's center. Only for terminals in the vicinity of the cell edge a frequency reuse factor greater than one applies. PFR schemes can be sub-divided into fractional frequency reuse (FFR) and soft frequency reuse (SFR) schemes. FFR [117] splits the given bandwidth into an inner and an outer part. The inner

part is completely reused by all base stations; the outer part is divided among the base stations with a frequency reuse factor greater than one. The assignment of terminals to reuse 1 or reuse 3 areas can be done based on the terminal's distance to the base station, or on the instantaneous SNIR situation. With soft frequency reuse [118], the overall bandwidth is shared by all base stations (*i. e.,* a reuse factor of one is applied), but for the transmission on each sub-carrier the base stations are restricted to a certain power bound. Again, the decision on the part of the frequency band on which the individual terminals are scheduled can be based on its location or interference situation.

The PFR approaches, as well as HFR can be described in terms of cell-specific *power profiles* over the system bandwidth. A power profile prescribes the fraction of the maximum transmit power that the base station may use depending at the part of the spectrum. In Figure 6.2 a scenario of three neighboring cells is assumed. In the case of HFR (Figure 6.2b), the power profiles block all but one third of the spectrum. In the fractional frequency reuse example (Figure 6.2c), the power profile for the first half of the spectrum is uniform, and the second half corresponds to a condensed version of the hard reuse case. Figure 6.2d illustrates the soft frequency reuse case.

Adaptive Frequency Reuse The application of soft frequency reuse schemes can significantly improve cell-edge throughput in cellular systems [119]. Initially, power profiles have been assumed to be of static nature and hence do not adapt to the current traffic situation. Furthermore, radio network engineers need to coordinate the power profiles such that the cell-edge users receive little interference from neighboring cells. Lately, *e. g.,* [120], the possibility is explored to dynamically adapt power profiles in the context of a self-organizing network (SON). The process of adapting power profiles is commonly referred to as *adaptive* soft frequency reuse, or short *adaptive frequency reuse (AFR)*. Mitigating co-channel interference in OFDMA networks by applying AFR is currently under investigation by different groups.

Promising AFR approaches have been presented in [121] and [122] (here referred to as *Self-organizing soft reuse*). While in the first paper, voice over IP (VoIP) constant bit rate (CBR) traffic is assumed, the latter considers power profile adaptation for best effort (BE) traffic. In both cases, the power level per sub-band is adjusted with respect to terminal utility in algorithms that are executed per cell. While in the CBR case, dynamic sub-carrier allocation is not employed (interference-averaging by sub-carrier hopping is applied instead), in the BE case opportunistic proportional fair scheduling is applied. Two different algorithmic approaches are considered: a sector autonomous and a co-operational case, in which sectors exchange some information about how their utility changes depending on modifications in the power profile. It is shown that both approaches achieve significant cell-edge data performance gains.

The authors of [123] suggest a scheme, where power profiles are adapted based on the buffer fill standings observed at each base station. The authors of [124] show benefits for adaptive over static fractional reuse schemes. In [125], the authors propose a linear model as well as an iterative procedure to compute the power profiles based on achievable capacity in a distributed way. Both approaches are analyzed in combination with a max sum rate and a fairer equal max rate scheduling algorithm, which improve system and cell-edge user performance, respectively.

Beamforming based scheduling

To supply a terminal with data via antenna array generated beams, beams can either be steered to the terminal's position, or beams that point to the vicinity of the terminal's position are selected out of a static set of beams. For both alternatives a short review of a representative example is given below:

Interference-graph coloring Necker introduces an interference-graph based global interference coordination scheme with full system knowledge in [126]. It requires beamforming antennas that can be steered towards each terminal with an accuracy of 1 degree, and it is assumed that all terminals can be tracked ideally. Interference-graph based scheduling generally consists of two steps: (1) creation of an interference graph (a graph is created based on the interference relations between the terminals, where the nodes represent the terminals, and the edges represent critical interference relations), and (2) resource assignment (a global scheduler assigns the resources to the terminals, taking the constraints of the interference graph into account; in particular, terminals which are connected must not be served using the same beams and/or frequency regions). The interference graph is constructed by evaluating the interference that a transmission to one terminal causes to any other terminal. For each terminal, the total interference is computed and then the largest interferers are set into relation via an edge in the graph. Within the resource allocation process, terminals in each cell are scheduled in a round robin fashion, where each terminal has highest scheduling priority in exactly one of a group of J_c consecutive TTIs (where J_c is the number of terminals in cell c). In the respective following TTI, it has second highest priority, and so on. The global scheduler starts the resource assignment process at a randomly chosen cell (sector). All available resources in this sector are assigned to the terminal with the highest priority. In the randomly selected second cell, all non-interfering resources are assigned to the highest priority terminal of that cell. This process is repeated for all cells. In a second round, in each cell the remaining resources are assigned to the terminal of second highest priority in a similar way. The process is repeated for the low priority terminals in descending order until all resources in all sectors have been assigned, or if no more assignments are possible due to conflicts in the interference graph.

In later works [127], Neckar presents a combined approach of interference-graph coloring with FFR techniques to realize local scheduling subject to spatially limited interference coordination (*i. e.,* missing global interference knowledge), as well as means to split the graph coloring problem into an outer part, which is solved by a central coordinator, and an inner part, which is solved per sector by each base station individually using simple genetic algorithms [128]. It is shown by means of simulation that in all investigated scenarios, the application of beamforming antennas significantly improves the throughput performance compared to legacy HFR3 based systems. In the latest version of Neckar's graph coloring approach, a 100% cell-edge throughput gain is possible compared to such systems.

Spatial mode switching Schellmann *et al.* present an instantaneous fairness based algorithm in [129, 130]. It requires a fixed grid of beams provided by the base station. A terminal is either scheduled exclusively on one beam (single stream (ss) mode), or its data is multiplexed

and transmitted in parallel on multiple beams (multistream (ms) mode) that are shared among a group of terminals. Each terminal measures the SNIR and determines the achievable rates for both modes per beam, and signals this information combined with the id of the best beam (ss) / best beam set (ms) back to the BS. Note that all information is RB specific. The information is collected at the BS and combined into a separate list per transmission mode. The lists are ordered and evaluated, such that for each RB the terminals with their preferred beams can be ranked against all other terminals. Then, the selection process decides on the terminals to be scheduled and the mode to be chosen: for each beam either the terminal owning the average best rank over all RBs, or the group of terminals sharing multiple beams having the best average rank is selected, depending on the currently more effective mechanism. It is shown by means of emulation and a proof-of-concept implementation that using this algorithm, even cell-edge users can participate in the multi-stream mode, leading to an increased cell-edge throughput compared to the non-coordinated case.

6.1.3 Coordinated multi-point transmissions

The general idea of coordinated multi-point transmission (CoMP) is to enable the base stations of neighboring cells to coordinately conduct their transmissions [131]. In contrast to the interference aware scheduling approaches described above, with CoMP base stations do not only attend to the transmissions of each other, but jointly transmit to the terminals located in the area between them (*cf.* Figure 6.1c). In other words, base stations act as inputs of a generalized MIMO system. As it is the case with most CCI mitigation approaches, CoMP techniques can be classified into two different groups: *centralized* and *distributed* CoMP approaches. In the case of centralized CoMP, the base station antennas are connected to a single point (referred to as *central entity (CE)*) performing joint signal processing for the terminals served by multiple cells. The CE pre-computes all waveforms and sends them to the coordinated base stations. Accordingly, the centralized approach has a high signaling effort. Moreover, it is computationally expensive to comply with the tight latency requirements. Thus, centralized CoMP is hardly a candidate for real world applications. In the distributed CoMP case, a limited set of base stations transmit data jointly to multiple terminals in their cells without being externally coordinated. For each terminal, the serving base station coordinates the data flow by exchanging data and CSI with the related base stations. As transferring these data is less costly than transferring waveforms and IQ samples, signaling overhead is reduced compared to the centralized approach. As the waveforms are computed locally at each base station, the latency requirements can easier be met. In both cases, the antennas of the terminals act as outputs of the MIMO system. In the case of optimal coordination, a performance similar to non-interference impacted performance of an isolated cell (*cf.* Section 4) can be achieved [132]. Optimal coordination, however, requires that both the desired and the interfering signals from different base stations arrive at each terminal simultaneously, which is fundamentally impossible in distributed CoMP systems. The base stations cannot align the signals mainly due to the different propagation times between base stations and terminals. Accordingly, unintended signals, referred to as asynchronous MIMO interference, arrive and interfere at the receiver.

Zhang *et al.* were the first ones to develop a detailed mathematical framework for base station cooperation in a multi-user MIMO cellular network that explicitly accounts for the asyn-

chronous interference [133]. The authors develop three linear pre-coding methods, which significantly reduce the impact of the asynchronicity in interference. They even show, by means of simulation, that asynchronous interference may even improve system performance, instead of degrading it, if the receivers are aware of its time structure and are capable of exploiting it. Based on these findings, Jungnickel *et al.* implemented a first real-time distributed CoMP proof-of-concept implementation using real-world channels [134]. To enable a distributed CoMP architecture on their testbed, the authors use a tight clock synchronization (for outside terminals utilizing the clock reference signal provided by the Global Positioning System (GPS)), cell-specific and precoded pilots, a special highly efficient and reliable protocol for inter base station communication, as well as a limited CSI feed-back channel exploiting correlation in the time and frequency domain. The authors have tested their system in a multi-cell scenario based on measured outdoor channels. They show that already with a simple linear CoMP precoder, a significant fraction of the capacity in an isolated cell can be realized while facing substantial co-channel interference. These observations, however, are made under the assumption of unlimited backhaul and negligible delay.

6.1.4 Summary

Three different CCI mitigation approaches have been introduced above: hard frequency reuse, interference aware scheduling, and coordinated multi-point transmission. While HFR is the most easy to implement among the three approaches, the spectral efficiency of HFR based systems is very low. HFR is, thus, not a serious candidate for next generation cellular systems. Using CoMP techniques, a significant fraction of the capacity of an isolated cell can be achieved. The requirements in computational power, backhaul capacity, processing delay and synchronization to enable CoMP transmissions, however, are very high. The ability to mitigate CCI of interference aware scheduling mechanisms is lower than for the CoMP approaches, but these mechanisms come with a significantly smaller set of requirements.

Among the interference aware scheduling scheduling approaches presented above, partial frequency reuse based scheduling is easier to implement than the beamforming based scheduling approaches, as no special hardware (antenna grids) is required, transmissions do not need to be orientated, and terminal positions are only of collateral importance. Still, they promise a significant reduction on CCI impact, especially to those terminals located at the cell border. At the moment, they are, thus, the most heavily discussed alternative to provide CCI protection to the upcoming LTE systems. Motivated by the 'good performance at low complexity' property of PFR, in this chapter, PFR is studied as CCI mitigation candidate for a system following the system model presented in Section 3.2. More particular, soft frequency reuse is studied as a means to combat CCI in combination with the dynamic OFDMA techniques presented in Chapter 4.

6.2 The Application of SFR in dynamic OFDMA based Systems

In the previous section, soft frequency reuse has been introduced as computationally cheap yet efficient CCI mitigation technique for cellular systems. SFR was originally designed in connec-

tion with simple scheduling strategies, such as round robin. By slightly modifying the RR policy, e.g. by scheduling three instead of one terminal in each TTI and placing the transmission of the weakest terminal in the high power, the transmissions of strongest one in the low power, and the remaining ones in the medium power region of the power profile (*cf.* Figure 6.2), immense gains in cell-edge user throughput performance have been reported for idealized symmetric system models [119].

In this section, first, the usage of applying soft frequency reuse in combination with simple scheduling policies to improve cell-edge user performance is explored for a system following the system model introduced in Section 3.2. Then, the application of SFR with the more sophisticated carrier-wise proportional fair scheduler (introduced in Section 4.1.1) policy is under consideration. A slightly modified SFR aware carrier-wise PF scheduling algorithm is introduced as a consequence of the original PF not being able to exploit the benefits of SFR power profiling.

6.2.1 The impact of SFR power-profile application

In the following, multiple cells of an orthogonal frequency division multiple access based cellular system are considered. The impact of co-channel interference for terminal j on sub-carrier s is modeled as the fraction of the received power that is transmitted by other sources than the base station $\check{c}$ terminal j is connected to:

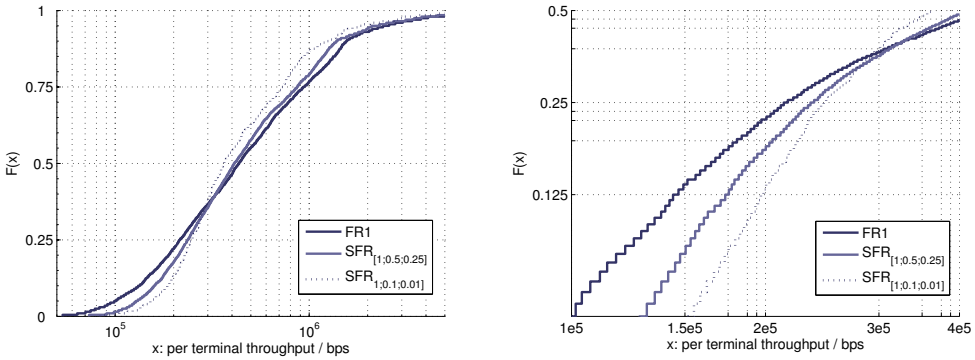
$$\sigma^2_{\xi,\mathrm{cci},j,s} = \sum_{\substack{c=1 \\ c\neq\check{c}}}^{C} h^{(t)}_{c,j,s} p^{(t)}_{c,s} \tag{6.1}$$

where $h^{(t)}_{c,j,s}$ is the attenuation that the interfering signal experiences on its way from interfering cell c to terminal j. If CCI is assumed to be the only kind of interference to be present, the signal-to-noise plus interference ratio experienced by terminal j is:

$$\gamma^{(t)}_{\check{c},j,s} = \frac{h^{(t)}_{\check{c},j,s} p^{(t)}_{\check{c},s}}{\sum\limits_{\forall c\neq\check{c}} h^{(t)}_{c,j,s} p^{(t)}_{c,s} + \sigma^2_{\eta}} \,. \tag{6.2}$$

Recall that assuming the inter-symbol interference and inter-channel interference not to impact the perceived SNIR is valid in OFDMA systems, as such systems are explicably designed to suppress them.

SFR power profiling Section 6.1.2 explains how the transmit power per sub-carier in an SFR enabled system is described by a power profile (*cf.* Figure 6.2). For cell $\check{c}$ and resource block n within the system bandwidth the power profile of $\check{c}$ concerning n is identified by $\pi_{\check{c},n} \in [0,1]$. The value denotes the fraction of the total available output power $p_{\max}$. On resource block n, cell $\check{c}$ thus transmits with a power of $p_{\max} \cdot \pi_{\check{c},n}$. Sticking to this notation, the SNIR at the receiving terminal side is is computed by:



(a) per terminal throughput cdf

(b) weaker half of terminals throughput cdf

Figure 6.3: the gain of applying soft frequency reuse on top of round robin scheduling

$$\gamma_{\check{c},j,n}^{(t)} = \frac{h_{\check{c},j,n}^{(t)} \pi_{\check{c},n} p_{\max}}{\sum\limits_{\forall c \neq \check{c}} h_{c,j,n}^{(t)} \pi_{c,n} p_{\max} + \sigma_\eta^2} \,. \tag{6.3}$$

If, as in this case, static soft frequency reuse is assumed, the power profile configuration does not adapt to current interference and load situation. Thus, a generic profile that is believed to be useful in most situations has to be applied. Throughout this chapter, in the static case SFR3 profiling is assumed, *i. e.,* each static power profile features three different power levels: high, middle and low (as shown in Figure 6.2d). Dividing the bandwidth in three equally large parts has widely been done in HFR scenarios, as it is the most bandwidth efficient among the possible HFR configurations. With SFR3, each cell uses one third of the spectrum with each power level. Two different standard SFR power profiles configurations known from literature (*e. g.,* [123]) are considered:

- $\text{SFR}_{[1;0.5;0.25]}$:
  the low power level equals one forth and the medium equals half of the high power level.

- $\text{SFR}_{[1;0.1;0.01]}$:
  the low power level equals one hundredth and the medium level equals one tenth of the high power level.

In order to stick to a consistent notation, for the FR1 case, in which the power is uniformly distributed over the resource blocks, a uniform power profile is assumed: $\pi_{c,n} = p_{\max}/N$ for all n and c (where N is the number of available resource blocks). In the following, these power profiles are applied to an OFDMA system following the system model of Section 3.2 with a simple round robin scheduling strategy in the trace-based simulation scenario presented in Section 3.3.3.

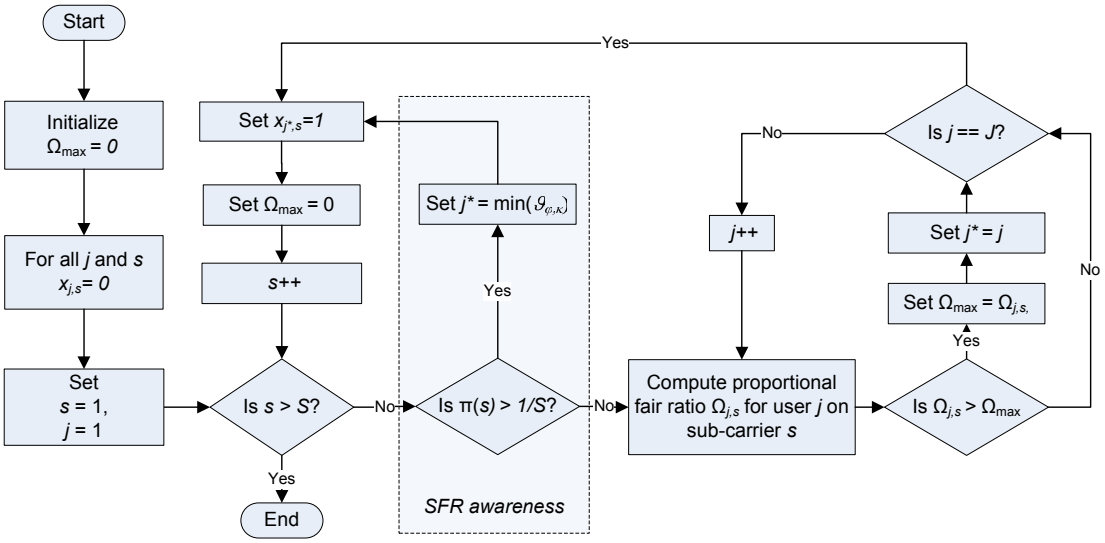
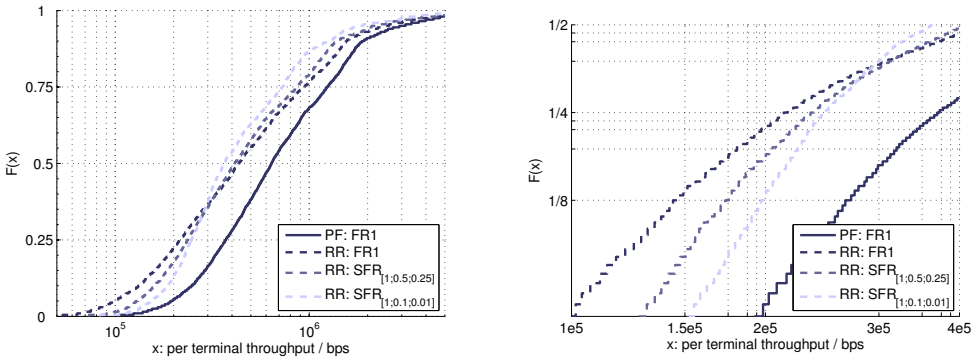


Figure 6.4: SFR aware proportional fair algorithm

SFR and RR scheduling

Figure 6.3 shows the impact of using soft frequency reuse on top of round robin scheduling. To enable the usage of SFR power profiles, the RR scheduler has been modified to select the next three terminals to be scheduled in the upcoming frame instead of selecting just one. It places the weakest terminal (identified by the least average throughput $\bar{\vartheta}_{j,k}^{(t-1)}$ over the last k TTIs) in the SFR profile's highest power region, the strongest terminal in the lowest power region, and the third in the medium power region.

The individual terminal throughput cumulative distribution function given in Figure 6.3a nicely shows the general impact of applying SFR power profiling: weaker terminal throughput is traded-off against stronger terminal throughput. A larger difference between the power-profile levels yields a larger shift towards weaker user performance. Figure 6.3b points out this effect by showing the throughput performance of the weaker half of terminals on a double-logarithmic scale: for the weakest terminals, a throughput increase of 50% is achievable with the $\text{SFR}_{[1;0.1;0.01]}$ power-profile. For the weakest 12.5% of all terminals an increase of more than 33% is possible. A significant increase is present for all terminals than belong to the weakest individual throughput quantile. A throughput decrease is present for the terminals that do not belong to the weaker approx. 37% of all terminals in the system. These numbers acknowledge the fact that soft frequency reuse is a strong technique for interference mitigation in round robin based OFDMA systems. It comes, however, at the cost of decreased *stronger terminal performance*, and, thus overall system spectral efficiency. Next, the question, whether SFR shows an equally strong performance in dynamic OFDMA systems is investigated.



(a) per terminal throughput cdf

(b) weaker half of terminals throughput cdf

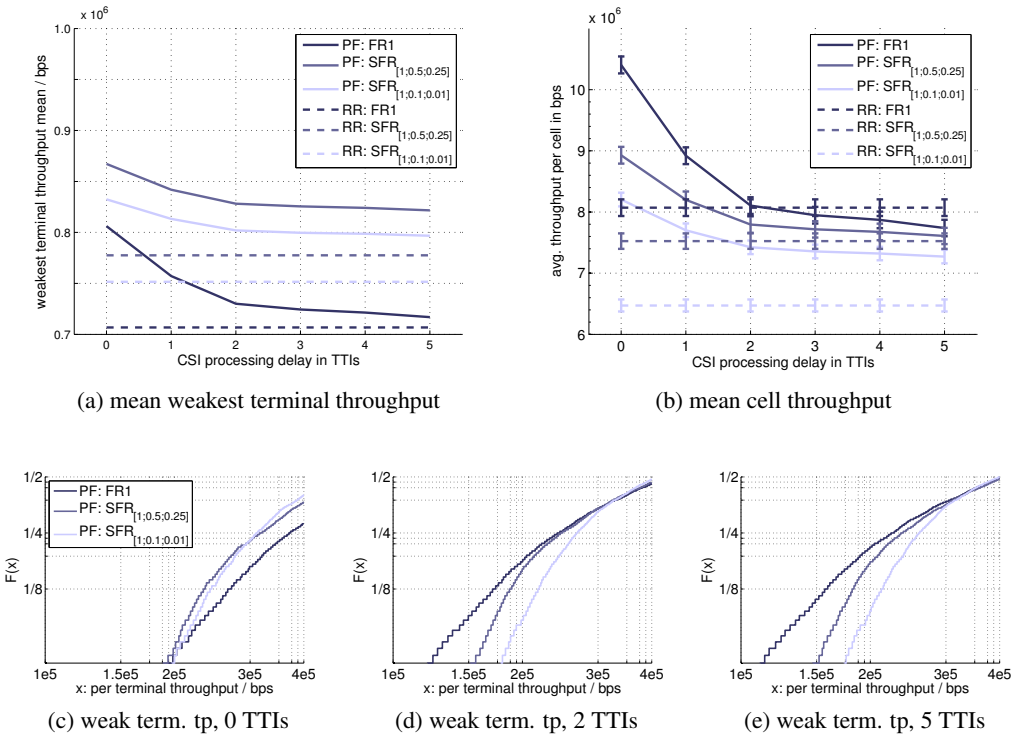
Figure 6.5: soft frequency reuse enabled round robin vs. proportional fair scheduling

SFR and PF scheduling

To explore the usage of SFR in the context of dynamic OFDMA systems, the carrier-wise proportional fair approach (4.5) is modified to be SFR aware. A flowchart of the resulting algorithm is shown in Figure 6.4: the scheduler in cell c determines for each sub-carrier s the fraction of transmission power that is radiated on it, which is determined by its power-profile $\pi_{c,s}$. If it is greater than the average fraction of transmission power, the sub-carrier is assigned to the currently weakest terminal j, again identified by the least average throughput $\bar{\vartheta}_{j,k}^{(t-1)}$ over the last k TTIs.

First, the performance of plain carrier-wise proportional fair is compared to SFR enabled round robin in the cumulative distribution function (CDF) shown in Figure 6.5. The superiority of the dynamic scheme over the static RR scheme is obvious, as PF scheduling provides a throughput performance gain to all terminals. On the throughput CDF of the weaker half terminals shown in Figure 6.5b it can be seen that for the weak terminals gains of up to 100% are present. The gains, however, are achieved assuming perfect instantenous channel knowledge at the base station. Figure 6.6 shows the impact of the CSI processing delay ΔT_{csi} on the performance of the carrier-wise PF approach, as has already been done for the single cell scenario in Section 4.1.1. Figure 6.6a shows the throughput of the average weakest user over all cells and TTIs. Again, PF performance is compared to the RR approach.

The application of soft frequency reuse power profiling on top of proportional fair scheduling shows some gains in weakest user performance. The gains, however, differ from the gains obtainable in a round robin based system in their significance. It is, thus, not clear, whether it is at all a good idea to utilize power profiles in dynamic OFDMA systems. Thus, before an attempt is made to improve power-profiles to better suit the dynamic OFDMA transmission approach, in the following section the general potential of power profiling in dynamic OFDMA systems is explored by formulating an optimization problem that finds the optimal power and sub-carrier assignment with respect to the current CCI situation and solving it in the basic multi-cell reference scenario.



(a) mean weakest terminal throughput

(b) mean cell throughput

(c) weak term. tp, 0 TTIs

(d) weak term. tp, 2 TTIs

(e) weak term. tp, 5 TTIs

Figure 6.6: the impact of the CSI processing delay ΔT_{csi} on system and weak terminal performance enabled round robin vs. carrier-wise proportional fair scheduling

6.3 Optimal Interference-Limiting Dynamic OFDMA Assignments

The investigations of this section aim in a comparison between the performance of systems employing standard hard frequency reuse and soft frequency reuse CCI mitigation techniques and the performance of a globally optimal working dynamic power allocation and sub-carrier assignment based system. In order to determine the globally optimal sub-carrier assignments, in the following an according non-linear optimization problem is formulated and solved for the basic multi-cell scenario.

6.3.1 Dynamic OFDMA based CCI mitigation approaches

As a consequence of the fairly complex co-channel interference aware non-linear resource allocation optimization, most contributions to this area have been focusing on specific system assumptions. Analogous to the single-user and single-cell scenarios, computational less complex algorithms have been considered to cut down the computational overhead. First approaches considered only power-allocation in the uplink of an OFDM-TDMA multi-cell scenario [135], [136], [137]. In general, these algorithms try to find a trade-off between small frequency-reuse factors (in order to keep the spectral efficiency high) and interference impacts. Hence, inside

a cell those sub-carriers are excluded that exceed certain interference-thresholds. The remaining sub-carriers are integrated into a *transmission group* on which a water-pouring algorithm is applied. Thus, this approach can be referred to as *distributed water-pouring algorithm*.

Regarding the down-link direction in a multi-cell environment featuring dynamic sub-carrier assignment and bit loading only a few approaches have been suggested. Due to the complexity of the related optimization problems only subproblems are considered. None of them considers both, bit loading and dynamic sub-carrier assignment. Mostly special schemes for the latter are presented. Li *et al.* suggest a radio resource control (RRC) scheme where dynamic resource control is realized at the access point as well as at a central radio network controller (RNC), which allows for sub-carrier reuse in neighboring cells [138]. However, the authors regard only one (the most dominant) co-channel signal from neighboring cells as interference. All signals from other interfering cells are treated as background noise. Also, they assume that the dominant interfering access point is known for each cell in the system and that the achievable rates per subcarrier/terminal combination are known in both cases (with and without interference from the dominant cell). Under these assumptions two optimization problems are formulated: One initial sub-carrier allocation problem depending on the interference impacts for the RNC, and one for the access points that takes the actual amount of data to be sent to each terminal into account when assigning the sub-carriers to the terminals located inside its cell. Since both problems are difficult to solve, the authors present heuristic algorithms for both problems. The RNC algorithm searches for each sub-carrier in each cell the terminal inside that cell, which achieves the highest throughput gain value when utilizing the sub-carrier. If no terminal can be found that achieves a positive throughput gain value, the sub-carrier is not assigned to that cell. On the basis of the RNC assignments, the access points try to further maximize the throughput by checking if the interference-based assignments done by the RNC match the rate requirements of the terminals. If not, at the access point a fast algorithm is applied, where the access point searches for each allocated sub-carrier the user that has the highest rate requirement, which currently can be realized depending on the assignments the RNC has selected. Performance evaluations show that the system utilization can be increased by up to 70% compared to the static round robin sub-carrier allocation scheme.

Another pseudo-dynamic approach is suggested by Kim *et al.* [139]. Again, this approach does not consider one of the original optimization problems presented in Section 2.4 for modification to suit a multi-cell scenario. However, among the traceable approaches for throughput maximization in a multi-cell downlink scenario, this work is one of the most interesting approaches: An optimal sub-carrier allocation scheme which allows cell coordination to assign sub-carrier reuse factors and modulation schemes for each sub-carrier is presented. The authors define a sub-channel to be a transmission unit consisting of a group of OFDMA sub-carriers. The sub-carriers for one sub-channel are chosen in a way such that with equal power allocation equal channel qualities can be achieved among the sub-channels. Consequently, the transmittable data rate of each sub-channel for a user can be assumed to be the same. Hence, sub-carriers in one sub-channel have to represent all frequency range specific characteristics of the bandwidth and thus are chosen to be rather widely spaced than neighboring. Hence, the selection process of sub-carriers for one sub-channel follows a comb-pattern, where the comb spans the complete bandwidth. Once the sub-carriers are distributed among the sub-channels, the distribution is

fixed and sub-channels are allocated among the users only. Under these assumptions, a throughput maximization optimization problem is stated along with a suite of sub-optimal heuristic algorithms.

So far, there is no comprehensive approach to formulate a mathematical problem to deliver optimal interference-limiting dynamic OFDMA assignments. Thus, in the following section, an according global CCI-aware OFDMA sub-carrier and power assignment optimization problem formulation is developed.

6.3.2 Global optimization

The aim of the global optimization approach is to dynamically distribute the resources such that the system's SNIR performance is maximized, where the optimal resource scheduling decision at time t is made at a single point (again referred to as *central entity*). It is assumed that at this point unlimited computational power and ideal system-wide knowledge is available. The central entity is in charge of determining the optimal real-valued power levels $y_{\check{c},s}^{(t)}$ for each sub-carrier in each cell c out of the set of active cells C, as well as for finding the optimal binary terminal/sub-carrier assignments $x_{\check{c},j,s}^{(t)}$ at time t. In the following, the non-linear optimization problem is formulated step by step.

Globally optimal power distribution

At first, the sub-carriers are considered to be statically assigned to different terminals per cell, *i. e.,* in this case dynamic sub-carrier assignment is not employed. In order to obtain the power distribution per base station and sub-carrier that maximizes the overall system throughput, the finite tones water filling problem (2.19) presented in Section 2.4.1 needs to be extended to consider the overall system and be aware of CCI:

$$\max_{\mathbf{Y}^{(t)}} \quad \sum_{\check{c}=1}^{C} \sum_{s=1}^{S} F\left(\frac{h_{\check{c},j,s}^{(t)} y_{\check{c},s}^{(t)}}{\sigma_\eta^2 + \sum\limits_{\forall c \neq \check{c}} h_{c,j,s}^{(t)} y_{c,s}^{(t)}} \right) \tag{6.4a}$$

$$\text{s. t.} \qquad \sum_{s=1}^{S} y_{c,s}^{(t)} \leq p_{\max} \qquad \forall\, c\,, \tag{6.4b}$$

where j is the terminal to which sub-carrier s is assigned in cell $\check{c}$, optimization variable $y_{c,s}^{(t)}$ is the real-valued power assignment that determines the amount of power radiated by cell c on sub-carrier s at time t, $\mathbf{Y}^{(t)}$ is the $C \times S$ matrix of power assignment variables $|y_{c,s}^{(t)}|$ at time t, and $F(\gamma)$ is a function that determines the throughput depending on the instantaneous SNIR (*e. g.,* the truncated Shannon capacity function described in Section 3.2.4). As in the single cell finite water filling case, constraint (6.4b) guarantees that in none of the cells the maximum power threshold is exceeded. Solving the *global finite water filling problem* (6.4) yields the set of power assignments $\{y_{c,s}^{(t)}\}$ that maximizes the system throughput. Note, that obtaining the solution is much more difficult than in the single cell case, as a better power distribution in one

cell might lead to a worse situation in another cell. This instance is reflected as a non-linearity in problem formulation (6.4): optimization variable $y_{c,s}^{(t)}$ is present in the nominator, as well as in the denominator of optimization goal (6.4a). Accordingly, finding optimal power allocation in a multi-cell OFDMA environment already is a very complex, non-linear task.

Globally optimal sub-carrier assignments

In the next step, the power allocations per sub-carrier are assumed to be fixed, and the terminal/sub-carrier assignments $x_{c,j,s}^{(t)}$ need to be optimized. In accordance with the optimal power distribution step above, this step aims in finding the set of terminal/sub-carrier assignments that maximizes the overall system throughput. In order to do so, max sum rate problem (2.22) is extended to consider the whole system and be aware of CCI:

$$\max_{\mathbf{X}^{(t)}} \quad \sum_{\check{c}=1}^{C} \sum_{j=1}^{J_c} \sum_{s=1}^{S} x_{c,j,s}^{(t)} F\left(\frac{h_{\check{c},j,s}^{(t)} p_{\check{c},s}^{(t)}}{\sigma_\eta^2 + \sum\limits_{\forall c \neq \check{c}} h_{c,j,s}^{(t)} p_{c,s}^{(t)}} \right) \tag{6.5a}$$

$$\text{s. t.} \qquad \sum_{j=1}^{J} x_{c,j,s}^{(t)} \leq 1 \qquad \forall\, c, s\ , \tag{6.5b}$$

where optimization variable $x_{c,j,s}^{(t)}$ is the binary terminal/sub-carrier assignment per cell for which holds

$$x_{c,j,s}^{(t)} = \begin{cases} 1, & \text{if in cell } c \text{ sub-carrier } s \text{ is assigned to terminal } j \text{ at } t \\ 0, & \text{if in cell } c \text{ sub-carrier } s \text{ is not assigned to terminal } j \text{ at } t. \end{cases}$$

As in the single cell max sum rate case, constraint (6.5b) guarantees that in none of the cells a single sub-carrier is assigned to more than one terminal. Solving the *global max sum rate* (6.5) optimization problem yields the set of terminal/sub-carrier assigments $\{x_{c,j,s}^{(t)}\}$ that maximizes the system throughput. Note, that obtaining the solution requires a comparable effort as in the single cell case, as the terminal/sub-carriers assignment decisions in one cell have no impact on the decisions in another cell. Accordingly, problem (6.5) is a linear optimization problem. Recall that the power shares per sub-carrier are assumed to be fixed.

Integrated optimization approach

Integrating the optimal power allocation and sub-carrier assignment problems above into a single optimization problem yields the non-linear global OFDMA resource allocation formulation:

$$\max_{\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}} \quad \sum_{\check{c}=1}^{C} \sum_{j=1}^{J_c} \sum_{s=1}^{S} x_{c,j,s}^{(t)} F\left(\frac{h_{\check{c},j,s}^{(t)} y_{\check{c},s}^{(t)}}{\sigma_\eta^2 + \sum\limits_{\forall c \neq \check{c}} h_{c,j,s}^{(t)} y_{c,s}^{(t)}} \right) \tag{6.6a}$$

$$\text{s. t.} \qquad \sum_{s=1}^{S} y_{c,s}^{(t)} \leq p_{\max} \qquad \forall\, c \tag{6.6b}$$

$$\sum_{j=1}^{J} x_{c,j,s}^{(t)} \leq 1 \qquad \forall\, c, s\ . \tag{6.6c}$$

Note, that there is a twofold non-linearity property in optimization goal (6.6a): the optimal power assignment non-linearity inherited from the *global finite water filling problem* (6.4), and a non-linearity due to the multiplication of the terminal/sub-carrier optimization variable $x_{\check{c},j,s}^{(t)}$ and the power allocation optimization variable $y_{\check{c},s}^{(t)}$ dependent $F\left(\right)$ term.

In order to account for intra-cell fairness, the *max throughput per terminal* constraint (6.6d), known from the single-cell case in Section 4.2 is added to form the complete *global constrained max sum rate problem*:

$$\sum_{s=1}^{S} x_{j,s}^{(t)} F\left(\frac{h_{\check{c},j,s}^{(t)} y_{\check{c},s}^{(t)}}{\sigma_\eta^2 + \sum\limits_{\forall c \neq \check{c}} h_{c,j,s}^{(t)} y_{c,s}^{(t)}}\right) \leq \vartheta_{\mathrm{req},j}\ , \tag{6.6d}$$

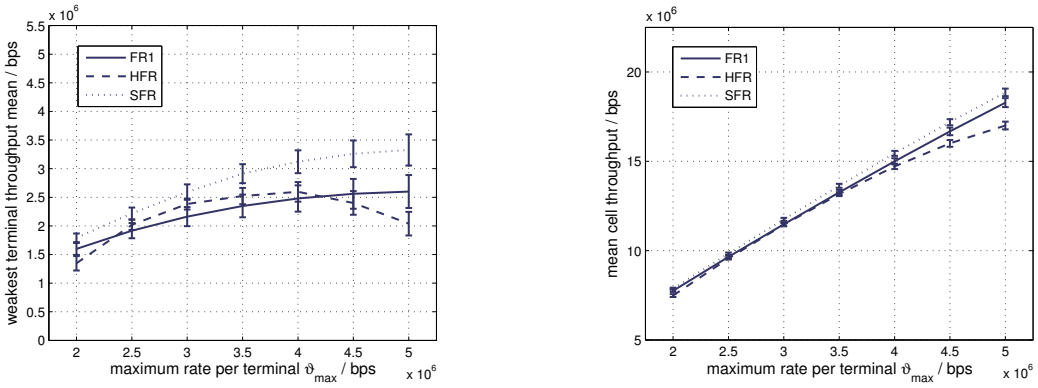
where $\vartheta_{\mathrm{req},j}$ is terminal j's maximum throughput share per TTI.

Due to its mixed integer and non-linear properties, *global constrained max sum rate* problem (6.6) is very hard to solve. Nonetheless, in the context of this thesis several thousand problem instances belonging to a simple scenario have been solved using appropriate non-linear simulation and optimization problem solving software. In the following, the simulation parameterization, methodology and results are presented.

6.3.3 Local optimization employing standard CCI mitigation techniques

In the case of local optimization, the base station schedulers of each cell make individual scheduling decisions per cell, *i. e.,* no central entity is involved. As in the single cell case (*cf.* Section 4) the schedulers rely on local information only. All sub-carrier assignment decisions are made by solving the according instance of the *constrained max sum rate* problem (4.12). Since merely local information is available at each scheduler, there is no opportunity to optimize the power allocation per sub-carrier in order to mitigate CCI, as it is done in the global optimization approach. Instead, standard techniques are used to suppress CCI: hard frequency reuse and soft frequency reuse.

While HFR restricts each base station to use a limited part of the system bandwidth only, SFR allows frequency reuse 1 bandwidth usage, but each base station is bound to different power restrictions at different parts of the available spectrum. Recall that the mode of action of HFR as well as SFR can be described in terms of cell-specific *power profiles* over the system bandwidth (*cf.* Figure 6.2). A power profile prescribes the fraction of the maximum transmit power that the base station may use depending on the part of the spectrum. In the following, the power



(a) mean throughput (bps) of weakest terminal

(b) mean cell throughput (bps)

Figure 6.7: the gain in throughput due to applying HFR and SFR

profile of cell $\check{c}$ is denoted by $\pi_{\check{c},s} \in [0,1]$. It denotes the fraction of the total available output power $p_{\max}$. On sub-carrier s, cell $\check{c}$ thus transmits with a power of $p_{\max}\pi_{\check{c},s}$. Considering power profiles, the SNIR experienced by terminal j on sub-carrier s in cell $\check{c}$ can be written as

$$\gamma_{\check{c},j,s}^{(t)} = \frac{\pi_{\check{c},s} p_{\max} h_{\check{c},j,s}^{(t)}}{\sum\limits_{\substack{c=1 \\ c\neq\check{c}}}^{C} \pi_{c,s} p_{\max} h_{c,j,s}^{(t)} + \sigma_{\eta,s}^2} \,. \tag{6.7}$$

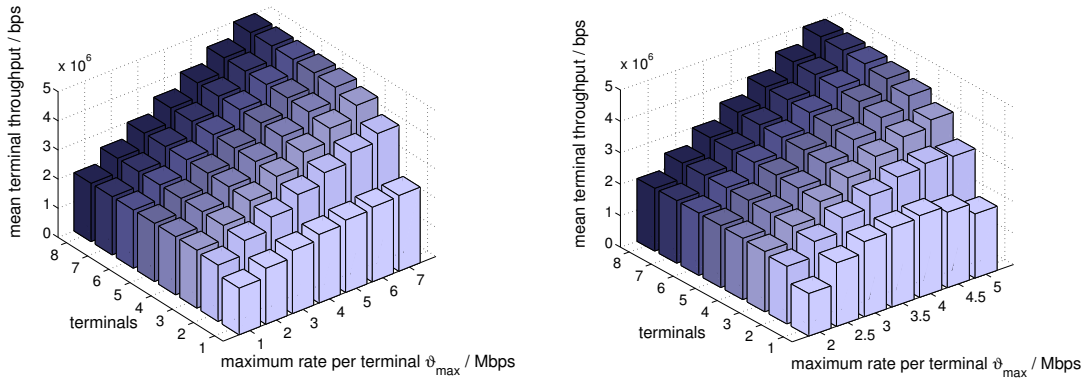
The above denominator sums up the co-channel interference from concurrently transmitting base stations $c \neq \check{c}$ and the noise power $\sigma_{\eta,s}^2$.

6.3.4 Results

In the following the results of applying global optimization in a system according to the system model from Section 3.2 working in the basic multi-cell scenario presented in Section 3.3.2 are shown. The mean cell throughput, the mean throughput of the weakest terminal (cell-edge terminal) of the system, as well as the individual per terminal throughputs are presented. The error bars display confidence intervals with a confidence level of 90 %.

Standard CCI mitigation techniques

Figure 6.7a shows a weakest terminal throughput performance comparison of FR1, HFR and SFR, where FR1 is the basic frequency reuse 1 scenario, in which no CCI mitigation techniques are applied. Comparing first HFR to FR1, the graphs show that the application of HFR hardly improves the cell edge terminal performance, even though up to a maximum required rate of 4Mbps HFR performs slightly better. The small advantage traces back to the fact that there is zero interference from the neighbor cell, and, thus, the channel states of the cell edge terminals are generally better than in the frequency reuse 1 case. Due to the limited resources in the hard



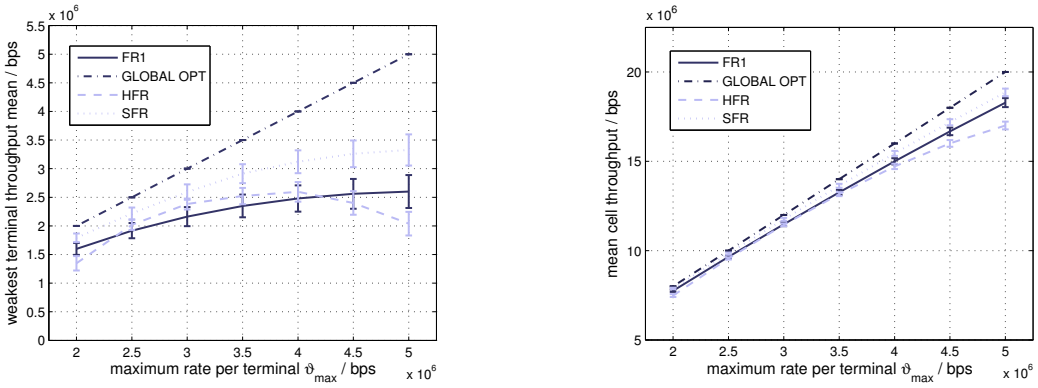
(a) frequency reuse 1

(b) hard frequency reuse

Figure 6.8: individual mean terminal throughput in the basic FR1 and the HFR case

frequency reuse case, however, the weakest terminals cannot take advantage of the increasing maximum rate above that 4Mbps threshold. This is mainly because the cell edge terminals hardly get any resources at all, if the stronger terminals are allowed to consume resources for such high rates. Accordingly, their mean throughput decreases with the increasing max rate after that turning point. Relating to the overall system performance shown in Figure 6.7b in terms of mean per cell throughput, it must be stated that HFR does not provide any system performance gain. On the contrary, for large maximum rate per terminal values $\vartheta_{\mathrm{req},}$, the system performance even decreases slightly. The reason for this can be seen in Figure 6.8. Not only drops the cell edge terminal performance for max rates larger than 4Mbps, but also the performance for stronger terminals. While in the FR1 scenario the stronger half of the terminals (terminals 5 through 8) achieves its maximum rate share up to 5Mbps, in the HFR case, due to the limited resources available, only the strongest two terminals get it. This is mainly due to the fact that the stronger terminals (which are located close to the base station and are, thus, less susceptible to interference) do not gain from HFR's zero interference advantage, but suffer from the limited amount of resources just like the weaker terminals. With the performance of the stronger terminals the overall system performance goes down.

Altogether, the application of HFR in the basic reference scenario achieves only performance gains under special circumstances. This result underlines the determination to use more advanced CCI mitigation schemes in cellular OFDMA networks. Due to this reason, in the following HFR is ignored. More significant performance gains can be achieved, if SFR is applied. Figure 6.7a shows a performance gain in cell edge terminal throughput over FR1 of up to 30 %. This is an immense gain, considering the fact that the gain solely stems from masking the resource block power levels. Note that the weakest terminal's throughput gain is present in all chosen maximum rate per terminal cases, and that it constantly grows with an increasing maximum rate share. Thus, in contrast to the HFR case, it can be stated that if SFR is applied all terminals gain from a large maximum allowed rate per terminal. This fact is also reflected in Figure 6.10a that gives an overview of the indivdual per terminal mean throughput values. Note



(a) mean throughput (bps) of weakest terminal

(b) mean cell throughput (bps)

Figure 6.9: the gain in throughput due to global resource allocation optimization

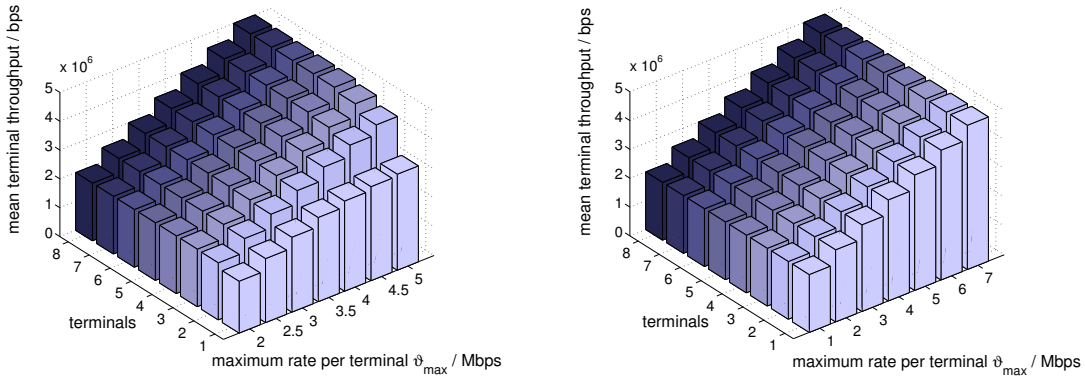
that the strong terminals do not suffer from the restricting the power levels on certain parts of the frequency band, as for the strong terminals the limited power suffices to achieve their maximum allowed throughput. Accordingly, the system througput performance (shown in Figure 6.7) is in all max rate per terminal cases larger than in the HFR case, as well as in the basic FR1 case. The difference is marginal, but the intersting fact is that the overall system does not suffer from improving the cell edge-terminal performance.

Altogether, the application of SFR in the basic reference scenario achieves significant cell-edge terminal performance gains along with marginal system throughput gains. This result underlines the decision to use SFR as a means to control CCI in cellular OFDMA networks.

Global optimization

Figure 6.9a shows the weakest terminal throughput performance comparison of the basic frequency reuse 1 scenario, and the scenario, in which global optimization according to Equation (6.6) is performed. What immediately catches the eye when looking at the curves is the fact that the optimal weakest user throughput grows linearly with the maximum rate per terminal value. For a maximum admitted throughput per terminal of $\vartheta_{\mathrm{req},}$ =5Mbps the weakest terminal's throughput is 5Mbps, *i. e.,* that all terminals achieve the maximum achievable throughput. This fact is also reflected in Figure 6.10b that shows the individual mean terminal throughput in the case of global optimization. Moreover, since the 99% confidence intervals in Figure 6.9a have zero spread, this fact is true for all simulation runs. Accordingly, the system throughput performance achieves its maximum per cell throughput of 20Mbps.

Altogether, the application of global optimization in the basic reference scenario achieves significant cell-edge terminal throughput performance gains of up to 100% and system throughput gains of more than 10%. This result underlines the assumption that dynamic resource allocation in terms of dynamic power and sub-carrier assignment techniques bears a high potential to mitigate CCI in cellular OFDMA networks.



(a) soft frequency reuse

(b) global optimization

Figure 6.10: individual mean terminal throughput in the SFR and the global optimization case

From the results above, two major reasonings can be derived for the further investigations of this thesis: first, under the idealistic assumptions of the basic reference scenario the application of dynamic sub-carrier assignments in combination with SFR is beneficial to improve cell edge terminal, as well as the system throughput; and, second, these performance gains are still significantly behind the gains that are possible if the power can be optimally distributed to mitigate CCI. Thus, in the following, a novel adaptive frequency reuse is introduced as a means to approach the optimal results with a complexity comparable to the one of SFR power-profiling.

6.4 Interference Coupling Based Adaptive Frequency Reuse

So far in this chapter, power profiles have been assumed to be of static nature and hence do not adapt to the current traffic situation. Furthermore, radio network engineers would need to coordinate the power profiles such that the cell-edge terminals receive little interference from neighboring cells. As an alternative to static soft frequency reuse, adaptive frequency reuse has been suggested. An overview of current state of the art AFR approaches is given at the beginning of this Chapter in Section 6.1.2. Among the cited approaches, the two approaches by Stolyar *et al.* [121] and [122] are closest to the following AFR approach.

This section presents the first scheme to adapt power profiles in OFDMA networks that are explicitly derived from the *interference coupling* among cells, which depends on radio and traffic conditions. Particularly, a mechanism to adapt power profiles (*i. e.,* the power profile coefficients $\pi_{\check{c},n}$) according to the current condition in the cell and its neighborhood is presented. In a nutshell, adapted power profiles shall be found based on two things. Namely, by reserving throughput for the terminals on the available resource blocks, and by computing the power that is required to provide this throughput. Therefore an effective method to calculate the power levels that can support a given throughput under the current channel conditions is needed (as the interactions among terminals through CCI have to be considered). To address this topic, the present section develops *linear interference coupling equations*. For computing the terminals'

contribution to the respective coupling matrix, average path loss values are used. These values could, for example, be taken from measurements over some number of recent frames. The resulting power values will finally serve as the power profile that is handed over to the scheduling units in the respective cells.

In the following, first an interference coupling model for LTE like systems is developed. Then, the above idea is formalized into an optimization model, and an algorithm is presented, which finds reasonable solutions to the optimization model. The problem is nonlinear and known to be not solvable to optimality in reasonable time by standard solver software. Finally the results of the interference coupling based AFR approach are compared to the non-AFR results obtained in the previous section.

6.4.1 Interference coupling model

The basic tool for interference coupling based adaptive power profile generation is the computation of the interference matrices $\mathbf{O}^{(j)}$ that are derived from according equation systems. Linear interference coupling equation systems have first been introduced in the context of power-controlled systems such as UMTS. Their original purpose is to reduce the dimension [140] of the equation systems that determine cell powers under perfect power control. The fact that the coupling equation systems describe the system at cell level (rather than at terminal level) is also exploited for large-scale network planning methods, see, *e. g.,* [141, 142] and the references contained therein.

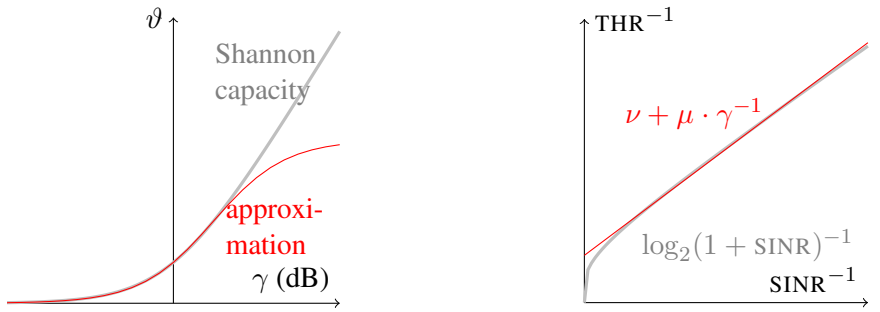
In this section, the concept of an interference coupling model for OFDMA based systems that apply adaptive modulation and coding to handle interference rather than power control is developed. An according approach for the case of WiMAX systems has already been done [143]. This approach largely parallels this work and comprises three steps: the approximation of Shannon's capacity formula (3.10) that is instrumental to obtaining a linear model is described, a description of a coupling model for a generic single channel, *i. e.,* a single resource block, and, finally, a formulation of the complete interference coupling model for LTE like systems.

Inverse linearization of Shannon capacity

Key to obtaining our linear model is the linearization of the relationship between inverse throughput and inverse SNIR. Shannon's law (3.10) is depicted as the top function graph in Figure 6.11a. It is approximated by the following linearization in inverse space:

$$\vartheta^{-1} = \nu + \mu \cdot \gamma^{-1}\,, \tag{6.8}$$

where ν and μ are the linearization parameters. The approximation is illustrated in the lower graph in Figure 6.11a; the actual linearization in inverse space is depicted in Figure 6.11b. The approximation corresponds to approximating the logarithm function in (3.10) by a hyperbola. As can be seen in Figure 6.11a, this works well for small SNIR values, while for large SNIR values (*i. e.,* exceptionally good channel conditions) the error can be arbitrarily large. Basically, while with Shannon's formula (3.10), infinite transmit power results in infinite throughput, the



(a) Approximation of SNIR-vs.-throughput relationship

(b) Underlying linearization in inverse space

Figure 6.11: Inverse linearization of Shannon capacity

substitute (6.8) implies an asymptotic bound on throughput. As the truncated Shannon bound (that does not allow infinite throughput for infinite power either) is used as the link level performance model in this thesis, the error due to the linearization approach can be assumed to me rather small.

The linearization is not restricted to the idealized Shannon capacity, but works equally well if realistic coding and modulation are used [143]. Throughout this section work (and, in particular, in Figure 6.11) the parameters have been chosen as $\nu = 0.25$ and $\mu = 0.75$. Depending on the available coding schemes and predominant channel conditions, a better parameterization may be achievable.

Coupling equation system for a generic single channel

Assume that terminal j is served by its connected cell $\check{c}$ on a generic channel with bandwidth B and obtains an (average) throughput of ϑ_j^*. Let

$$\nu_j := \nu(\vartheta_j^*/B)$$
$$\mu_j := \mu(\vartheta_j^*/B)$$

Cell $\check{c}$'s transmit power is denoted by $p_{\check{c}}$ and the (interfering) powers of cells $c \neq \check{c}$ by p_c. Furthermore, the average channel gain between terminal j and cell c by $h_{c,j}$ and the total noise experienced by terminal j by σ_η^2. Assume that the channel is used by several terminals, which are scheduled sequentially and obtain an average throughput based on the respective average SNIR. For terminal j to obtain an average throughput of ϑ_j^*, the average activity α_j of terminal j on the channel, *i. e.,* the fraction of the time that the terminal needs to access the channel, has to be

$$\alpha_j = \nu_j + \mu_j \left(\sum_{c \neq \check{c}} \alpha_c \frac{h_{c,j}}{h_{\check{c},j}} \frac{p_c}{p_{\check{c}}} + \frac{\sigma_\eta^2}{h_{\check{c},j} p_{\check{c}}} \right) \tag{6.9}$$

according to (6.8). The expression in brackets corresponds to the inverse SNIR, and the contributions of interfering cells $c \neq \check{c}$ are scaled with their respective total activity α_c. By summing up (6.9) for all terminals served by cell $\check{c}$ over time, the following expression for the total activity of cell $\check{c}$, *i. e.,* the total fraction of the time the cell transmits on the channel, is obtained:

$$\begin{aligned}\alpha_{\check{c}} &= \sum_{j=1}^{J} \alpha_j \\ &= \sum_{j=1}^{J} \nu_j + \sum_{\substack{c=1 \\ c\neq\check{c}}}^{C} \left(\alpha_c \sum_{j=1}^{J} \mu_j \frac{h_{c,j}}{h_{\check{c},j}} \right) \frac{p_c}{p_{\check{c}}} + \sum_{j=1}^{J} \mu_j \frac{\sigma_\eta^2}{h_{\check{c},j} p_{\check{c}}}\end{aligned}$$

This can be rearranged into the following coupling equation relating powers and activity values in the different cells:

$$\alpha_{\check{c}} p_{\check{c}} = \overbrace{\sum_{j=1}^{J} \nu_j}^{:=o_{\check{c}\check{c}}} \; p_{\check{c}} + \sum_{\substack{c=1 \\ c\neq\check{c}}}^{C} \overbrace{\left(\sum_{j=1}^{J} \mu_j \frac{h_{c,j}}{h_{\check{c},j}} \right)}^{:=o_{\check{c}c}} \; \alpha_c \, p_c + \overbrace{\sum_{j=1}^{J} \mu_j \frac{\sigma_\eta^2}{h_{\check{c},j}}}^{:=p_{\check{c}}^{(\mathrm{FIX})}}$$

The *coupling elements* $o_{\check{c}c}$ defined in the above equation form a *coupling matrix*

$$\mathbf{O} := (o_{\check{c},c}) \ . \tag{6.10}$$

The *noise-compensation power* $p_{\check{c}}^{(\mathrm{FIX})}$ represents the power that the cell would need to overcome noise alone, *i. e.,* without any inter-cell interference.

Coupling model for LTE like systems

There is a separate coupling equation system for each resource block n, because the resource blocks are separated in frequency. The coupling matrix for resource block n is denoted by $\mathbf{O}^{(\mathrm{TOT},n)}$ and the associated noise power vector by $\mathbf{p}^{(\mathrm{FIX},n)}$. Under the assumption that the cell transmits continuously on each resource block, it holds $\alpha_c = 1$ for the activity values of all cells c. On each resource block n, the coupling equation then reads:

$$\mathbf{p}^{(n)} = \mathbf{O}^{(\mathrm{TOT},n)} \mathbf{p}^{(n)} + \mathbf{p}^{(\mathrm{FIX},n)} \tag{6.11}$$

The solution to (6.11) describes the power vector $\mathbf{p}^{(n)} = (p_{\check{c}}^{(n)})$ required to serve the terminals on resource block n, if the noise compensation power vector of $\mathbf{p}^{(\mathrm{FIX},n)} = (p_{\check{c}}^{(\mathrm{FIX},n)})$ is assumed.

There is a positive solution p_n to the Equation system (6.11) if and only if the spectral radius of the matrix $\mathbf{O}^{(\mathrm{TOT},n)}$ is smaller than one, *e. g.,* [144]. The solution is then given by

$$\mathbf{p}^{(n)} = \left(I - \mathbf{O}^{(\mathrm{TOT},n)} \right)^{-1} \mathbf{p}^{(\mathrm{FIX},n)} = \sum_{k=0}^{\infty} \left(\mathbf{O}^{(\mathrm{TOT},n)} \right)^k \mathbf{p}^{(\mathrm{FIX},n)} \tag{6.12}$$

where, I denotes the unit matrix. The first equation simply follows by rearranging (6.11) under the assumption that the inverse of $I - \mathbf{O}^{(\text{TOT},n)}$ exists. The alternative infinite series form of this inverse on the right-hand side of (6.12) is called *Neumann series* and is used in (6.14b).

6.4.2 Optimization model

For each terminal j and each resource block n, an assignment variable $w_{nj} \in [0,1]$ is introduced, which denotes the fraction of the maximum terminal throughput $\vartheta_{\text{req},j}$ terminal j receives on resource block n. The matrix of all such variables is denoted by $W = (w_{nj})_{n\in N, j\in J}$. The w_{nj} variables can be seen as the time-average of the throughput assigned to the respective resource block and terminal in the scheduling problem (4.12).

The maximum total contribution of terminal j to the coupling matrices is determined by the square matrix $\mathbf{O}^{(j)}$. This matrix has nonzero entries only on the row corresponding to the serving cell (the coupling model and equations are derived in Section 6.4.1). Similarly, $p^{(\text{FIX},j)}$ denotes a vector with all zero components except for a power for terminal j (whose value is also determined in Section 6.4.1) in the coordinate corresponding to the serving cell. Given a complete assignment W, the coupling matrix and noise-compensation power vector for resource block n are then

$$\mathbf{O}^{(\text{TOT},n)} = \sum_{j=1}^{J} w_{nj}\mathbf{O}^{(j)} \tag{6.13a}$$

$$p^{(\text{FIX},n)} = \sum_{j=1}^{J} w_{nj}p^{(\text{FIX},j)} \tag{6.13b}$$

The task of finding optimal power profiles is represented by the following *nonlinear optimization model:*

$$\max_{\mathbf{y}^{(n)}} \quad \sum_{n=1}^{N}\sum_{j=1}^{J} w_{nj} \tag{6.14a}$$

$$\text{s.t.} \quad \mathbf{y}^{(n)} = \sum_{k\geq 0}\left(\sum_{j=1}^{J} w_{nj}\mathbf{O}^{(j)}\right)^{k}\left(\sum_{j=1}^{J} w_{nj}p^{(\text{FIX},j)}\right) \qquad \forall n \tag{6.14b}$$

$$\sum_{n=1}^{N} \mathbf{y}^{(n)} \leq p_{\max} \tag{6.14c}$$

$$0 \leq \sum_{n=1}^{N} w_{nj} \leq 1 \qquad \forall\, j \tag{6.14d}$$

Objective function (6.14a) corresponds to maximizing the overall throughput. The power vector optimization variable $\mathbf{y}^{(n)}$ to serve the reserved throughput is calculated in constraint (6.14b). Note that the model is nonlinear because of the *Neumann series* in this constraint. The W variables determine the contributions of the terminals to the individual coupling systems as stated in (6.13). Constraint (6.14c) limits the total power emitted per cell to the maximum available

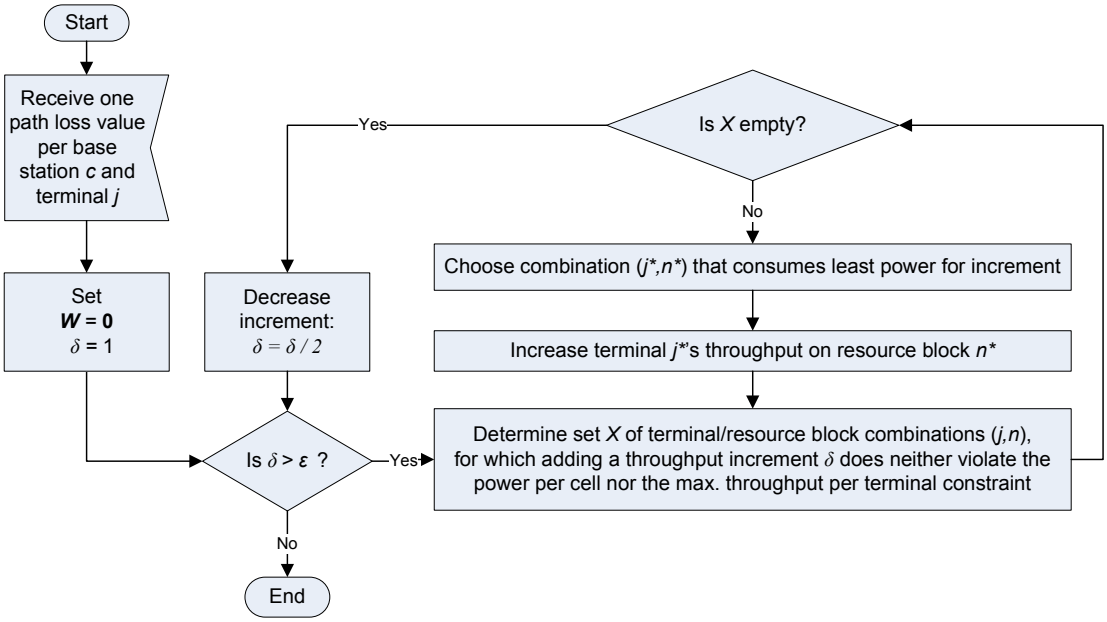


Figure 6.12: power profile adaptation algorithm

power $p_{\max}$. Finally, constraint (6.14d) ensures that each terminal obtains at most its maximum desired throughput.

6.4.3 Heuristic

A central entity is assumed to be responsible for the calculation and distribution of the individual power profiles. To find reasonable solutions to the optimization model (6.14), a simple method that combines the principles of binary search and greedy algorithms [145] is executed at the central entity.

Binary search finds the desired value by successively halving the search interval (assuming a convex or concave function on the search space). The optimal value can be found up to a fixed precision after a number of steps that is logarithmic in the precision. In this context, the amount of throughput that a terminal receives on a given resource block is the desired value to be optimized. Throughout the algorithm, the throughput for some terminal j on some resource block n is increased by an increment δ. If no terminal can be assigned an additional throughput of δ anymore, then δ is halved and the inner loop is started anew.

Algorithms employing the *greedy principle* subsequently take decisions that currently seem best from a local point of view and never reconsider a decision. Greedy algorithms produce optimal solutions for a special class of optimization problems [145], but they remain suboptimal in general. Nonetheless, greedy algorithms are popular because they are easy to implement and often produce reasonable results. Here, the greedy principle is applied to decide which terminal should receive additional throughput on which resource block: To steer clear of the power constraint (6.14c), which ultimately limits the amount of cell throughput to be distributed,

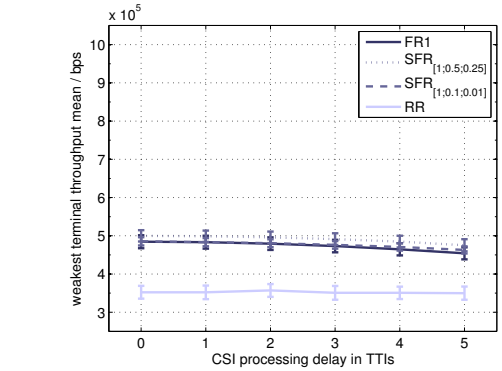
1 $W \leftarrow \mathbf{0}$
2 $\delta \leftarrow 1$
while $\delta > \epsilon$ **do**
  while $true$ **do**
3 $X \leftarrow \{(n,j) \mid \sum_{n=1}^{N} w_{nj} + \delta \leq 1 \wedge\ p^{(\mathrm{TOT})}(W + \delta \mathbf{e}_{nj}) \leq p_{\max}\}$
4 **if** $X = \varnothing$ **then**
      break
    end
5 $(n^*, j^*) \leftarrow \arg\max_{(n,j)\in X} \left\| p^{(\mathrm{TOT})}(W + \delta \mathbf{e}_{nj}) - p_{\max} \right\|_{\infty}$
6 $W \leftarrow W + \delta \mathbf{e}_{n^* j^*}$
  end
7 $\delta \leftarrow \delta / 2$
end
8 **return** $(p_n(W)/p_{\max})_n$

Pseudo-Code Box 4: greedy binary search for finding power profiles

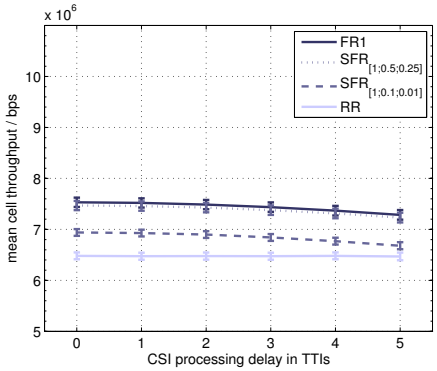
each additional throughput increment is assigned such that the slack in this constraint is least reduced.

Figure 6.12 shows the general structure of the algorithm. Note that, in order to enable the central entity to do the power profile calculations, all base stations periodically report *average* channel state information for all of their terminals. That is, for each terminal a vector is reported, which contains average CSI on the strongest received base stations. Thereafter, the actual power profile calculation algorithm is executed. Pseudo-code Box 4 lists the procedure in detail. In Steps 1 and 2, the throughput values are initialized to zero and the throughput increment is initialized to one. The inner loop, spanning Steps 3–6, assigns an additional throughput of δ to as many terminals as possible. To this end, first the currently feasible set X of pairs resource blocks/terminal pairs (n, j) are computed in Step 3. Such a pair is feasible if the corresponding throughput increment does neither violate the maximum terminal throughput constraint (6.14d) nor the maximum cell power constraint (6.14c). Shorthand $W + \delta \mathbf{e}_{nj}$ is used to denote the throughput assignment that is obtained by taking the assignment W and adding δ units of throughput to terminal j on block n. Moreover, shorthand $p^{(\mathrm{TOT})}(W)$ denotes the total power $\sum_n p_n$ as computed in (6.14b) for the throughput specified by W.

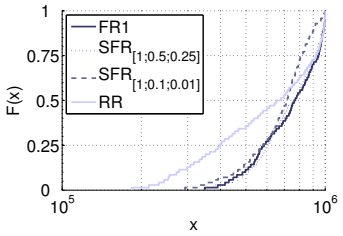
Step 4 checks whether set X is empty. In this case, no terminal can receive an additional throughput of δ, and the algorithm continues with the outer loop after halving the current throughput increment in Step 7. Otherwise, the increment with the largest power gap is selected in Step 5 and added to the current values of W in Step 6. The outer loop is repeated until the current throughput increment δ is smaller than some threshold value $\epsilon > 0$. Finally, the algorithm returns the ratios of the transmit powers for the current throughput assignment W on the resource blocks over the maximum power as the adapted power profiles in Step 8.



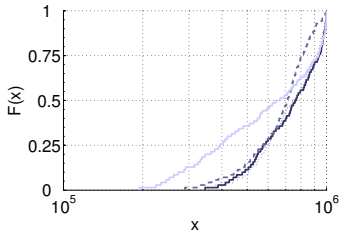
(a) mean throughput (bps) of weakest terminal



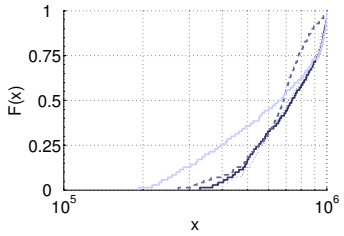
(b) mean cell throughput (bps)



(c) term. throughput cdf, 0 TTIs



(d) term. throughput cdf, 2 TTIs



(e) term. throughput cdf, 5 TTIs

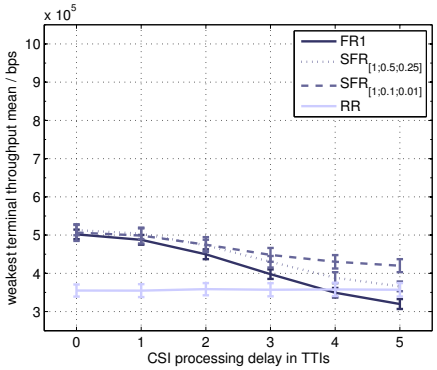
Figure 6.13: extended multi-cell reference scenario I: the gain of applying SFR on top of optimal SCA, *stationary terminals*

6.4.4 Results

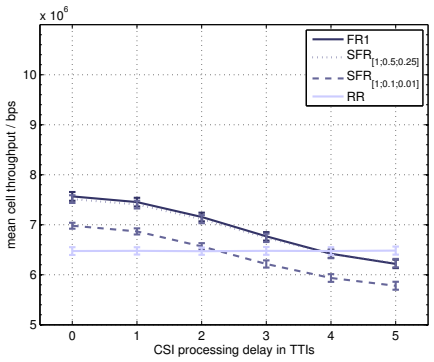
In the following computational tests, the performance of a system applying adaptive frequency reuse power profiles as introduced above is compared to the performance of a system applying the standard static soft frequency reuse power profiles known from Section 6.2 and to a basic frequency reuse 1 based system that uniformly distributes the power over all sub-carriers. The present results are achieved by simulation following the system model introduced in Section 3.2 and the extended hexagonal reference scenario parameterization given in Section 3.3.3. Constraint max sum rate resource scheduling optimization problem (4.12) is solved independently by each base station in each TTI applying either the static SFR or adapted AFR profiles.

The local optimization performance results in terms of weakest terminal, per cell and individual terminal throughput are shown in Figures 6.13 through 6.16.

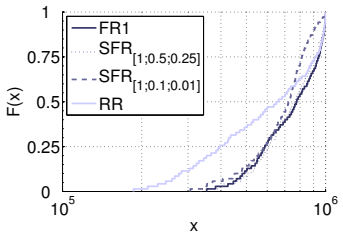
FR1 (uniform power profiles) Figure 6.13 shows the difference in throughput performance between locally optimal sub-carrier assignments and the static benchmark case for stationary terminals, Figure 6.14 shows the results for terminals moving at a speed of $v = 10m/s$. As expected, the performance gain due to optimally assigning the sub-carriers decreases with an increasing CSI processing delay ΔT_{csi}. While the optimal curve decreases rather slowly in the



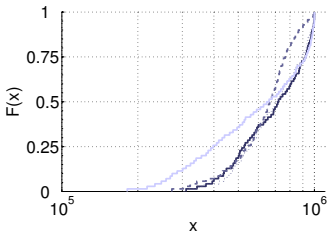
(a) mean throughput (bps) of weakest terminal



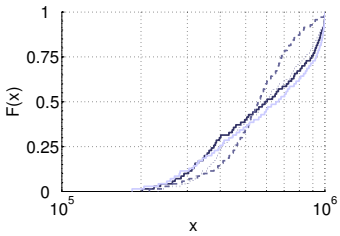
(b) mean cell throughput (bps)



(c) term. throughput cdf, 0 TTIs



(d) term. throughput cdf, 2 TTIs

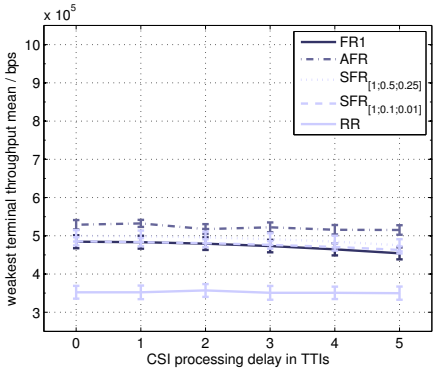


(e) term. throughput cdf, 5 TTIs

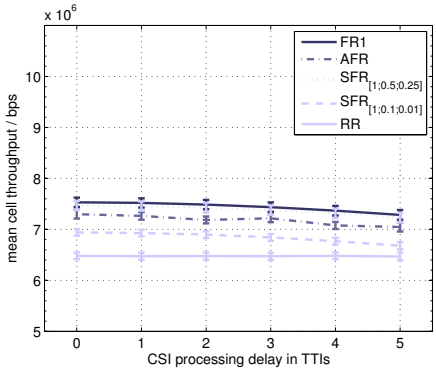
Figure 6.14: extended multi-cell reference scenario I: the gain of applying SFR on top of optimal SCA, *moving terminals at speed* $v = 10\frac{m}{s}$

case of stationary terminals, it decreases fast in the case of moving terminals and is completely consumed if the channel information has an age of 4 TTIs (Figure 6.14b). This is due to the fact that the differences in path loss, shadowing and fading add up to a large difference in SNIR for the moving terminals, outdating the information about the channel quite fast. Bearing in mind that the sub-carrier assignments are optimally chosen, it must be stated that dynamically assigning sub-carriers in a frequency reuse 1 scenario does not pay off in systems that feature mostly moving terminals and are unable to deliver adequate CSI values in a timely manner.

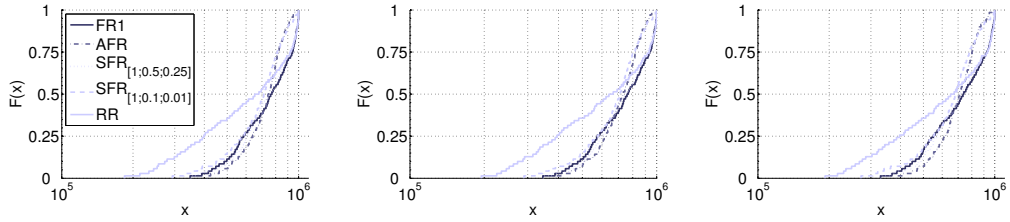
For systems that can be assumed to encounter CSI processing delays of 2 to 3 TTIs, however, a large gain is achievable also with the delayed CSI values. This holds not only for the experienced per cell throughput, but - more importantly - also for the individual per terminal throughput as can be seen in Figures 6.13c–6.13e and 6.14c–6.14e. The figures show the empirical per terminal throughput cumulative distribution function for the stationary and the moving terminal cases with 0, 2, and 5 TTIs delay respectively. Large gains in weak terminal throughput performance can be observed for the 0 and 2 TTI CSI processing delay cases. In all these cases, the throughput of the first quartile (weakest 25%) of the terminals is increased by 30-50%. For stationary terminals, the gain is achievable even when using CSI with an age of 5 TTIs. It can, thus, be stated that dynamically assigning sub-carriers among terminals according to the current channel situation in an FR1 scenario has a high potential to boost the performance of weak ter-



(a) mean throughput (bps) of weakest terminal



(b) mean cell throughput (bps)



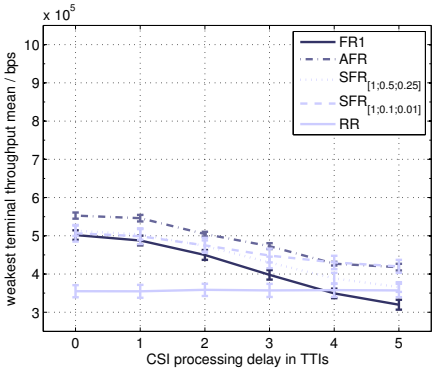
(c) terminal throughput cdf, 0 TTIs (d) terminal throughput cdf, 2 TTIs (e) terminal throughput cdf, 5 TTIs

Figure 6.15: extended multi-cell reference scenario I: the gain of applying AFR on top of optimal SCA, *stationary terminals*

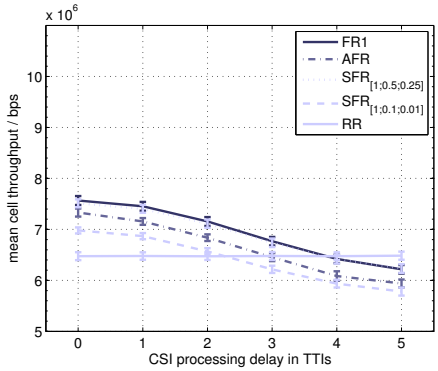
minals without harming the overall system performance, if the CSI is available within two TTIs. Note that for LTE systems, a CSI delivery time of 2 TTIs is assessed.

SFR power profiling The benefit of applying SFR on top of optimal SCA assignments in case of stationary terminals, however, is small. For both power profiles the cell-edge terminal throughput gain is smaller than 3%, independent of the CSI age. This is mainly due to the fact that the additional channel gain diversity, which is achieved by applying the power profiles, is not significant compared to the original frequency diversity of the channel. On the other hand, when using SFR power profiles the losses to the overall cell throughput are large (up to 10% for the $[1; 0.1; 0.01]$ profile configuration). This is because applying power profiles shifts power from the good situated terminals to the weak terminals, which generally accounts for overall system performance losses. This fact can also be studied when looking at the terminal throughput CDFs in Figures 6.13c–6.13e: for all four CSI ages studied, the gain for the first quartile (weakest 25%) of the terminal throughput is small, while the loss for the last quartile is large. Note that the losses for the $[1; 0.1; 0.01]$ profile are larger, as up 25 times more power is shifted toward the weaker terminals.

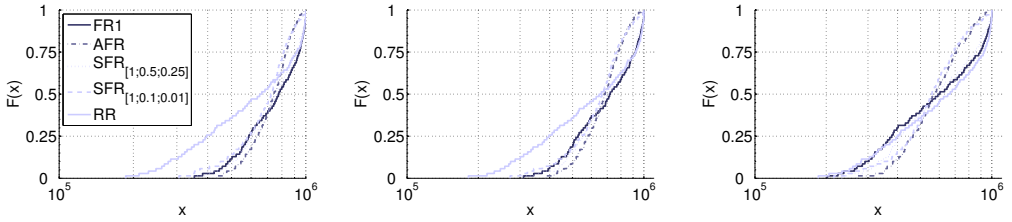
In the case of moving terminals an *artificial* reliability of the interference situation is established: resource blocks that fall into the low power sub-band do not provide better channel



(a) mean throughput (bps) of weakest terminal



(b) mean cell throughput (bps)



(c) terminal throughput cdf, 0 TTIs (d) terminal throughput cdf, 2 TTIs (e) terminal throughput cdf, 5 TTIs

Figure 6.16: extended multi-cell reference scenario I: the gain of applying AFR on top of optimal SCA, *moving terminals at speed* $v = 10\frac{m}{s}$

conditions than high power resource blocks. Accordingly, even if the quantitative ratio between arbitrary resource block changes, the qualitative ratio is roughly stable. The SFR cases, thus, make the sub-carrier assignment less susceptible to outdated or even erroneous CSI. At a CSI age of 5 TTIs, still a gain in cell-edge terminal performance of about 20% compared to the round robin case can be observed, when using the $[1; 0.1; 0.01]$ profile. Recall that, at this point, solely optimizing the sub-carrier assignments leads to a cell-edge terminal performance degradation of about 10%. The increased stability can also be seen in the terminal throughput CDF at CSI age 5 TTIs (Figure 6.13e): while optimizing the sub-carrier assignments in the FR1 case does not show any improvement in the moving terminals case, a throughput increase of approximately 25% is possible for the first quantile of terminals.

This gain, however, comes at the cost of large losses in per-cell throughput performance (Figure 6.14b). Note that for cases in which the CSI processing delay exceeds 2 TTIs the system performance drops beneath the performance of the static assignment case. At a CSI age of 5 TTIs, the performance loss is 10%. In general, both, the stabilizing, and the system performance decreasing effect are present in the performance results of the $[1; 0.5; 0.25]$ profile with a lower impact.

AFR power profiling In general, it can be stated that applying AFR increases the cell-edge terminal performance much more significantly than SFR, while at the same time it achieves are smaller decrease in the overall cell throughput performance. Particularly, looking at the AFR cell-edge terminal throughput performance in Figures 6.15a and 6.16a, there is a 10% increase in cell-edge terminal performance compared to the uniform power profile, independent of the CSI age. On the other hand, the average cell throughput goes down less than 5% for AFR with increasing CSI age. Moreover, according to the curves in the CDFs, the gain for the first quartile of the terminals is about 10%. The loss for the strongest quartile is comparable to that with the $[1; 0.1; 0.01]$ profile.

From the curves concerning the moderately moving terminals in Figure 6.16, another general insights can be derived: the significance of applying AFR power profiles increases with decreasing CSI accuracy. Particularly, the stability against fast changing channels that has been observed for SFR in the previous section, holds true with an even more significant impact for AFR. An additional gain can be achieved that is independent of the CSI age. At an age of two TTIs – recall that this is believed to be a reasonable assumption for the CSI processing delay in systems such as LTE – the performance gain in cell-edge terminal throughput compared to the SFR cases is around 5%, compared to the FR1 case around 10%, and compared to the round robin assignment is about 30%. In all cases, also in the case of AFR, the gain in cell-edge terminal performance comes at the cost of a decreased overall cell performance. Compared to SFR with the $[1; 0.1; 0.01]$ profile, the losses for AFR are significantly smaller. Compared to SFR with the $[1; 0.5; 0.25]$ profile, the losses are larger, but SFR does not achieve a significant increase of cell-edge terminal performance.

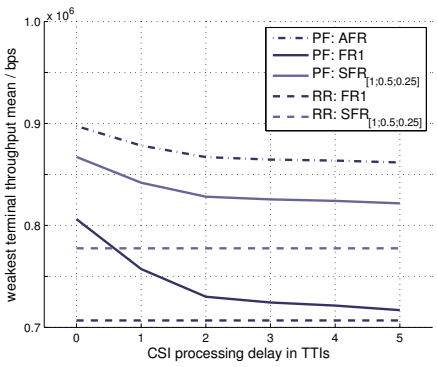
6.5 The application of AFR in PF scheduling based systems

In Section 6.2 it is shown that the application of soft frequency reuse power profiling on top of proportional fair scheduling shows some gains in weakest user performance. The gains, however, come at the cost of decreased system throughput performance. Section 6.3 shows that this decrease in system performance is not a mandatory consequence of power profiling and that – on the contrary – system throughput performance can be increased jointly with the weakest user performance, if the power levels are optimally adapted to the individual user requirements. In the last section, a power profile adaptation mechanism based on interference coupling matrices has been introduced and tested under optimal scheduling assumptions.

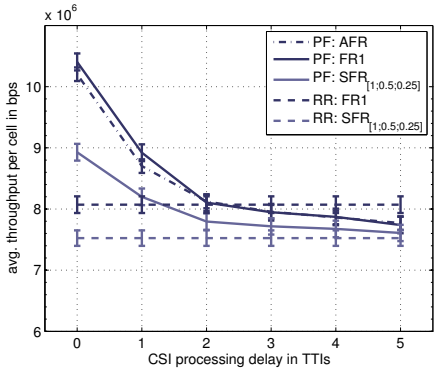
In this section, the interference coupling based AFR approach is tested under more realistic assumptions, namely a power-profile aware heuristic proportional fair scheduler (according to the flow chart in Figure 6.4) per base station in the trace file based extended multi-cell reference scenario of Section 3.3.3. Particularly, the AFR approach is tested under the same conditions as SFR in Section 6.2, and the results are compared to each other.

6.5.1 Results

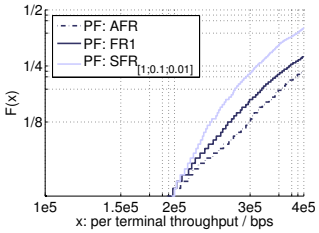
The results under realistic conditions show a similar tendency as the optimal scheduling results in the last section: the application of AFR profiles further improves the weakest terminal



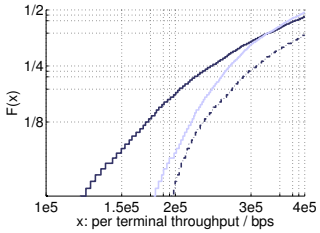
(a) mean weakest terminal throughput



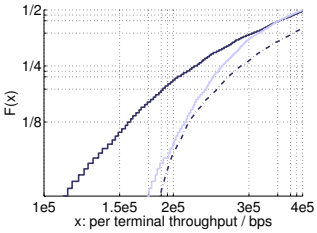
(b) mean cell throughput



(c) weak term. tp, 0 TTIs



(d) weak term. tp, 2 TTIs



(e) weak term. tp, 5 TTIs

Figure 6.17: the impact of applying AFR on top of proportional fair scheduling

throughput performance, while at the same time not significantly impacting the overall system throughput performance. Just like the application of static SFR power profiles, protection against outdated CSI information is provided, such that weakest user gains are also achievable after the maximum tested CSI processing delay of $\Delta T_{\text{csi}} = 5$TTIs. In Figures 6.17c-6.17e it is shown that a performance gain is not only present for the weakest terminal per cell (cell-edge users), but for at least the weakest quartile of terminals in the system.

6.6 Conclusions

There is a need for co-channel interference mitigation in cellular OFDMA systems. Especially terminals located at the cell border largely suffer from the power radiated by the base station of neighboring cells in their communication band. As an alternative to standard interference coordination techniques, integrated resource allocation and interference mitigation has been suggested to allow each base station the usage of the complete spectrum in each cell.

In this chapter an answer to the consequent question '*Can any performance gain be expected, if neighboring cells jointly optimize their usage of transmission resources?*' has been found. The approach of having neighboring cells jointly optimize their transmission resource usage in a frequency reuse 1 scenario provides large potential gains over the classical approach that treats interference mitigation and resource allocation per base station as separate tasks. This

has been shown by solving several instances of an according non-linear optimization problem formulation. Throughput gains of up to 100% can be achieved for the cell-edge users, which are highly impacted by co-channel interference from surrounding cells, while at the same time the overall system performance can be improved. Motivated by this fact, a heuristic algorithm based on power profiling and interference-coupling matrices has been developed, in order to achieve these gains at low computational complexity. This heuristic approach combines sub-carrier assignment at cell-level with a procedure to coordinate transmit power profiles across neighboring cells. The impact of adapting power profiles on cell throughput and cell-edge user throughput has been studied in computational experiments.

The results confirm that solely optimizing the sub-carrier assignment per cell can already considerably increase the average cell as well as the cell-edge user throughput performance. The application of soft frequency reuse on top of that increases cell-edge user performance, but decreases the overall cell performance. The presented heuristic method for adaptive frequency reuse, in comparison, increases the cell-edge user performance even slightly more, but with a lesser reduction of overall cell throughput performance. Furthermore, the aging of channel state information proves to be a serious obstacle to taking good decisions. Outdated information in a hard frequency reuse scenario, for example, may lead to a system performance that is worse than a random assignment of the sub-carriers. In contrast, both, soft and adaptive frequency reuse reduce the dependency on up-to-date information on channel conditions for optimizing the sub-carrier selection. Again, the adaptive scheme proves favorable to the static soft frequency reuse schemes analyzed here.

Chapter 7

Conclusions and Outlook

THIS THESIS considers the application of dynamic resource allocation mechanisms in the downlink of packet-oriented OFDMA based cellular systems. OFDMA systems split the available system bandwidth into several low rate sub-channels, on which data of different users is transmitted in parallel. Each active terminal is assigned a distinct subset of the sub-channels. In general, dynamic resource allocation mechanisms can improve the performance of such systems in terms of throughput, robustness and/or power efficiency by adapting the transmission parameters, such as the sub-channel transmission power, the sub-channel modulation type and coding rate, the user/sub-channel assignments, and/or the number of sub-channels per subset to the individual user needs and instantaneous sub-channel quality conditions. These performance gains, however, come at the cost of an increased computational complexity at the transmitter and the receiver and significant signaling overhead that needs to be delivered on a separate control channel.

The thesis answers the research questions that arise when applying dynamic OFDM mechanisms that have originally been developed for streaming systems or connection oriented point-to-point systems to packet-oriented cellular systems as described above. The major differences between the first and the latter lie in (1) the packet nature of the traffic, (2) the quality and quantity of the required control signaling data, and (3) the particular interference situation of cellular systems.

Conclusions The following answers to the related research questions have been found and discussed within the scope of this thesis:

1. *Are special packet-aware resource allocation techniques necessary in packet-oriented OFDMA systems?* Special packet-aware resource allocation techniques are not urgently necessary in packet oriented OFDMA systems. Standard resource allocation techniques used in systems with continuous data flows cannot directly be applied to such systems, but already the combination of such schemes with simple heuristics yields good results, such that the far more complex packet aware resource optimization techniques do not perform significantly better. Still, it is evident that by applying specialized packet resource scheduling algorithms, as the one developed from the packet-aware optimization

problem in Chapter 4, more attention can be payed to the QoS requirements of the individual packet streams, resulting in slightly improved quality of *e. g.,* voice transmissions. In general, however, slightly modified standard dynamic mechanisms do perform well in packet-oriented cellular OFDMA systems, such that techniques that adapt the transmission resources on the packet granularity level are only necessary for special purposes.

2. *Does reliable control data transmission require a disproportional split of the available transmission resources in favor of the control channel redundancy?* Reliable control data transmission in a cellular OFDMA based system is possible in a frequency reuse 1 scenario, even if only a relatively small part of the transmission resources is reserved to implement the control channel. This has been shown in Chapter 5 for a standard inband control signaling scheme, where the first few OFDM symbols in each frame are reserved for control data signaling. The split of the resources can be optimized in favor of the data channel capacity, if dynamic mechanisms are used to allocate the control data resources. In this context, two dynamic algorithms have been developed that enable dynamic power loading and dynamic coding on the control channel, respectively. By trading off the control channel redundancy against the transmission resource split in favor of the user data channel, the mechanisms optimize the system throughput subject to control data reliability. Particularly, if these dynamic means are used to optimize the control data/user data resource split, the control data signaling effort using an inband control channel can be reduced to a minimum (*i. e.,* a single symbol per frame).

3. *Can any performance gain be expected, if neighboring cells jointly optimize their usage of transmission resources?* The approach of having neighboring cells jointly optimize their transmission resource usage in a frequency reuse 1 scenario provides large potential gains over the classical approach that treats interference mitigation and resource allocation per base station as separate tasks. This has been shown by solving several instances of an according non-linear optimization problem formulation in Chapter 6. Throughput gains of up to 100% can be achieved for the cell-edge users, which are highly impacted by the co-channel interference from surrounding cells, while at the same time the overall system performance can be improved. Motivated by this fact, a heuristic algorithm based on power profiling and interference-coupling matrices has been derived, in order to achieve these gains at low computational complexity. The latter, however, shows a rather small increase in cell edge user performance at the cost of overall system performance. Nevertheless, it provides an important opportunity to increase system fairness by shifting transmission resources in favor of the weaker users.

In summary, this thesis demonstrates the potential of using dynamic mechanisms to allocate the resources of the user data, and the control data channel, as well as to realize interference mitigation by inter-cell coordination in packet-oriented OFDMA based systems. In addition to the optimization models that show the gains achievable under ideal conditions, efficient means to implement these techniques in real world systems have been presented. The latter show reduced gains in favor of significantly reduced computational complexity. By answering the above research questions, the thesis faces all significant cellular down-link related dynamic OFDMA research challenges.

Outlook Still, several issues remain as future work. First, the potential of gaining performance by having neighboring cells jointly optimizing their transmission resources is much larger than the gain observed when applying the presented heuristic based on power profiling and interference-coupling. Hence, there might exist low-complexity techniques that further approach the optimal results presented in Chapter 6. Also, combinations of inter-cell resource allocation optimization with other interference mitigation techniques (such as beamforming or coordinated multi-point transmission techniques presented in Section 6.1) are possible.

While in principle all presented mechanisms are applicable to the up- and down-link direction of such systems, the up-link direction faces additional significant challenges that are not within the scope of this thesis, such as synchronization issues, CSI signaling, and problems related to the peak-to-average ratio of the up-link transmission signal. Thus, the evaluation of the presented schemes in the up-link direction remains as a future work issue, which seems to be an easy task for researchers with the according up-link know how and this thesis at hand.

Finally, an implementation of the presented efficient means on a real-world testbed would provide the final proof of the significant performance gains achievable in according systems. For example, Eurecom's CardBus MIMO platform, which has been presented in Section 3.4 as the hardware basis with which the utilized channel traces have been recorded, could be considered for this purpose.

Appendix A

Channel Trace Fading Profile Analysis

On the following pages, an analysis of the fading profiles that have been recorded in trace files using the CardBus MIMO OFDMA hardware platform introduced in Section 3.4 and used as part of the extended multi cell simulation model of Section 3.3.3 is provided. The particular interest lies in the fading channel's

1. coherence bandwidth,
2. coherence time, and
3. the particular difference (range) between the highest and lowest attenuated sub-carrier per user and TTI.

In order to determine the coherence bandwidth of measured channel data, the according channel data frequency correlation function needs to be determined. The coherence bandwidth can then be defined as the bandwidth over which the frequency correlation function is above 0.9; a more relaxed definition requires the function to be above 0.5 [13]. The time correlation function can be used in a similar way in order to obtain the channel's coherence time.

Fading profile, 3D correlation and attenuation range histogram plots The figures on the following pages provide the desired information about the fading channel characteristics. First, Figure 7.1 presents the fading profiles (a 3D illustration of the attenuation over time and frequency) of three selected users at two different points in time. It nicely shows the channel's selectivity in frequency and time – and space, as the three users are located at different positions. The plots show the channel's fading holes, as well as some correlation in time and frequency. In order to further investigate these correlations and derive the channel's coherence bandwidth and time, the correlation coefficients of the fading profiles are calculated and depicted in Figures 7.2 and 7.3 respectively. The ranges for which the values are above 0.5 and 0.9 are determined and pointed out as the individual user's coherence bandwidth/time. Then, in Figure 7.4 the range between the weakest and strongest attenuated sub-carrier in dB per user is determined in each TTI and summarized in according histograms.

Summarizing the results, it can be stated that the measured coherence bandwidth and time suit well the LTE resource block bandwidth and TTI duration respectively.

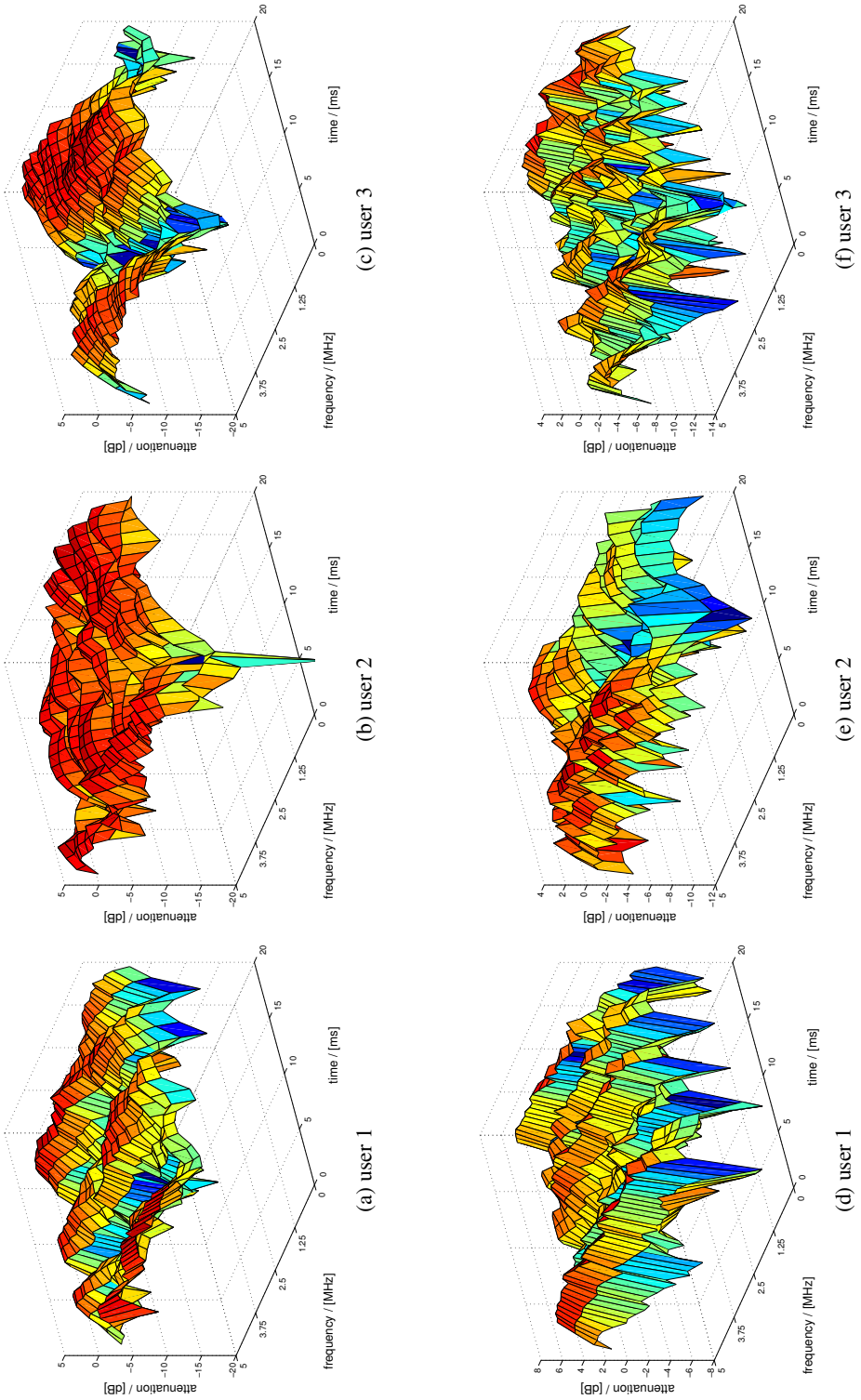
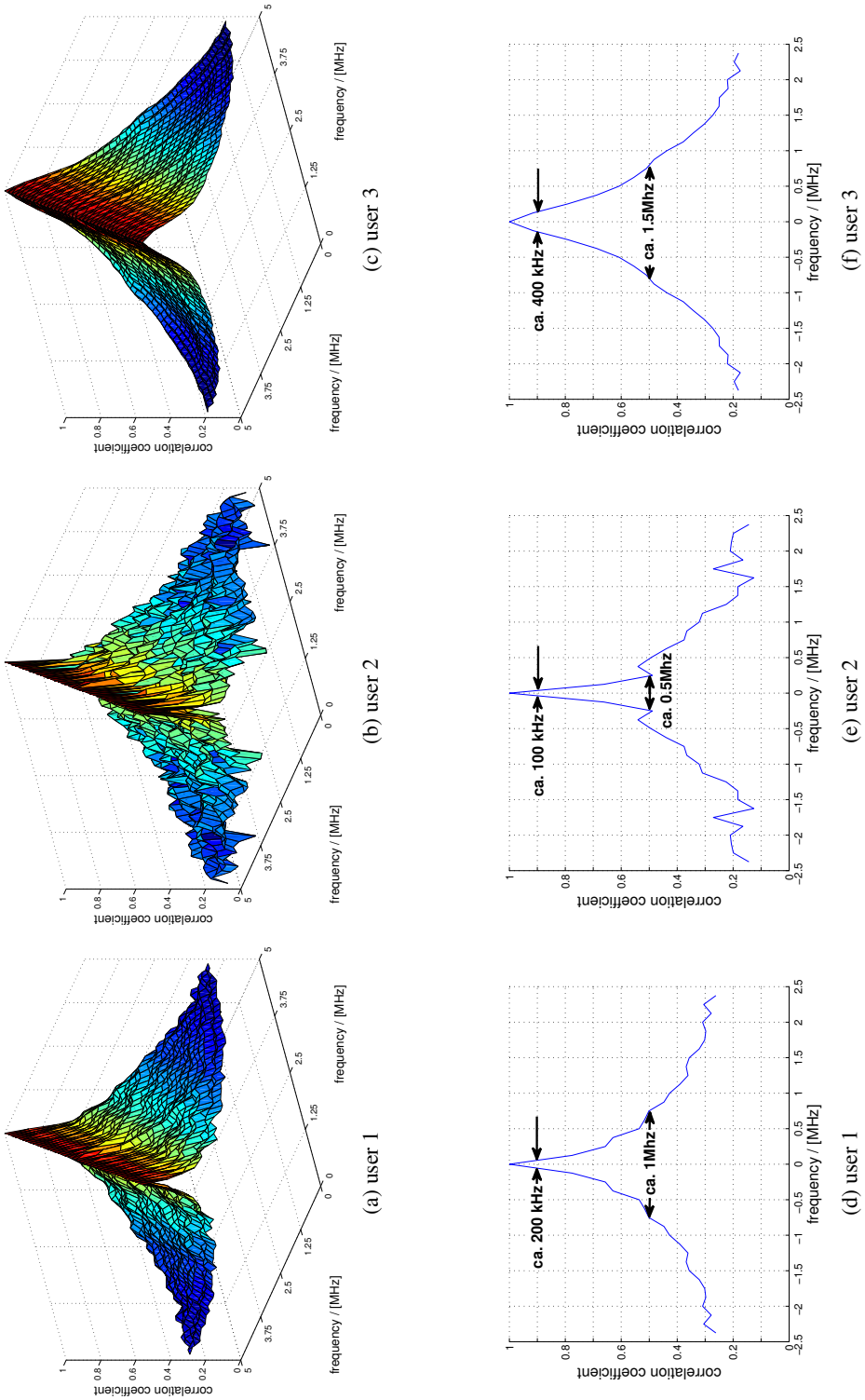


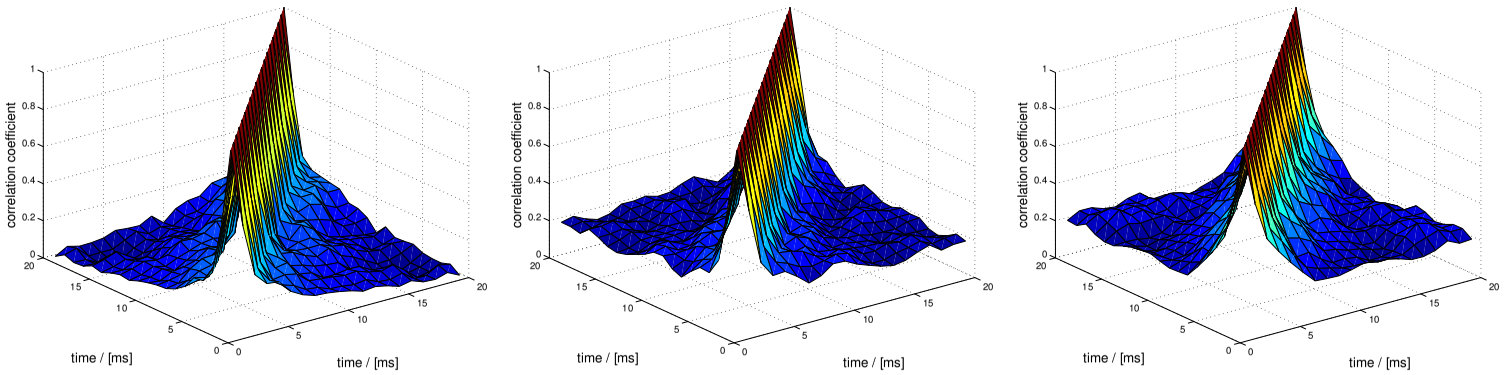
Figure 7.1: fading profiles of different users at t=1ms..20ms and t=8001ms..8020ms



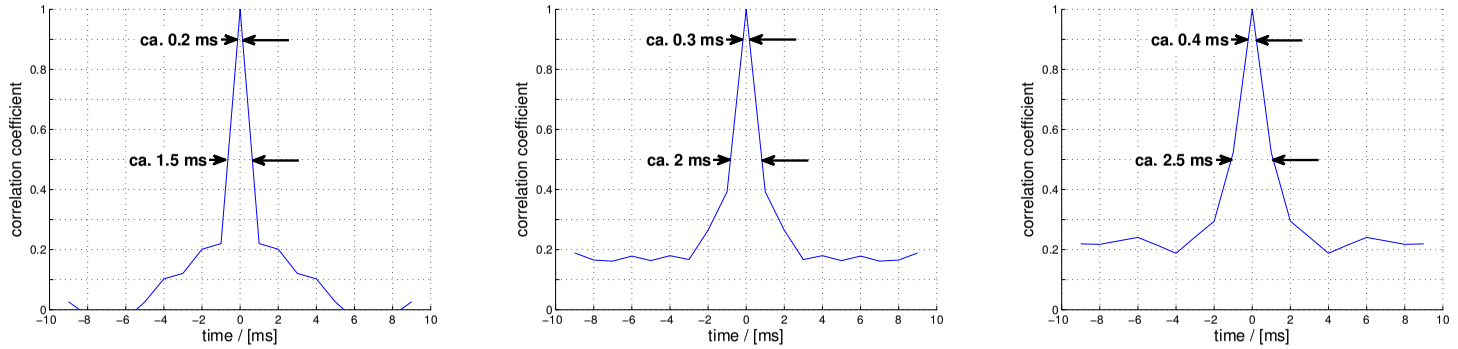
(a) user 1 (b) user 2 (c) user 3

(d) user 1 (e) user 2 (f) user 3

Figure 7.2: frequency correlation of different users in 3d and 2d (middle sub-carrier)



(a) user 1 (b) user 2 (c) user 3



(d) user 1 (e) user 2 (f) user 3

Figure 7.3: time correlation of different users on in 3d and 2d

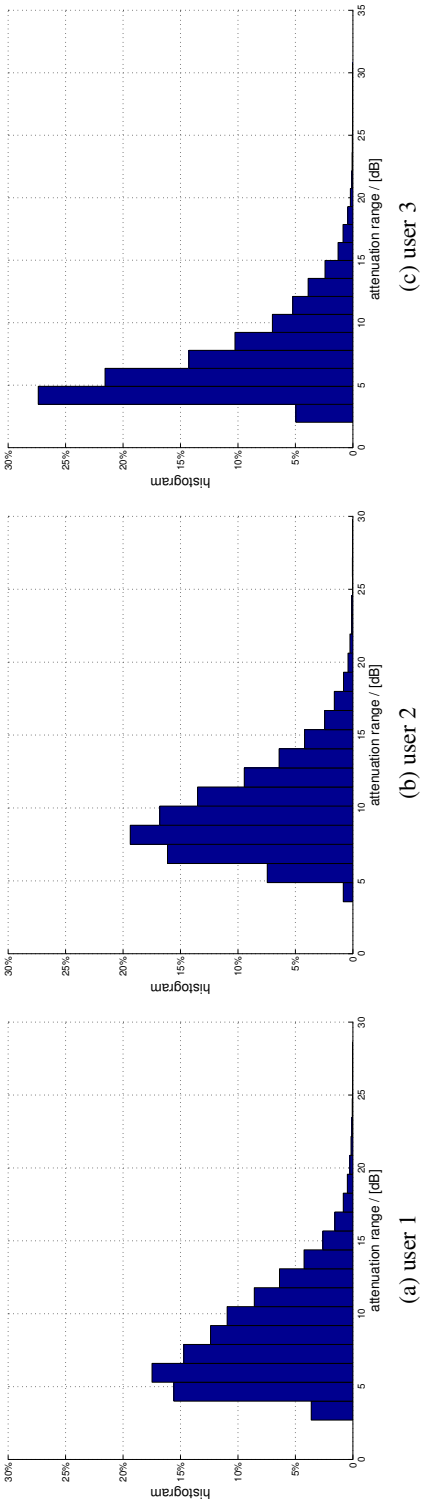


Figure 7.4: attenuation range histograms of different users

Appendix B

Publications

Journals/Magazines

- M. Bohge, J. Gross, and A. Wolisz
  ”Optimal Soft Frequency Reuse and Dynamic Sub-carrier Assignments in Cellular OFDMA Networks”
  Wiley European Transactions on Telecommunications
  pp. 704–713, vol. 21, no. 8, ISSN: 1541-8251
  August 2010.

- M. Bohge, J. Gross, M. Meyer, and A. Wolisz
  ”Dynamic Resource Allocation in OFDM Systems: An Overview of Cross-Layer Optimization Principles and Techniques”
  IEEE Network Magazine
  Special Issue: ”Evolution toward 4G wireless networking”
  pp. 53–59, vol. 21, no. 1, ISSN: 0016-1136
  Januar/Februar 2007.

- M. Bohge, J. Gross, M. Meyer, and A. Wolisz
  ”A New Optimization Model for Dynamic Power and Sub-Carrier Allocations in Packet-Centric OFDMA Cells”
  Frequenz, Journal of RF-Engineering and Telecommunications
  pp. 35–38, vol. 61, no. 1-2, ISSN: 0890-8044
  Januar/Februar 2007.

Bookchapter

- A. de Baynast, M. Bohge, D. Willkomm, and J. Gross
  Chapter 9: ”Physical Layer Modeling”
  in ”Modeling and Tools for Network Simulation”
  Editors: K. Wehrle, M. Günes, and J. Gross
  pp. 135–172, ISBN: 978-3-642-12330-6
  June 2010.

Conference Proceedings

- M. Bohge and A. Wolisz
  "The Impact of Channel State Information Processing Delay in Optimally Scheduled OFDMA Networks" *in* Proc. of the International OFDM Workshop (InOWo'09)
  pp. 282–232, Hamburg, Germany, September 2009.

- M. Bohge, J. Gross, and A. Wolisz
  "Optimal Power Masking in Soft Frequency Reuse based OFDMA Networks" *in* Proc. of the European Wireless Conference (EW '09)
  pp. 162–166, ISBN: 978-3-8007-3167-1, Aalborg, Denmark, May 2009
  BEST PAPER AWARD.

- D. Willkomm, M. Bohge, D. Hollós, J. Gross, and A. Wolisz
  "Double Hopping: A new Approach for Dynamic Frequency Hopping in Cognitive Radio Networks" *in* Proc. of the 19th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008)
  pp. 1–6, ISBN: 978-1-4244-2643-0, Cannes, France, September 2008.

- M. Bohge, F. Naghibi, and A. Wolisz
  "The use of guard bands to mitigate multiple access interference in the OFDMA uplink" *in* Proc. of he 13th International OFDM-Workshop (InoWo'08)
  pp. 75–79, Hamburg, Germany, August 2008.

- M. Bohge, A. Wolisz, A. Furuskär, and M. Lundevall
  "Multi-User OFDM System Performance Subject to Control Channel Reliability in a Multi-Cell Environment" *in* Proc. of the IEEE International Conference on Communications (ICC'08)
  pp. 3647–3652, ISBN: 978-1-4244-2075-9, Bejing, China, May 2008.

- M. Bohge and M. Renwanz
  "A Realisitic VoIP Traffic Generation and Evaluation Tool for OMNeT++" *in* Proc. of the 1. International Workshop on OMNeT++
  pp. 1–4, ISBN: 978-963-9799-45-5, Marseille, France, March 2008.

- M. Bohge and A. Wolisz
  "A Generic MAC-PHY Interface for Supporting Multi-User Dynamic OFDM Implementations on Software Defined Radios" *in* Proc. of the 5th Workshop on Software Radios (WSR'08)
  pp. 1–5, Karlsruhe, Germany, March 2008.

- M. Bohge, J. Gross, and A. Wolisz
  "A New Optimization Model for Dynamic Power and Sub-Carrier Allocations in Packet-Centric OFDMA Cells" *in* Proc. of he 11th International OFDM-Workshop (InoWo '06)
  pp. 21–25, Hamburg, Germany, August 2006.

- M. Bohge, J. Gross, and A. Wolisz
  "The Potential of Dynamic Power and Sub-carrier Assignments in Multi-User OFDM-FDMA Cells" *in* Proc. of IEEE Global Communications Conference (Globecom'05)
  pp. 2932–2936, ISBN 0-7803-9414-3, St. Louis, MO, USA, November 2005.

- A. Hess, M. Bohge, and G. Schäefer
  "Towards Distributed Network Intrusion Prevention with Respect to QoS Requirements" *in* Poster Session of 9th IFIP/IEEE International Symposium on Integrated Network Management (IM 2005)
  pp. 1-4, ISBN 0-7803-9088-1, Nice, France, May 2005.

- M. Bohge and W. Trappe
  "TESLA Certificates: An Authentication Tool for Networks of Compute-Constrained Devices" *in* Proc. of 6th international symposium on wireless personal multimedia communications (WPMC '03)
  pp. 1–5, Yokosuka, Kanagawa, Japan, October 2003.

- M. Bohge and W. Trappe
  "An Authentication Framework for Hierarchical Ad Hoc Sensor Networks" *in* Proc. of 2003 ACM workshop on Wireless Security (WiSE '03)
  pp. 79-87, ISBN: 1-58113-769-9, San Diego, CA, USA, August 2003.

Technical Reports

- D. Willkomm, M. Bohge, D. Hollos, and J. Gross
  "Double Hopping: A new Approach for Dynamic Frequency Hopping in Cognitive Radio Networks"
  TKN Technical Report Series TKN-08-001, Telecommunication Networks Group, Technische Universität Berlin, Jan. 2008.

- M. Bohge, A. Furuskär, and M. Lundevall
  "Adaptive Coding as a Means toIincrease Multi-user OFDM Control Channel Reliability"
  TKN Technical Report Series TKN-07-005, Telecommunication Networks Group, Technische Universität Berlin, Dec. 2007.

- M. Bohge, and J. Gross
  "PHY-MAC Interface Definitions for Dynamic OFDMA Systems"
  TKN Technical Report Series TKN-07-003, Telecommunication Networks Group, Technical Universität Berlin, Sept. 2007.

- M. Bohge, J. Gross, and M. Meyer
  "Optimal Packet-centric Down-link Scheduling in Cellular OFDMA"
  TKN Technical Report Series TKN-06-006, Telecommunication Networks Group, Technische Universität Berlin, Sept. 2006.

- J. Gross, and M. Bohge
  ”Dynamic Mechanisms in OFDM Wireless Systems: A Survey on Mathematical and System engineering contributions”
  TKN Technical Report Series TKN-06-001, Telecommunication Networks Group, Technische Universität Berlin, May 2006.

Acronyms

3GPP 3rd Generation Partnership Project

ACM adaptive coding and modulation

ADC analog-to-digital converter

AFR adaptive frequency reuse

ASK amplitude shift keying

AWGN additive white gaussian noise

BCH broadcast data channel

BE best effort

BEP bit error probability

BLEP block error probability

BLER block error rate

BPSK binary phase shift keying

CCE control channel element

CCI co-channel interference

CCIC co-channel interference coordination

CBR constant bit rate

CDF cumulative distribution function

CDMA code division multiple access

CE central entity

CMSR constrained max sum rate

CoMP coordinated multi-point transmission

CP cyclic prefix

CRC cyclic redundancy check

C-RNTI cell radio network temporary identity

CSI channel state information

DAB digital audio broadcasting

DAC digital-to-analog converter

DEMOD demodulator

DFT discrete Fourier transform

DL downlink

DSL Digital Subscriber Line

DVB digital video broadcasting

EMOS Eurecom MIMO Openair Sounder

FDD frequency division duplex

FDM frequency division multiplexing

FDMA frequency division multiple access

FFT fast Fourier transform

FFR fractional frequency reuse

FIFO first-in first-out

FR1 frequency reuse 1

FSK frequency shift keying

FTP file transfer protocol

GPS global positioning system

GPRS general packet radio service

HARQ hybrid automatic repeat request

HFR hard frequency reuse

HTTP hyper text transfer protocol

IC integrated circuit

ICI inter-channel interference

ICIC inter-cell interference coordination

IDFT inverse discrete Fourier transform

IFCO interference coordination

IFFT inverse fast Fourier transform

IP integer programming

ISI inter-symbol interference

ISM Industrial Scientific and Medical

GNIR gain-to-noise and interference ratio

GPS Global Positioning System

GSM Global System for Mobile Communications

HFR hard frequency reuse

LAN local area network

LOS line of sight

LP linear programming

LTE Long Term Evolution

M-ASK multilevel amplitude shift keying

M-FSK multilevel frequency shift keying

MAC medium access control layer

MAN metropolitan area network

MCM multi carrier modulation

MIESM mutual information effective SNIR metric

MIMO multiple input multiple output

MOD modulator

MOS mean opinion score

OFDM orthogonal frequency division multiplexing

OFDMA orthogonal frequency division multiple access

OOK on off keying

PDCCH physical downlink control channel

PESQ perceptual evaluation of speech quality

PF proportional fair

PFR partial frequency reuse

PHY physical layer

PPBP Poisson Pareto burst process

PSC parallel-to-serial converter

PSK phase-shift keying

QAM quadrature amplitude modulation

QPSK quadrature phase shift keying

QoS quality of service

RB resource block

RE resource element

RF radio frequency

RNC radio network controller

RR round robin

RRC radio resource control

SCA sub-carrier assignment

SCH synchronization symbol

SCM single carrier modulation

SDR software defined radio

SEP symbol error probability

SFR soft frequency reuse

SMTP simple mail transfer protocol

SNR signal-to-noise ratio

SNIR signal-to-noise plus interference ratio

SON self-organizing network

SPC serial-to-parallel converter

SPM sub-carrier/packet matching

TBS transport block size

TD timely delivery

TDD time division duplex

TDMA time division multiple access

TTI transmission time interval

UMTS Universal Mobile Telecommunications System

VLSI very large scale integration

VoIP voice over IP

WiMAX Worldwide interoperability for Microwave Access

WLAN wireless local area network

Bibliography

[1] F. Halsall, *Data Communications, Computer Networks and Open Systems, 4th Edition*, Addison-Wesley Publishing Company, 1995.

[2] J. Proakis, *Digital Communications*, McGraw-Hill, 1995.

[3] R. Steele and L. Hanzo, *Mobile Radio Communications*, John Wiley & Sons, 1999.

[4] R. Steele, *Mobile Radio Communications*, Pentech Press, 1992.

[5] W. C. Jakes, *Microwave Mobile Communications*, IEEE Press, Wiley Interscience, 1994.

[6] J. Cavers, *Mobile Channel Characteristics*, Kluwer Academic, 2000.

[7] A. Neskovic, N. Neskovic, and G. Paunovic, "Modern approches in modeling of mobile radio systems propagation envorinment," IEEE Communications Surveys and Tutorials, Third Quarter 2000.

[8] U. Sorger, Ed., *Script zur Vorlesung Mobile Kommunikation an der Technischen Hochschule Darmstadt,FB 19*, Institut für Netzwerk- und Signaltheorie, 2004.

[9] A. Aguiar and J. Gross, "Wireless channel models," Tech. Rep. TKN-03-007, Telecommunication Networks Group, Technische Universität Berlin, Apr. 2003.

[10] 3GPP; Technical Specification Group GSM/EDGE Radio Access Network, "Mobile radio interface layer 3 specification; radio resource control protocol (release 1999)," TS-04.18, June 2006.

[11] B. Walke, M.P. Althoff, and P.Seidenberg, *UMTS - A Comprehensive Introduction*, vol. 0, John Wiley & Sons, Chichester, New York, Weinheim, Brisbane, Singapore, Toronto, 350 pages, ISBN: 0470845570, Aug 2002.

[12] 3GPP: Technical Specification Group Radio Access Network, "Utran overall description (release 8)," TS-25.401, Dec. 2008.

[13] Theodore S. Rappaport, *Wireless Communications*, Prentice Hall, 1999.

[14] R. Chang, "Synthesis of band limited orthogonal signals for multichannel data transmission," *Bell Systems Technical Journal*, vol. 45, pp. 1775–1796, Dec. 1966.

[15] R. Chang and R. Gibby, "A theoretical study of performance of an orthogonal multiplexing data transmission scheme," *IEEE Transactions on Communication Technology*, vol. 16, pp. 529–540, Aug. 1968.

[16] ANSI T1.413-1995, *Network and customer installation interfaces - Asymetric digital subscriber line (ADSL) metallic interface*, Mar. 1995.

[17] ETSI ETS 300 401, *Digital Audio Broadcasting (DAB); DAB to mobile portable and fixed receivers*, Feb. 1995.

[18] ETSI EN 300 744, *Digital Video Broadcasting (DVB); Framing structure, channel coding, and modulation for digital terrestial television*, Aug. 1997.

[19] IEEE Computer Society, "IEEE 802.11-2007: IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks; specific requirements - part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications," June 2007.

[20] ETSI, *BRAN HIPERLAN Type 2, DLC Layer, Part 1: Basic Data Transport Functions*, TS 101-761-1 edition, Dec. 2001.

[21] IEEE, "Standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands and corrigendum 1," Feb. 2006, IEEE Std 802.16e-2005.

[22] 3GPP; Technical Specification Group Radio Access Network, "Physical layer aspects for evolved UTRA (release 7)," TR-25.814, Oct. 2006, Version 1.0.3.

[23] R. van Nee and R. Prasad, *OFDM Wireless Multimedia Communications*, Artech House, 2000.

[24] R. Mosier and R. Clabaugh, "Kinplex, a bandwidth efficient binary transmission system," *AIEE Transactions*, vol. 76, pp. 723–728, Jan. 1958.

[25] S. B. Weinstein and Paul M. Ebert, "Data transmission by frequency-division multiplexing using the discrete fourier transform," *IEEE Transactions on Communication Technology*, vol. 19, no. 5, pp. 628–634, Oct. 1971.

[26] B. Hirosaki, "An orthogonally multiplexed QAM system using the discrete fourier transform," *IEEE Transactions on Communications*, pp. 982–989, July 1981.

[27] W.Y. Zou and Y.Wu, "COFDM: an overview," *IEEE Transactions on Broadcasting*, vol. 41, no. 1, pp. 1–8, Mar. 1995.

[28] L. Hanzo, M. Münster, B.J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*, J. Wiley & Sons, Inc., 2003.

[29] R.E. Blahut, *Fast Algorithms for Digital Signal Processing*, MA: Addison-Wesley, 1985.

[30] T. Cover and J. Thomas, *Elements of Information Theory*, J. Wiley & Sons, Inc., 1991.

[31] H. Liu and G. Li, *OFDM-Based Broadband Wireless Networks*, John Wiley & Sons, Inc., 2005.

[32] I. Kalet, "The multitone channel," *IEEE Transactions on Communications*, vol. 37, no. 2, pp. 119–124, Feb. 1989.

[33] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, 2 edition, 1999.

[34] A. Schrijver, *Combinatorial Optimization*, Springer, 2003.

[35] P. Chow, J. Cioffi, and J. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Transactions on Communications*, vol. 43, no. 2, Feb. 1995.

[36] R. Fischer and J. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. of the IEEE Global Telecommunications Conference (Globecom '01)*, Nov. 1996, vol. 1, pp. 724–728.

[37] B. Krongold, K. Ramchandran, and D. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 23–27, Jan. 2000.

[38] J. Campello, "Practical bit loading for DMT," in *Proc. of the IEEE International Conference on Communications (ICC'99)*, June 1999, vol. 2, pp. 801–805.

[39] R. Sonalkar and R. Shively, "An efficient bit-loading algorithm for DMT applications," *IEEE Communications Letters*, vol. 4, no. 3, pp. 80–82, Mar. 2000.

[40] S. Lai, R. Cheng, K. Letaief, and C. Tsui, "Adaptive tracking of optimal bit and power allocation for OFDM systems in time-varying channels," in *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC '99)*, Sept. 1999, vol. 2, pp. 776–780.

[41] D. Hughes-Hartogs, *Ensemble Modem Structure for Imperfect Transmission Media*, U.S. Patents 4,679,227 (July 1987); 4,731,816 (March 1988); 4,833,706 (May 1989).

[42] J. Gross and M. Bohge, "Dynamic mechanisms in OFDM wireless systems: A survey on mathematical and system engineering contributions," TKN Technical Report Series TKN-06-001, Telecommunication Networks Group, Technische Universität Berlin, May 2006.

[43] J. Jang and K.B. Lee, "Transmit power adaption for multiuser OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[44] I. Kim, H. Lee, B. Kim, and Y. Lee, "On the use of linear programming for dynamic subchannel and bit allocation in multiuser OFDM," in *Proc. of the IEEE Global Telecommunications Conference (Globecom '01)*, Nov. 2001, vol. 6, pp. 3648–3652.

[45] H. Yin and H. Liu, "An efficient multiuser loading algorithm for OFDM-based broadband wireless systems," in *Proc. of the IEEE Global Telecommunications Conference (Globecom '00)*, Nov. 2000, vol. 1.

[46] W. Rhee and J. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. of the IEEE Vehicular Technology Conference (VTC Spring '00)*, May 2000, vol. 2, pp. 1085–1089.

[47] Z. Shen, J. Andrews, and B. Evans, "Optimal power allocation in multiuser OFDM systems," in *Proc. of the IEEE Global Telecommunications Conference (Globecom '03)*, Dec. 2003, vol. 1, pp. 337–341.

[48] C.Y. Wong, R.S. Cheng, K.B. Letaief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit and power allocation," *IEEE Journal on Selected Areas of Communications*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[49] C.Y. Wong, C.Y. Tsui, R.S. Cheng, and K.B. Letaief, "A real-time sub-carrier allocation scheme for multiple access downlink OFDM transmission," in *Proc. of the IEEE Vehicular Technology Conference (VTC Fall '99)*, Sept. 1999, vol. 2, pp. 1124–1128.

[50] D. Kivanc and H. Liu, "Subcarrier allocation and power control for OFDMA," in *Proc. of the 34th Asilomar Conference on Signals, Systems and Computers (ACSSC '00)*, Oct. 2000, vol. 1, pp. 147–151.

[51] S. Pfletschinger, G. Muenz, and J. Speidel, "Efficint subcarrier allocation for multiple access in OFDM systems," in *Proc. of the 7th Int. OFDM Workshop, Hamburg, Germany*, Sept. 2002.

[52] G. Kulkarni and M. Srivastava, "A channel assignment scheme for FDMA based wireless ad hoc networks in rayleigh fading environments," in *IEEE Vehicular Technology Conference*, Sept. 2002, vol. 2, pp. 1090–3038.

[53] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1150–1158, Nov. 2003.

[54] L. Xiaowen and Z. Jinkang, "An adaptive subcarrier allocation algorithm for multiuser OFDM system," in *Proc. of the IEEE Vehicular Technology Conference (VTC Fall '03)*, Oct. 2003, vol. 3, pp. 1502–1506.

[55] Y.-F. Chen, J.-W. Chen, and C.-P. Li, "A fast suboptimal subcarrier, bit, and power allocation algorithm for multiuser OFDM-based systems," in *Proc. of the IEEE International Conference on Communications (ICC'04)*, June 2004, vol. 6, pp. 3212–3216.

[56] H.S. Kim, J.S. Kwak, J.M. Choi, and J.H. Lee, "Efficient subcarrier and bit allocation algorithm for OFDMA system with adaptive modulation," in *Proc. of the IEEE Vehicular Technology Conference (VTC Spring '04)*, May 2004, vol. 3, pp. 1816–1820.

[57] A. Czylwik, "OFDM and related methods for broadband mobile radio channels," in *Proc. of Broadband Communications*, Feb. 1998, pp. 91–98.

[58] H. Rohling and R. Gruenheid, "Performance of an OFDM-TDMA mobile communication system," in *Proc. of the IEEE Vehicular Technology Conference (VTC Spring '96)*, May 1996, vol. 3, pp. 1589–1593.

[59] A. Czylwik, "Adaptive OFDM for wideband radio channels," in *Proc. of the IEEE Global Telecommunications Conference (Globecom '96)*, Nov. 1996, vol. 2, pp. 713–718.

[60] A. Barreto and S. Furrer, "Adaptive bit loading for wireless OFDM systems," in *Proc. of the IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '01)*, Sept. 2001, vol. 2, pp. 88–92.

[61] C. Mutti, D. Dahlhaus, T. Hunziker, and M. Foresti, "Bit and power loading procedures for OFDM systems with bit-interleaved coded modulation," in *Proc. of the International Conference on Telecommunications (ICT '03)*, Feb. 2003, vol. 2, pp. 1422–1427.

[62] T. Hunziker and D. Dahlhaus, "Optimal power adaptation for OFDM systems with ideal bit-interleaving and hard-decision decoding," in *Proc. of the IEEE International Conference on Communications (ICC'03)*, May 2003, vol. 5, pp. 3392–3397.

[63] M. Bohge, J. Gross, and A. Wolisz, "The potential of dynamic power and sub-carrier assignments in multi-user OFDM-FDMA cells," in *Proc. of the IEEE Global Telecommunications Conference (Globecom '05)*, Nov. 2005, vol. 5, pp. 2932–2936.

[64] M. Bohge, J. Gross, M. Meyer, and A. Wolisz, "A new optimization model for dynamic power and sub-carrier allocations in packet-centric OFDMA cells," *Frequenz, Journal of RF-Engineering and Telecommunications, "Special Issue on Selected Papers of the 2006 International OFDM Workshop"*, vol. 61, no. 1, pp. 35–38, Jan. 2007.

[65] R. Gruenheid and H. Rohling, "Performance comparison of different multiple access schemes for the downlink of an OFDM communication system," in *Proc. of the IEEE Vehicular Technology Conference (VTC Spring '97)*, May 1997, vol. 3, pp. 1365–1369.

[66] J. Gross, I. Paoluzzi, H. Karl, and A. Wolisz, "Throughput study for a dynamic OFDM-FDMA system with inband signaling," in *Proc. of the IEEE Vehicular Technology Conference (VTC Spring '04)*, May 2004, vol. 3, pp. 1787–1791.

[67] W. Wang, T. Ottosson, M. Sternad, A. Ahlen, and A. Svensson, "Impact of multiuser diversity and channel variability on adaptive OFDM," in *Proc. of the IEEE Vehicular Technology Conference (VTC Fall '03)*, Oct. 2003, vol. 1, pp. 547–551.

[68] T. Alen, A. Madhukumar, and F. Chin, "Capacity enhancement of a multi-user OFDM system using dynamic frequency allocation," in *Proc. of the IEEE Vehicular Technology Conference (VTC Fall '03)*, Oct. 2003, vol. 4, pp. 344–353.

[69] J. Gross, H. Geerdes, H. Karl, and A. Wolisz, "Performance analysis of dynamic OFDMA systems with inband signaling," *IEEE Journal on Selected Areas in Communications, Special Issue on 4G Wireless Systems*, vol. 24, no. 3, pp. 427–436, Mar. 2006.

[70] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks – Part I," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, Mar. 2005.

[71] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks – Part II," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[72] M. Bohge, J. Gross, M. Meyer, and A. Wolisz, "Dynamic resource allocation in ofdm systems: An overview of cross-layer optimization principles and techniques," *IEEE Network Magazine, Special Issue: "Evolution toward 4G wireless networking"*, vol. 21, no. 1, pp. 53–59, Jan. 2007.

[73] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," *IEEE Communications Surveys Tutorials*, vol. 11, no. 3, pp. 92–106, Mar. 2009.

[74] D. Astely, E. Dahlman, P. Frenger, R. Ludwig, M. Meyer, S. Parkvall, P. Skillermark, and N. Wiberg, "A future-radio-access framework," *IEEE Journal on Selected Areas in Communications, Special Issue on 4G Wireless Systems*, Mar. 2006.

[75] R.H. Clarke, "A statistical theory of mobile radio-reception," *Bell System Technical Journal*, pp. 957–1000, July 1968.

[76] T. Aulin, "A modified model for the fading signal at a mobile radio channel," *IEEE Transactions on Communication*, vol. 51, pp. 920–928, June 2003.

[77] A. Eisenblätter, H.-F. Geerdes, T. Koch, and U. Türke, "Momentum data scenarios for radio network planning and simulation," in *Proc. of the 2nd IEEE Conference on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'04)*, Mar. 2004.

[78] Eurecom, "Open Air Interface – project homepage," http://www.open-air-interface.org, 2009.

[79] F. Kaltenberger, M. Kountouris, R. Knopp, and D.Gesbert, ," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008)*, Sept. 2008, pp. 1–5.

[80] 3GPP; Technical Specification Group Radio Access Network, "Universal terrestrial radio access (e-utra); radio frequency (rf) system scenarios," TS-36.942, 2008, Version 8.1.0.

[81] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, Oct. 1948.

[82] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," in *Proc. of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '05)*, Sept. 2005, vol. 4, pp. 2306–2311.

[83] M. Zukerman, T.D. Neame, and R.G. Addie, "Internet traffic modeling and future technology implications," in *Proc. of the 22nd IEEE International Conference on Computer Communications (Infocom'03)*, Apr. 2003, vol. 1, pp. 587–596.

[84] W. Willinger amd M.S. Taqqu, R. Sherman, and D.V. Wilson, "Self-similarity through high-variability: statistical analysis of ethernet LAN traffic at the source level," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 71–86, Feb. 1997.

[85] R. Jain, *The art of computer systems performance analysis*, J. Wiley & Sons, Inc., 1991.

[86] Telecommunication Standardizatoin Sector of ITU (ITU-T), "Perceptual evaluation of speech quality (PESQ) tool, available at: http://www.itu.int/rec/t-rec-p.862-200102-i/en [retrieved:12/07/2007]," .

[87] Telecommunication Standardization Sector of ITU (ITU-T), "Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective mehtod for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Feb. 2001.

[88] M. Bohge and M. Renwanz, "A realisitic voip traffic generation and evaluation tool for omnet++," in *Proc. of the 1st International Workshop on OMNeT++*, Mar. 2008, pp. 1–6.

[89] The International Telegraph and Telephone Consultative Committee (CCITT), "Recommendation g.726: 40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)," Oct. 1990.

[90] M. Sternad and D. Aronsson, "Channel estimation and prediction for adaptive ofdm downlinks [vehicular applications]," in *Proc. of the 58th IEEE Vehicular Technology Conference (VTC-Fall'03)*, Oct. 2003, vol. 2, pp. 1283–1287.

[91] S. Gifford, C. Bergstrom, and S. Chuprun, "Adaptive and linear prediction channel tracking algorithms for mobile ofdm-mimo applications," in *Proc. of the IEEE Military Communications Conference (MILCOM'05)*, Oct. 2005, vol. 2, pp. 1298–1302.

[92] I.C. Wong and B.L. Evans, "Sinusoidal modeling and adaptive channel prediction in mobile OFDM systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1601–1615, Apr. 2008.

[93] András Varga, *OMNeT++ User Manual 3.2*, Mar. 2005.

[94] S.A. ILOG, *ILOG CPLEX 10.0 - User's Manual*, Jan. 2006.

[95] LINDO Systems Inc., *LINGO 10 - User's Guide*, Chicago, Illinois, 2008.

[96] Eurecom, ," http://www.openairinterface.org.

[97] F.P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[98] F.P. Kelly, A.K. Maulloo, and D.K.H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[99] H. Kim and Y. Han, "A proportional fair scheduling for multicarrier transmission systems," *IEEE Communications Letters*, vol. 9, no. 3, pp. 210–212, Sept. 2005.

[100] M. Kaneko, P. Popovski, and J. Dahl, "Proportional fairness in multi-carrier system: Upper bound and approximation algorithms," *IEEE Communications Letters*, vol. 10, no. 6, pp. 462–464, June 2006.

[101] G. Song and Y. Li, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 127–134, Dec. 2005.

[102] H. Kim, K. Kim, Y. Han, and S. Yun, "A proportional fair scheduling for multicarrier transmission systems," in *Proc. of the 60th IEEE Vehicular Technology Conference (VTC-Fall'04)*, Sept. 2004, vol. 1, pp. 409–413.

[103] Q. Wang, J. Xu, and Z. Bu, "Proportional-fair bit and power adaptation in multi-user OFDM systems," in *Proc. of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, Sept. 2006, pp. 1–4.

[104] M. Khedr, I. El-Rube, Y. Hanafy, and H. Abou-zeid, "Subcarrier opportunistic proportional fair scheduling for OFDMA systems," in *Proc. of the IEEE/IFIP International Conference on Internet (ICI'08)*, Sept. 2008, pp. 1–5.

[105] C. Liu, L. Shi, and B. Liu, "Utility-based bandwidth allocation for triple-play services," in *4th European Conference on Universal Multiservice Networks (ECUMN'07)*, feb 2007, pp. 1–6.

[106] T. Harks and T. Poschwatta, "Utility fair congestion control for real-time traffic," in *Proc. of the 25th IEEE International Conference on Computer Communications (Infocom'06)*, apr 2006, pp. 1–6.

[107] M. Bohge, J. Gross, and A. Wolisz, "A new optimization model for dynamic power and sub-carrier allocations in packet-centric OFDMA cells," in *Proc. of the 11th International OFDM-Workshop 2006 (InOWo'06)*, Aug. 2006.

[108] P. Parag, S. Bhashyam, and R. Aravind, "A subcarrier allocation algorithm for OFDMA using buffer and channel state information," in *Proc. of the 62nd IEEE Vehicular Technology Conference (VTC'05-Fall)*, Sept. 2005, vol. 1, pp. 622–625.

[109] J. Gross, J. Klaue, H. Karl, and A. Wolisz, "Cross-layer optimization of OFDM transmission systems for MPEG-4 video streaming," *Computer Communications*, vol. 27, pp. 1044–1055, July 2004.

[110] J. Gross, *Dynamic Algorithms in Multi-User OFDM Wireless Cells*, Ph.D. thesis, TU Berlin, 2006.

[111] Z. Diao, D. Shen, and V. Li, "CPLD-PGPS scheduler in wireless OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 5, no. 10, pp. 2923–2931, Oct. 2006.

[112] Y. Zhang and K. Letaief, "Adaptive resource allocation and scheduling for multiuser packet-based OFDM networks," in *Proc. of the IEEE International Conference on Communications (ICC'04)*, June 2004, vol. 5, pp. 2249–2953.

[113] 3GPP; Technical Specification Group Radio Access Network, "Physical channels and modulation (release 8)," TS-36.211, June 2007, Version 1.2.0.

[114] 3GPP:Ericsson, Samsung LGE, Panasonic, Nokia, NEC, Huawei, Nortel, Qualcomm, Motorola, NTT DoCoMo, Mitsubishi, and Alcatel-Lucent, "DL control channel stucture," R1-071820, Apr. 2007.

[115] G. Fodor, C. Koutsimanis, A. Rácz, N. Reider, A. Simonsson, and W. Müller, "Inter-cell interference coordination in OFDMA networks and in the 3GPP long term evolution system," *Journal of Communications*, vol. 4, no. 7, pp. 445–453, 2009.

[116] M.C. Necker, "Interference coordination in cellular OFDMA networks," *IEEE Network*, vol. 22, no. 6, pp. 12–19, Nov. 2008.

[117] M. Sternad, T. Ottoson, A. Ahlén, and A. Svensson, "Attaining both coverage and high spectral efficiency with adaptive OFDM downlinks," in *Proc. of the IEEE Vehicular Technology Conference (VTC Fall '03)*, Oct. 2003, vol. 4, pp. 2486–2440.

[118] 3GPP; Huawei, "Soft frequency reuse scheme for UTRAN LTE," R1-050507, May 2005.

[119] 3GPP; Huawei, "Further analysis of soft frequency reuse scheme," R1-050841, Sept. 2005.

[120] N. Scully, S. Thiel, R. Litjens, L. Jorgueski, R. Nascimento, O. Linnell, K. Zetterberg, M. Amirijoo, C. Blondia, K. Spaey, I. Moerman, I. Balan, T. Kürner, A. Hecker, T. Jansen, J. Oszmianski, and L.C. Schmelz, "D2.1 Use Cases for Self-Organising Networks," Tech. Rep., INFSO-ICT-216284 SOCRATES, 2008.

[121] A.L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse in OFDMA systems," in *Proc. of the 27th IEEE Conference on Computer Communications (Infocom '08)*, Apr. 2008, pp. 691–699.

[122] A.L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination," in *Proc. of the 28th IEEE Conference on Computer Communications (Infocom '09)*, Apr. 2009.

[123] K. Doppler, X. He, C. Witjing, and A. Sorri, "Adaptive soft reuse for relay enhanced cells," in *Proc. of the IEEE Vehicular Technology Conference (VTC Spring '07)*, Apr. 2007, pp. 758–762.

[124] A. Hernandez, I. Guio, and A. Valdovinos, "Interference management through resource allocation in multi-cell OFDMA networks," in *Proc. of the IEEE Vehicular Technology Conference (VTC Spring '09)*, April 2009.

[125] V. Corvino, D. Gesbert, and R. Verdone, "A novel distributed interference mitigation technique using power planning," in *Proc. of Wireless Communications and Networking Conference (WCNC '09)*, April 2009, vol. 2.

[126] M.C. Necker, "Towards frequency reuse 1 cellular fdm/tdm systems," in *Proc. of the 9th ACM International symposium on Modeling analysis and simulation of wireless and mobile systems (MSWiM'06)*, 2006, pp. 338–346.

[127] M.C. Necker, "Local interference coordination in cellular ofdma networks," in *Proc. of the 66th IEEE Vehicular Technology Conference (VTC'07-Fall)*, Oct. 2007, pp. 1741–1746.

[128] M.C. Neckar, "A graph-based scheme for distributed interference coordination in cellular OFDMA networks," in *Proc. of the 67th IEEE Vehicular Technology Conference (VTC'08-Spring)*, May 2008, pp. 713–718.

[129] M. Schellmann, L. Thiele, V. Jungnickel, and T. Haustein, "A fair score-based scheduler for spatial transmission mode selection," in *Proc. of the 41st Asilomar Conference on Signals, Systems and Computers (ACSSC'07)*, Nov. 2007, pp. 1961–1966.

[130] V. Jungnickel, M. Schellmann, L. Thiele, T. Wirth, T. Haustein, O. Koch, W. Zirwas, and E. Schulz, "Interference aware scheduling in the multiuser MIMO-OFDM downlink," *IEEE Communications Magazine*, vol. 47, no. 6, pp. 56–66, June 2009.

[131] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *Eurasip - Journal on Wireless Communications and Networkin (JWCN)*, , no. 2, pp. 222–235, 2004.

[132] H. Huang and V. Sivarama, "Asymptotic downlink capacity of coordinated cellular networks," in *Proc. of the 38th Asilomar Conference on Signals, Systems and Computers (ACSSC'04)*, Nov. 2004, vol. 1, pp. 850–855.

[133] Hongyuan Zhang, N.B. Mehta, A.F. Molisch, Jin Zhang, and Huaiyu Dai, "On the fundamentally asynchronous nature of interference in cooperative base station systems," in *Proc. of the IEEE International Conference on Communications (ICC'07)*, June 2007, pp. 6073–6078.

[134] V. Jungnickel, L. Thiele, T. Wirth, T. Haustein, S. Schiffermuller, A. Forck, S. Wahls, S. Jaeckel, S. Schubert, H. Gabler, C. Juchems, F. Luhn, R. Zavrtak, H. Droste G., Kadel,

W. Kreher, J. Mueller, W. Stoermer, and G. Wannemacher, "Coordinated multipoint trials in the downlink," in *Proc. of the IEEE Global Telecommunications Conference (Globecom '09)*, Dec. 2009, pp. 1–7.

[135] H.-J. Su and E. Geraniotis, "A distributed power allocation algorithm with adaptive modulation for multi-cell OFDM systems," in *Proc. of the 5th IEEE International Symposium on Spread Spectrum Technologies and Applkications*, September 1998, vol. 2, pp. 474–478.

[136] C. Yih and E. Geraniotis, "Adaptive modulation, power allocation and control for OFDM wireless networks," in *Proc. of the IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '00)*, Sept. 2000, vol. 2, pp. 809–813.

[137] C. Yih and E. Geraniotis, "Centralized power allocation algorithms for OFDM cellular networks," in *Proc. IEEE MILCOM 2003*, Oct. 2003, vol. 2, pp. 1250–1255.

[138] G. Li and H. Liu, "Downlink dynamic resource allocation for multi-cell OFDMA system," in *Proc. of the IEEE Vehicular Technology Conference (VTC Fall '03)*, Oct. 2003, vol. 3, pp. 1698–1702.

[139] H. Kim, Y. Han, and J. Coo, "Optimal subchannel allocation scheme in multicell ofdma systems," in *IEEE Conference on Vehicular Technology '04 (VTC 2004-Spring), Milano, Italy*, May 2004.

[140] L. Mendo and J.M. Hernando, "On dimension reduction for the power control problem," *IEEE Transactions on Communication*, vol. 49, no. 2, pp. 243–248, Feb. 2001.

[141] U. Türke, *Efficient Methods for WCDMA Radio Network Planning and Optimization*, Ph.D. thesis, Universität Bremen, 2006.

[142] H.-F. Geerdes, *UMTS Radio Network Planning: Mastering Cell Coupling for Capacity Optimization*, Ph.D. thesis, TU Berlin, 2008.

[143] K. Majewski, U. Türke, X. Huang, and B. Bonk, "Analytical cell load assessment in OFDM radio networks," in *Proc. of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, Sept. 2007, pp. 1–5.

[144] S. Stańczak, M. Wiczanowski, and H. Boche, *Resource Allocation in Wireless Networks*, LNCS. Springer, 2006.

[145] T.H. Cormen, C.E. Leiserson, and R.L. Rivest, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 1990.

Index